# Markov chain Monte Carlo methods for stochastic volatility models

Siddhartha Chib[a,*], Federico Nardari[b], Neil Shephard[c]

[a]*John M. Olin School of Business, Washington University, Campus Box 1133, 1 Brookings Dr., St. Louis, MO 63130, USA*
[b]*Department of Finance, Arizona State University, Tempe, AZ, USA*
[c]*Nuffield College, Oxford OX1 1NF, UK*

## Abstract

This paper is concerned with simulation-based inference in generalized models of stochastic volatility defined by heavy-tailed Student-$t$ distributions (with unknown degrees of freedom) and exogenous variables in the observation and volatility equations and a jump component in the observation equation. By building on the work of Kim, Shephard and Chib (Rev. Econom. Stud. 65 (1998) 361), we develop efficient Markov chain Monte Carlo algorithms for estimating these models. The paper also discusses how the likelihood function of these models can be computed by appropriate particle filter methods. Computation of the marginal likelihood by the method of Chib (J. Amer. Statist. Assoc. 90 (1995) 1313) is also considered. The methodology is extensively tested and validated on simulated data and then applied in detail to daily returns data on the S&P 500 index where several stochastic volatility models are formally compared under different priors on the parameters. © 2002 Elsevier Science B.V. All rights reserved.

*JEL classification:* C1; C4

*Keywords:* Bayes factor; Markov chain Monte Carlo; Marginal likelihood; Mixture models; Particle filters; Simulation-based inference; Stochastic volatility

## 1. Introduction

Stochastic volatility models have gradually emerged as a useful way of modeling time-varying volatility with significant potential for applications, especially in finance (Taylor, 1994; Shephard, 1996; Ghysels et al., 1996) for a discussion of the models and

---

the related literature). In this paper we consider two versions of SV models and extend existing work on more restricted models to develop efficient and fast Bayesian Markov chain Monte Carlo (MCMC) estimation algorithms. We develop a straightforward procedure for computing the marginal likelihood and Bayes factors for SV models. This procedure combines a simulation-based filter for estimating the likelihood ordinate with the method of Chib (1995) for estimating the posterior density of the parameters. The marginal likelihood procedure is tested on several problems and it is shown that the approach is capable of correctly choosing between various competing models. As a by-product of our work, we provide a method for filtering the current value of the unobserved volatility using contemporaneous data. This method is a simpler alternative to the reprojection method proposed by Gallant and Tauchen (1998).

The simplest formulation of the SV model, labelled $SV_0$, is given by

$$y_t = \exp(h_t/2)u_t,$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma\eta_t, \quad t \leqslant n,$$

where $y_t$ is the response variable, $h_t$ is the unobserved log-volatility of $y_t$ and the errors $u_t$ and $\eta_t$ are Gaussian white noise sequences. This model has been heavily analyzed in the literature. The first Bayesian analysis was provided by Jacquier et al. (1994) where the posterior distribution of the parameters was sampled by MCMC methods using the distributions $h_t|y, h_{(-t)}, \mu, \phi, \sigma(t \leqslant n)$; $\mu|y, h, \phi, \sigma$; $\phi|y, h, \mu, \sigma$ and $\sigma|y, h, \mu, \phi$, where $h = (h_1, \ldots, h_n)$ and $h_{(-t)}$ denotes all the elements of $h$ excluding $h_t$. Although this algorithm is conceptually simple it is not particularly efficient from a simulation perspective, as is shown by Kim et al. (1998) who develop an alternative, more efficient, MCMC algorithm for the above model. The efficiency gain in the Kim, Shephard and Chib algorithm arises from the joint sampling of $(\phi, \sigma)$ in one block marginalized over both $\{h_t\}$ and $\mu$, followed by the sampling of $\{h_t\}$ in one block conditioned on everything else in the model.

In this paper, we are concerned with two extensions of the basic SV model. The first of these models, which we label the $SVt$ model, is defined by Student-$t$ observation errors, level effect in the volatility and covariates in the volatility evolution. The model is given by the specification

$$y_t = x_t'\beta + w_t^\gamma \exp(h_t/2)u_t,$$

$$h_t = \mu + z_t'\alpha + \phi(h_{t-1} - \mu) + \sigma\eta_t, \quad t \leqslant n, \tag{1}$$

where $x_t$, $w_t$ and $z_t$ are covariates, $\gamma$ denotes the level effect and $u_t$ is distributed as a Student-$t$ random variable with mean zero, variance $v/(v-2)$ and $v > 2$ degrees of freedom. By exploiting the well-known fact that the Student-$t$ distribution can be expressed as a particular scale mixture of normals, we write $u_t = \lambda_t^{-1/2}\varepsilon_t$ where $\varepsilon_t$ is standard normal N(0,1) and $\lambda_t$ is i.i.d. Gamma($v/2, v/2$). In the SV context, Student-$t$ error-based models were used by Harvey et al. (1994), while Mahieu and Schotman (1998) discuss the use of a mixture distribution. Recently, Jacquier et al. (1999) have computed the posterior density of the parameters of a Student-$t$-based SV model. We assume that the degrees of parameter $v$ of the $t$-distribution is unknown and is estimated

from the data. Finally, $\{w_t\}$ is a non-negative process, such as the lag of the interest rate (see, for example, Andersen and Lund (1997) and the references contained within).

As motivation we should mention that the above model can be thought of as an Euler discretization of a Student-*t*-based Lévy process with additional stochastic volatility effects. The latter models are being actively studied in the continuous-time mathematical options and risk assessment literature. Leading references include Eberlein (2002), Prause (1999) and Eberlein and Prause (2002), while an introductory exposition is given in Barndorff-Nielsen and Shephard (2002, Chapter 2). The extension to allow for stochastic volatility effects is discussed in Eberlein and Prause (2002) and Eberlein et al. (2001).

The second model we discuss is similar to the *SVt* model except that it contains a jump component in the observation equation to allow for large, transient movements. This model, which we call the *SVt* plus jumps model (*SVJt*), is defined as

$$y_t = x_t'\beta + k_t q_t + w_t^\gamma \exp(h_t/2)u_t,$$

$$h_t = \mu + z_t'\alpha + \phi(h_{t-1} - \mu) + \sigma\eta_t, \quad t \leqslant n, \tag{2}$$

where $q_t$ is a Bernoulli random variable that takes the value one with unknown probability $\kappa$ and the value zero with probability $1 - \kappa$. The time-varying variable $k_t$ represents the size of the jump when a jump occurs and is assumed to a priori follow the distribution

$$\log(1 + k_t) \sim \mathrm{N}(-0.5\delta^2, \delta^2), \tag{3}$$

following Andersen et al. (2001). Taken together $k_t q_t$ can be viewed as a discretization of a finite activity Lévy process. Jump models are quite popular in continuous time models of financial asset pricing (for example, Merton, 1976; Ball and Torous, 1985; Bates, 1996; Duffie et al., 2000; Barndorff-Nielsen and Shephard, 2001a). Recent econometric work on jump-type SV models includes Barndorff-Nielsen and Shephard (2001b), Chernov et al. (2000) and Eraker et al. (2002). Our innovation is to provide an efficient and complete Bayesian tool-kit for parameter estimation, model checking, volatility estimation via filtering and model comparison.

The rest of the paper is organized as follows. In Section 2 we suggest an MCMC approach for fitting the *SVt* model and discuss methods for doing filtering and model choice. Section 3 presents a similar analysis for the *SVJt* model. Section 4 reports on an extensive Monte Carlo experiment in which the proposed estimation methods and model choice criterion are tested and validated. Then, the methodology is applied in detail to daily returns data on the S&P 500 index where several stochastic volatility models are formally compared under different priors on the parameters. Concluding remarks are presented in Section 5.

## 2. *SVt* model and Bayesian inference

A key feature of the *SVt* model (2) is that its likelihood function is not available easily. To see this difficulty, let $\psi = (\mu, \phi, \sigma, \alpha, \gamma, \beta, \nu)$ denote the parameters of the

model. Then, by the law of total probability, it follows that the density of the data $y = (y_1, \ldots, y_n)$ given $\psi$ can be expressed as

$$f(y|\psi) = \prod_{t=1}^{n} f(y_t|\mathcal{F}_{t-1}, \psi)$$

$$= \prod_{t=1}^{n} \int f(y_t|h_t, \psi) f(h_t|\mathcal{F}_{t-1}, \psi) \, dh_t,$$

where $\mathcal{F}_{t-1}$ denotes the history of the observation sequence up to time $(t-1)$. For model (2)

$$f(y_t|h_t, \psi) = \mathrm{St}(y_t|x_t'\beta, w_t^{2\gamma} \exp(h_t), v)$$

is the Student-$t$ density function with mean $x_t'\beta$, dispersion $w_t^{\gamma} \exp(h_t)$ and $v$ degrees of freedom. The source of the problem is that the density $f(h_t|\mathcal{F}_{t-1}, \psi)$ cannot be expressed in the closed form. Furthermore, $h_t$ appears in the variance structure of the Student-$t$ density and this precludes direct marginalization given any non-degenerate distribution of $h_t$.

Inspite of this apparent difficulty in the computation of the likelihood, it is possible to estimate the likelihood by taking recourse to a simulation method, called particle filtering, that for each $t$ delivers a sample of draws on $h_t$ from $f(h_t|\mathcal{F}_{t-1}, \psi)$. We discuss that method in the next sub-section because in the Bayesian context parameter estimation can be done without computing the likelihood.

## 2.1. Prior–posterior analysis

To complete the Bayesian model we formulate a prior distribution on the parameters $\psi$. We first assume that each parameter is a priori independent. Next, following Kim et al. (1998), we assume that for some choice of hyperparameters $\phi^{(1)}, \phi^{(2)} > 0.5$, the prior density of $\phi$ is a scaled beta distribution

$$\pi(\phi) = 0.5 \frac{\Gamma(\phi^{(1)} + \phi^{(2)})}{\Gamma(\phi^{(1)})\Gamma(\phi^{(2)})} \{0.5(1 + \phi)\}^{\phi^{(1)}-1} \{0.5(1 - \phi)\}^{\phi^{(2)}-1}, \tag{4}$$

with support on the stationary region $(-1, 1)$. Under this prior, the prior mean of $\phi$ is $2\phi^{(1)}/(\phi^{(1)} + \phi^{(2)}) - 1$. Third, we represent our information about $\gamma$ by a uniform distribution on the range $(0, 2)$ so as to cover the values that have been considered in the literature. Fourth, we let our prior on $\sigma$ for some hyperparameters $(\sigma_0, \Sigma_0)$ be given by a log-normal distribution. Thus, $\sigma \sim \mathrm{LN}(\sigma|\sigma_0, \Sigma_0)$. Fifth, for $\mu$, $\beta$ and $\alpha$ we assume independent normal priors $\mathrm{N}(\mu|\mu_0, M_0)$, $\mathrm{N}(\beta|\beta_0, B_0^{-1})$, and $\mathrm{N}(\alpha|\alpha_0, A_0^{-1})$, respectively, where the hyperparameters $(\delta_0, v_0, \mu_0, M_0, \beta_0, B_0, \alpha_0, A_0)$ are specified to reflect the available prior information. Finally, for the degrees of freedom $v$ we assume that the prior is uniform over the range $(2, 128)$. We mention that the precise distributional assumptions made above can be changed without affecting most of the details of the analysis.

Table 1
Parameters of seven-component Gaussian mixture to approximate the distribution of $\log \chi_1^2$

| $s_t$ | $q$ | $m_{s_t}$ | $v_{s_t}^2$ |
|---|---|---|---|
| 1 | 0.00730 | $-11.40039$ | 5.79596 |
| 2 | 0.10556 | $-5.24321$ | 2.61369 |
| 3 | 0.00002 | $-9.83726$ | 5.17950 |
| 4 | 0.04395 | 1.50746 | 0.16735 |
| 5 | 0.34001 | $-0.65098$ | 0.64009 |
| 6 | 0.24566 | 0.52478 | 0.34023 |
| 7 | 0.25750 | $-2.35859$ | 1.26261 |

Given this prior information the goal of the Bayesian analysis is to learn about the parameters $\psi$ from the augmented posterior distribution

$$\pi(\psi, h | y) \propto \pi(\psi) \prod_{t=1}^{n} f(y_t | h_t, \psi) f(h_t | h_{t-1}, \psi)$$

$$\propto \pi(\psi) \prod_{t=1}^{n} St(y_t | x_t'\beta, w_t^\gamma \exp(h_t), v) N(h_t | \mu + z_t'\alpha + \phi(h_{t-1} - \mu), \sigma^2).$$

Inference on the parameters is conducted by producing a sample $\{\psi^{(g)}, h^{(g)}\}$ from this density by a Markov chain Monte Carlo (MCMC) procedure (see, e.g., Chib and Greenberg (1996)). The draws $\{\psi^{(g)}\}$ thus produced are automatically from the posterior density of $\psi$ marginalized over $h$. We now discuss how the augmented posterior density can be sampled efficiently. Our methods are extensions of those in Kim et al. (1998). Wong (2000) recently has compared several existing and new algorithms for fitting the basic SV model and reported that the latter approach is both fast and reliable.

We start by noting that the *SVt* model can be converted into a conditionally Gaussian state space model. Let $u_t = \lambda_t^{-1/2}\varepsilon_t$, where $\varepsilon_t$ is a standard normal and $\lambda_t$ is i.i.d. Gamma$(v/2, v/2)$. Now, conditioned on $\psi$ and $\{\lambda_t\}$ the model may be reexpressed as

$$y_t^* = \gamma \log(w_t^2) + h_t + z_t, \tag{5}$$

$$h_t = \mu + z_t'\alpha + \phi(h_{t-1} - \mu) + \sigma\eta_t, \tag{6}$$

where

$$y_t^* = \log(y_t - x_t'\beta)^2 + \log \lambda_t$$

and $z_t = \log \varepsilon_t^2$. Following Kim et al. (1998) we approximate the distribution of $z_t$ by a seven-component mixture of normal densities with the representation

$$z_t | s_t \sim N(m_{s_t}, v_{s_t}^2),$$

$$Pr(s_t = i) = q_i, \quad i \leqslant 7, t \leqslant n, \tag{7}$$

where $s_t \in (1, 2, \ldots, 7)$ is an unobserved mixture component indicator with probability mass function $q = \{q_i\}$ and the parameters $\{q, m_{s_t}, v_{s_t}^2\}$ are as reported in Table 1.

