Computational Statistics and Data Analysis 54 (2010) 2883–2898

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions

C.A. Abanto-Valle [a,*], D. Bandyopadhyay [b], V.H. Lachos [c], I. Enriquez [d]

[a] *Department of Statistics, Federal University of Rio de Janeiro, Brazil*

[b] *Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina, Charleston, SC, USA*

[c] *Department of Statistics, Campinas State University, Brazil*

[d] *Department of Statistics, São Paulo State University, Brazil*

ARTICLE INFO

Article history:
Received 26 November 2008
Received in revised form 22 June 2009
Accepted 23 June 2009
Available online 26 June 2009

ABSTRACT

A Bayesian analysis of stochastic volatility (SV) models using the class of symmetric scale mixtures of normal (SMN) distributions is considered. In the face of non-normality, this provides an appealing robust alternative to the routine use of the normal distribution. Specific distributions examined include the normal, student-t, slash and the variance gamma distributions. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) algorithm is introduced for parameter estimation. Moreover, the mixing parameters obtained as a by-product of the scale mixture representation can be used to identify outliers. The methods developed are applied to analyze daily stock returns data on S&P500 index. Bayesian model selection criteria as well as out-of-sample forecasting results reveal that the SV models based on heavy-tailed SMN distributions provide significant improvement in model fit as well as prediction to the S&P500 index data over the usual normal model.

1. Introduction

The stochastic volatility (SV) model was introduced by Tauchen and Pitts (1983) and Taylor (1982) as a way to describe the time-varying volatility of asset returns. It has emerged as an alternative to generalized autoregressive conditional heteroscedasticity (GARCH) models of Bollerslev (1986), because it is directly connected to the type of diffusion processes used in asset-pricing theory in finance (Melino and Turnbull, 1990) and captures the main empirical properties often observed in daily series of financial returns (Carnero et al., 2004) in a more appropriate way.

The SV model with a conditional normal distribution for the returns has been extensively analyzed in the literature. From a Bayesian standpoint, several MCMC-based algorithms have been suggested for the estimation of the SV model. For example, Jacquier et al. (1994) use the single-move Gibbs sampling within the Metropolis–Hastings algorithm to sample from the log-volatilities. Kim et al. (1998) and Mahieu and Schotman (1998), among others, approximate the distribution of log-squared returns with a discrete mixture of several normal distributions, allowing jointly drawing on the components of the whole vector of log-volatilities. Shephard and Pitt (1997) and Watanabe and Omori (2004) suggested the use of random blocks containing some of the components of the log-volatilities in order to reduce the autocorrelation effectively. However, in all of these, the normal distribution was assumed as the basis for parameter inference.

Unfortunately, normality assumption is too restrictive and suffers from the lack of robustness in the presence of outliers, which can have a significant effect on the model-based inference. Thus, various generalizations of the standard SV model have emerged and their model fittings have been investigated. It has been specifically pointed out that asset returns data

* Corresponding address: Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970, RJ, Brazil.
E-mail address: cabantovalle@im.ufrj.br (C.A. Abanto-Valle).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2009.06.011

have heavier tails than those of normal distribution. See for instance, Mandelbrot (1963), Fama (1965), Liesenfeld and Jung (2000), Chib et al. (2002), Jacquier et al. (2004) and Chen et al. (2008). In this context, the SV model with Student-t errors (SV-t) is one of the most popular basic models to account for heavier-tailed returns. In this paper, we extend the SV model by assuming the flexible class of scale mixtures of normal (SMN) distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993; Fernández and Steel, 2000; Chow and Chan, 2008). Interestingly, this rich class contains as proper elements the SV model with normal (SV-N), student-t (SV-t), slash (SV-S) and variance gamma (SV–VG) distributions. All these distributions have heavier tails than the normal one, and thus can be used for robust inference in these types of models. We refer to this generalization as SV–SMN models. Our work is motivated by the fact that the daily stock returns data on S&P500 index seems to exhibit significant heavy-tail behavior as shown in Yu (2005). Inference in the class of SV–SMN models is performed under a Bayesian paradigm via MCMC methods, which permits to obtain the posterior distribution of parameters by simulation starting from reasonable prior assumptions on the parameters. We simulate the log-volatilities and the shape parameters by using the block sampler algorithm (Shephard and Pitt, 1997; Watanabe and Omori, 2004) and the Metropolis–Hastings sampling, respectively.

