Ann Inst Stat Math (2014) 66:527–552
DOI 10.1007/s10463-014-0456-y

SPECIAL ISSUE: BAYESIAN INFERENCE AND STOCHASTIC COMPUTATION

Simulated likelihood inference for stochastic volatility models using continuous particle filtering

Michael K. Pitt · Sheheryar Malik · Arnaud Doucet

Received: 30 March 2013 / Revised: 15 January 2014 / Published online: 4 April 2014

Abstract Discrete-time stochastic volatility (SV) models have generated a considerable literature in financial econometrics. However, carrying out inference for these models is a difficult task and often relies on carefully customized Markov chain Monte Carlo techniques. Our contribution here is twofold. First, we propose a new SV model, namely SV–GARCH, which bridges the gap between SV and GARCH models: it has the attractive feature of inheriting unconditional properties similar to the standard GARCH model but being conditionally heavier tailed. Second, we propose a likelihood-based inference technique for a large class of SV models relying on the recently introduced continuous particle filter. The approach is robust and simple to implement. The technique is applied to daily returns data for S&P 500 and Dow Jones stock price indices for various spans.

Keywords Stochastic volatility · Particle filter · Simulated likelihood · State space · Leverage effect · Jumps

1 Introduction

Statistical models for time-varying conditional volatility fall broadly within two competing categories: (i) autoregressive conditional heteroscedasticity (ARCH) model,

M. K. Pitt
Department of Economics, University of Warwick, Coventry CV4 7AL, UK

S. Malik
Banque de France, 31 rue Croix des Petits Champs, 75001 Paris, France

A. Doucet (✉)
Department of Statistics, University of Oxford, Oxford OX1 3TG, UK
e-mail: doucet@stats.ox.ac.uk

originally proposed by Engle (1982) and the generalized version (GARCH) surveyed by Bollerslev et al. (1992); and (ii) stochastic volatility (SV) models as considered by Harvey et al. (1994) and Jacquier et al. (1994). Whereas the former category of models make conditional variance a deterministic function of past squared returns, SV models allow the variance to evolve according to some latent stochastic process. These are natural discrete-time versions of continuous-time models on which much of modern financial economics relies, see for example Hull and White (1987). It can also be intuitively more appealing to consider information flow, especially at higher frequencies as being governed by a stochastic process. In a similar vein, the rapidly increasing usage of high frequency intraday data for constructing so-called, realized volatility measures is intimately linked to the SV framework in financial economics (see Barndorff-Nielsen and Shephard 2002).

A major reason for the popularity of the ARCH family of models in describing the dynamics of financial market volatility is the fact that the likelihood of parameters can be explicitly written. Estimation of SV models is however greatly complicated by the stochastic evolution of volatility which implies that, unlike ARCH counterparts, the likelihood cannot be obtained in closed form. There have been different methodologies proposed in the context of parameter estimation for such models. Harvey et al. (1994) advocate a quasi maximum likelihood estimation (QMLE) procedure, whereas Jacquier et al. (1994) propose a Markov chain Monte Carlo (MCMC) method to construct a Markov chain that can be used to draw directly from the posterior distributions of the model parameters and unobserved volatilities (see also Shephard and Pitt 1997).

Over recent years, numerous SV models have been proposed to extend the standard SV model by including leverage and jumps components. A leverage effect refers to the increase in future expected volatility following bad news. The underlying reasoning is that bad news tends to decrease price, thus leading to an increase in debt-to-equity ratio (i.e. financial leverage). The firms are hence riskier and this translates into an increase in expected future volatility as captured by a negative relationship between volatility and return. In the finance literature, empirical evidence supportive of a leverage effect has been provided by Black (1976) and Christie (1982). Jumps can basically be described as rare events: large, infrequent movements in returns which are an important feature of financial markets (see Merton 1976). These have been widely documented to be important in characterizing the non-Gaussian tail behaviour of conditional distributions of returns. However, conducting inference in the resulting SV leverage jump models is an intricate task and requires the design of sophisticated MCMC schemes; see for example Eraker et al. (2003) and Omori et al. (2007).

We contribute to this literature in two ways. From a modelling viewpoint, we introduce a new SV model which is characterized by a non-linear non-Gaussian state-space form. The essential point is that the proposed hybrid model, namely SV–GARCH, attempts to bridge elements of SV and GARCH specifications. This model nests the standard GARCH model as a special case. It has the attractive feature of inheriting the same extensively well-documented unconditional properties of the standard GARCH model, but being conditionally heavier tailed. From a computational viewpoint, likelihood-based inference in non-linear non-Gaussian state-space models can be performed using particle filtering as first proposed by Kitagawa (1993, 1996). However, the resulting simulated likelihood function is not continuous, which hinders its maximization. We design here a continuous particle filter in the spirit of Malik and Pitt (2011) for a general class of SV models to obtain a continuous simulated likelihood function. The approach is simple to implement and relatively fast on a standard PC or laptop. We demonstrate the speed and robustness of the methodology by examining simulated data arising from the specified data-generating process. The generality of the method is highlighted by the fact that the standard SV or SV with leverage specifications is nested within the SV leverage with jumps model, and can thus straightforwardly be recovered imposing restrictions on the latter complete model. We also show how diagnostics, filtered volatilities, quantile plots of filtered volatilities and filtered probability of jumps can be easily estimated. It is demonstrated how simulated likelihood via particle filtering can be employed to estimate this model. Its robustness to jumps/outliers relative to GARCH is demonstrated and we also investigate its performance relative to the other three SV models mentioned, which have a comparatively deeper theoretical underpinning in the financial econometrics literature.

The structure of the paper is as follows. In Sect. 2 we describe the standard SV model, the SV with leverage model, the SV with leverage and jumps model and the SV–GARCH model. In Sect. 3 we first describe how parameter estimation can be carried out using particle filters generally, and then specifically in the context of the SV with leverage and jumps model. This methodology of course allows for no jumps or leverage as special cases. We also describe the relevant diagnostic tests for the general case. Section 4 provides results for simulation experiments testing estimator performance in the case of both SV with leverage and jumps model and SV–GARCH. Section 5 provides empirical examples using daily returns data for S&P500. We conclude in Sect. 6.

2 Volatility models

The four models to be considered in this paper are detailed in Sects. 2.1 and 2.2. There are three models in Sect. 2.1 which are the standard stochastic volatility (SV) model, the SV model with leverage (SVL) and the SV model with leverage and jumps (SVLJ). The SVL model nests the SV model, and the SVLJ model nests the other two models subject to restrictions on the parameters. In Sect. 2.2, we introduce the SV–GARCH model which nests both the standard SV model and the standard GARCH model as special cases.

2.1 Stochastic volatility specifications

The standard stochastic volatility (SV) model, see Taylor (1986), with uncorrelated measurement and state equation disturbances is given by

$$\begin{aligned} y_t &= \epsilon_t \exp(h_t/2) \\ h_{t+1} &= \mu(1-\phi) + \phi h_t + \sigma_\eta \eta_t, \quad t = 1, \ldots, T, \end{aligned} \tag{1}$$

where the shocks to returns and log-volatility are standard Gaussian so that $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $\eta_t \overset{\text{i.i.d.}}{\sim} N(0,1)$. Here, y_t is the observed return, $\{h_t\}$ are the unobserved log-volatilities, μ is the drift in the state equation, σ_η^2 is the volatility of log-volatility and ϕ is the persistence parameter. Typically, we would impose that $|\phi| < 1$, so that $h_0 \sim N\{0, \sigma_\eta^2/(1-\phi^2)\}$ yields a stationary process. This is in fact the Euler–Maruyama discretization of the continuous-time Orstein–Uhlenbeck (log-OU) process. Within the financial econometrics literature, this model is seen as a generalization of the Black–Scholes model for option pricing that allows for volatility clustering in returns.

We can take the standard SV model just described and adapt it to incorporate a leverage effect, the SVL model. We retain the form of (1), but allow for the disturbances to be correlated as

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{2}$$

Due to the timing of the disturbances, the typically negative correlation in disturbances does not affect the unconditional distribution of y_t (see Yu 2005). For example, the unconditional skewness of the returns, y_t, remains zero. Note that we can write $\eta_t = \rho\,\epsilon_t + \sqrt{1-\rho^2}\xi_t$, where $\xi_t \overset{\text{i.i.d.}}{\sim} N(0,1)$ so the state equation can then be reformulated as

$$h_{t+1} = \mu(1-\phi) + \phi h_t + \sigma_\eta \rho \epsilon_t + \sigma_\eta \sqrt{1-\rho^2}\xi_t. \tag{3}$$

We note that ϵ_t in (3) is defined as $\epsilon_t = y_t \exp(-h_t/2)$ and so the evolution of the state is non-linear. This complicates the procedure for inference using many techniques, including MCMC, but is straightforward to address using particle filter methods; see Sect. 3. It is a particularly effective scheme based on the evolution given by (3). In particular, it will be seen that the approach performs increasingly well as $|\rho|$ becomes close to unity.

The SV model with leverage which allows for jumps (SVLJ) in the returns process is now described. This is a simple extension where

$$\begin{aligned} y_t &= \epsilon_t \exp(h_t/2) + J_t \varpi_t \\ h_{t+1} &= \mu(1-\phi) + \phi h_t + \sigma_\eta \eta_t, \quad t = 1, \ldots, T \end{aligned} \tag{4}$$

with $(\epsilon_t\ \eta_t)$ as in (2). Here, $J_t \in \{0, 1\}$ is the time-t jump arrival modelled as a Bernoulli random variable with parameter p where $\varpi_t \sim N(0, \sigma_J^2)$ dictates the jump size when $J_t = 1$[1]. The jump formulation, without leverage, has been proposed by Eraker et al. (2003) who use MCMC techniques to perform inference.