In the MCMC context, it is helpful to use this approximation because the minor approximation error can be removed at the conclusion of the posterior sampling by a reweighting procedure, as discussed in Kim et al. (1998). This strategy of working with an efficient approximating model, and then reweighting the posterior sample ex-post, is a useful method of dealing with complicated models.

Another point to note is that if we let $s = \{s_1, \ldots, s_n\}$ and $\mathscr{F}_t^* = (y_1^*, \ldots, y_t^*)$, then the density of $y_t^*|s, \psi$ only depends upon a subset of parameters $\theta = (\mu, \phi, \sigma, \alpha, \gamma)$. In particular, the density of $y_t^*|s, \psi$ can be expressed as

$$f(y^*|s, \theta) = \prod_{t=1}^{n} f(y_t^*|\mathscr{F}_{t-1}^*, s, \theta), \tag{8}$$

where each one-step ahead density $f(y_t^*|\mathscr{F}_{t-1}^*, s, \theta)$ can be derived from the output of the Kalman filter recursions (adapted to the differing components, as indicated by the component vector $s$). These two points prove important in the formulation of our MCMC algorithm.

Our MCMC algorithm is based on the four vector blocks $\{\beta, [\theta, h], s \text{ and } [v, \lambda]\}$, where $\lambda = \{\lambda_1, \ldots, \lambda_n\}$ and the notation $[\theta, h]$ means that $\theta$ and $h$ are sampled in one block, conditioned on the remaining blocks. Note that the parameters $v$ and $\lambda$ are also sampled in one block conditioned on the other blocks. Extensive experimentation has shown that these steps are important for reducing the serial dependence in the MCMC output. The empirical sections below address this point in detail.

We summarize our algorithm as follows.

**Algorithm 1: MCMC algorithm for the *SVt* model.**

1. Initialize $\beta, s, \lambda$ and $v$
2. Sample $\theta$ and $h$ from $\theta, h|y, s, \lambda, \beta$ by drawing

(a) $\theta$ from $\theta|y^*, s, \beta$ and
(b) $h$ from $h|y^*, s, \beta, \theta$

3. Sample $\beta$ from $\beta|y, h, s, \lambda, \gamma$
4. Sample $s_t$ from $s_t|y_t^*, h_t, \gamma$
5. Sample $v$ and $\lambda$ from $v, \lambda|y, h, \psi$ by drawing

(a) $v$ from $v|y, h, \psi$ and
(b) $\lambda_t$ from $\lambda_t|y_t, h_t, \psi, v$.

6. Goto 2.

Steps 2a and 5a as mentioned above are important. We implement Step 2a by using the Metropolis–Hastings (M–H) algorithm (see, for example, Chib and Greenberg (1995) for a detailed account of the algorithm) by making a proposal draw $\theta^i$ from a tailored multivariate-$t$ density $f_T(\theta|m, V, \xi)$ with $\xi$ degrees of freedom, where $m$ is the value that maximizes the density $\log f(y^*|s, \theta)$ defined in (8) and $V$ is minus the inverse Hessian matrix of the objective function evaluated at $m$. This approach for

specifying the proposal density was introduced by Chib and Greenberg (1994). The proposal value generated from this density is then accepted or rejected according to the Metropolis–Hastings algorithm.

Step 2b is implemented using the simulation smoother algorithm described in Kim et al. (1998). Step 3 follows from the update of a regression model with heteroskedastic errors. Step 4 corresponds to the sampling of $s_t$ from a seven-point discrete distribution in which the prior weights $\Pr(s_t)$ are updated to $\Pr(s_t)N(y_t^*|\gamma \log(w_t^2) + h_t + m_{s_t}, v_{s_t}^2)$ and then normalized. Finally, Step 5a involves the sampling of the degrees of freedom by a Metropolis–Hastings step from the reduced conditional density of $v$ (given by the product of Student-$t$ densities in Eq. (2)) and Step 5b is a drawing from updated gamma distributions. Full details of this algorithm are given in Appendix A.

We note that while it may appear reasonable to sample $\gamma$ and the remaining parameters of $\theta$ as separate blocks, we have found that the resulting sampler is much less efficient due to the strong correlation between $\gamma$ and $\phi$. It is well known that strongly correlated components should be simulated as one block to minimize the serial dependence of the MCMC output. We sample parameters $(\mu, \alpha, \phi, \sigma)$ and $h$ as one block for a similar reason.

## 2.2. Filtering and likelihood estimation

We mentioned above that it is possible to estimate the likelihood function by simulation. In this section, we show how that can be done for a given value of the parameters $\psi$. The general approach relies on a particle filtering method that recursively delivers sequences of draws of $\{h_t\}$ from the filtered distributions

$$f(h_1|\mathscr{F}_1, \psi), \ldots, f(h_t|\mathscr{F}_t, \psi), \ldots, f(h_n|\mathscr{F}_n, \psi).$$

As part of this particle filtering procedure we obtain draws of $h_t$ from $f(h_t|\mathscr{F}_{t-1}, \psi)$. These draws from the one-step ahead predictive distribution of $h_t$ allow us to estimate the one-step ahead density of $y_t$

$$f(y_t|\mathscr{F}_{t-1}, \psi) = \int St(y_t|x_t'\beta, w_t^{2\gamma} \exp(h_t), v) f(h_t|\mathscr{F}_{t-1}, \psi) \, dh_t$$

by simple Monte Carlo averaging of $St(y_t|x_t'\beta, w_t^{2\gamma} \exp(h_t), v)$ over the draws of $h_t$ from $f(h_t|\mathscr{F}_{t-1}, \psi)$. We should note that the problem of filtering is important in its own right (independent of its use for calculating the likelihood function) because it may be used in the on-line forecasting of volatility and in the construction of model diagnostics.

The problem of filtering is solved by utilizing the *particle filter* method (see, for example, Gordon et al., 1993; Kitagawa, 1996; Berzuini et al., 1997; Isard and Blake, 1996; Pitt and Shephard, 1999a; and the booklength survey by Doucet et al., 2001). Kim et al. (1998) discuss a simple particle filter for the SV model but their algorithm has to be modified before it can be applied to the *SVt* model.

It may be mentioned that an alternative approach would be to use an importance sampling method as discussed by Danielsson (1994) and Sandmann and Koopman

(1998). Although this method has some theoretical advantages over particle filtering, it is not easy to implement in the SV context. Shephard (2000) presents evidence that typical importance samplers for the basic SV model may not possess a variance (and so would not obey a standard central limit theorem). Further experiments along these lines suggest that this problem becomes more severe when we deal with more involved SV models like the *SVt* model.

From Bayes theorem,

$$f(h_t | \mathcal{F}_t, \psi) \propto St(y_t | x_t'\beta, w_t^{2\gamma} \exp(h_t), v) f(h_t | \mathcal{F}_{t-1}, \psi), \tag{9}$$

where

$$f(h_t | \mathcal{F}_{t-1}, \psi) = \int \mathrm{N}\{h_t | \mu + \phi(h_{t-1} - \mu) + z_{t-1}'\alpha, \sigma^2\} f(h_{t-1} | \mathcal{F}_{t-1}, \psi) \, \mathrm{d}h_{t-1}.$$

Let us suppose that we have a sample $h_{t-1}^1, \ldots, h_{t-1}^M \sim f(h_{t-1} | \mathcal{F}_{t-1}, \psi)$, then it follows that $f(h_t | \mathcal{F}_{t-1}, \psi)$ can be approximated as

$$f(h_t | \mathcal{F}_{t-1}, \psi) \simeq \frac{1}{M} \sum_{j=1}^{M} \mathrm{N}\{h_t | \mu + \phi(h_{t-1}^j - \mu) + z_{t-1}'\alpha, \sigma^2\},$$

whence, approximately,

$$f(h_t | \mathcal{F}_t, \psi) \dot{\propto} St(y_t | x_t'\beta, w_t^{2\gamma} \exp(h_t), v) \frac{1}{M} \sum_{j=1}^{M} \mathrm{N}\{h_t | \mu + \phi(h_{t-1}^j - \mu) + z_{t-1}'\alpha, \sigma^2\}.$$

To sample $h_t$ from the latter density we work with the *auxiliary particle filter* introduced in Pitt and Shephard (1999a). This filter requires a first stage in which proposal values $h_t^{*1}, \ldots, h_t^{*R}$ are created. These values are then reweighted to produce draws $\{h_t^1, \ldots, h_t^M\}$ that correspond to draws from the target distribution. Typically, we take $R$ to be five or ten times larger than $M$. In the examples, we set the latter to be equal to 20,000. We now summarize the steps involved for the filter in period $t$.

**Algorithm 2: Auxiliary particle filter for *SVt* model.**

1. Given $\{h_{t-1}^1, \ldots, h_{t-1}^M\}$ from $(h_{t-1} | \mathcal{F}_{t-1}, \psi)$ calculate

$$\hat{h}_t^{*j} = \mu + \phi(h_{t-1}^j - \mu) + z_{t-1}'\alpha,$$

$$w_j = St\{y_t | x_t'\beta, w_t^{2\gamma} \exp(\hat{h}_t^{*j}), v\}, \quad j = 1, \ldots, M$$

   and sample $R$ times the integers $1, 2, \ldots, M$ with probability proportional to $\{w_j\}$. Let the sampled indexes be $k_1, \ldots, k_R$ and associate these with $\hat{h}_t^{*k_1}, \ldots, \hat{h}_t^{*k_R}$.

2. For each value of $k_j$ from Step 1 simulate

$$h_t^{*j} \sim \mathrm{N}(\mu + \phi(h_{t-1}^{k_j} - \mu) + z_{t-1}'\alpha, \sigma^2), \quad j = 1, \ldots, R$$

3. Resample $\{h_t^{*1}, \ldots, h_t^{*R}\}$ $M$ times with probabilities proportional to

$$\frac{St\{y_t | x_t'\beta, w_t^{2\gamma} \exp(h_t^{*j}), v\}}{St\{y_t | x_t'\beta, w_t^{2\gamma} \exp(\hat{h}_t^{*k_j}), v\}}, \quad j = 1, \ldots, R,$$

to produce the filtered sample $\{h_t^1, \ldots, h_t^M\}$ from $(h_t | \mathscr{F}_t, \psi)$.

Once we have a sample from $(h_t | \mathscr{F}_t, \psi)$ it is possible to produce many econometrically interesting quantities. The mean of the sample draws provides an estimate of $E(h_t | \mathscr{F}_t, \psi)$ which is the minimum mean square estimator of the log-volatility. It may be noted that our estimate of the log-volatility provides an alternative approach to the Gallant and Tauchen (1998) reprojection algorithm. We can also estimate the volatility directly by working with the exponentiated sample. We can also estimate $\Pr(Y_t \leqslant y_t | \mathscr{F}_{t-1}, \psi)$ which can be used to perform diagnostic checking, as discussed in this context by Shephard (1994) and Kim et al. (1998) and in other contexts by Diebold et al. (1997) and Gerlach et al. (1999). Finally, the particle filtering steps can be used to estimate the likelihood ordinate as summarized by the following steps.

**Algorithm 3: Likelihood function of *SVt* model.**

1. Set $t = 1$, initialize $\psi$ and obtain a sample of $h_{t-1}^{(g)}$ $(g \leqslant M)$.
2. For each value of $h_{t-1}^{(g)}$ sample

$$h_t^{(g)} \sim N(\mu + \phi(h_{t-1}^{(g)} - \mu) + z_{t-1}'\alpha, \sigma^2).$$

3. Estimate the one-step ahead density as

$$\hat{f}(y_t | \mathscr{F}_{t-1}, \psi) = \frac{1}{M} \sum_{g=1}^{M} St\{y_t | x_t'\beta, w_t^{2\gamma} \exp(h_t^{(g)}), v\}.$$

4. Apply the filtering procedure in Algorithm 2 to obtain $h_t^1, \ldots, h_t^M$ from $h_t | \mathscr{F}_t, \psi$.
5. Increment $t$ to $t + 1$ and goto Step 2.
6. Return the log likelihood ordinate

$$\log f(y | \psi) = \sum_{t=1}^{n} \log \hat{f}(y_t | \mathscr{F}_{t-1}, \psi).$$

### 2.3. Marginal likelihood

A central goal of any fitting exercise is the comparison of alternative models that may be conceived for the data at hand. In our context, it is of some importance to judge if the modeling assumptions, such as the assumption of Student-$t$ distributions, level effects or other features are in fact supported by the data. In the Bayesian context, such model comparisons may be done by computing the model marginal likelihood for each model that is proposed and fit to the data. The marginal likelihood is defined as the integral of the likelihood function with respect to the prior density and this quantity

can be computed by the approach of Chib (1995) which requires little more than the likelihood and posterior ordinates, both estimated at a single high density point $\psi^*$.