The rest of the paper is structured as follows. Section 2 gives a brief description of SMN distributions. Section 3 outlines the general class of the SV–SMN models as well as the Bayesian estimation procedure using MCMC methods. Additionally, we discuss some technical details about Bayesian model selection and out-of-sample forecasting of aggregated squared returns. Section 4 is devoted to application and model comparison among particular members of the SV–SMN models using the S&P500 index data set. Some concluding remarks as well as future developments are deferred to Section 5.

2. SMN distribution

Scale mixtures of normal distributions, which play a very important role in statistical modeling, are derived by mixing a normally distributed random variable (Z) with a non-negative scale random variable (λ), as follows

$$Y = \mu + \kappa^{1/2}(\lambda) Z,$$

where μ is a location parameter, λ is a positive mixing random variable with probability density function (pdf) $h(\lambda|\mathbf{v})$, independent of $Z \sim \mathcal{N}(0, \sigma^2)$, where $\mathbf{v}$ is a scalar or parameter vector indexing the distribution of λ and $\kappa(.)$ is a positive weight function. As in Lange and Sinsheimer (1993) and Chow and Chan (2008), we restrict our attention to the case in that $\kappa(\lambda) = 1/\lambda$ in this paper. Thus, given λ, $Y|\lambda \sim \mathcal{N}(\mu, \lambda^{-1}\sigma^2)$ and the pdf of Y is given by

$$f(y|\mu, \sigma^2, \nu) = \int_0^\infty \mathcal{N}(y|\mu, \lambda^{-1}\sigma^2) h(\lambda|\mathbf{v}) \mathrm{d}\lambda. \tag{1}$$

From a suitable choice of the mixing density $h(.|\mathbf{v})$, a rich class of continuous symmetric and unimodal distribution can be described by the density given in (1) that can readily accommodate a thicker-than-normal process. Note that when $\lambda = 1$ (a degenerate random variable), we retrieve the normal distribution. Apart from the normal model, we explore 3 different types of heavy-tailed densities based on the choice of the mixing density $h(.|\mathbf{v})$. These are as follows.

- *The student-t distribution*, $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$

  The use of the student-t distribution as an alternative robust model to the normal distribution has frequently been suggested in the literature (Little, 1988; Lange et al., 1989). For the student-t distribution with location μ, scale σ and degrees of freedom ν, the pdf can be expressed in the following SMN form:

$$f(y|\mu, \sigma, \nu) = \int_0^\infty \mathcal{N}\left(y|\mu, \frac{\sigma^2}{\lambda}\right) \mathcal{G}\left(\lambda|\frac{\nu}{2}, \frac{\nu}{2}\right) \mathrm{d}\lambda. \tag{2}$$

  where $\mathcal{G}(.|a, b)$ is the Gamma density function of the form

$$\mathcal{G}(\lambda|a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp(-b\lambda), \quad \lambda, a, b > 0, \tag{3}$$

  and $\Gamma(a)$ is the gamma function with argument $a > 0$. That is, $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$ is equivalent to the following hierarchical form:

$$Y|\mu, \sigma^2, \nu, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{G}(\nu/2, \nu/2). \tag{4}$$