We note that for the SVLJ model, the transition density of the state process h_t can be expressed as

$$f(h_{t+1}|h_t; y_t) = \int f(h_{t+1}|h_t; \epsilon_t) f(\epsilon_t|h_t, y_t)\, d\epsilon_t. \tag{5}$$

[1] This model can be considered a discrete-time counterpart to a general, continuous-time jump-diffusion model (see Duffie et al. 2000). In brief, assume log of stock price $y(t)$ and the underlying state variable,

where $f(h_{t+1}|h_t;\epsilon_t)$ is given by (3). In the leverage model without jumps, we have $f(\epsilon_t|h_t, y_t) = \delta_{y_t \exp(-h_t/2)}(\epsilon_t)$, whereas in the presence of jumps $f(\epsilon_t|h_t, y_t)$ is a mixture of this Dirac-delta mass and of a regular density; see (14) in Sect. 3.3. The rather involved transition density of (5) makes this model complicated to estimate using MCMC but the proposed particle filter scheme of Sect. 3 is straightforward to apply as it is only necessary to be able to simulate forward from this transition density.

2.2 SV–GARCH

In the spirit of studying heavier tailed volatility models we propose a new model for volatility, the SV–GARCH. If we denote the observed return y_t, and lagged conditional variance $\sigma_t^2 \equiv v_t$, then the generalized ARCH (GARCH) model as put forth by Bollerslev (1986) can be written as:

$$\begin{aligned} y_t &= \sqrt{v_t}\epsilon_t \\ v_{t+1} &= \gamma + \alpha v_t + \beta y_t^2, \quad t = 1, \ldots, T, \end{aligned} \tag{6}$$

where $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0, 1)$. Parameter restrictions $\gamma > 0$, $\alpha \geq 0$, $\beta \geq 0$ are set to ensure that conditional variances are uniformly positive, and for the existence of stationarity of the process we require the condition $\alpha + \beta < 1$ to hold. The initial condition is typically given by the unconditional expectation of the variance process

$$v_1 = \gamma/(1 - \alpha - \beta).$$

The GARCH specification implies that the conditional variance depends on the previous squared return, i.e. $y_t^2 = v_t \epsilon_t^2$. Let us define a disturbance term ζ_t as

$$\zeta_t = \varphi\epsilon_t + \sqrt{1 - \varphi^2}\xi_t \quad \text{where } \xi_t \overset{\text{i.i.d.}}{\sim} N(0, 1). \tag{7}$$

Replacing ϵ_t^2 by ζ_t^2 in the GARCH specification yields the non-linear transition equation:

Footnote 1 continued
i.e. the volatility $X(t)$, jointly solve:

$$\mathrm{d}y(t) = a^y(X(t))\mathrm{d}t + \sigma^y(X(t))\mathrm{d}B(t) + \mathrm{d}\left(\sum_{n=1}^{N_t^y} Z_n^y\right),$$

$$\mathrm{d}X(t) = g^x(X(t))\mathrm{d}t + \sigma^x(X(t))\mathrm{d}W(t) + \mathrm{d}\left(\sum_{n=1}^{N_t^x} Z_n^x\right).$$

Here, $B(t)$ and $W(t)$ are correlated Brownian motions, and N_t^y and N_t^X are homogenous (or non-homogenous) Poisson processes with Z_n^y and Z_n^x being the jump sizes for stock returns and volatility, respectively. The functions $a^y(.)$, $\sigma^y(.)$, $g^x(.)$ and $\sigma^x(.)$ are general functions subject to certain constraints.

$$\begin{aligned} v_{t+1} &= \gamma + \alpha v_t + \beta v_t \zeta_t^2 \\ &= \gamma + \alpha v_t + \beta v_t \left(\varphi \epsilon_t + \sqrt{1-\varphi^2} \xi_t \right)^2. \end{aligned} \tag{8}$$

Here, as in GARCH, parameter restrictions $\gamma > 0,\ \alpha \geq 0,\ \beta \geq 0, \alpha + \beta < 1$ apply and additionally $|\varphi| \leq 1$. In the case of $\varphi = 1$, the model collapses to the standard GARCH specification with linear transition function as in (6), whereas $\varphi = 0$ yields a specification which is 'stochastic' in nature, in that the feedback effect via the observed standardized return $\epsilon_t = y_t v_t^{-\frac{1}{2}}$ is eliminated.

The SV–GARCH model has some attractive features in that it inherits all the same unconditional properties of the well-established standard GARCH model, i.e. skewness, kurtosis and autocorrelation structure (see Bollerslev 1986), but the stochastic nature of the transition equation (8) renders the conditional distribution of returns a mixture,

$$f(y_{t+1}|Y_t) = \int f(y_{t+1}|v_{t+1}) f(v_{t+1}|Y_t) dv_{t+1}, \tag{9}$$

where $Y_t = \{y_1, \ldots, y_t\}$. The implication of this is that the model displays conditional leptokurtosis, so long as $\varphi \neq 1$. In the standard GARCH, the predictive density $f(v_{t+1}|Y_t)$ would be (degenerate) with Dirac-delta mass concentrated upon a single value. This suggests that in principle the SV–GARCH model is more robust to jumps/outliers relative to conditionally Gaussian counterparts. Authors such as Bollerslev (1987) have assumed heavier-tailed distributions such as standardized Student's t (GARCH-t) and generalized error distributions (GED), respectively, to provide robustness to outliers. The advantage of employing the SV–GARCH approach in incorporating heavier-tailed behaviour is that, unlike GARCH-t and GED, which postulate (fixed) heavier-tailed unconditional (and conditional) distribution for the returns process, this formulation with a latent stochastic process driving volatility is far less dependent on possible misspecification brought about by assuming a fixed distribution. Essentially, the path of SV–GARCH volatility can thus adjust after encountering an outlier, since in essence it remains centred on the GARCH volatility path in normal times. This feature also enables us to quantify the contribution to volatility of deviations brought about by abnormal (jumps) returns.

3 Likelihood inference via particle filtering

All the SV models described in the previous section can be formulated as non-linear state-space models. In this context, online state inference relies on the so-called filtering density $f(h_t|Y_{t,})$, $t = 1, \ldots, T$ where Y_t is contemporaneously available information. For linear Gaussian state-space models the density is Gaussian and its statistics can be computed using the Kalman filter. In the SV context, we cannot obtain a closed form expression for the required conditional density and it needs to be approximated numerically. A powerful deterministic numerical approach for non-linear state-space models is provided by Kitagawa (1987). This shares some similarities with the methodology advocated in this paper, as the aim is again to provide an approximation to the likelihood which is continuous as a function of the parameters. We focus here on particle filters which is a powerful class of simulation-based methods introduced by Gordon et al. (1993) and Kitagawa (1993, 1996).

The great advantage of the basic particle filtering scheme discussed in Gordon et al. (1993) and Kitagawa (1993, 1996) is that it only requires having to simulate forward in time from the transition density of the unobserved states. This is typically straightforward, whereas Bayesian imputation via Markov chain Monte Carlo (MCMC) is usually much more complicated. In the presence of highly informative measurements, this basic scheme can be inefficient and various improved sampling strategies have been proposed in Pitt and Shephard (1999) and Doucet et al. (2000). However the benefits of using these sophisticated techniques in the SV context is limited as observations are not typically individually very informative.

We begin by providing a description of a particle filter and then describe how this framework can be adapted to facilitate parameter estimation for a variety of SV models. In particular, we use a scheme which results in a likelihood estimator which is continuous as a function of the parameters. This allows simulated maximum likelihood methods (SML) to be employed.

3.1 Particle filtering algorithm

The basic particle filter, known as the Bootstrap filter, requires the ability to simulate from the transition density $f(h_{t+1}|h_t; y_t)$ and compute the measurement density $f(y_t|h_t)$. Suppose we have a set of random samples, 'particles', $h_t^1, \ldots, h_t^M$ with associated discrete probability masses $\lambda_t^1, \ldots, \lambda_t^M$, approximating the density $f(h_t|Y_t)$. The principle of Bayesian updating implies that the density of the state conditional on all available information can be constructed by combining a prior with a likelihood, recursive implementation of which forms the basis for particle filtering. The particle filtering algorithm thus propagates and updates these particles to yield a sample which is approximately distributed as $f(h_{t+1}|Y_{t+1})$; i.e. the true filtering density:

$$f(h_{t+1}|Y_{t+1}) \propto f(y_{t+1}|h_{t+1}) \int f(h_{t+1}|h_t; y_t) f(h_t|Y_t) dh_t. \tag{10}$$

The basic SIR algorithm is outlined below. We start at $t = 0$ with samples from $h_0^i \sim f(h_0),\ i = 1, \ldots, M$ which is generally the stationary distribution, if it exists.

Algorithm: particle filter (PF) *for* $t = 0, \ldots, T-1$:

We have samples $h_t^i \sim f(h_t|Y_t)$ *for* $i = 1, \ldots, M$.