Let $\psi$ denote the parameters of a given generalized stochastic volatility model with likelihood function $f(y|\psi)$ and prior density $\pi(\psi)$, where the likelihood function is computed using the particle filtering algorithm given above. Then, the Chib method exploits the fact that the marginal likelihood can be written as

$$m(y) = \frac{f(y|\psi)\pi(\psi)}{\pi(\psi|y)},$$

where each term on the right-hand side appears with its normalizing constant. This expression, which is called the basic marginal likelihood identity, can be evaluated at any appropriately selected high density point $\psi^*$ (say). If $\pi(\psi^*|y)$ is an estimate of the posterior ordinate at $\psi^*$, then the marginal likelihood on the log scale can be estimated as

$$\ln m(y) = \ln f(y|\psi^*) + \ln \pi(\psi^*) - \ln \hat{\pi}(\psi^*|y), \tag{10}$$

where $\ln f(y|\psi^*)$ is found by the particle filter method while $\ln \pi(\psi^*)$ is available directly.

To estimate the posterior ordinate, we decompose the posterior ordinate as

$$\pi(\psi^*|y) = \pi(v^*|y)\pi(\theta^*|y, v^*)\pi(\beta^*|y, v^*, \theta^*) \tag{11}$$

and estimate each ordinate in turn. First, to estimate $\pi(v^*|y)$ we apply the method of kernel smoothing (in one dimension) to the draws $v^{(g)}$ from the full MCMC run. Next, to estimate the reduced conditional ordinate $\pi(\theta^*|y, v^*)$ we adopt the method of Chib and Jeliazkov (2001). If we let $\alpha(\theta, \theta^*|y, s, v)$ denote the probability of move in the Metropolis–Hastings step given above and $q(\theta|y, s, v)$ the multivariate-$t$ proposal density, then it can be shown that

$$\pi(\theta^*|y, v^*) = \frac{\int \alpha(\theta, \theta^*|y, v^*, \beta, s)q(\theta, \theta^*|y, v^*, \beta, s)\pi(\theta, \beta, s|y)\,d\theta\,d\beta\,ds}{\int \alpha(\theta^*, \theta|y, v^*, \beta, s)q(\theta^*, \theta|y, v^*, \beta, s)\pi(\beta, s|y, v^*, \theta^*)\,d\theta\,d\beta\,ds}. \tag{12}$$

To estimate each of the integrals we can proceed as follows. For the numerator fix $v$ at $v^*$ and continue the MCMC simulations for $M$ cycles sampling all parameters and state variables except $v$. This is referred to as a "reduced run" because one parameter is fixed in the sampling. For the denominator fix $v$ at $v^*$ and additionally $\theta$ at $\theta^*$ and conduct a second reduced run sampling all parameters and state variables except $v$ and $\theta$. In this reduced run, we also draw

$$\theta^{(j)} \sim q(\theta|y, \beta^{(j)}, s^{(j)}), \quad j \leqslant M$$

from the multivariate-$t$ proposal density. Given these draws, the reduced ordinate is estimated as

$$\hat{\pi}(\theta^*|y, v^*) = \frac{M^{-1}\sum_{g=1}^{M} \alpha(\theta^{(g)}, \theta^*|y, v^*, \beta^{(g)}, s^{(g)})q(\theta^{(g)}, \theta^*|y, v^*, \beta^{(g)}, s^{(g)})}{M^{-1}\sum_{j=1}^{M} \alpha(\theta^*, \theta^{(j)}|y, v^*, \beta^{(j)}, s^{(j)})}. \tag{13}$$

Table 2
Jeffreys' Bayes factor scale with $B_{jk}$ denoting the Bayes factor for model $j$ versus model $k$

| $B_{jk}$ | Evidence against model $k$ |
| --- | --- |
| 1–3.2 | Not worth more than a bare mention |
| 3.2–10 | Substantial |
| 10–100 | Strong |
| > 100 | Decisive |

Finally, to estimate $\pi(\beta^*|y,\theta^*,v^*)$, kernel smoothing is applied to the $\{\beta\}$ draws from the reduced run that was used to produce the draws in the denominator of (13). The sum of the log of these conditional posterior ordinates provides an estimate of $\log\pi(\psi^*|y)$.

Given the marginal likelihood estimate for each competing model, we can find the Bayes factor for pairs of competing models. The strength of evidence in favor of model $j$ versus model $k$ is evaluated according to the Bayes factor scale in Table 2, first proposed by Jeffreys (1961). Bayes factors are the appropriate quantities for comparing models when the prior odds on the models are one, which we have assumed. Under the latter assumption the Bayes factor and the posterior odds are equal.

## 3. *SVt* plus jumps model

We now turn to an analysis of the *SVJt* model given in (2) and (3). We show how the methods described in the previous section can be adapted for this setting. Recall that the model is specified as

$$y_t = x_t'\beta + k_t q_t + w_t^\gamma \exp(h_t/2)u_t,$$

$$h_t = \mu + z_t'\alpha + \phi(h_{t-1} - \mu) + \sigma\eta_t, \quad t \leqslant n, \tag{14}$$

where $q_t$ is a Bernoulli random variable that takes the value 1 with probability $\kappa$ and the value 0 with probability $1 - \kappa$ and the jump size $k_t$ is distributed as $\log(1+k_t) \sim$ N$(-0.5\delta^2, \delta^2)$.

To deal with this model, it is necessary to think about suitable prior distributions on the probability $\kappa$ and the size of the jump $\delta$. In practice, at least when $y_t$ represents a return on a financial asset measured daily (say), it may be expected that a jump will occur once every few months (say 100 days) in which case the implied value of $\kappa$ is about $10^{-2}$. This belief about $\kappa$ can be built into the prior distribution. A similar type of reasoning is required to fix the likely values of $\delta$. In the analysis below we shall assume that our prior information about these two parameters is embodied by a Beta$(\kappa|u_0, n_0)$ density and that on $\delta$ by a log-normal LN$(\delta|\delta_0, \Delta_0)$ density where the hyperparameters $(u_0, n_0, \delta_0, \Delta_0)$ are specified depending on the context. For the remaining parameters in the model, we can assume that our prior information is represented by the same distributional forms given in the previous section.

Another point to note is that we reparameterize $k_t$ by letting

$$\psi_t \equiv \log(1 + k_t)$$

in which case the term $k_t q_t$ in the observation equation of the model becomes $(\exp(\psi_t) - 1)q_t$. In the situations, where $k_t$ is small (as in financial applications where return is measured in decimals, not percentages), $\exp(\psi_t)$ may be approximated quite closely by $1 + \psi_t$, implying that the $k_t q_t$ term can be expressed as $\psi_t q_t$. This reparameterization proves quite important in our MCMC analysis of this model because when we sample $\delta$ we can do this marginalized over $\psi_t$. This is possible because the density of $y_t$ marginalized over $\psi_t$ is

$$y_t | h_t, q_t, \lambda_t, \psi, \delta \sim \mathrm{N}(x_t'\beta - 0.5\delta^2 q_t, \delta^2 q_t^2 + \sigma_t^2),$$

where

$$\sigma_t^2 = w_t^{2\gamma} \exp(h_t) \lambda_t^{-1}.$$

### 3.1. Prior–Posterior analysis

With the prior information and the reparameterization on hand, the goal of the analysis is to summarize the augmented posterior density $\pi(h, \{q_t\}, \{\psi_t\}, \psi, \kappa, \delta | y)$ which, up to a constant of proportionality, is given by the product of the prior on the parameters $\pi(\psi)\pi(\kappa)\pi(\delta)$ and the function

$$\prod_{t=1}^{n} St(y_t | x_t'\beta + k_t q_t, w_t^{2\gamma} \exp(h_t), v) \mathrm{N}(h_t | \mu + z_t'\alpha + \phi(h_{t-1} - \mu), \sigma)$$

$$Ber(q_t | \kappa) \mathrm{N}(\psi_t | -0.5\delta^2, \delta^2).$$

To sample this rather high-dimensional distribution, it is necessary to set up the MCMC sampler in an efficient way. The general idea is to again use an approximating model which conditioned on $(\beta, \{\lambda_t\}, \{\psi_t\}, q_t)$ allows us to write the jump model as

$$y_t^* = \gamma \log(w_t^2) + h_t + z_t, \tag{15}$$

$$h_t = \mu + z_t'\alpha + \phi(h_{t-1} - \mu) + \sigma\eta_t, \tag{16}$$

where now

$$y_t^* = \log(y_t - x_t'\beta - (\exp(\psi_t) - 1)q_t)^2 + \log\lambda_t$$

and the distribution of $z_t = \log\varepsilon_t^2$ is approximated as before by a seven-component mixture of normals. This means that we can sample $(\{s_t\}, [\theta, \{h_t\}])$ where $\theta = (\mu, \phi, \sigma, \gamma, \alpha)$ in exactly the same manner as in the non-jump model (with the new definition of $y_t^*$). In the remaining steps we sample $\beta$ in one block, $[v, \{q_t\}, \{\lambda_t\}]$ in another block and finally $(\delta, \{\psi_t\})$ and $\kappa$ in two separate blocks. Formally, our MCMC strategy is based around the six blocks $(\beta, \{s_t\}, [\theta, \{h_t\}], [v, \{q_t\}, \{\lambda_t\}], [\delta, \{\psi_t\}]$ and $\kappa)$. This blocking strategy is vital for producing a well-behaved sampler. We now give a general description of the entire algorithm.

**Algorithm 4: MCMC algorithm for the *SVJt* model.**

1. Initialize $v, \{q_t\}, \{\lambda_t\}, \{\psi_t\}, \gamma$ and $\{h_t\}$.
2. Sample $\beta$ from $\beta | y, \{h_t\}, v, \{q_t\}, \{\lambda_t\}, \gamma$.
3. Sample $\{s_t\}$ independently from $s_t | y_t^*, h_t, \gamma$.
4. Sample $\theta$ and $\{h_t\}$ from $\theta, \{h_t\} | y, \{s_t\}, \{q_t\}, \{\lambda_t\}, \beta, \{\psi_t\}$ by drawing

(a) $\theta$ from $\theta | y^*, \{s_t\}$ and
(b) $h_t$ from $\{h_t\} | y^*, \{s_t\}, \theta$.

5. Sample $v, \{q_t\}, \{\lambda_t\} | y, \{h_t\}, \gamma, \beta, \{\psi_t\}$ by drawing

(a) $v$ from $v | y, \{h_t\}, \gamma, \beta, \{q_t\}, \{\psi_t\}$,
(b) $q_t$ from $\{q_t\} | y, \{h_t\}, \gamma, \beta, v, \kappa$ and
(c) $\lambda_t$ from $\{\lambda_t\} | y, \{h_t\}, \gamma, \beta, \{q_t\}, \{\psi_t\}, v$.

6. Sample $\delta, \{\psi_t\} | y, \{h_t\}, \gamma, \beta, \{q_t\}, \{\lambda_t\}, v$ by drawing

(a) $\delta$ from $\delta | y, \{h_t\}, \gamma, \beta, \{\lambda_t\}, v, \{q_t\}$
(b) $\psi_t$ from $\{\psi_t\} | y, \{h_t\}, \gamma, \beta, \{\lambda_t\}, v, \{q_t\}, \delta$.

7. Sample $\kappa | \{q_t\}$.
8. Goto 2.

Full details are given in Appendix A.


### 3.2. Filtering and likelihood estimation

We now discuss how the filtering and likelihood/marginal likelihood computations can be extended to the jump model beginning with the computations for the filtering of $\{h_t\}$. It turns out that the computations are similar to those for the non-jump model except for the addition of a step in which the $\{\psi_t\}$ are generated from their prior normal distribution. An important point is that since $q_t$ is a Bernoulli random variable we marginalize it out of the observation equation, conditioned on $(h_t, \psi_t, \psi)$, and work with the density

$$f(y_t | h_t, \psi_t, \psi) = \sum_{q=0}^{1} \kappa^q (1 - \kappa)^{1-q} \, St(y_t | x_t' \beta + (\exp(\psi_t) - 1)q, w_t^{2\gamma} \exp(h_t), v) \quad (17)$$

in the filtering step.

**Algorithm 5: Auxiliary particle filter for *SVJt* model.**

1. Given $\{h_{t-1}^1, \ldots, h_{t-1}^M\}$ from $(h_{t-1} | \mathcal{F}_{t-1}, \psi)$ set $\hat{\psi}_t^{*j} = 0$, calculate

$$\hat{h}_t^{*j} = \mu + \phi(h_{t-1}^j - \mu) + z_{t-1}' \alpha$$

and

$$w_j = St(y_t|x_t'\beta, w_t^{2\gamma}\exp(\hat{h}_t^{*j}), v), \quad j = 1, \ldots, M$$

and sample $R$ times the integers $1, 2, \ldots, M$ with probability proportional to $\{w_j\}$. Let the sampled indexes be $k_1, \ldots, k_R$ and associate these with

$$\{\hat{h}_t^{*k_1}, \hat{\psi}_t^{*k_1}\}, \ldots, \{\hat{h}_t^{*k_R}, \hat{\psi}_t^{*k_R}\}.$$

2. For each $k_j$ from Step 1 simulate

$$h_t^{*j} \sim N(\mu + \phi(h_{t-1}^{k_j} - \mu) + z_{t-1}'\alpha, \sigma^2),$$
$$\psi_t^{*j} \sim N(-0.5\delta^2, \delta^2), \quad j \leqslant R.$$

3. Resample $(h_t^{*1}, \psi_t^{*1}), \ldots, (h_t^{*R}, \psi_t^{*R})$ $M$ times with probabilities proportional to

$$\frac{\sum_{q=0}^1 \kappa^q (1-\kappa)^{1-q} St(y_t|x_t'\beta + (\exp(\psi_t^{*j}) - 1)q, w_t^{2\gamma}\exp(h_t^{*j}), v)}{St(y_t|x_t'\beta, w_t^{2\gamma}\exp(\hat{h}_t^{*k_j}), v)}, \quad j \leqslant R.$$

Discard the $\{\psi_t^j\}$ to leave $\{h_t^1, \ldots, h_t^M\}$ from $(h_t|\mathscr{F}_t, \psi)$.