- *The slash distribution*, $Y \sim \mathcal{S}(\mu, \sigma^2, \nu)$, $\nu > 0$

  This distribution presents heavier tails than those of the normal distribution and it includes the normal case when $\nu \uparrow \infty$. Its pdf is given by

$$f(y|\mu, \sigma, \nu) = \nu \int_0^1 \lambda^{\nu-1} \mathcal{N}\left(y|\mu, \frac{\sigma^2}{\lambda}\right) \mathrm{d}\lambda. \tag{5}$$

  where the density of λ is given by

$$h(\lambda|\nu) = \nu\lambda^{\nu-1}\mathbb{I}_{(0,1)}. \tag{6}$$

Thus, the slash distribution is equivalent to the following hierarchical form:

$$Y|\mu,\sigma^2,\nu,\lambda \sim N\left(\mu,\frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{B}e(\nu,1), \tag{7}$$

where $\mathcal{B}e(.,.)$ denotes the beta distribution. The slash distribution has been mainly used in simulation studies because it represents some extreme situations depending on the value of ν, see for example Andrews et al. (1972), Gross (1973), Morgenthaler and Tukey (1991) and Wang and Genton (2006).

- *The variance gamma distribution*, $Y \sim \mathcal{VG}(\mu,\sigma^2,\nu)$, $\nu > 0$

  The symmetric variance gamma (VG) distribution was first proposed by Madan and Seneta (1990) to model share market returns. The VG distribution is controlled by the shape parameter $\nu > 0$, presents heavier tails than those of the normal distribution and has a similar SMN density representation to the student-t distribution. It can be shown that the VG density can be expressed as

$$f(y|\mu,\sigma,\nu) = \int_0^\infty N\left(y|\mu,\frac{\sigma^2}{\lambda}\right)\mathcal{IG}\left(\lambda|\frac{\nu}{2},\frac{\nu}{2}\right)\mathrm{d}\lambda. \tag{8}$$

Thus, the VG distribution is equivalent to the following hierarchical form:

$$Y|\mu,\sigma^2,\nu,\lambda \sim N\left(\mu,\frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{IG}\left(\frac{\nu}{2},\frac{\nu}{2}\right), \tag{9}$$

where $\mathcal{IG}(a,b)$ is the inverse gamma distribution with pdf

$$\mathcal{IG}(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{-(a+1)}\exp\left(-\frac{b}{\lambda}\right).$$

When $\nu = 2$, the VG distribution is the Laplace distribution.

3. The heavy-tailed stochastic volatility model

Among the variants of the SV models, Taylor (1982, 1986) formulated the discrete-time SV model given by

$$y_t = \mathrm{e}^{\frac{h_t}{2}}\varepsilon_t, \tag{10a}$$

$$h_t = \alpha + \phi h_{t-1} + \sigma_\eta \eta_t, \tag{10b}$$

where y_t and h_t are respectively the compounded return and the log-volatility at time t. The innovations ε_t and η_t are assumed to be mutually independent and normally distributed with mean zero and unit variance.

In this article, we modify the basic specification (the SV-N model) in order to capture heavy-tailed features in the marginal distribution of random errors, by replacing the normality assumption of ε_t by the SMN class of distributions as follows:

$$\varepsilon_t \sim SMN(0,1,\nu), \qquad \eta_t \sim \mathcal{N}(0,1), \tag{11}$$

ε_t and η_t assumed to be independent. We refer to this generalization as SV–SMN models. It follows from (1) that the setup defined in (10a), (10b) and (11) can be written hierarchically as

$$y_t = \mathrm{e}^{\frac{h_t}{2}}\lambda_t^{-\frac{1}{2}}\epsilon_t, \tag{12a}$$

$$h_t = \alpha + \phi h_{t-1} + \sigma_\eta \eta_t, \tag{12b}$$

$$\lambda_t \sim p(\lambda_t), \qquad \epsilon_t \sim \mathcal{N}(0,1), \qquad \eta_t \sim \mathcal{N}(0,1). \tag{12c}$$