1. *For* $i = 1 : M$, *sample* $\widetilde{h}_{t+1}^i \sim f\left(h_{t+1}|h_t^i; y_t\right)$.
2. *For* $i = 1 : M$ calculate normalized weights,

$$\lambda_{t+1}^i = \frac{\omega_{t+1}^i}{\sum_{k=1}^M \omega_{t+1}^k}, \quad \text{where } \omega_{t+1}^i = f\left(y_{t+1}|\widetilde{h}_{t+1}^i\right).$$

3. *For* $i = 1 : M$, *sample (from the mixture)* $h_{t+1}^i \sim \sum_{k=1}^M \lambda_{t+1}^k \delta_{\widetilde{h}_{t+1}^k}(h_{t+1})$.

This will yield an approximation of the desired posterior density, $f(h_{t+1}|Y_{t+1})$, as t varies. Here, $\delta(.)$ is a Dirac-delta function. Sampling in **Step 3** can be done using a multinomial sampling scheme and is computationally $O(M)$. However, a stratified scheme is typically much more efficient at **Step 3**, see Kitagawa (1996).

3.2 Likelihood evaluation

Assume that the model is indexed by a vector of fixed parameters θ. To carry out parameter estimation we estimate the log-likelihood function, which is given by

$$\log L(\theta) = \log f(y_1, \dots, y_T|\theta) = \sum_{t=1}^{T} \log f(y_t|\theta; Y_{t-1}), \tag{11}$$

where

$$f(y_{t+1}|\theta; Y_t) = \int f(y_{t+1}|h_{t+1}; \theta) f(h_{t+1}|Y_t; \theta) dh_{t+1}. \tag{12}$$

As the particle filter delivers samples $\{\widetilde{h}^i_{t+1}\}$ from $f(h_{t+1}|Y_t; \theta)$ after **Step 1** of **Algorithm : PF**, we may estimate the predictive density (12) as

$$\widehat{f}(y_{t+1}|\theta; Y_t) = \frac{1}{M}\sum_{k=1}^{M} f\left(y_{t+1}|\widetilde{h}^k_{t+1}; \theta\right) = \frac{1}{M}\sum_{k=1}^{M} \omega^k_{t+1}.$$

The terms ω^k_{t+1} are simply the unnormalized weights computed in **Step 2** of **Algorithm : PF**. The estimation of the likelihood is therefore a by-product of a single run of the particle filter. The estimator for the log-likelihood would therefore be

$$\log \widehat{L}_M(\theta) = \sum_{t=1}^{T} \log \widehat{f}(y_t|\theta; Y_{t-1}) = \sum_{t=1}^{T} \log\left(\frac{1}{M}\sum_{k=1}^{M} \omega^k_t\right). \tag{13}$$

This was first proposed by Kitagawa (1993, 1996) which uses it to perform (approximate) maximum likelihood parameter estimation. One drawback of this approach is that the estimated likelihood function will not be continuous as a function of θ. This hinders the maximization of the associated simulated likelihood function and the computation of standard errors using conventional techniques. This non-continuity problem arises because of the resampling step, i.e. **Step 3** of the **Algorithm : PF**. Even if we generate the same uniforms at each time step, the resampled particles will not be close as we are sampling from the following discontinuous empirical distribution function,

$$\widehat{F}(h_{t+1}) = \sum_{k=1}^{M} \lambda^k_{t+1} I\left(h_{t+1} - \widetilde{h}^k_{t+1}\right),$$

where $I(\bullet)$ is an indicator function which takes a value of unity when the argument is positive and zero otherwise. An alternative proposed recently consists of using the continuous resampling procedure described in Malik and Pitt (2011). We simply replace this empirical distribution function by an approximation $\widetilde{F}(h_{t+1})$, which is essentially constructed to be continuous in h_{t+1}. It is important to note that as $M \to \infty$, $\widetilde{F}(h_{t+1}) \to \widehat{F}(h_{t+1}) \to F(h_{t+1}|Y_t)$, with $F(h_{t+1}|Y_t)$ being the true predictive distribution function. In practice the difference between $\widetilde{F}(h_{t+1})$ and $\widehat{F}(h_{t+1})$ becomes negligible for moderate M. For the precise form of $\widetilde{F}(h_{t+1})$ and details about the method of continuous resampling from this distribution function, we refer the reader to Malik and Pitt (2011).[2]

The computational overhead is in $O(T \times M \times \log M)$ due to the necessary sorting of the sampled h_{t+1}. The random numbers (or equivalently the random numbers seed) used in **Step 1** of **Algorithm : PF** are fixed. If it is possible to sample from $f(h_{t+1}|h_t; y_t, \theta)$ continuous way, then we obtain a continuous, in θ, estimator of the likelihood function which can be maximized numerically. We now describe the method for simulating from $f(h_{t+1}|h_t; y_t, \theta)$ in a continuous manner for the models of Sect. 2.

3.3 Implementation of stochastic volatility with leverage and jumps model

Given the replacement of the resampling step (**Step 3**) of the basic SIR algorithm with a continuous resampling scheme, implementing the particle filter for parameter estimation in the context of the standard SV model (see Sect. 2) is straightforward. Our method only requires sampling from expression (5). The simpler models, standard SV and SV with leverage, may of course be estimated in exactly the same way imposing the necessary restrictions. In **Step 1** of **Algorithm : PF**, we sample from $f\ (h_{t+1}|\ h_t;\ y_t)$ given in (5) as follows:

$$\textbf{Step1}.\begin{cases} \textbf{(1a)}\ \ For\ i = 1 : M,\ \ sample\ \ \epsilon_t^i \sim f(\epsilon_t|h_t^i, y_t). \\ \textbf{(1b)}\ \ For\ i = 1 : M,\ \ sample\ \ \widetilde{h}_{t+1}^i \sim f(h_{t+1}|h_t^i; \epsilon_t^i). \end{cases}$$

The density of **Step (1a)** is a mixture of the form

$$f(\epsilon_t|h_t, y_t) = \Pr(J_t = 0|h_t, y_t) \times \delta_{y_t \exp(-h_t/2)}(\epsilon_t) \\ +\Pr(J_t = 1|h_t, y_t) \times N(\epsilon_t|\upsilon_{\epsilon_1}, \sigma_{\epsilon_1}^2) \tag{14}$$

with a singular component at $y_t\ \exp(-h_t/2)$, corresponding to no jump, and a regular component corresponding to a jump. The expressions of υ_{ϵ_1}, $\sigma_{\epsilon_1}^2$, $\Pr(J_t = 1|h_t, y_t)$ are given in Appendix A. This distribution function can be inverted easily allowing simple continuous simulation by using fixed uniform random variates. The simulation from the density $f(h_{t+1}|h_t^i; \epsilon_t^i)$ for **Step (1b)** may be performed straightforwardly by

[2] The generality and robustness of the methodology described in Malik and Pitt (2011) have been demonstrated by Duan and Fulop (2009) on credit risk models and Christoffersen et al. (2010) on affine and non-affine volatility models.

applying (3). The non-normalized weights for **Step 2** in the SIR algorithm are of the form:

$$f\left(y_{t+1}|\widetilde{h}^i_{t+1}\right) = (1-p)\left\{2\pi\exp\left(\widetilde{h}^i_{t+1}\right)\right\}^{-\frac{1}{2}}\exp\left(-\frac{1}{2}y^2_{t+1}\exp\left(-\widetilde{h}^i_{t+1}\right)\right)$$
$$+\ p\left\{2\pi\left[\exp(\widetilde{h}^i_{t+1})+\sigma^2_J\right]\right\}^{-\frac{1}{2}}\exp\left(-\frac{1}{2}y^2_{t+1}\left[\exp(\widetilde{h}^i_{t+1})+\sigma^2_J\right]^{-1}\right).$$

3.4 Diagnostics

Standard approaches involved in specification analysis of time-series models is to investigate the properties of residuals in terms of their dynamic structure and unconditional distributions. This is infeasible given the latent dimension of the model under consideration. Alternatively, therefore, to test the hypothesis that the prior and model are true, we require the distribution function

$$u_t = F(y_t|Y_{t-1}) = \int F(y_t|h_t)f(h_t|Y_{t-1})dh_t.$$

In the specific case of SV with leverage and jumps, the distribution function can be estimated by

$$\widehat{u}_t = (1-p)\left\{\frac{1}{M}\sum_{i=1}^{M}\Phi\left(y_t\exp\left(-\widetilde{h}^i_t/2\right)\right)\right\}$$
$$+p\ \left\{\frac{1}{M}\sum_{i=1}^{M}\Phi\left(y_t\left[\exp(\widetilde{h}^i_t)+\sigma^2_J\right]^{-1/2}\right)\right\},$$

where $\Phi(.)$ denotes the standard normal distribution function and $\widetilde{h}^i_t$ arise from **Step 1b** of **Algorithm : PF**. If the parameters and model are true, then the estimated distribution functions should be independently uniformly distributed through time, so $\widehat{u}_t \sim UID(0,1)$, for $t = 1,\ldots,T$, as $M \to \infty$ [see Rosenblatt (1952)].