Given the above procedure for filtering $\{h_t\}$ it becomes possible to estimate the likelihood ordinate by the decomposition $f(y|\psi) = \prod_{t=1}^n f(y_t|\mathscr{F}_{t-1}, \psi)$ where now $f(y_t|\mathscr{F}_{t-1}, \psi)$ is the integral of the density in (17) with respect to the distribution of $(h_t, \psi_t)$ given $(\mathscr{F}_{t-1}, \psi)$. This integration is achieved by a simple Monte Carlo average over the draws based on the particle filter. We summarize the steps as follows.

**Algorithm 6: Likelihood function of the *SVJt* model.**

1. Set $t = 1$ and obtain a sample of $h_{t-1}^{(g)}$ $(g \leqslant M)$.
2. For each $h_{t-1}^{(g)}$ sample

$$h_t^{(g)} \sim N(\mu + \phi(h_{t-1}^{(g)} - \mu^*) + z_{t-1}'\alpha, \sigma^2)$$

$$\psi_t^{(g)} \sim N(-0.5\delta^2, \delta^2)$$

3. Estimate the one-step ahead density as

$$\hat{f}(y_t|\mathscr{F}_{t-1}, \psi) = \frac{1}{M} \sum_{g=1}^M \sum_{q=0}^1 (\kappa^q(1-\kappa)^{1-q} St\, y_t|x_t'\beta$$

$$+ (\exp(\psi_t^{(g)}) - 1)q, w_t^{2\gamma}\exp(h_t^{(g)}), v)$$

4. Apply the filtering procedure in Algorithm 5 to obtain $h_t^{(1)}, \ldots, h_t^{(M)}$ from $h_t|\mathscr{M}, \mathscr{F}_t, \psi$.
5. Increment $t$ to $t + 1$ and goto Step 2.

6. Return the log likelihood ordinate

$$\log f(y|\psi) = \sum_{t=1}^{n} \log \hat{f}(y_t | \mathscr{F}_{t-1}, \psi).$$

### 3.3. Marginal likelihood calculation

As we have done for the *SVt* model, we discuss how the marginal likelihood of the jump model can be computed by the Chib method. Let $\psi^* \equiv (v, \theta^*, \delta^*, \kappa^*, \beta^*)$ denote a particular high density point, where $\theta = (\mu, \phi, \sigma, \gamma, \alpha)$, then the marginal likelihood is computed by the expression

$$\log m(y) = \log \hat{f}(y|\psi^*) + \log \pi(\psi^*) - \log \hat{\pi}(\psi^*|y), \tag{18}$$

where $\hat{\pi}(\psi^*|y)$ is an estimate of the posterior density at $\psi^*$ and $\hat{f}(y|\psi^*)$ is the estimate of the likelihood function computed using the particle filter.

To compute the posterior ordinate, we write

$$\pi(\psi^*) = \pi(v^*|y)\pi(\theta^*|y, v^*)\pi(\delta^*|y, v^*, \theta^*)\pi(\kappa^*|v^*, \theta^*, \delta^*)\pi(\beta^*|y, v^*, \theta^*, \delta^*, \kappa^*) \tag{19}$$

and utilize the method of kernel smoothing to the draws $v^{(g)}$ from the full MCMC run to estimate the univariate ordinate $\pi(v^*|y)$. To estimate $\pi(\theta^*|y, v^*)$ we note from Chib and Jeliazkov (2001) that the ordinate can be expressed as

$$\pi(\theta^*|y, v^*)$$
$$= \frac{\int \alpha(\theta, \theta^*|y^*, \{s_t\})q(\theta, \theta^*|y^*, \{s_t\})\,\mathrm{d}\pi(\theta, \{s_t\}, \{q_t\}, \{\lambda_t\}, \beta, \{\psi_t\}|y, v^*)}{\int \alpha(\theta^*, \theta|y^*, \{s_t\})\,\mathrm{d}q(\theta^*, \theta|y^*, \{s_t\})\,\mathrm{d}\pi(\{s_t\}, \{q_t\}, \{\lambda_t\}, \beta, \{\psi_t\}|y, v^*)} \tag{20}$$

where the M–H probability of move $\alpha(\theta, \theta^*|y^*, \{s_t\})$ and proposal density $q(\theta, \theta^*|y^*, \{s_t\})$ are defined in Appendix A. To estimate each of the integrals we can proceed as follows. For the numerator fix $v$ at $v^*$ and continue the MCMC simulations for $M$ cycles sampling all the state variables and parameters except $v$. The draws obtained from this run are used to average the product $\alpha(\theta, \theta^*|y^*, \{s_t\})q(\theta, \theta^*|y^*, \{s_t\})$. For the denominator, fix $v$ at $v^*$, $\theta$ at $\theta^*$ and conduct a second reduced run for $M$ cycles sampling all the state variables and parameters except $v$ and $\theta$. Given the sampled draw on $y^*, \{s_t\}$ in this reduced run, we also draw $\theta$ from the proposal density $q(\theta^*, \theta|y^*, \{s_t\})$. The draws from this run are used to average the function $\alpha(\theta^*, \theta|y^*, \{s_t\})$. The ratio of these averages is the estimate of $\pi(\theta^*|y, v^*)$.

Next, an estimate of $\pi(\delta^*|y, \theta^*, v^*)$ is done by applying kernel smoothing (in one dimension) to the draws from the reduced run used to estimate the denominator in the preceding paragraph. The ordinate $\pi(\kappa^*|y, v^*, \theta^*, v^*, \delta^*)$ is also estimated by kernel smoothing but this time to draws on $\kappa$ obtained from a third reduced run in which the parameters $(v, \theta, \delta)$ are fixed at the starred values. A final reduced run is conducted with $(v, \theta, \delta, \kappa)$ all fixed and the draws on $\beta$ from this run are used to estimate the last ordinate by kernel smoothing.

## 4. Examples

In this section, we present empirical results based on both simulated and real datasets. The purpose of these examples is to illustrate the efficacy of the algorithm along several dimensions. First, we measure the observed serial correlation in the sampled output. This is summarized by the inefficiency factors of the estimation of the posterior mean. Recall the *inefficiency factor* is defined as

$$\left[1 + 2 \sum_{k=1}^{\infty} \rho(k)\right],$$

where $\rho(k)$ is the autocorrelation at lag $k$ for the parameter of interest. In Geweke (1992), the quantity *numerical efficiency* is used which is the inverse of the inefficiency factor. In practice, the higher order terms in this infinite sum are downweighted according to a window-based estimator. The inefficiency factor is a useful quantity that may be interpreted as the ratio of the *numerical variance* of the posterior mean from the MCMC chain to the variance of the posterior mean from hypothetical independent draws. It serves to quantify the relative loss from using correlated draws, in comparison with hypothetical uncorrelated draws, for computing the posterior mean. Second, we document the ability of the suggested model comparison criterion in selecting the correct model. Third, we evaluate the accuracy of the parameter estimates in terms of how accurately they reproduce the true values. Finally, we conduct a robustness analysis of the proposed methodology to the choice of the prior distributions and/or the hyperparameters.

### 4.1. Example 1: A simulation study

In order to assess the performance of the proposed estimation procedure and model selection criterion, we first utilize simulated data. In the simulation design, datasets are generated from the following two models:

- Model $SVt$: The SV model with $t$-errors. Student-$t$ errors with eight degrees of freedom in the measurement equation, a constant and the lagged return of the response variable as covariates in the measurement equation, no covariates in the evolution equation and no jumps. This is the model in (1) with $x_t = (1, y_{t-1})$, $\beta = (a, b)$, $\gamma = 0$ and $\alpha = 0$.
- Model $SVJ$: The SV model with jumps. Gaussian errors in the measurement equation, same covariates as in $SVt$, and binomial jumps in the measurement equation. This is the model in (2) with $u_t \sim N(0, 1)$, $x_t = (1, y_{t-1})$, $\beta = (a, b)$, $\gamma = 0$ and $\alpha = 0$.

For each model we simulate 50 data series of 1500 and 3000 observations, respectively. In all cases, the data are simulated using the following values for the volatility parameters: $\mu = -10$, $\phi = 0.985$, $\sigma = 0.12$. When jumps are present, the jumps parameters are $\delta = 0.04$, $\lambda = 0.02$. All these values are meant to be representative of typical features of high frequency financial series, as described below in the application to stock market data.

Four models are fit to these datasets: $SV_0$, $SVt$, $SVJ$ and $SVJt$, where the $SV_0$ and $SVJt$ models are specified as follows:

- Model $SV_0$: The Basic SV model. Gaussian errors in the measurement equation, same covariates as in $SVt$, no jumps in either equation.
- Model $SVJt$: The SV model with $t$-errors and jumps. Same as $SVt$ but with jumps in the measurement equation.

Thus, for each true model and sample size four models are each fit 50 times. In the estimation, the following prior distributions are used:

$$\beta \sim N_2(0, 0.2I_2),$$

$$\mu \sim N(-8, 25),$$

$$\phi^* \sim Beta(20, 1.5),$$

$$\log(\sigma) \sim N(-2.49, 0.73),$$

$$v \sim U[2, 128],$$

$$\log(\delta) \sim N(-3.07, 0.149),$$

$$\kappa \sim Beta(2, 100),$$

where $\phi^* \equiv (\phi + 1)/2$. The normal priors on the transformed parameters $\log(\sigma)$ and $\log(\delta)$ imply log-normal prior distributions for the actual parameters with the following means and standard deviations: $\sigma : 0.2, 0.125$, $\delta : 0.05, 0.02$. The hyperparameters in the prior of $\kappa$ imply a mean of 0.0196 and a standard deviation of 0.0136. The marginal likelihood for each model and each simulated data set is based on $G = 6000$ full Gibbs iterations of which 5000 are collected. They are followed by reduced runs of 5000 iterations each. In calculating the likelihood ordinate, the particle filter algorithm is based on $M = 20,000$ and $R = 200,000$.

On a Pentium III 700 MHz laptop running Windows 2000, 5000 Gibbs loops in the 3000 observation data set require $< 20$ min for the $SV_0$ model, about 25 min for the $SVJ$ model and 40 min for the $SVJt$ model. These figures and the constant advances in computing power make the proposed simulation approach fully practical for even very large datasets.

Tables 3 and 4 contain summaries of the parameter estimates across simulated samples. Specifically, for each true model ($SVt$ and $SVJ$) the grand average of the posterior means and their standard deviation are reported for each parameter. Overall, the accuracy of the adopted MCMC scheme is remarkable for both $SVt$ and $SVJ$: all parameters are estimated very precisely and with negligible variations across simulated samples. Expectedly, the precision of the estimates increases and their dispersion decreases as the sample size grows. The largest discrepancy is observed for the $v$ parameter. Results are, nonetheless, fully satisfactory for the smaller samples as well. Given that sample sizes even larger than those used here are quite common in empirical work, we expect the proposed method to be very reliable in terms of estimation accuracy.

Table 3
Summary output for 50 replications from simulated and fitted $SVt$ models*

|  | Truth: $SVt$ | Prior | | Properties of posterior means | | | |
|  | Estimated: $SVt$ | | | $T = 1500$ | | $T = 3000$ | |
|  |  | Mean | SD | Mean | SD | Mean | SD |
| $a$ | 0.0005 | 0.00 | 0.20 | 0.0005 | 0.0002 | 0.0005 | 0.0001 |
| $b$ | 0.1500 | 0.00 | 0.20 | 0.1470 | 0.0244 | 0.1502 | 0.0165 |
| $\mu$ | $-10.00$ | $-8.00$ | 5.00 | $-10.00$ | 0.2304 | $-9.9727$ | 0.2720 |
| $\phi$ | 0.9850 | 0.86 | 0.10 | 0.9760 | 0.0117 | 0.9812 | 0.0055 |
| $\sigma$ | 0.1200 | 0.20 | 0.13 | 0.1446 | 0.0240 | 0.1299 | 0.0211 |
| $v$ | 8.00 | 65.00 | 36.37 | 9.6648 | 4.4184 | 8.8129 | 1.3155 |

*Reported are the sample mean and standard deviation of the posterior mean estimates across 50 samples. Based on 6000 Gibbs draws, discarding the first 1000.

Table 4
Summary output for 50 replications from the $SVJ$ model and the $SVJt$ model is fitted*

|  | Truth: $SVJ$ | Prior | | Properties of posterior means | | | |
|  | Estimated: $SVJt$ | | | $T = 1500$ | | $T = 3000$ | |
|  |  | Mean | SD | Mean | SD | Mean | SD |
| $a$ | 0.0005 | 0.00 | 0.40 | 0.0005 | 0.0002 | 0.0005 | 0.0001 |
| $b$ | 0.1000 | 0.00 | 0.40 | 0.0974 | 0.0197 | 0.0998 | 0.0164 |
| $\mu$ | $-10.00$ | $-8.00$ | 5.00 | $-10.09$ | 0.2051 | $-10.04$ | 0.1690 |
| $\phi$ | 0.9850 | 0.86 | 0.10 | 0.9781 | 0.0119 | 0.9839 | 0.0058 |
| $\sigma$ | 0.1200 | 0.20 | 0.13 | 0.1186 | 0.0252 | 0.1087 | 0.0125 |
| $\delta$ | 0.0400 | 0.05 | 0.02 | 0.0431 | 0.0053 | 0.0422 | 0.0046 |
| $\kappa$ | 0.0200 | 0.02 | 0.02 | 0.0192 | 0.0050 | 0.0191 | 0.0038 |
| $v$ | 64.00 | 65.00 | 36.37 | 67.212 | 10.501 | 68.810 | 12.684 |

*Reported are the mean and standard deviation of the posterior mean estimates across 50 samples. Based on 6000 Gibbs draws, discarding the first 1000.