As depicted in Section 2, this class of models includes the SV with student-t (SV-t), with slash (SV-S) and with variance gamma distributions (SV–VG) as special cases. All these distributions have heavier tails than the normal density and thus provide an appealing robust alternative to the usual Gaussian process in SV models. The SV-t, SV-S and SV–VG models are obtained chosen the mixing density as: $\lambda_t \sim \mathcal{G}(\frac{\nu}{2},\frac{\nu}{2})$, $\lambda_t \sim \mathcal{B}e(\nu,1)$ and $\lambda_t \sim \mathcal{IG}(\frac{\nu}{2},\frac{\nu}{2})$ respectively, where $\mathcal{G}(.,.)$, $\mathcal{IG}(.,.)$ and $\mathcal{B}e(.,.)$ denote the gamma, inverse gamma and beta distributions respectively. Under a Bayesian paradigm, we use MCMC methods to conduct the posterior analysis in the next subsection. Conditionally to λ_t, some derivations are common to all members of the SV–SMN family (see Appendix for details).

3.1. Parameter estimation via MCMC

A Bayesian approach to parameter estimation in the SV–SMN class of models defined by Eqs. (12a)–(12c) relies on MCMC techniques. We propose to construct a novel algorithm based on MCMC simulation methods to make the Bayesian analysis feasible.

Let $\boldsymbol{\theta}$ be the entire parameter vector of the entire class of SV–SMN models, $\mathbf{h}_{0:T} = (h_0, h_1, \ldots, h_T)'$ be the vector of the log-volatilities, $\boldsymbol{\lambda}_{1:T} = (\lambda_1, \ldots, \lambda_T)'$ the mixing variables and $\mathbf{y}_{1:T} = (y_1, \ldots, y_T)'$ is the information available up to time T. The Bayesian approach for estimating the parameters in the SV–SMN models uses the data augmentation principle, which considers $\mathbf{h}_{0:T}$ and $\boldsymbol{\lambda}_{1:T}$ as latent parameters. By using the Bayes theorem, the joint posterior density of parameters and latent variables can be written as

$$p(\mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}, \boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T} \mid \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T})p(\mathbf{h}_{0:T} \mid \boldsymbol{\theta})p(\boldsymbol{\lambda}_{1:T} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{13}$$

where

$$p(\mathbf{y}_{1:T} \mid \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T}) \propto \prod_{t=1}^{T} \lambda_t^{1/2} e^{-\frac{h_t+\lambda_t y_t^2 e^{-h_t}}{2}}, \tag{14}$$

$$p(\mathbf{h}_{0:T} \mid \boldsymbol{\theta}) \propto e^{-\frac{1-\phi^2}{2\sigma_\eta^2}(h_0-\frac{\alpha}{1-\phi})^2} \prod_{t=1}^{T} e^{-\frac{1}{2\sigma_\eta^2}(h_t-\alpha-\phi h_{t-1})^2}, \tag{15}$$

$$p(\boldsymbol{\lambda}_{1:T} \mid \boldsymbol{\theta}) = \prod_{t=1}^{T} p(\lambda_t), \tag{16}$$

where $p(\boldsymbol{\theta})$ is the prior distribution. For the common parameters of the SV–SMN class, the prior distributions are set as: $\alpha \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2)$, $\phi \sim \mathcal{N}_{(-1,1)}(\bar{\phi}, \sigma_\phi^2)$, and $\sigma_\eta^2 \sim \mathcal{IG}(\frac{T_0}{2}, \frac{M_0}{2})$, where $\mathcal{N}_{(a,b)}(.,.)$ denotes the truncated normal distribution in the interval (a, b).

Since the posterior density $p(\mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}, \boldsymbol{\theta} \mid \mathbf{y}_{0:T})$ does not have closed form, we first sample the parameters $\boldsymbol{\theta}$, followed by the latent variables $\boldsymbol{\lambda}_{1:T}$ and $\mathbf{h}_{0:T}$ using Gibbs sampling. The sampling scheme is described by the following algorithm:

Algorithm 3.1.