3.5 Model comparison

We have concentrated on maximum likelihood approaches for inference. However, we can also conduct model comparison, in a Bayesian context, by computing marginal likelihoods of competing models. We have the marginal likelihood,

$$f(y|M_k) = \int f(y|\theta_k; M_k)f(\theta; M_k)\mathrm{d}\theta_k,$$

where $f(y|\theta_k; M_k)$ is our likelihood approximation via the particle filter for model M_k $(k = 1,\ldots,K)$ given the model specific maximum likelihood estimate of the parameter vector θ_k resulting from the optimization of the likelihood function. We may express this as

$$f(y|M_k) = \int \frac{f(y|\theta_k; M_k) f(\theta_k; M_k)}{g(\theta_k|y, M_k)} \quad g(\theta_k|y, M_k) \mathrm{d}\theta_k,$$

where $g(\theta_k|y, M_k)$ is a multivariate Gaussian or t-distribution centred at the maximum likelihood estimate [or the mode of $f(y|\theta_k; M_k) f(\theta_k; M_k)$] with the variance given by the inverse of the observed information matrix. This importance sampling scheme leads to an approximation:

$$\widehat{f}(y|M_k) = \frac{1}{S} \sum_{j=1}^{S} \frac{\widehat{f}\left(y|\theta_k^j; M_k\right) f\left(\theta_k^j; M_k\right)}{g\left(\theta_k^j|y, M_k\right)},$$

where $\theta_k^j \sim g(\theta_k|y, M_k)$. In practice this may only take a small number of draws as the posterior may be close to being log-quadratic (asymptotically under the usual assumptions this will be the case). Once the appropriate prior density $f(\theta_k; M_k)$ is selected, this model comparison scheme based on the ratios of marginal likelihoods between competing models can be implemented. Given the fact that we integrate out the parameter vector and the states, through particle filtering, when computing the marginal likelihoods, we do not suffer from the nuisance parameter problem encountered in similar contexts using likelihood ratio tests.

3.6 Implementation of SV–GARCH model

We apply the same general methodology described above for the estimation for the SV–GARCH model. The procedure is similar and is conducted relatively straightforwardly, within the standard **Algorithm: PF** framework. In this case, we require only to simulate forward from the transition equation (8) in conjunction with continuous resampling at **Step 3.** Thus, no additional modifications are required, as in the case of SV with leverage and jumps. As before, output such as filtered volatilities, quantiles and diagnostics are again obtained as a by-product of the procedure. This is briefly described in Appendix B.

We note the particle filter method has particular advantages for the initialization at time $t = 0$ for the SV–GARCH process. In **Algorithm : PF** the state (variance) at time $t = 0$ is denoted as v_0. Rather than setting v_0 in line with the unconditional expectation of the variance which is $\gamma/(1 - \alpha - \beta)$, see Sect. 2.2, we can attempt to draw v_0 from the unconditional distribution of v_t. For the initialization of **Algorithm: PF**, we require that at $t = 0$, we have $v_0^i \sim f(v_0)$, where $i = 1, \ldots, M$ for a particular parameter ordinate θ. To do this, we can simply start by simulating (or deterministically choosing) values v_{-l}^i for $i = 1, \ldots, M$. That is, we start with M values at l lags in the past and simply iterate through the transition equation, in this case given by (2.8), to obtain v_{-l+1}^i then v_{-l+2}^i until we obtain M values $v_0^i \sim f(v_0)$, where $i = 1, \ldots, M$. The forgetting property in the GARCH model (if stationarity is satisfied) will lead to v_0^i arising from the invariant distribution of the time series, provided that the number of lags l is sufficiently large. By choosing the random variates (standard Gaussian variables) to be constant as θ changes, the continuity of the resulting estimated likelihood is preserved. The particle filter approach therefore provides a simple and numerically fast solution for the initialization problem for stationary time series.

4 Simulation experiments

4.1 Stochastic volatility with leverage and jumps

Now, we investigate parameter estimation in the case of SV with leverage and jumps model. We run the smooth particle filter and maximize the estimated log-likelihood with respect to the parameter vector $\theta = (\mu, \phi, \sigma_\eta^2, \rho, \sigma_J^2, p)$.We begin by simulating two time series of length 1000 and 2000, setting parameters $\theta = (\mu, \phi, \sigma_\eta^2, \rho, \sigma_J^2, p) = (0.5, 0.975, 0.02, -0.8, 10, 0.10)$. These values for parameters are in line with those that have been adopted in similar contexts in the literature. The smooth particle filter is run 50 times using a different random number seed, but keeping the dataset fixed. The estimated log-likelihood is maximized with respect to θ for each run. In Table 1, the average of the resulting 50 maximum likelihood estimates ($\overline{\text{ML}_\text{S}}$) and 50 variance estimates ($\overline{\text{Var}}$), along with the variance for the sample of maximum likelihood estimates $\text{Var}(\text{ML}_\text{S})$, are reported for different cases considered. The variance covariance matrix is again estimated using the OPG estimator.

We examine the ratio of the variance of the maximum likelihood estimates to the variance of each parameter with respect to the data. These are, for $M =$

Table 1 Fixed dataset

	$\overline{\text{ML}_\text{S}}$	$\overline{\text{Var}} \times 10^2$	$\text{Var}(\text{ML}_\text{S}) \times 10^2$		$\overline{\text{ML}_\text{S}}$	$\overline{\text{Var}} \times 10^2$	$\text{Var}(\text{ML}_\text{S}) \times 10^2$
$M = 300,\ T = 1{,}000$				$M = 300,\ T = 2{,}000$			
μ	0.5595	3.0020	0.0602	μ	0.4770	1.2653	0.03098
ϕ	0.9648	0.0103	0.0002	ϕ	0.9680	0.00522	0.00013
σ_η^2	0.0458	0.0186	0.0002	σ_η^2	0.0338	0.00661	0.00012
ρ	−0.7072	1.0326	0.0162	ρ	−0.7419	0.7275	0.01352
σ_J^2	10.176	813.98	6.9054	σ_J^2	7.7568	207.71	1.19598
p	0.0769	0.0754	0.0012	p	0.11263	0.0659	0.00079
$M = 600,\ T = 1{,}000$				$M = 600,\ T = 2{,}000$			
μ	0.5650	2.9623	0.03853	μ	0.4830	1.2760	0.01097
ϕ	0.9648	0.0103	0.00013	ϕ	0.9681	0.0052	0.00005
σ_η^2	0.0461	0.0192	0.00012	σ_η^2	0.0338	0.0067	0.00008
ρ	−0.7026	1.0333	0.00665	ρ	−0.7396	0.7425	0.00622
σ_J^2	10.174	823.13	2.5625	σ_J^2	7.7929	216.21	0.87021
p	0.0764	0.0771	0.00045	p	0.1115	0.0667	0.00047

Performance of the smooth particle filter for the stochastic volatility model with leverage and jumps for two cases, $T = 1{,}000$ and 2,000, considering $M = 300, 600$ for each case

Table 2 50 different datasets. Analysis of the maximum likelihood estimator for stochastic volatility with leverage and jumps model for cases, $M = 200$, 500 and 900. $T = 2{,}000$ in all cases

	$\overline{\text{ML}_\text{S}}$	$\overline{\text{Var}} \times 10^2$	$\text{Var}(\text{ML}_\text{S}) \times 10^2$
$M = 200$			
μ	0.49151	2.0908	1.7937
ϕ	0.97101	0.0140	0.0181
σ_η^2	0.02211	0.0087	0.0072
ρ	−0.8468	1.3943	1.1835
σ_J^2	9.8470	954.42	621.81
p	0.10458	0.1300	0.0699
$M = 500$			
μ	0.5000	2.2045	1.5714
ϕ	0.9719	0.0153	0.0107
σ_η^2	0.0224	0.0097	0.0065
ρ	−0.8371	1.4793	1.1215
σ_J^2	9.8013	1018.7	637.60
p	0.1036	0.1367	0.0631
$M = 900$			
μ	0.4972	2.1724	1.6280
ϕ	0.9720	0.0146	0.0100
σ_η^2	0.0225	0.0090	0.0075
ρ	−0.8450	1.5008	1.1664
σ_J^2	9.8524	1007.0	648.20
p	0.1037	0.1350	0.0653

300, $T = 1000$:(0.0201, 0.0209, 0.0108, 0.01578, 0.0085, 0.0159); $M = 600$, $T = 1000$:(0.0131, 0.0132, 0.0062, 0.0064, 0.0032, 0.0059); $M = 300$, $T = 2000$:(0.0245, 0.0251, 0.0186, 0.0186, 0.0058, 0.0121) and $M = 600$, $T = 2000$:(0.0086, 0.0095, 0.0121, 0.0084, 0.0040, 0.0070). These ratios suggest that the variance of the simulated estimates is small in comparison to the variance induced by the data.

Next, we generate 50 different time series each of length $T = 2000$, setting values of parameters $\theta = (\mu, \phi, \sigma_\eta^2, \rho, \sigma_J^2, p) = (0.5, 0.975, 0.02, -0.8, 10, 0.10)$. Keeping the random number seed fixed, we run the smooth particle filter in turn for each of the time series and maximize the estimated log-likelihood with respect to θ for each run. The average of 50 maximum likelihood estimates ($\overline{\text{ML}_\text{S}}$) and 50 variance estimates ($\overline{\text{Var}}$) along with mean-squared errors $\text{Var}(\text{ML}_\text{S})$ are reported in Table 2, for each of the three cases considered. Variance estimates are computed using the OPG estimator for the variance covariance matrix.