Next, we assess the validity of the proposed model selection criterion. Tables 5 and 6 report the results of our model comparisons. When the data are simulated from $SVt$ (Table 6) and the sample size is 1500, the Bayes factor selects the correct model 98% of the times against $SV_0$, 90% of the times against $SVJ$ and 82% against $SVJt$. In most cases, the evidence is either substantial or decisive. With the larger sample size of 3000 observations, the performance of the model selection criterion improves further: $SVt$ comes out on top of both $SV_0$ and $SVJ$ in all instances and in 96% of the cases is preferred to $SVJt$. It is important to notice that the marginal likelihood criterion is able to support the more parsimonious true model.

Next consider the case where the data are simulated from the jump model with Gaussian errors. Our results are summarized in Table 6. In this case, the $SVJ$ model is always picked over the $SV_0$ model, and picked over the $SVt$ model at least 98% of the times. These findings are robust across sample sizes. The results are mixed, however, in the comparison between $SVJ$ and $SVJt$. The posterior odds favor $SVJ$ in about half of the samples (44% with $T = 1500$, 54% with $T = 3000$). In general, these two

Table 5
Frequency distribution (percentage) of Bayes factors across 50 simulated replications

|  | True model: $SV_t$ | | | | |
|---|---|---|---|---|---|
|  | 1–3.2 | 3.2–10 | 10–100 | > 100 | Total > 1 |
| ($T = 1500$) |  |  |  |  |  |
| $SVt/SV_0$ | 4 | 2 | 28 | 64 | 98 |
| $SVt/SVJ$ | 10 | 16 | 42 | 22 | 90 |
| $SVt/SVJt$ | 24 | 22 | 34 | 2 | 82 |
| ($T = 3000$) |  |  |  |  |  |
| $SVt/SV_0$ | 0 | 0 | 2 | 98 | 100 |
| $SVt/SVJ$ | 0 | 0 | 18 | 82 | 100 |
| $SVt/SVJt$ | 16 | 16 | 46 | 18 | 96 |

Table 6
Frequency distribution (percentage) of Bayes factors across 50 simulated replications

|  | True model: $SVJ$ | | | | |
|---|---|---|---|---|---|
|  | 1–3.2 | 3.2–10 | 10–100 | > 100 | Total > 1 |
| ($T = 1500$) |  |  |  |  |  |
| $SVJ/SV_0$ | 0 | 0 | 0 | 100 | 100 |
| $SVJ/SVt$ | 0 | 0 | 2 | 98 | 100 |
| $SVJ/SVJt$ | 14 | 10 | 6 | 14 | 44 |
| ($T = 3000$) |  |  |  |  |  |
| $SVJ/SV_0$ | 0 | 0 | 2 | 98 | 100 |
| $SVJ/SVt$ | 0 | 0 | 2 | 98 | 100 |
| $SVJ/SVJt$ | 12 | 6 | 18 | 18 | 54 |

models are relatively difficult to tell apart unless there are many jumps or the sample size is very large. Further discussion of this issue is provided below.

## 4.2. Example 2: Stock market data

In this section, we consider applications of the proposed method to a dataset that has been extensively analyzed in the finance literature. The data series for this study comes from the Center for Research on Security Prices (CRSP) files and consists of daily continuously compounded returns, $y_t$, in decimals, on the S&P 500 index (computed without considering dividends) from July 3, 1962 through August 26, 1997, for a total of 8849 observations. The model specification adopted here is given by

$$y_t = x_t'\beta + k_t q_t + \exp(h_t/2)u_t,$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma\eta_t,$$

where $x_t = (1, y_{t-1})$ and $\beta = (a, b)$. Within this setup, four models are fit to the data:

- Model $SV_0$: Basic SV Model with Gaussian errors in both equations and no jumps;
- Model $SVt$: SV Model with Student $t$-errors with unknown degrees of freedom parameter $v$ in the measurement equation and no jumps;
- Model $SVJ$: SV Model with Gaussian errors and jumps in the measurement equation;
- Model $SVJt$: SV Model with both Student $t$-errors with unknown degrees of freedom parameter $v$ and jumps in the measurement equation.

Notice that in all these models there is no level effect: the volatility does not depend on the magnitude of the observable variable and the $\gamma$ parameter is accordingly set to zero. Nardari (1999) focuses on applications to interest rates dynamics where the level effect is present for theoretical and empirical reasons.

### 4.2.1. Prior distributions

Before turning to the data analysis, we complete the model by specifying the parameters of the prior distributions given in Sections 2.1 and 3. For $\phi$ we choose the values $\phi^{(1)} = 20$ and $\phi^{(2)} = 1.5$ which implies that our prior guess of $\phi$ is 0.86. This reflects the high degree of persistence in volatility commonly found in high frequency financial series where it is quite common to find estimates of $\phi$ between 0.95 and 0.99 for weekly and/or daily data. For $\sigma$ we set a mean of 0.20 and a standard deviation of 0.125. These parameters also accommodate values typically reported in the empirical studies of daily return series.

In the jump model, we assume that the jump size $k_t$ has mean of zero and a standard deviation of $\sqrt{(\exp(\delta^2) - 1)}$. As the $\delta$ parameter is directly related to the variability of jump sizes, we choose the log-normal prior with mean 0.05 and standard deviation of 0.02, so that jumps are mostly expected to fall in the $\pm 10\%$ range. For daily equity index returns, this appears to be a fairly adequate characterization. For the jump intensity, $\kappa$, we let $u_0 = 2$ and $n_0 = 100$: these parameters represent an average jump probability of 1.97% per day, with a standard deviation of 1.36%. In other words, the prior assumes that jumps are expected to occur about 50 trading days apart. The magnitude of the standard deviation relatively to the mean allows, however, for significantly different jump intensities.

For the $\mu$ and $\beta$ parameters we assume independent normal priors with means and standard deviations that are reported in Tables 7 and 8. Finally, for the degrees of freedom $v$ we assume that the prior is uniform over the range $(2, 128)$. These are reasonable priors as they impose some structure but are not particularly informative. As such, they are not expected to affect the posterior analysis in any substantive way. Nonetheless, in order to assess the impact of the priors on the posterior inference, we run several experiments in Section 4.2.4 below with alternative prior distributions and/or hyperparameters.

### 4.2.2. Results

Tables 7 and 8 report posterior quantities and inefficiency factors for the four models described above. The posterior quantiles are computed from 5000 draws of the MCMC algorithm, collected after an initial burn-in period of 1000 iterations. In order to estimate the posterior ordinates in Eqs. (10) and (18) the full Gibbs run is followed by reduced

Table 7
Summary output for the S&P 500 data using the $SV_0$ and $SVt$ models*

|  | Prior | | $SV_0$ model Posterior | | | | $SVt$ model Posterior | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mean | SD | Mean | Lower | Upper | INEF | Mean | Lower | Upper | INEF |
| $a$ | 0.00 | 0.20 | 0.0004 | 0.0003 | 0.0006 | 1.60 | 0.0004 | 0.0003 | 0.0005 | 1.56 |
| $b$ | 0.00 | 0.20 | 0.1466 | 0.1251 | 0.1689 | 1.38 | 0.1381 | 0.1167 | 0.1592 | 1.71 |
| $\mu$ | −10.00 | 5.00 | −9.9478 | −10.1573 | −9.7427 | 1.62 | −10.0879 | −10.3453 | −9.8400 | 2.21 |
| $\phi$ | 0.86 | 0.10 | 0.9846 | 0.9787 | 0.9897 | 5.75 | 0.9903 | 0.9858 | 0.9942 | 5.39 |
| $\sigma$ | 0.20 | 0.13 | 0.1459 | 0.1253 | 0.1671 | 9.86 | 0.1105 | 0.0930 | 0.1304 | 9.99 |
| $v$ | 65.00 | 36.37 |  |  |  |  | 12.528 | 9.7057 | 16.599 | 14.78 |

*"Lower" and "Upper" denote the 2.5th percentile and the 97.5th percentile, respectively, and INEF denotes the inefficiency factor. Based on 6000 draws, discarding first 1000.

Table 8
Summary output for the S&P 500 data using the $SVJ$ and $SVJt$ models*

|  | Prior | | $SVJ$ model Posterior | | | | $SVJt$ model Posterior | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mean | SD | Mean | Lower | Upper | INEF | Mean | Lower | Upper | INEF |
| $a$ | 0.00 | 0.20 | 0.0004 | 0.0003 | 0.0006 | 1.55 | 0.0004 | 0.0003 | 0.0005 | 1.54 |
| $b$ | 0.00 | 0.20 | 0.1448 | 0.1238 | 0.1659 | 1.43 | 0.1391 | 0.1182 | 0.1599 | 1.69 |
| $\mu$ | −10.00 | 5.00 | −9.9603 | −10.1910 | −9.7212 | 1.72 | −10.0676 | −10.3272 | −9.7920 | 2.35 |
| $\phi$ | 0.86 | 0.10 | 0.9886 | 0.9839 | 0.9927 | 4.60 | 0.9906 | 0.9863 | 0.9943 | 6.23 |
| $\sigma$ | 0.20 | 0.13 | 0.1213 | 0.1045 | 0.1397 | 9.57 | 0.1087 | 0.0921 | 0.1267 | 11.63 |
| $\delta$ | 0.05 | 0.02 | 0.0393 | 0.0195 | 0.0722 | 22.49 | 0.0344 | 0.0182 | 0.0642 | 7.01 |
| $\kappa$ | 0.02 | 0.01 | 0.0037 | 0.0012 | 0.0087 | 16.77 | 0.0019 | 0.0004 | 0.0047 | 13.37 |
| $v$ | 65.00 | 36.37 |  |  |  |  | 15.3828 | 10.9116 | 23.1821 | 13.68 |

*"Lower" and "Upper" denote the 2.5th percentile and the 97.5th percentile, respectively, and INEF denotes the inefficiency factor. Based on 6000 draws, discarding the first 1000.

runs of 5000 draws each. The likelihood function which is in an input into the marginal likelihood computation is computed using the particle filter with parameters $M = 20,000$ and $R = 200,000$.

The magnitudes of the inefficiency factors clearly indicate that the simulation procedure implemented here is efficient for all models considered: the inefficiency factors are below ten for the Basic SV Models and below 15 for the $SVt$ and $SVJt$ models. Only the inefficiency measure for the $\delta$ parameter in the $SVJ$ model is above 20. It should be noticed that in their original contribution, Kim et al. (1998) report inefficiency factor between 10 and 15 for the Basic SV model.

A key point is that the low inefficiency factors are possible from our method because of Steps 2 and 5 of Algorithm 1 and Steps 4, 5 and 6 of Algorithm 4. If, instead, the latent variables are not integrated out the loss of efficiency is dramatic. In particular, inefficiency factors for $\delta$ and $\kappa$ reach values close to 100, while the inefficiency measure for $v$ not marginalized over $\{\lambda_t\}$ surpasses 50. The tables with the inefficiency factors

Table 9
Model selection criteria for the S&P 500 index return*

| Model | SVt | SVJ | SVJt |
|---|---|---|---|
| $SV_0$ | 10.75 | 5.18 | 9.76 |
| SVt | — | −5.57 | −0.99 |
| SVJt | — | — | 4.58 |

*The entries in the table are Log (base 10) of Bayes factors for row model against column model (see text for a definition of the four models).

for these suboptimal cases are not reported. We would like to emphasize that these low inefficiency factors are obtained both by the blocking schemes that we have used and the careful tailoring the proposal density in the M–H steps to the target densities. The numerical optimization used to find the location and spread of the proposal density appears to be crucial in this context, especially for the $\delta$ and $v$ parameters. Extensive experimentation shows that, without such careful implementation of the M–H algorithm, the inefficiency factors are much higher.

In summary, our results based on accurate tailoring and marginalization suggest that a few thousands sweeps of the Gibbs sampler should be adequate in generating representative samples from the posterior distributions of interest. In the next subsection further evidence will be provided in order to reinforce this conclusion.

Next, we turn our attention to the posterior distributions of the parameters and to formal model comparison using the marginal likelihood/Bayes factor criterion. The drift parameters, $a$ and $b$, in the mean equation are tightly estimated and virtually identical across models. They are consistent with the well-known autocorrelation patterns in daily (or weekly) S&P 500 returns.

The distributions of the volatility parameters, $(\mu, \phi, \sigma)$ are also concentrated around their means. The values of $\phi$ close to one confirm a strong daily volatility persistence, in accordance with typical estimates reported in the literature. Notice that such persistence increases slightly as the fat-tailed error distribution is introduced in the mean equation ($\phi = 0.9902$ and 0.9906 for the SVt and SVJt models, respectively). More importantly, allowing for Student-$t$ distributed errors greatly enhances the performance of the model. The strength of evidence in favor of a model is based on the Jeffreys' scale given in Section 2.3. The Bayes factor of SVt versus $SV_0$ in Table 9 is 14.22 on the log 10 scale and this indicates decisive evidence in favor of the SVt model. In the SVt model, the distribution of the degrees of freedom parameter $v$ is centered around 12.5 with a standard deviation of 1.8. These values confirm that the assumption of conditionally Gaussian errors is untenable for S&P daily data.

Interestingly, our estimate of $v$ is higher than those previously reported in studies using the S&P 500 series. For example, Sandmann and Koopman (1998) report $v = 7.6$ whereas Jacquier et al. (1999) compute a posterior mean around 11. Both articles, however, utilize data spanning a shorter time period (1980–1987). When we reestimate the SVt model over the same sample period as these studies the posterior mean of $v$ drops to 8.9. It is possible that extreme and rare events like the market crash of October 1987 affect inference about this parameter.