1. Set $i = 0$ and get starting values for the parameters $\boldsymbol{\theta}^{(i)}$, the states $\boldsymbol{\lambda}_{1:T}^{(i)}$ and $\mathbf{h}_{0:T}^{(i)}$
2. Draw $\boldsymbol{\theta}^{(i+1)} \sim p(\boldsymbol{\theta} \mid \mathbf{h}_{0:T}^{(i)}, \boldsymbol{\lambda}_{1:T}^{(i)}, \mathbf{y}_{1:T})$
3. Draw $\boldsymbol{\lambda}_{1:T}^{(i+1)} \sim p(\boldsymbol{\lambda}_{1:T} \mid \boldsymbol{\theta}^{(i+1)}, \mathbf{h}_{0:T}^{(i)}, \mathbf{y}_{1:T})$
4. Draw $\mathbf{h}_{0:T}^{(i+1)} \sim p(\mathbf{h}_{0:T} \mid \boldsymbol{\theta}^{(i+1)}, \boldsymbol{\lambda}_{1:T}^{(i+1)}, \mathbf{y}_{1:T})$
5. Set $i = i + 1$ and return to 2 until convergence is achieved.

As described by Algorithm 3.1, the Gibbs sampler requires to sample parameters and latent variables from their full conditionals. Sampling the log-volatilities $\mathbf{h}_{0:T}$ in Step 4 is the more difficult task due to the nonlinear setup in the mean equation (12a). In order to avoid the higher correlations due to the Markovian structure of the h_t's, we develop a multi-move sampler (Shephard and Pitt, 1997; Watanabe and Omori, 2004; Omori and Watanabe, 2008; Abanto-Valle et al., 2009) in the next subsection to sample the $\mathbf{h}_{0:T}$ by blocks. Multi-move algorithms are computationally efficient and convergence is achieved much faster than using a single move (Carter and Kohn, 1994; Frühwirth-Schnater, 1994; de Jong and Shephard, 1995). Details on the full conditionals of $\boldsymbol{\theta}$ and the latent variable $\boldsymbol{\lambda}_{1:T}$ are given in the Appendix, some of them are easy to simulate from.

3.2. Multi-move algorithm

In order to simulate $\mathbf{h}_{0:T}$, we consider a two-step process: first, we simulate h_0 conditional on $\mathbf{h}_{1:T}$, next $\mathbf{h}_{1:T}$ conditional on h_0. In our block sampler, we divide $\mathbf{h}_{1:T}$ into $K+1$ blocks, $\mathbf{h}_{k_{i-1}+1:k_i-1} = (h_{k_{i-1}+1}, \dots, h_{k_i-1})'$ for $i = 1, \dots, K+1$, with $k_0 = 0$ and $k_{K+1} = T$, where $k_i - 1 - k_{i-1} \geq 2$ is the size of the ith block. Following Shephard and Pitt (1997) and Omori and Watanabe (2008), the K knots $(k_1, \dots, k_K)$ are generated randomly using

$$k_i = \text{int}[T \times \{(i + u_i)/(K + 2)\}], \quad i = 1, \dots, K, \tag{17}$$

where the u_i's are independent realizations of the uniform random variable on the interval (0, 1) and int[x] represents the floor of x. A suitable selection of K is important to obtain an efficient sampler that can reduce the correlation imposed by the model in the sampling process. If K is too large the sampler will be slow because of rejections; if K is too small it will be correlated because of the structure of the model.