In testing for bias we find very encouraging results. We find that all parameters, except the leverage parameter ρ (which is estimated with slight bias), are either within or on the boundary of their 95 % confidence limits. It should be pointed out that unbiasedness is an asymptotic property associated with the likelihood and there is no reason for us not to expect some degree of bias given a time series of moderate length, such as what we are considering for purposes of our experiments. The results are stable

Table 3 50 different datasets. Analysis of the maximum likelihood estimator for stochastic volatility with leverage and jumps model

We set parameter values $\mu = 0.25$, $\phi = 0.975$, $\sigma_\eta^2 = 0.025$, $\rho = -0.8$, $\sigma_J^2 = 0.5$ and $p = 0.10$. $M = 500$ and $T = 2{,}000$

	Small jump–high intensity		
	$\overline{\text{ML}}_\text{S}$	$\overline{\text{Var}} \times 10^2$	$\text{Var}(\text{ML}_\text{S}) \times 10^2$
μ	0.21240	3.4545	2.7098
ϕ	0.97290	0.0066	0.0073
σ_η^2	0.02917	0.0132	0.0148
ρ	−0.85641	0.7031	0.6688
σ_J^2	0.63322	95.170	60.103
p	0.23544	4.3614	6.8037

Table 4 50 different datasets. Analysis of the maximum likelihood estimator for stochastic volatility with leverage and jumps model

We set parameter values $\mu = 0.25$, $\phi = 0.975$, $\sigma_\eta^2 = 0.025$, $\rho = -0.8$, $\sigma_J^2 = 10$ and $p = 0.01$. $M = 500$ and $T = 2{,}000$

	Large jump–low intensity		
	$\overline{\text{ML}}_\text{S}$	$\overline{\text{Var}} \times 10^2$	$\text{Var}(\text{ML}_\text{S}) \times 10^2$
μ	0.25359	1.9024	1.3926
ϕ	0.97293	0.0063	0.0074
σ_η^2	0.02673	0.0067	0.0071
ρ	−0.82253	0.5556	0.4225
σ_J^2	9.6201	2162.1	3884.2
p	0.01325	0.0756	0.0202

across different values of M. We note that the settings for this experiment were one of a large jump variance σ_J^2 with very high intensity, p. One would expect the additional noise induced by these settings to render the estimation of the SV components less accurate (see Eraker et al. 2003). Our findings suggest that in spite of having large jumps with high intensity, our procedure delivers highly reliable estimates for all the parameters.[3]

We proceed to investigate how the error in estimation is affected by varying the intensity and jump size. The results in Table 3 suggest that having smaller jumps occurring with high intensity induces a slight amount of bias is estimating σ_η^2, ρ and p. In sharp contrast, if large jumps occur at a very low frequency, i.e. setting $p = 0.01$, the accuracy of our estimates is greatly enhanced; see Table 4. In this case, all parameters fall well within their 95 % confidence limits. Using simulated data generated with large jump–low intensity calibration for θ, we provide the diagnostic check (see Sect. 3.4) for the SV with leverage and jumps model in addition to a plot of the data, filtered standard deviation and filtered jump probabilities in Fig. 1.[4] The diagnostic test illustrated by the QQ plot and autocorrelation function (acf) indicate the prior and model are correct.

[3] $E(\widehat{\theta}) - \theta = Bias \backsim N(0, \frac{MSE}{50})$ where the mean squared error (MSE) is $E[(\widehat{\theta} - \theta)^2]$.

[4] Note that the plots in each of these figures illustrate output generated by a single run of the smooth particle filter.

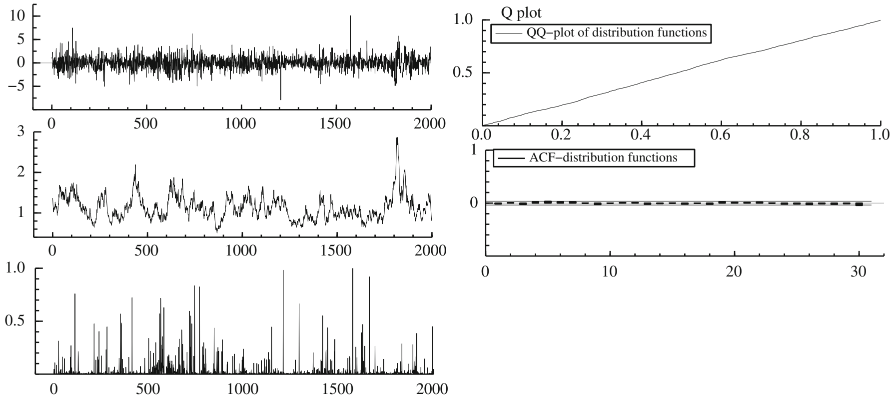

Fig. 1 Fixed simulated dataset. Parameters $\mu = 0.25$, $\phi = 0.975$, $\sigma_\eta^2 = 0.025$, $\rho = -0.8$, $\sigma_J^2 = 10$, $p = 0.01$ and a single run of the continuous particle filter. *Left panel* (i) Plot of data, (ii) filtered standard deviation, (iii) estimated jump probabilities. *Right panel* (i) QQ-plot of estimated distribution functions, $\widehat{u}_t$ (ii) correlogram of $\widehat{u}_t$. $M = 500$, $T = 2{,}000$

Table 5 50 different datasets. Analysis of the maximum likelihood estimator for SV–GARCH model

	$\overline{\text{ML}}_\text{S}$	$\overline{\text{Var}} \times 10^2$	$\text{Var}(\text{ML}_\text{S}) \times 10^2$		$\overline{\text{ML}}_\text{S}$	$\overline{\text{Var}} \times 10^2$	$\text{Var}(\text{ML}_\text{S}) \times 10^2$
$\varphi = 0.05$				$\varphi = 0.50$			
μ	0.0132	0.0034	0.0041	μ	0.0138	0.0033	0.0049
α	0.9210	0.0253	0.0231	α	0.9208	0.0243	0.0204
β	0.0707	0.0178	0.0138	β	0.0701	0.0179	0.0141
φ	0.0785	4.4531	1.4143	φ	0.4208	8.1521	7.5707
$\varphi = 0.10$				$\varphi = 0.90$			
μ	0.0114	0.0027	0.0033	μ	0.0131	0.0033	0.0036
α	0.9252	0.0221	0.0249	α	0.9218	0.0198	0.0189
β	0.0674	0.0164	0.0184	β	0.0697	0.0159	0.0141
φ	0.1529	9.988	4.2387	φ	0.8695	5.2917	3.2639

$M = 500$ and $T = 2{,}000$. Parameters $\mu = 0.010$, $\alpha = 0.925$, $\beta = 0.069$ and φ is varied

4.2 SV–GARCH

We now consider the performance of the estimator in the case of the SV–GARCH model. We generated 50 different time series each of length $T = 2000$. Keeping the random number of seed fixed, we run the smooth particle filter in turn for each of the time series and maximize the estimated log-likelihood with respect to $\theta = (\mu, \alpha, \beta, \varphi)$ for each run. We conduct four different experiments keeping the values of μ, α, β fixed at 0.010, 0.925 and 0.069, respectively, and taking the values of $\varphi \in \{0.05, 0.10, 0.50, 0.90\}$. The average of 50 maximum likelihood estimates ($\overline{\text{ML}}_\text{S}$) and 50 variance estimates ($\overline{\text{Var}}$) along with mean-squared errors $\text{Var}(\text{ML}_\text{S})$ are reported in Table 5 for each of the four cases considered. In all cases we find that biases are not significantly different from zero and the true values of the parameters lie well within their 95 % confidence limits. In unreported results, we repeated this experiment taking $M = 300$ and 600. There was no substantial variability in the results and the finding of unbiasedness remained unaltered an all cases.

5 Empirical examples

We now employ the described methodology to estimate four models: (i) stochastic volatility (SV), (ii) stochastic volatility with leverage (SVL), (iii) stochastic volatility with leverage and jumps (SVLJ) and (iv) SV–GARCH model, using daily returns S&P 500 over three different spans. Returns are continuously compounded and scaled by 100; holidays and weekends are excluded. This is a prominent index with actively traded futures and European option contracts. The spans we consider cover the well-documented episodes of market stress, October 1987, October 1997, late summer–fall 1998 as well as the most recent episode in fall 2008. For each of the series, the parameter estimates and standard errors, log-likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for these four specifications are reported in Tables 6, 7, 8 and 9; see Sakamoto et al. (1986) for a review. The AIC and BIC values are employed for purposes of model comparison given that the specifications are characterized by differing levels of complexity. We illustrate the actual returns data, along with the quantiles of filtered standard deviation and filtered jump probabilities for SVLJ specification for the spans considered in Figs. 2, 3, 4 and 5. These figures suggest that the path of the estimated filtered standard deviation captures adequately the underlying volatility of the returns process in addition to identifying periods which may be described as market stress, i.e. short periods of time with clusters of large movements in returns. In addition, the filtered probabilities adequately identify jump times.

Estimates of the jump probabilities (times) and average jumps size allow us to better understand the contribution of these components to volatility, especially during periods of market stress. Understanding this contribution is extremely important because jump risk can typically not be hedged away and thus investors demand higher premia to carry this risk.[5] From our estimates of the jump components (p and σ_J^2), it is revealed that jumps over the three spans considered can indeed be rare events which occur (approximately) between 1.3 and 2 times per year. The average jumps sizes across the three spans in Tables 6, 7 and 8 do tend to differ, with the largest being over the span containing the October 19, 1987 crash (Table 6). The diagnostics do not reveal any evidence of potential misspecification of the SVLJ model for any series. For the spans considered in Tables 7 and 8, we find that the magnitude of the estimated leverage parameter is high, $\rho > |0.8|$. In contrast we find a relatively lower estimate of $\rho = -0.33$ for the earlier span (see Table 6). Furthermore, it is for this span that we find the highest gain of SVLJ in log-likelihood terms over the SVL model. We find that the inclusion of a leverage effect in general is extremely important when modelling SV. This is indicated by the substantial gain in the log-likelihood over the standard SV model in all cases.