Consider now the SV model with Gaussian errors and jumps (*SVJ*). First, this model also provides a much more accurate description of the data than the Basic SV Model. The Bayes factor in Table 3 (*SVJ* versus $SV_0$) is 8.66 on the log 10 scale showing decisive evidence in favor of the *SVJ* model. Second, having established the superiority of this model, the posterior distributions of the jump parameters are of particular interest. This is especially true given the limited and/or fairly recent empirical evidence available in the literature. In general, from an option pricing and risk management perspective it is of relevance to learn about the probability that a jump occurs over a given time period (or jump intensity) and the magnitude of such jumps.

One general observation is that the estimates of the jump parameters with real data are not as precise as those found for other parameters, as can be seen by looking at the spread of the posterior distributions of $\delta$ and $\kappa$. The variability of the jump size is characterized by the $\delta$ coefficient which implies a standard deviation of about 4%. Consequently, under the *SVJ* specifications, daily jumps around 8% (positive and negative) are not uncommon. Regarding the jump intensity, the posterior distribution of $\kappa$ shows that the average probability of observing a jump is about 0.37% on a daily basis. This roughly implies that a jump may occur every 270 trading days in the sample considered here. The data seem to induce a substantial departure of the jump probability from its prior distribution.

Notice that the average jump intensity estimated here is significantly lower than that reported by Andersen et al. (2001): their estimates suggest that jumps occur every 50–60 days but those estimates are based on a data set that starts prior 1962. Andersen et al. (2001) also analyze for the period 1980–1996. Our refitted posterior mean over this period is 0.0178 which is close to their point estimate of 0.0192. In addition, we have validated the accuracy of our estimation procedure with the simulation study reported in Section 4.1.

We next compare the *SVt* and the *SVJ* models since both appear to be improvements over the Basic SV model. Both models capture the excess kurtosis in the data that is left unexplained by the standard Gaussian SV model. It is an empirically relevant question whether the more extreme realizations are due to the tail behavior of a non-Gaussian distribution or, instead, to the superimposition of a jump component on a Gaussian diffusion. The Bayes factor in Table 9 tends to favor the *SVt* model over *SVJ*. Thus, the excess kurtosis in the data seems to be better characterized by a process that allows for large innovations as opposed to a Gaussian model with a jump component. One further possibility is that both forces are at work within the same data generating process. A priori, it could well be that some big swings in S&P 500 returns need to be characterized as jumps as they are too large even for a Student-*t* error distribution with low degrees of freedom. To explore this possibility we compare the *SVt* model with the *SVJt* model, a specification that incorporates both fat-tails and jumps in the mean equation.

Fig. 1 shows the estimated posterior densities for the degrees of freedom parameter *v* using the two models *SVt* and *SVJt*. The figure shows that in the jump model, the density of *v* is shifted to the right, with the mean increasing from around 12 to 15. As expected, once jumps are permitted, a slightly less fat-tailed error distribution suffices
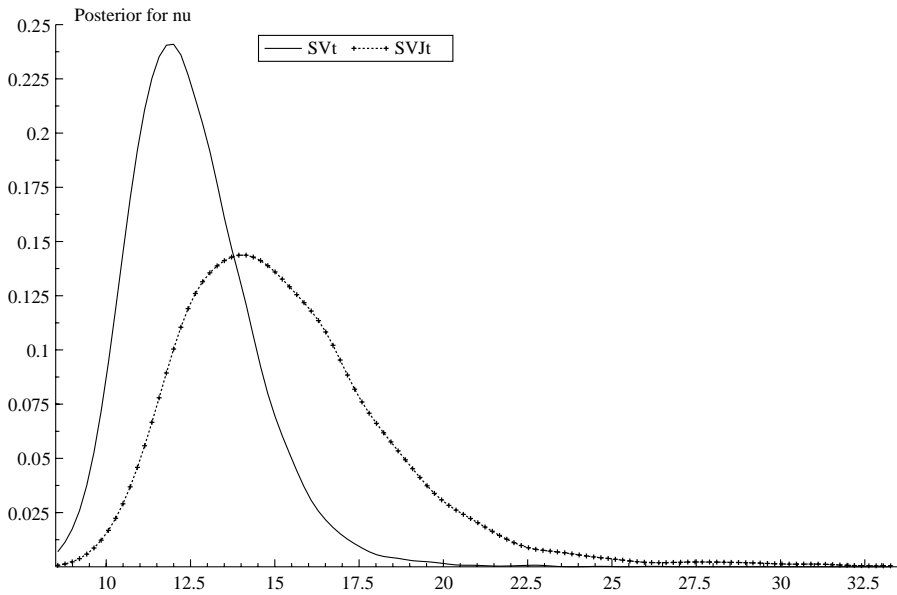
Fig. 1. Estimated posterior densities for ν for the SVt and SVJt models.

for capturing the kurtosis in the data. Further, the spread of the ν distribution is wider under the jump model, which suggests that this parameter is harder to estimate in the presence of jumps.

More formally, looking at the Bayes factors table, SVt is still the better model. Although the evidence in favor of the SVt model may not appear especially strong, it is reasonable to conclude that the more involved SVJt parameterization does not outperform SVt for the dataset at hand. The posterior distributions of the jump intensity, given in Fig. 2, are consistent with this conclusion. Notice that the posterior mean of κ in SVJt is almost half of what it is in SVJ (0.00195 versus 0.003369) whereas the standard deviations are much closer (0.00192 versus 0.00114).

Jumps are, in other words, much less frequent since many of the realizations seen as jumps in the SVJ model are, instead, just considered as tail realizations in the SVt model. Besides, the high dispersion of the κ posterior distribution is perhaps a symptom of overparameterization or just a consequence of the fact that jumps, if rarely observed, are difficult to estimate. For example, in this instance, the posterior mean of κ implies that on average a jump occurs every 500 days. Thus, even with the long data series analyzed here this means that there are < 20 observations from which we can learn about the jump process.

Further insights about the differences across models can be gained by analyzing the estimated latent volatility process under each model. The Gibbs sampler produces draws from the unobservable log-variance process. Exponentiating the draws and taking an average across iterations yields the smoothed estimate of the instantaneous volatility
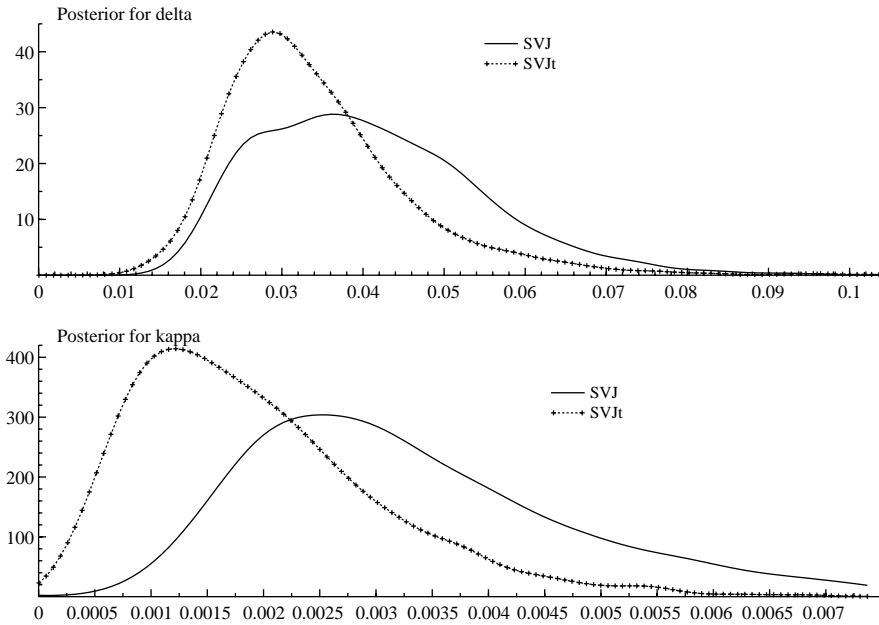
Fig. 2. Estimated posterior densities for $\delta$ and $\kappa$ for the $SVJ$ and $SVJt$ models.

at each point in time. Figs. 3 and 4 display the differences in these estimates for pairs of model. The volatilities are annualized before computing their differences. The annualized volatility estimates are computed using 252 trading days per year as

$$\widehat{\sigma}_t = \frac{1}{G} \sum_{g=1}^{G} \exp\left( 252\, \frac{h_t^{(g)}}{2} \right).$$

Fig. 3 shows the comparison between the $SV_0$ model and the rest. First, the differences are quite pronounced as they fluctuate between 2% and 5% with a few spikes close to 10%. The largest gap, over 20%, corresponds to the crash of October 1987. It is clear that the $SV_0$ model yields higher volatility estimates than the other models. These differences seem to widen in periods of market turbulence. The differences are greatest when the $SV_0$ volatilitiy estimates are compared with those from the $SVt$ and $SVJt$ models, and smallest when compared with the $SVJ$ model. A general interpretation is that big movements in returns in the basic Gaussian model are attributed almost exclusively to sharp changes in the volatility level whereas they are filtered differently, either as jumps ($SVJ$) or as tail realizations ($SVt$), or both ($SVJt$), by other models. As a consequence, the level and the variance of volatility are expected to be higher for $SV_0$ than for the other models. Indeed, this is the case as one can see quantitatively from the posterior estimates of $\mu$ (volatility level) and $\sigma$ (volatility of volatility) in Tables 7 and 8: the respective distributions are shifted towards lower values as one moves to models with fat tails and jumps.
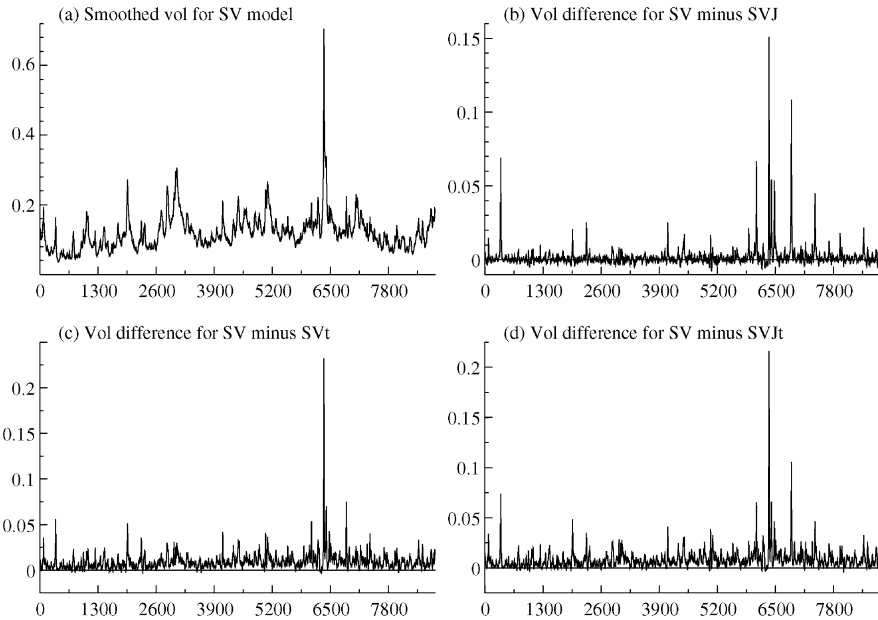
Fig. 3. Smoothed estimate of annualized standard deviation. The table reports the volatility estimates for the basic SV model (panel (a)) and the difference in estimated volatilities between the basic SV model and each of the extended specifications (panels (b)–(d)).

Next, Fig. 4 compares the smoothed volatility sequences from the three extended models. The estimates from *SVJ* are almost always higher than for the other two models with differences reaching 2–3% even in periods of relative market tranquility. Finally, notice that the volatility process extracted from the *SVJt* is similar to that estimated from the *SVJt* model. This provides additional support to the conclusion, obtained from the marginal likelihood comparison above, that the *SVt* and *SVJt* models are more or less equivalent in this application.

The differences in the estimated volatility processes suggests that the models under consideration should produce different option pricing and hedging implications.

### 4.2.3. Robustness check 1: Simulation size

The low inefficiency factors from the MCMC output suggest that a few thousand sweeps of our samplers should be adequate for the models and data set considered here. In this section, we provide additional evidence of this fact. The four models estimated above on the S&P 500 data are reestimated with a MCMC sample size of 50,000. The results are shown in Table 10. For ease of comparison, the table includes the results discussed above from the 5000 sample size simulation. We see that for most parameters the posterior mean and standard deviation are unchanged up to the second or third decimal place. In addition, the marginal likelihood and Bayes
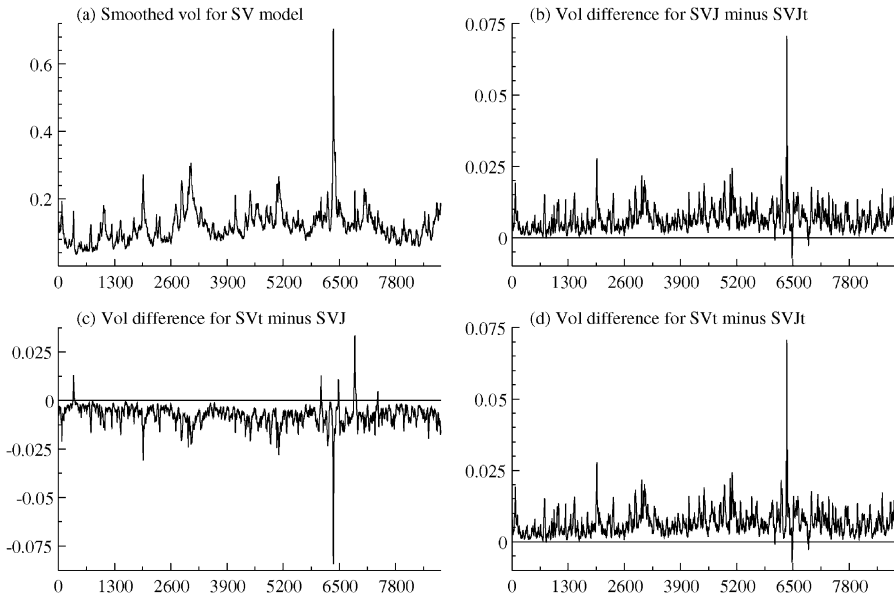
Fig. 4. Smoothed estimate of annualized standard deviation. The table reports the volatility estimates for the SV model (panel (a)) and the pairwise difference in estimated volatilities between the extended specifications (panels (b)–(d)).

factors estimates from the Chib method, given in the bottom row of Tables 10 and 11, respectively, do not change in any appreciable way as the simulation sample size is increased.