We sample the block of disturbances $\boldsymbol{\eta}_{k_{i-1}+1:k_i-1} = (\eta_{k_{i-1}+1}, \dots, \eta_{k_i-1})'$ instead of $\mathbf{h}_{k_{i-1}+1:k_i-1} = (h_{k_{i-1}+1}, \dots, h_{k_i-1})'$, exploring the fact that the innovations η_t are *i.i.d.* with $\mathcal{N}(0, 1)$ distribution. Suppose that $k_{i-1} = t$ and $k_i = t+k+1$ for the ith block, such that $t + k < T$. Then $\boldsymbol{\eta}_{t+1:t+k} = (\eta_{t+1}, \dots, \eta_{t+k})'$ are sampled at once from their full conditional distribution $f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k}, \boldsymbol{\theta})$, which is expressed in the log scale as

$$\begin{aligned} &\log f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k}, \boldsymbol{\theta}) \\ &\quad = \text{const} - \frac{1}{2\sigma_\eta^2}\sum_{r=t+1}^{t+k}\eta_r^2 + \sum_{r=t+1}^{t+k} l(h_r) - \frac{1}{2\sigma_\eta^2}(h_{t+k+1} - \alpha - \phi h_{t+k})^2. \end{aligned} \tag{18}$$

We denote the first and second derivatives of $l(h_r) = \log p(y_r \mid \lambda_r, h_r)$ with respect to h_r by l' and l''. As $f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k}, \boldsymbol{\theta})$ does not have a closed form, we use the Metropolis–Hastings acceptance–rejection algorithm (Tierney, 1994; Chib, 1995). To obtain the proposal density g, we propose to use an artificial Gaussian state space model for simulating $\boldsymbol{\eta}_{t+1:t+k}$. Applying a second order Taylor series expansion to $\sum_{r=t+1}^{t+k} l(h_r)$ in Eq. (18) around some preliminary estimate of $\boldsymbol{\eta}_{t+1:t+k}$, denoted by $\hat{\boldsymbol{\eta}}_{t+1:t+k}$, we thus have

$$\log f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \boldsymbol{\theta}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k})$$
$$\approx \text{const} - \frac{1}{2\sigma_\eta^2}\sum_{r=t+1}^{t+k}\eta_r^2 - \frac{1}{2\sigma_\eta^2}(h_{t+k+1}-\alpha-\phi h_{t+k})^2 + \sum_{r=t+1}^{t+k}\left\{l(\hat{h}_r) + (h_r-\hat{h}_r)l'(\hat{h}_r) + \frac{1}{2}(h_r-\hat{h}_r)^2 l''(\hat{h}_r)\right\}, \quad (19)$$

where $\hat{\mathbf{h}}_{t+1:t+k}$ is the estimate of $\mathbf{h}_{t+1:t+k}$ corresponding to $\hat{\boldsymbol{\eta}}_{t+1:t+k}$.

After some simple but tedious algebra, we have the resulting normal density as our proposal, g, defined by:

$$\log f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k}, \boldsymbol{\theta}) \approx \text{const} - \frac{1}{2\sigma_\eta^2}\sum_{r=t+1}^{t+k}\eta_r^2 + \frac{1}{2}\sum_{r=t+1}^{t+k-1} l''(\hat{h}_r)\left(\hat{h}_r - \frac{l'(\hat{h}_r)}{l''(\hat{h}_r)} - h_r\right)^2$$
$$- \frac{\phi^2 - l''(\hat{h}_{t+k})\sigma_\eta^2}{2\sigma_\eta^2}\left\{\frac{\sigma_\eta^2}{\phi^2 - l''(\hat{h}_{t+k})}\left(l'(\hat{h}_{t+k}) - l''(\hat{h}_{t+k})\hat{h}_{t+k} + \frac{\phi-\alpha}{\sigma_\eta^2}h_{t+k+1}\right) - h_{t+k}\right\}^2$$
$$= \log g. \quad (20)$$

From (20), we define auxiliary variables d_r and $\hat{y}_r$ for $r = t+1, \dots, t+k-1$ as follows:

$$d_r = -\frac{1}{l''(\hat{h}_r)},$$
$$\hat{y}_r = \hat{h}_r + d_r l'(\hat{h}_r). \quad (21)$$