[5] Evidence of large jump risk premia is found by Pan (2002) (see also Eraker et al. 2003).

Table 6 Parameter estimates for S&P500 daily returns data for the period 02/02/1982–29/12/1989

	ML estimate	Standard error
SV: log-lik value = −2654.6, AIC = 5312.2, BIC = 5332.0		
μ	−0.2476	0.1052
ϕ	0.9492	0.0120
σ_η^2	0.0639	0.0099
SVL: log-lik value = −2645.4, AIC = 5298.8, BIC = 5321.2		
μ	−0.1781	0.1018
ϕ	0.9436	0.0098
σ_η^2	0.0693	0.0093
ρ	−0.3170	0.0647
SVLJ: log-lik value = −2621.1, AIC = 5254.2, BIC = 5287.8		
μ	−0.1376	0.1363
ϕ	0.9804	0.0064
σ_η^2	0.0147	0.0043
ρ	−0.3315	0.0957
σ_J^2	34.749	15.037
p	0.0061	0.0026
SV–GARCH: log-lik value = −2632.5, AIC = 5273.0, BIC = 5295.4		
γ	0.0703	0.0123
α	0.6744	0.0345
β	0.2568	0.0280
φ	0.0565	0.2235

$M = 500$. GARCH: log-lik $= -2738.1$, AIC $= 5482.2$, BIC $= 5499.0$

In Table 9, a longer time series from 31/03/1987 to 13/01/2011 is considered, which covers all the spans analysed above including the financial turmoil in fall 2008. We find that the SVLJ model describes well the evolution of S&P 500 volatility over this longer span with 6,000 observations. The estimate of leverage $\rho = -0.67$ falls between those found in the previous examples. Although jumps occur with roughly the same frequency found in the smaller samples, the average jump size is higher than that found in Tables 7 and 8, but lower as compared to that in Table 6. As a comparison we also fit the model to daily Dow Jones Composite returns (Fig. 6). We find that the leverage effect is of a comparable magnitude to that found for S&P 500 over the same span, but jumps arrive 1.5 times less often and tend to be larger, i.e. the average jumps size was found to be greater by a factor of 1.7 relative to S&P 500.

In terms of model comparison, it is found that for the spans considered in Tables 7 and 8, SVL is the preferred specification given that it yields the lowest BIC values. In terms of AIC, the SVLJ appears to be the preferred specification. This disagreement in model selection is driven by the fact that the BIC places a much larger penalty on additional parameters in models in comparison to the AIC. In contrast, the cases considered in Table 6 and Table 9 indicate that the SVLJ specification is the one where both the AIC and BIC values are minimized. Focusing on SV–GARCH, we find that it is generally outperformed by the SVL (indicated by an approximately 30

Table 7 Parameter estimates for S&P 500 daily returns data for the period 16/05/1995–24/04/2003

	ML estimate	Standard error
SV: log-lik value = −3044.7, AIC = 6095.4, BIC = 6112.2		
μ	0.1318	0.1819
ϕ	0.9821	0.0059
σ_η^2	0.0226	0.0048
SVL: log-lik value = −2994.0, AIC = 5996.0, BIC = 6018.4		
μ	0.2424	0.0977
ϕ	0.9737	0.0046
σ_η^2	0.0304	0.0049
ρ	−0.8106	0.0435
SVLJ: log-lik value = −2991.5, AIC = 5995.0, BIC = 6028.6		
μ	0.2548	0.1000
ϕ	0.9765	0.0040
σ_η^2	0.0269	0.0047
ρ	−0.8288	0.0432
σ_J^2	6.1967	0.4483
p	0.0089	0.0035
SV–GARCH: log-lik value = −3045.5, AIC = 6099.0, BIC = 6121.4		
γ	0.0098	0.0033
α	0.8878	0.0123
β	0.1041	0.0110
φ	0.0112	0.8464

$M = 500$. GARCH: log-likelihood = −3074.5, AIC = 6155.0, BIC = 6171.8

log-likelihood point gain and according to the information criteria) when leverage is relatively high, whereas interestingly it outperforms SVL for the example of the span considered in Table 6. In this case, we found a relatively less pronounced estimate of leverage and significantly larger contribution of incorporating jump components. In all cases, SV–GARCH decisively gains over the standard GARCH model. This is reinforced furthermore by the finding that φ is found to be close to zero in all cases, thus suggesting a stochastic as opposed to GARCH-type evolution for volatility. For the span considered in Table 6, we illustrate the returns, filtered standard deviation paths and quantiles in Fig. 7. Moreover, the robustness of SV–GARCH to jumps/outliers relative to the standard GARCH model is demonstrated in terms of the log-likelihood error which captures the predictive gain of the SV–GARCH when such events occur.

6 Conclusion

This paper has attempted to provide a unified particle filter-based methodology to conduct likelihood-based inference on the unknown parameters of discrete-time SV models, incorporating both a leverage effect and jumps in the returns process. An advantage of this unified methodology over MCMC is that it delivers the filtered path of the states, jump probabilities (i.e. in the case of SV with leverage and jumps) and output

Table 8 Parameter estimates for S&P 500 daily returns data for the period 19/12/2000–12/12/2008

	ML estimate	Standard error
SV: log-lik value = −2866.0, AIC = 5738.0, BIC = 5754.0		
μ	0.5006	0.4821
ϕ	0.9937	0.0027
σ_η^2	0.0168	0.0035
SVL: log-lik value = −2806.5, AIC = 5621.0, BIC = 5643.4		
μ	0.5858	0.1473
ϕ	0.9878	0.00386
σ_η^2	0.0229	0.00228
ρ	−0.8438	0.03773
SVLJ: Log-lik value = −2800.2, AIC = 5612.4, BIC = 5646.0		
μ	0.5852	0.14901
ϕ	0.9877	0.00248
σ_η^2	0.0245	0.00422
ρ	−0.8634	0.03825
σ_J^2	3.8493	0.03825
p	0.0079	0.00309
SV–GARCH: log-lik value = −2844.3, AIC = 5696.6, BIC = 5718.0		
γ	0.0076	0.0018
α	0.8963	0.0129
β	0.1009	0.0126
φ	0.0099	0.8884

$M = 500$. GARCH: log-lik = −2909.1, AIC = 5824.2, BIC = 5841.06

Table 9 Left: Parameter estimates for S&P 500 daily returns for 31/03/1987–13/01/2011. GARCH log-lik = −8318, AIC = 16642, BIC = 16662. Right:Parameter estimates for Dow Jones Composite daily returns for 3 1/03/1987–13/01/2011. GARCH log-lik = −8134, AIC = 16273, BIC = 16293. M = 500

	ML estimate	Standard error		ML estimate	Standard error
SV: log-lik = −8,163, AIC = 16,331, BIC = 16,351			SV: log-lik = −7,970, AIC = 15,946, BIC = 15,966		
μ	−0.0994	0.1488	μ	−0.2153	0.1112
ϕ	0.9864	0.0025	ϕ	0.9802	0.0033
σ_η^2	0.0283	0.0022	σ_η^2	0.0337	0.0026
SVL: log-lik = −8,078, AIC = 16,164, BIC = 16,191			SVL: log-lik = −7,905, AIC = 15,818, BIC = 15844		
μ	−0.0088	0.0840	μ	−0.1080	0.0751
ϕ	0.9771	0.0024	ϕ	0.9725	0.0030
σ_η^2	0.0377	0.0027	σ_η^2	0.0380	0.0026
ρ	−0.6381	0.0324	ρ	−0.5616	0.0374
SVLJ: log-lik = −8,046.1, AIC = 16,104, BIC = 16,144			SVLJ: log-lik = −7,866, AIC = 15,743, BIC = 15,784		
μ	0.0112	0.0905	μ	−0.1080	0.0805
ϕ	0.9831	0.0020	ϕ	0.9725	0.0023
σ_η^2	0.0268	0.0027	σ_η^2	0.0233	0.0026
ρ	−0.6724	0.0327	ρ	−0.6794	0.0374

Table 9 continued

	ML estimate	Standard error		ML estimate	Standard error
σ_J^2	16.992	2.3341	σ_J^2	27.330	4.3646
p	0.0055	0.0016	p	0.0036	0.0011
SV–GARCH: log-lik $= -8{,}147$, AIC $= 16{,}305$, BIC $= 16{,}332$			SV–GARCH: log-lik $= -7{,}952$, AIC $= 15{,}912$, BIC $= 15{,}939$		
γ	0.0103	0.0013	γ	0.0157	0.0022
α	0.8620	0.0068	α	0.8369	0.0091
β	0.1338	0.0055	β	0.1543	0.0080
φ	0.0112	0.0155	φ	0.0012	0.2461