### 4.2.4. Robustness check 2: Prior sensitivity

In this section, we explore the implications of alternative prior distributions on the posterior estimates and model comparison results reported above. We conducted prior sensitivity checks on all parameters. In what follows we report the more interesting cases. In particular, we present the robustness of the findings to changes in the distributional form and in the hyperparameters of $\phi$, $v$, $\delta$ and $\kappa$. First, consider the prior density for $v$, which has been assumed to be uniform over the range $[2, 128]$ and the prior for $\phi$, originally specified as a Beta distribution. We reestimate the $SVt$ model with two different log-normal distributions on $v$ and a flat prior on $\phi$, as follows:

- Prior 2: $v \sim LN(8, 20)$, $\phi \sim U(-1, 1)$;
- Prior 3: $v \sim LN(64, 20)$,

where LN denotes the log-normal distribution. As for $v$, these two cases are seen as opposite extremes, pointing decisively towards a very fat-tailed error distribution or towards a Gaussian model, respectively. The flat prior on $\phi$ only imposes stationarity in the volatility process but is otherwise uninformative. The posterior summaries

Table 10
Posterior summaries obtained using $G$ draws[*]

| | $SV_0$ model | | $SVt$ model | | $SVJ$ model | | $SVJt$ model | |
|---|---|---|---|---|---|---|---|---|
| $G$ | 5000 | 50,000 | 5000 | 50,000 | 5000 | 50,000 | 5000 | 50,000 |
| $a$ | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| $b$ | 0.1466 | 0.1464 | 0.1381 | 0.1378 | 0.1448 | 0.1450 | 0.1391 | 0.1388 |
| | (0.0111) | (0.0110) | (0.0109) | (0.0107) | (0.0107) | (0.0109) | (0.010) | (0.0108) |
| $\mu$ | −9.9478 | −9.9495 | −10.0879 | −10.0934 | −9.9603 | −9.9641 | −10.0681 | −10.0728 |
| | (0.1056) | (0.1058) | (0.1321) | (0.1304) | (0.1194) | (0.1200) | (0.1326) | (0.1323) |
| $\phi$ | 0.9846 | 0.9845 | 0.9903 | 0.9903 | 0.9886 | 0.9886 | 0.9906 | 0.9906 |
| | (0.0028) | (0.0022) | (0.0022) | (0.0021) | (0.0022) | (0.0023) | (0.0020) | (0.0020) |
| $\sigma$ | 0.1459 | 0.1461 | 0.1105 | 0.1106 | 0.1213 | 0.1213 | 0.1087 | 0.1086 |
| | (0.0105) | (0.0105) | (0.0095) | (0.0093) | (0.0091) | (0.0094) | (0.0087) | (0.0089) |
| $v$ | | | 12.5281 | 12.3692 | | | 15.8059 | 15.4746 |
| | | | (1.8025) | (1.7858) | | | (3.2313) | (3.4512) |
| $\delta$ | | | | | 0.0393 | 0.0388 | 0.0355 | 0.0355 |
| | | | | | (0.0145) | (0.0140) | (0.0133) | (0.0135) |
| $\kappa$ | | | | | 0.0037 | 0.0038 | 0.0020 | 0.0019 |
| | | | | | (0.0019) | (0.0020) | (0.0011) | (0.0012) |
| Marg Lik −31,000 | 59.65 | 58.80 | 84.41 | 84.33 | 71.58 | 70.54 | 82.14 | 81.67 |

[*]Posterior mean and standard deviation (in parentheses) are reported. Marginal likelihood, minus 31,000, is reported on the natural log scale.

Table 11
Model selection criteria for the S&P 500 index return[*]

| | G = 5000 | | | | G = 50 000 | | |
|---|---|---|---|---|---|---|---|
| | $SVt$ | $SVJ$ | $SVJt$ | | $SVt$ | $SVJ$ | $SVJt$ |
| $SV_0$ | 10.75 | 5.18 | 9.76 | $SV_0$ | 11.95 | 5.10 | 9.93 |
| $SVt$ | — | −5.57 | −0.99 | $SVt$ | — | −6.86 | −2.02 |
| $SVJt$ | — | — | 4.58 | $SVJt$ | — | — | 4.83 |

[*]Results are based on different simulation sizes. $G$ denotes the number of Gibbs draws. The entries in the table are Log (base 10) of Bayes factors for row model against column model (see text for a definition of the four models).

corresponding to these different prior choices are reported in Table 12. A comparison with the results in Table 7 shows that Prior 2 yields a posterior distribution that is essentially unchanged. The flat prior on $\phi$ appears to have no effect on the posterior inference. The posterior mean of $v$ is still around 12.5 and the likelihood and posterior ordinates are close across priors. Only when the prior for $v$ is concentrated on large values of $v$ does the posterior distribution get shifted to the right, with the posterior mean increasing to about 20.

Table 12
Sensitivity to the prior*

| | SVt model | | SVJ model | | SVJt model | |
|---|---|---|---|---|---|---|
| | Prior 2 | Prior 3 | Prior 4 | Prior 5 | Prior 6 | Prior 7 |
| $a$ | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| $b$ | 0.1378 | 0.1408 | 0.1444 | 0.1447 | 0.1388 | 0.1425 |
| | (0.0106) | (0.0108) | (0.0109) | (0.0106) | (0.0107) | (0.0108) |
| $\mu$ | −10.0910 | −10.0317 | −9.9632 | −9.9654 | −10.0700 | −10.0100 |
| | (0.1211) | (0.1219) | (0.1173) | (0.1215) | (0.1323) | (0.1271) |
| $\phi$ | 0.9905 | 0.9885 | 0.9885 | 0.9886 | 0.9906 | 0.9895 |
| | (0.0021) | (0.0024) | (0.0022) | (0.0023) | (0.0021) | (0.0022) |
| $\sigma$ | 0.1101 | 0.1222 | 0.1217 | 0.1215 | 0.1087 | 0.1165 |
| | (0.0085) | (0.0100) | (0.0093) | (0.0093) | (0.0089) | (0.0095) |
| $v$ | 12.6145 | 21.7873 | | | 15.5117 | 39.2705 |
| | (1.7007) | (5.0180) | | | (3.0063) | (12.8780) |
| $\delta$ | | | 0.0383 | 0.0390 | 0.0346 | 0.0341 |
| | | | (0.0143) | (0.0142) | (0.0139) | (0.0140) |
| $\kappa$ | | | 0.0039 | 0.0038 | 0.0022 | 0.0033 |
| | | | (0.0020) | (0.0022) | (0.0014) | (0.0019) |
| Log Lik | 31109.36 | 31105.37 | 31098.68 | 31097.14 | 31113.15 | 31105.26 |
| Marg Lik | 31083.51 | 31076.15 | 31070.01 | 31070.99 | 31081.85 | 31075.37 |

*The characteristics of each prior are described in the main text. The posterior mean and standard deviation (in parentheses) are reported. The log likelihood ordinate is computed at the parameter posterior means and reported.

As for the effect of these priors on the marginal likelihood, what matters is its effect on the ranking of the models. On this ground, the ranking between the SVt and SVJ models is unchanged. Even in the case of Prior 3, the Bayes factor in favor of the SVt model is $> 100$. Notice that Prior 3 implies a Gaussian model without jumps: nonetheless, the data continue to indicate a preference for the SVt relative to the SVJ model.

Next, we turn our attention to the prior distribution for $\kappa$, the jump intensity parameter. Previously the prior mean was set at 0.02 implying a 2% average probability of a jump per day. With this prior, the SVJ model clearly outperforms the Basic SV model but it falls short against the SVt model. Table 12 also displays the results for the following priors on $\kappa$:

- Prior 4: $\kappa \sim$ Beta, mean $= 0.0385$, standard deviation $= 0.0264$;
- Prior 5: $\kappa \sim$ Beta, mean $= 0.0909$, standard deviation $= 0.0600$.

In both cases, the prior on $\delta$ is the same as in the previous section. Notice that Prior 4 predicts an average probability of a jump every 25–26 days, whereas Prior 5 implies an average probability of a jump every 11 days. The posterior estimates, however, still point towards a much lower jump frequency. In fact, the posterior moments are largely the same under the different prior specifications, as shown in Table 12. The posterior

mean of $\kappa$ is slightly below 0.4% indicating that jumps arrive, on average, at the rate of one per year. In addition, the marginal likelihoods are stable as one varies the prior on the jump intensity. The marginal likelihoods in Table 12 confirm that, for a variety of prior choices, the *SVt* model outperforms *SVJ* for the S&P 500 return series.

Finally, we want to assess the impact of changes in both the prior on $\kappa$ and the prior on $v$ in the *SVJt* model. We report results for two alternative sets of priors:

- Prior 6: $v \sim \mathrm{U}[2, 128]$, $\kappa \sim \mathrm{Beta}(2, 20)$;
- Prior 7: $v \sim \mathrm{LN}(64, 20)$, $\kappa \sim \mathrm{Beta}(2, 20)$.

The posterior summaries for *SVJt* corresponding to these prior choices are shown in Table 12. Once again, the prior of $\kappa$ does not materially affect the posterior or the model marginal likelihood. On the other hand, under Prior 7, where the prior for $v$ is concentrated on large values, the posterior mean of $v$ moves up to around 40 and its marginal posterior distribution becomes more dispersed. The jump intensity is almost twice as high. Both the likelihood ordinate and the marginal likelihood drop substantially which indicates that the *SVJt* model estimated with Prior 7 is not an especially adequate description of the data. Finally, the rankings of the models is not altered.

## 5. Concluding remarks

This paper has considered a class of generalized stochastic volatility models defined by heavy tails, a level effect on the volatility and covariate effects in the observation and evolution equations and a jump component in the observation equation. Two fast and efficient MCMC fitting algorithms have been developed for such models. The discussion has also considered the estimation of the volatility process and the comparison of alternative models via the marginal likelihood/Bayes factor criterion. The methodology is extensively tested and validated on simulated data and then applied in detail to real data where several models are fit and compared under different realistic priors on the parameters. Taken together the framework and results will be important for the practical analysis of high frequency data.

The analysis can be extended in a number of directions. First, one can consider generalized SV models in which the parameters are allowed to switch amongst a given number of states according to a hidden Markov process. The basic SV model under this assumption has been considered recently by So et al. (1998). The MCMC implementation follows from the procedures developed in Albert and Chib (1993) and Chib (1996). Second, one can fit continuous time analogues of the model discussed in this paper (Andersen and Lund, 1997; Gallant and Tauchen, 1998; Gallant et al., 1998). Such extensions can be handled in the MCMC context by amalgamating the approach of this paper with that of Elerian et al. (2001) or Eraker (2001). Another possible extension is to the generalized models of multivariate stochastic volatility of the type recently investigated in detail by Pitt and Shephard (1999b) and subsequently by Aguilar and West (2000). This extension is taken up by Chib et al. (1999).

## Appendix A. Simulation algorithm for *SVJt* model

In this section we provide full details of each step of the MCMC algorithm for the *SVJt* model. From this scheme one can easily obtain the algorithms for all the special cases (*SV*$_0$, *SVt* and *SVJ*) by omitting the appropriate steps and by deleting the variables that do not appear in those specifications. The simulation steps follow:

1. Initialize $v$, $\{\lambda_t\}$, $\{\psi_t\}$, $\{q_t\}$, $\gamma$ and $\{h_t\}$.
2. Sample $\beta$ from $\beta|y, \{h_t\}, \{\lambda_t\}, \gamma, v, \{\psi_t\}, \{q_t\}$, an $N(\hat{\beta}, B^{-1})$ distribution where

$$B = B_{10} + \sum_{t=1}^{n} \frac{x_t x_t'}{w_t^{2\gamma} \lambda_t^{-1} \exp(h_t)},$$

$$\hat{\beta} = B^{-1} \left( B_{10}\beta_{10} + \sum_{t=1}^{n} \frac{x_t(y_t - (\exp(\psi_t) - 1)q_t)}{w_t^{2\gamma} \lambda_t^{-1} \exp(h_t)} \right).$$

3. Compute

$$y_t^* = \log(y_t - x_t'\beta - (\exp(\psi_t) - 1)q_t)^2 + \log \lambda_t, \quad t = 1, \ldots, n$$

and sample $s_t$ from $s_t|y_t^*, h_t, \gamma$ Conditionally, the $\{s_t\}$ are independent with

$$\Pr(s_t|y_t^*, h_t, \gamma) \propto \Pr(s_t) N(y_t^*|\gamma \log(w_t^2) + h_t + \mu_{s_t}, v_{s_t}^2),$$

where $\mu_{s_t}$ and $v_{s_t}^2$ denote the mean and the variance of the mixture component at time $t$.