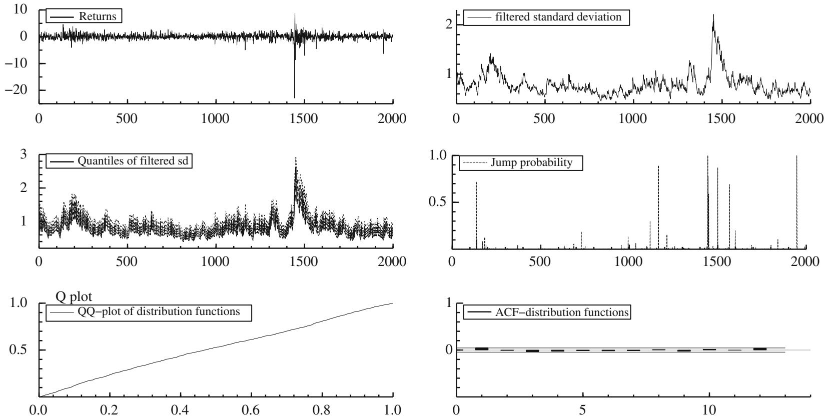

Fig. 2 Daily S&P 500 returns over the period 02/02/1982–29/12/1989. SV with leverage and jumps model. (i) Returns data, (ii) quantiles of filtered standard deviation and (iii) estimated jump probabilities (iv) QQ-plot of estimated distribution functions, $\widehat{u}_t$ and (v) associated correlograms of $\widehat{u}_t$. $M = 500$

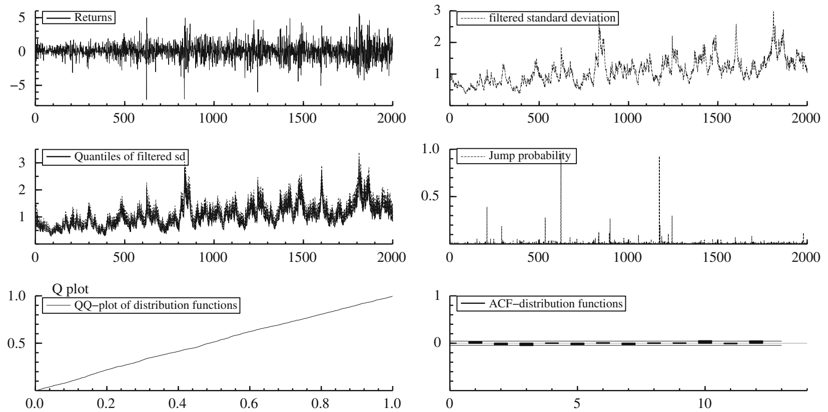

Fig. 3 Daily S&P 500 returns over the period 16/05/1995–24/04/2003. SV with leverage and jumps model. (i) Returns data, (ii) quantiles of filtered standard deviation, (iii) estimated jump probabilities, (iv) QQ-plot of estimated distribution functions, $\widehat{u}_t$ and (v) associated correlograms of $\widehat{u}_t$. $M = 500$

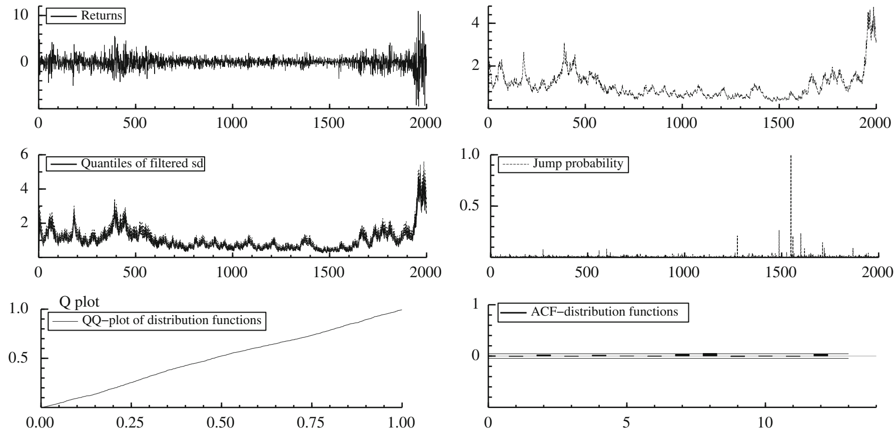

Fig. 4 Daily S&P 500 returns over the period 19/12/2000–12/12/2008. SV with leverage and jumps model. (i) Returns data, (ii) quantiles of filtered standard deviation, (iii) estimated jump probabilities, (iv) QQ-plot of estimated distribution functions, $\widehat{u}_t$ and (v) associated correlograms of $\widehat{u}_t$. $M = 500$

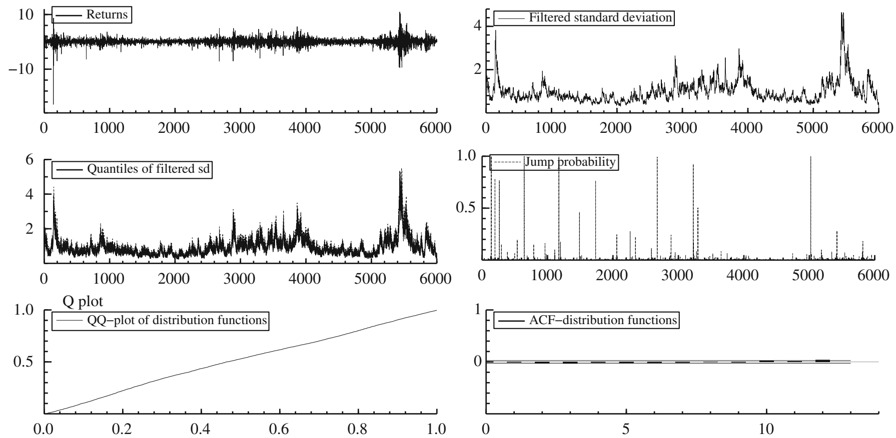

Fig. 5 Daily S&P 500 returns over the period 31/03/1987–13/01/2011. SV with leverage and jumps model. (i) Returns data, (ii) quantiles of filtered standard deviation and (iii) estimated jump probabilities, (iv) QQ-plot of estimated distribution functions, $\widehat{u}_t$ and (v) associated correlograms of $\widehat{u}_t$. $M = 500$

required to perform diagnostics. Implementation is easy and has the benefit of being both faster is terms of computation time and more general than many alternatives in the literature. With regard to generality, note that the standard SV and SV with leverage models (SVL) are restricted forms of the SV with leverage and jumps model (SVLJ). It was highlighted how the proposed methodology can easily facilitate parameter estimation for all three types of models without any alteration in the basic structure of the algorithm and as a consequence also allow for model comparison. The Monte Carlo experiments indicate that the method is both robust and statistically efficient. When examining finite sample bias in parameters, very encouraging results are found, even when considering very high jump intensity.

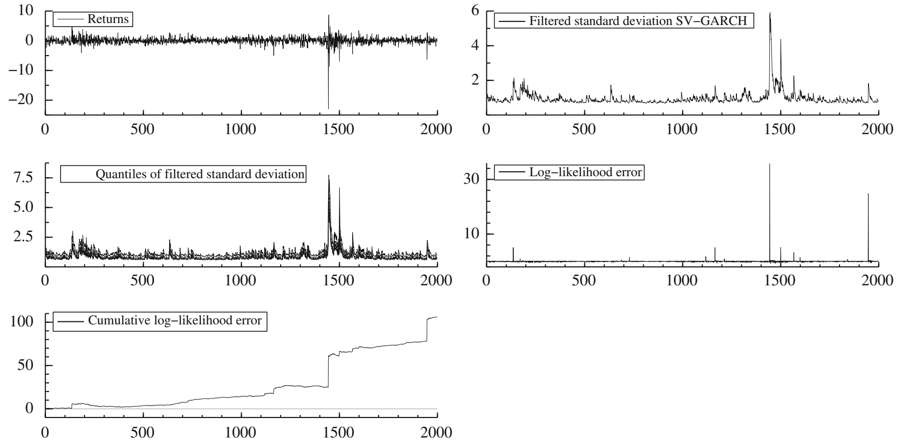

Fig. 6 Daily S&P 500 returns over the period 02/02/1982–29/12/1989. SV–GARCH model. (i) Returns data, (ii) filtered standard deviation, (iii) quantiles of filtered standard deviation,(iv) error in log-likelihood components between SV–GARCH and GARCH and (v) cumulative error. $M = 500$

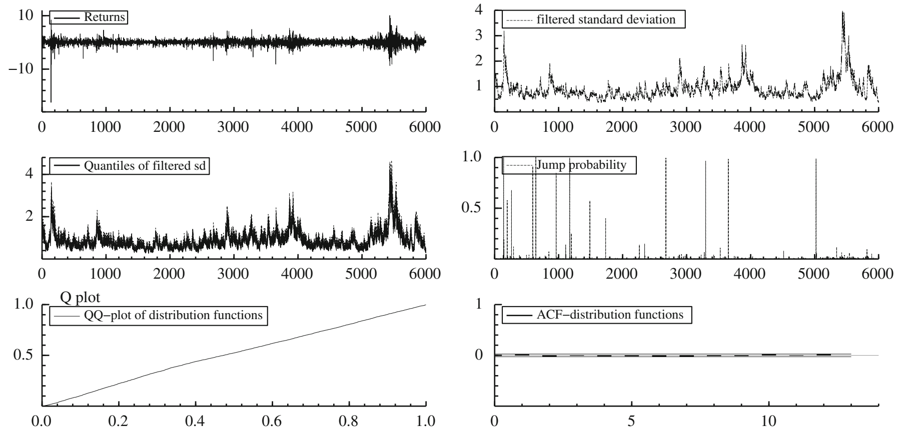

Fig. 7 Daily Dow Jones Composite returns over the period 31/03/1987–13/01/2011. SV with leverage and jumps model. (i) Returns data, (ii) quantiles of filtered standard deviation and (iii) estimated jump probabilities, (iv) QQ-plot of estimated distribution functions, $\widehat{u}_t$ and (v) associated correlograms of $\widehat{u}_t$. $M = 500$

The proposed methodology was used to estimate four models (SV, SVL, SVLJ and SV–GARCH) for daily S&P 500 returns and compare their relative performance over various time spans considering log-likelihood values and corresponding Akaike and Bayesian information criteria. The SVLJ model did very well in identifying jumps times and adequately detecting periods of market stress. Of particular interest in these applications was assessing how leverage, frequency of jumps and average jump size differed over the various spans. The inclusion of leverage was found to be very important in modelling stochastic volatility in all cases. The inclusion of the jump components provided a further gain in predictive ability which varied in magnitude over the different spans we considered. Moreover, considering a long span ($T = 6{,}000$), covering all the well-documented episodes of market stress (i.e. 1987, 1997, 1998 and 2008), it was found that allowing for jumps led to a substantial gain in excess of 30 log-likelihood points after having incorporated leverage. In the comparative example of Dow Jones, this gain was close to 40 log-likelihood points. The SVLJ thus comprehensively outperformed all the other competing models in this case and was favoured by both Akaike and Bayesian information criteria. Additionally, the SV–GARCH model consistently outperformed the standard GARCH and SV models. By considering the error in the predictive log-likelihood components, the robustness of the SV–GARCH model to outliers relative to GARCH was illustrated. It was found that the estimated value of φ was generally closer to zero than unity, which given the structure of the model would imply a more stochastic evolution for volatility rather than purely deterministic process implied by the boundary case of $\varphi = 1$.

7 Appendix A

Appendix A deals with the specific implementation of the particle filter for the case of leverage and jumps. This relates to Sect. 3.3. Specifically, we are concerned with **Step (1a)** of **Algorithm: PF**. We describe how to sample continuously (via inversion of the cumulative distribution function) from the mixture,

$$f(\epsilon_t|h_t, y_t) = \sum_{j=0}^{1} f(\epsilon_t|J_t = j; h_t, y_t) \Pr(J_t = j|h_t, y_t)$$

where the conditional probability of a jump is given by

$$\begin{aligned}\Pr(J_t = 1|h_t, y_t) &= \frac{\Pr(y_t|h_t, J = 1)\Pr(J = 1)}{\Pr(y_t|h_t, J = 1)\Pr(J = 1) + \Pr(y_t|h_t, J = 0)\Pr(J = 0)},\\ &= \frac{\mathrm{N}(y_t|0; \exp(h_t) + \sigma_J^2)p}{\mathrm{N}(y_t|0; \exp(h_t) + \sigma_J^2)p + \mathrm{N}(y_t|0; \exp(h_t))(1 - p)}.\end{aligned}$$

and

$$f(\epsilon_t|J = 1; h_t, y_t) \propto f(y_t|J = 1, h_t, \epsilon_t) f(\epsilon_t).$$

As we have $f(\epsilon_t|J = 1; h_t, y_t) \propto \mathrm{N}(y_t|\epsilon_t \exp(h_t/2); \sigma_J^2) \times \mathrm{N}(\epsilon_t|0; 1)$, it follows that

$$f(\epsilon_t|J_t = 1; h_t, y_t) = \mathrm{N}\left(\upsilon_{\epsilon_1}, \sigma_{\epsilon_1}^2\right) \quad \text{where } \upsilon_{\epsilon_1} = \frac{y_t \exp(h_t/2)}{\exp(h_t) + \sigma_J^2}$$

$$\text{and } \sigma_{\epsilon_1}^2 = \frac{\sigma_J^2}{\exp(h_t) + \sigma_J^2}.$$

If the process does not jump, there is a Dirac-delta mass at the point $\epsilon_t = y_t \exp(-h_t/2)$. We therefore have the expression (14). If we denote $p_t^* \equiv \Pr(J_t = j|h_t, y_t)$, then this mixture is

$$f(\epsilon_t|h_t, y_t) = (1 - p_t^*)\ \delta_{y_t \exp(-h_t/2)}(\epsilon_t) + p_t^*\ \mathrm{N}\left(\epsilon_t|\upsilon_{\epsilon_1}, \sigma_{\epsilon_1}^2\right). \tag{15}$$

We may invert the corresponding distribution function $F(\epsilon_t|h_t, y_t)$ straightforwardly allowing for draws which are continuous as a function of our parameters.

Assume we have generated a uniform random variate $U \sim \mathrm{UID}(0, 1)$. We show how to generate a single sample $\epsilon_t = F^{-1}(U|h_t, y_t)$ accordingly, where $\epsilon_t^* = y_t \exp(-h_t/2)$,

$$K = \Phi\left(\frac{\epsilon_t^* - \upsilon_{\epsilon_1}^1}{\sigma_{\epsilon_1}^1}\right) p_t^*,$$

$p_t^* \equiv \Pr(J_t = j|h_t, y_t)$ again, and $\Phi(.)$ denotes the standard normal distribution function. The following scheme is applied:

- If $U \leq K$, set $\epsilon_t = \upsilon_{\epsilon_1} + \sigma_{\epsilon_1}\Phi^{-1}\left(\frac{u}{p_t^*}\right)$.
- If $K < U \leq K + (1 - p_t^*)$, set $\epsilon_t = y_t \exp(-h_t/2)$.
- If $U > K + (1 - p_t^*)$, set $\epsilon_t = \upsilon_{\epsilon_1} + \sigma_{\epsilon_1}\Phi^{-1}\left(\frac{U-(1-p_t^*)}{p_t^*}\right)$.

The above probability integral transform procedure is repeated for each of the uniform $u_1, \ldots, u_M$ to obtain the required sample $\epsilon_t^i \sim f(\epsilon_t^i|h_t^i, y_t)$, $i = 1, \ldots, M$.

8 Appendix B

Particle filter estimation of SV–GARCH model

We start at $t = 0$ with samples from the stationary distribution of GARCH, $v_0^i \sim f(v_0)$, $i = 1, \ldots, M$.

Algorithm : PF *for* $t = 0, \ldots, T-1$:

We have samples $v_t^i \sim f(v_t|Y_t)$ *for* $i = 1, \ldots, M$.

1. *For* $i = 1 : M$, *sample* $\widetilde{v}_{t+1}^i \sim f(v_{t+1}|v_t^i)$.
2. *For* $i = 1 : M$ calculate normalized weights,

$$\lambda_{t+1}^i = \frac{\omega_{t+1}^i}{\sum_{k=1}^M \omega_{t+1}^k}, \quad \text{where } \omega_{t+1}^i = f\left(y_{t+1}|\widetilde{v}_{t+1}^i\right)$$
$$= \left\{2\pi\widetilde{v}_{t+1}^i\right\}^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{y_{t+1}^2}{\sqrt{\widetilde{v}_{t+1}^i}}\right).$$

3. *For* $i = 1 : M$, *sample* $v_{t+1}^i \sim \sum_{k=1}^M \lambda_{t+1}^k \delta_{\widetilde{v}_{t+1}^k}(v_{t+1})$.

As in the case of SV with leverage and jumps, we replace **Step 3** with the continuous resampling scheme described in Malik and Pitt (2011). Parameters of the SV–GARCH $\theta = (\mu, \alpha, \beta, \varphi)$ can be estimated by maximizing the simulated log-likelihood function.

References

Barndorff-Nielsen, O.-E., Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, *63*, 253–280.

Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association* (pp. 177–181). Business and Economic Statistics Section.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, *69*(3), 542–547.

Bollerslev, T., Chou, R. Y., Kroner, K. F. (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, *52*, 5–59.

Christie, A. A. (1982). The stochastic behaviour of common stock variances. *Journal of Financial Economics*, *10*, 407–432.

Christoffersen, P., Jacobs, K., Mimouni, K. (2010). Volatility dynamics for the S&P 500: Evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies*, *23*, 3141–3189.

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

Duan, J. C., Fulop, A. (2009). Estimating the structural credit risk model when equity prices are contaminated by trading noises. *Journal of Econometrics*, *150*, 288–296.

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

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Eraker, B., Johannes, M., Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, *58*(3), 1269–3000.

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

Harvey, A. C., Ruiz, E., Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, *61*, 247–264.

Hull, J., White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, *42*, 281–300.

Jacquier, E., Polson, N. G., Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, *12*, 371–389.

Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series. *Journal of the American Statistical Association*, *82*, 1032–1041.

Kitagawa, G. (1993). A Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space models. In *Proceedings of the 2nd U.S.-Japan Joint Seminar on Statistical Time Series, Analysis* (pp. 110–131).

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

Malik, S., Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, *165*, 190–209.

Merton, R. C. (1976). Option pricing when underlying stock returns and discontinuous. *Journal of Financial Economics*, *3*, 125–144.

Omori, Y., Chib, S., Shephard, N., Nakajima, J. (2007). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics*, *140*, 425–449.

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

Pitt, M. K., Shephard, N. (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association*, *94*, 590–599.
Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, *23*, 470–472.
Sakamoto, Y., Ishiguro, M., Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht/Tokyo: D. Reidel Publishing Company.
Shephard, N., Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, *84*, 653–667.
Taylor, S. (1986). *Modelling financial time series*. Chichester: John Wiley.
Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, *127*, 165–178.