4. Sample $\theta$ and $h$ from $\theta, \{h_t\}|y, \{s_t\}, \{\psi_t\}, \{q_t\}, \{\lambda_t\}, \beta$ by drawing

(a) First, we note that the target density $\pi(\theta|y^*, s)$ is proportional to

$$g(\theta) = \prod_{t=1}^{n} N(y_t^*|m_{s_t} + \gamma \log w^2 + \hat{h}_{t|t-1}, f_{t|t-1}),$$

where $\{\hat{h}_{t|t-1}, f_{t|t-1}\}$ are computed via the Kalman filter recursions for $t = 1, 2, \ldots, n$ as

$$\hat{h}_{t|t-1} = \mu + \phi(\hat{h}_{t-1|t-1} - \mu) + z_t'\alpha, \quad p_{t|t-1} = \phi^2 p_{t-1|t-1} + \sigma_\eta^2,$$

$$f_{t|t-1} = p_{t|t-1} + v_{s_t}^2, \quad k_t = p_{t|t-1} f_{t|t-1}^{-1},$$

$$\hat{h}_{t|t} = \hat{h}_{t|t-1} + k_t(y_t^* - m_{s_t} - \gamma \log w^2 - \hat{h}_{t|t-1}), \quad p_{t|t} = (1 - k_t)p_{t|t-1}. \quad (21)$$

The density $g(\theta)$ is, thus, evaluated using the prediction error decomposition

$$\log g(\theta) \propto \frac{1}{2} \sum_{t=1}^N \log f_{t|t-1} - \frac{1}{2} \sum_{t=1}^N \frac{e_t^2}{f_{t|t-1}},$$

where $e_t = y_t^* - m_{s_t} - \gamma \log w^2 - \hat{h}_{t|t-1}$.
We then calculate

$$m = \arg \max_{\theta} \log(g(\theta)),$$

$$V = \{-\partial^2 \log(g(\theta))/\partial\theta \, \partial\theta'\}^{-1}$$

the negative inverse of the hessian at $m$. These are found by numerical optimization, typically initializing at the current value of $\theta$. The proposal density for the M–H step (Chib and Greenberg, 1995) is based on $(m, V)$ and is specified as multivariate-$t$ with $\xi$ degrees of freedom. The M–H step is completed by proposing the value $\theta'$ from $f_T(\theta|m, V, \xi)$ and accepting it with probability

$$\alpha(\theta, \theta'|y^*, s) = \min\left\{\frac{g(\theta')}{g(\theta)} \frac{f_T(\theta|m, V, \xi)}{f_T(\theta'|m, V, \xi)}, 1\right\}.$$

If the proposal value is rejected, the current value $\theta$ is retained as the next draw.
(b) $\{h_t\}$ from $\{h_t\}|y^*, s, \theta$ in one block using the simulation smoother of de Jong and Shephard (1995). This involves running the Kalman filter (21), storing $\{e_t, f_{t|t-1}^{-1}, k_t\}$ followed by backward recursions, defining $n_t = f_{t|t-1}^{-1} + k_t^2 u_t$ and $d_t = f_{t|t-1}^{-1} e_t - r_t k_t$, where going from $t = n, \ldots, 1$ with $r_n = 0$ and $u_n = 0$,

$$c_t = v_{s_t}^2 - v_{s_t}^4 n_t, \qquad \zeta_t \sim N(0, c_t), \qquad b_t = v_{s_t}^2(n_t - \phi k_t u_t),$$

$$r_{t-1} = f_{t|t-1}^{-1} e_t + (\phi - k_t)r_{t-1} - b_t \frac{\zeta_t}{c_t}, \qquad u_{t-1} = f_{t|t-1}^{-1} + (\phi - k_t)^2 u_t + \frac{b_t^2}{c_t}.$$

Then, $h_t = y_t^* - m_{s_t} - v_{s_t}^2 d_t - \zeta_t$.

5. Sample $v, \{\lambda_t\}, \{q_t\}|y, \{h_t\}, \gamma, \beta, \{\psi_t\}$ by drawing

(a) $v|y, \{h_t\}, \gamma, \beta, \{\psi_t\}, \{q_t\}$ by the M–H algorithm.
The target density is

$$\pi(v|y, \{h_t\}, \gamma, \beta, \{\psi_t\}, \{q_t\})$$

$$\propto \pi(v) \prod_{t=1}^n St(y_t|x_t'\beta + (\exp(\psi_t) - 1)q_t, w_t^{2\gamma} \exp(h_t), v). \quad (22)$$

We sample this density by finding a proposal density that is tailored to the target $\pi(v|y, h, \gamma, \beta)$ and applying the Metropolis–Hastings algorithm in a manner analogous to the case of $\theta$.

(b) $\{q_t\}|y,\{h_t\},\gamma,\beta,v,\kappa$ directly, using the two-point discrete distribution

$$\Pr(q_t = 1|y,\{h_t\},\gamma,\beta,v,\kappa) \propto \kappa\, St(y_t|x_t'\beta + k_t, \sigma_t^{2*}, v),$$

$$\Pr(q_t = 0|y,\{h_t\},\gamma,\beta,v,\kappa) \propto (1 - \kappa)\, St(y_t|x_t'\beta, \sigma_t^{2*}, v),$$

where $\sigma_t^{2*} \equiv w_t^{2\gamma}\exp(h_t)$.

(c) $\{\lambda_t\}|y,\{h_t\},\gamma,\beta,\{\psi_t\},\{q_t\},v$ independently for $t \leqslant n$ from $\lambda_t|y_t, h_t, \gamma, \beta, \psi_t, q_t, v$ where

$$\lambda_t|y_t, h_t, \gamma, \beta, \psi_t, q_t, v$$

$$\sim \text{Gamma}\left(\frac{v+1}{2}, \frac{v + (y_t - x_t'\beta - (\exp(\psi_t) - 1)q_t)^2/\{w_t^{2\gamma}\exp(h_t)\}}{2}\right).$$

6. Sample $\delta, \{\psi_t\}|y,\{h_t\},\gamma,\beta,\{\lambda_t\},\{q_t\},v$

This is done in block under the assumption that $\psi_t$ is small in which case $\exp(\psi_t)-1$ may be approximated by $\psi_t$ and the distribution of $y_t$ marginalized over $\psi_t$ can be derived to be

$$y_t|h_t, q_t, \lambda_t, \delta, \gamma, \beta, v \sim \text{N}(x_t'\beta - 0.5\delta^2 q_t, \delta^2 q_t^2 + \sigma_t^2),$$

where $\sigma_t^2 = w_t^{2\gamma}\exp(h_t)\lambda_t^{-1}$. Now, sample $\delta$ and $\{\psi_t\}$ from $\delta, \{\psi_t\}|y,\{h_t\},\{q_t\},\{\lambda_t\}$, $v,\beta$ by drawing

- $\delta$ by M–H from

$$g(\delta) = \pi(\delta) \prod_{t=1}^{n} f(y_t|h_t, q_t, \lambda_t, \delta, \gamma, \beta, v) \tag{23}$$

using a tailored proposal density. Similarly to what is done for $\theta$ and $v$ the parameters of the proposal density are computed by maximizing (23) with respect to $\delta$.

- $\{\psi_t\}$ from $\psi_t|y, h_t, q_t, \lambda_t, \delta, \gamma, \beta, v$ as follows:

   Repeat for $t = 1, \ldots, N$:
   — if $q_t = 0$ simulate $\psi_t$ from the normal prior in (3),
   — if $q_t = 1$ then $\psi_t|y, h_t, q_t, \lambda_t, \delta, \gamma, \beta, v \sim \text{N}(\psi_t^*, \Psi^{-1})$ where $\Psi = 1/\delta^2 + q_t^2\sigma_t^{-2}$ and $\psi_t^* = \Psi^{-1}(-0.5 + q_t\sigma_t^{-2}y_t)$.

7. Sample $\kappa|\{q_t\}$ by drawing

$$\kappa|q_1 \ldots q_N \sim \text{Beta}(u_0 + n_1, u_1 + n_0),$$

where $n_0$ is the count of $q_t = 0$ and $n_1 = n - n_0$ is the count of $q_t = 1$.

8. Repeat Steps 2–7 using the most recent values of the conditioning variables.

# References

Albert, J., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business and Economic Statistics 11, 1–15.

Andersen, T.G., Lund, J., 1997. Estimating continuous-time stochastic volatility models of the short term interest rate. Journal of Econometrics 77 (2), 343–377.

Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance, forthcoming.

Aguilar, O., West, M., 2000. Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. Journal of Business and Economic Statistics 18, 338–357.

Ball, C., Torous, W., 1985. On jumps in common stock prices and their impact on call option pricing. Journal of Finance 40, 155–173.

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

Barndorff-Nielsen, O.E., Shephard, N., 2001b. Econometric analysis of realised volatility and its use in estimating stochastic volatility models, to appear. Journal of the Royal Statistical Society Series B, 64.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Lévy based dynamic models for financial economics. Unpublished book manuscript.

Bates, D.S., 1996. Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options. The Review of Financial Studies 9, 69–107.

Berzuini, C., Best, N.G., Gilks, W.R., Larizza, C., 1997. Dynamic conditional independence models and Markov chain Monte Carlo methods. Journal of the American Statistical Association 92, 1403–1412.

Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G., 2000. A new class of stochastic volatility models with jumps. Theory and estimation. Working Paper, Columbia University.

Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Chib, S., 1996. Calculating posterior distributions and modal estimates in markov mixture models. Journal of Econometrics 75, 79–97.

Chib, S., Greenberg, E., 1994. Bayes inference for regression models with ARMA($p, q$) errors. Journal of Econometrics 64, 183–206.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. American Statistician 49, 327–335.

Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in Econometrics. Econometric Theory 12, 409–431.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis Hastings output. Journal of the American Statistical Association 96, 270–281.

Chib, S., Nardari, F., Shephard, N., 1999. Analysis of high dimensional multivariate stochastic volatility models. Unpublished paper, John M. Olin School of Business, Washington University, St. Louis. Revised June 2001.

Danielsson, J., 1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. Journal of Econometrics 61, 375–400.

de Jong, P., Shephard, N., 1995. The simulation smoother for time series models. Biometrika 82, 339–350.

Diebold, F.X., Gunther, T., Tay, A., 1997. Evaluating density forecasts, with applications to financial risk management. International Economic Review 39, 863–883.

Doucet, A., deFreitas, N., Gordon, N., 2001. Sequential Monte Carlo Methods in Practice. Springer-Verlag, New York.

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

Eberlein, E., 2001. Application of generalized hyperbolic Lévy motion to finance. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S. (Eds.), Lévy Processes—Theory and Applications. Birkhauser, Boston, pp. 319–337.

Eberlein, E., Prause, K., 2002. The generalized hyperbolic model: financial derivatives and risk measures. In: Mathematical Finance—Bachelier Congress 2000. Springer Verlag, forthcoming.

Eberlein, E., Kallsen, J., Kristen, J., 2001. Risk management based on stochastic volatility. FDM Preprint 72, University of Freiburg.

Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed non-linear diffusions. Econometrica 69, 959–994.

Eraker, B., 2001. MCMC analysis of diffusion models with applications to finance. Journal of Business and Economic Statistics 19, 177–191.

Eraker, B., Johannes, M., Polson, N., 2002. The impact of jumps in volatility and returns. Journal of Finance, forthcoming.

Gallant, A.R., Tauchen, G., 1998. Reprojection partially observed systems with applications to interest rate diffusions. Journal of the American Statistical Association 93, 10–24.

Gallant, A.R., Hsu, C., Tauchen, G., 1998. Calibrating volatility diffusions and extracting integrated volatility. Unpublished paper, Duke University.

Gerlach, R., Carter, C.K., Kohn, R., 1999. Diagnostics for time series analysis. Journal of Time Series Analysis 21, 309–330.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of moments (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 4. Oxford University Press, Oxford, pp. 169–193.

Ghysels, E., Harvey, A.C., Renault, E., 1996. Stochastic volatility. In: Rao, C.R., Maddala, G.S. (Eds.), Statistical Methods in Finance. North-Holland, Amsterdam, pp. 119–191.

Gordon, N.J., Salmond, D.J., Smith, A.F.M., 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.

Isard, M., Blake, A., 1996. Contour tracking by stochastic propagation of conditional density. Proceedings of the European Conference on Computer Vision, Cambridge, Vol. 1, pp. 343–356.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal of Business and Economic Statistics 12, 371–417.

Jacquier, E., Polson, N.G., Rossi, P.E., 1999. Models and prior distributions for multivariate stochastic volatility. Draft paper, Chicago School of Business (first version circulated 1994).

Jeffreys, H., 1961. Theory of Probability, 3rd ed. Oxford, Clarendon Press.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

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

Mahieu, R., Schotman, P., 1998. An empirical application of stochastic volatility models. Journal of Applied Econometrics 16, 333–359.

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

Nardari, F., 1999. Comparing stochastic volatility models of the short term interest rate: a Bayesian approach. Manuscript, Arizona State University, Tempe, AZ, USA.

Pitt, M., Shephard, N., 1999a. Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association 94, 590–599.

Pitt, M., Shephard, N., 1999b. Time varying covariances: a factor stochastic volatility approach (with discussion). In: Bernardo, J., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 6. Oxford University Press, Oxford, pp. 547–570.

Prause, K., 1999. The generalised hyperbolic models: estimation, financial derivatives and risk measurement. Unpublished Ph.D. Thesis, Mathematics Faculty, Freiburg University.

Sandmann, G., Koopman, S.J., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. Journal of Econometrics 87, 271–301.

Shephard, N., 1994. Partial non-Gaussian state space. Biometrika 81, 115–131.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielson, O.E. (Eds.), Time Series Models with Econometric, Finance and Other Applications. Chapman & Hall, London, pp. 1–67.

Shephard, N., 2000. Discussion of Durbin and Koopman (2000). Journal of the Royal Statistical Society, Series B 62, 30–32.

So, M.K.P., Lam, K., Lee, W.K., 1998. A stochastic volatility model with Markov switching. Journal of Business and Economic Statistics 16, 244–253.

Taylor, S.J., 1994. Modelling stochastic volatility. Mathematical Finance 4, 183–204.

Wong, C.K., 2000. Estimation of latent factor models by simulation methods. Unpublished D. Phil. Thesis, Department of Economics, Oxford University.