For $r = t+k < T$

$$d_r = \frac{\sigma_\eta^2}{\phi - \sigma_\eta^2 l''(\hat{h}_{t+k})},$$
$$\hat{y}_r = d_r\left[l'(\hat{h}_r) - l''(\hat{h}_r)\hat{h}_r + \frac{(\phi-\alpha)}{\sigma_\eta^2}h_{r+1}\right], \quad (22)$$

and when $r = t+k = T$, we use (21) to define the auxiliary variables. From (12a), we have that $l(h_r) = \text{const} - \frac{h_r}{2} - \frac{\lambda_r}{2}y_r^2 e^{-h_r}$. It is easy to show that $l(h_r)$ is log-concave, so d_r is always positive.

The resulting normalized density in (20), defined as g, is a k-dimensional normal density, which is the exact density of $\boldsymbol{\eta}_{t+1:t+k}$ conditional on $\hat{\mathbf{y}}_{t+1:t+k}$ in the linear Gaussian state space model:

$$\hat{y}_r = h_r + \epsilon_r, \quad \epsilon_r \sim \mathcal{N}(0, d_r), \quad (23)$$
$$h_r = \alpha + \phi h_{r-1} + \sigma_\eta \eta_r, \quad \eta_r \sim \mathcal{N}(0, 1). \quad (24)$$

Applying the de Jong and Shephard's simulation smoother (de Jong and Shephard, 1995) to this model with the auxiliary $\hat{\mathbf{y}}_{t+1:t+k}$ defined above enables us to sample $\boldsymbol{\eta}_{t+1:t+k}$ from the density g. Since f is not bounded by g, we use the Metropolis–Hastings acceptance–rejection algorithm to sample from f as recommended by Chib (1995). In the SV-N case, we use the same procedure with $\lambda_t = 1$ for $t = 1, \dots, T$.

We select the expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ as follows. Once an initial expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ is selected, we can calculate the auxiliary $\hat{\mathbf{y}}_{t+1:t+k}$ by using Eqs. (21) and (22). In the MCMC implementation, the previous sample of $\mathbf{h}_{t+1:t+k}$ may be taken as an initial value of the $\hat{\mathbf{h}}_{t+1:t+k}$. Then, applying the Kalman filter and a disturbance smoother to the linear Gaussian state space model consisting of Eqs. (23) and (24) with the artificial $\hat{\mathbf{y}}_{t+1:t+k}$ yields the mean of $\mathbf{h}_{t+1:t+k}$ conditional on $\hat{\mathbf{h}}_{t+1:t+k}$ in the linear Gaussian state space model, which is used as the next $\hat{\mathbf{h}}_{t+1:t+k}$. By repeating the procedure until the smoothed estimates converge, we obtain the posterior mode of $\mathbf{h}_{t+1:t+k}$. This is equivalent to the method of scoring to maximize the logarithm of the conditional posterior density. Although, we have just noted that iterating the procedure achieves the mode, this will slow our simulation algorithm if we have to iterate this procedure until full convergence. Instead we suggest to use only five iterations of this procedure to provide reasonably good sequence $\hat{\mathbf{h}}_{t+1:t+k}$ instead of an optimal one.

3.3. Bayesian model selection

In this section, we describe two Bayesian model selection criteria: the deviance information criterion (Spiegelhalter et al., 2002; Berg et al., 2004; Celeux et al., 2006) and the Bayesian predictive information criterion (Ando, 2006, 2007).

3.3.1. Deviance information criterion

Spiegelhalter et al. (2002) introduced the deviance information criterion (DIC), defined as:

$$DIC = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})] + p_D. \quad (25)$$

The second term in (25) measures the complexity of the model by the effective number of parameters, p_D, defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$p_D = 2[\log L(\mathbf{y}_{1:T} \mid \bar{\boldsymbol{\theta}}) - E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})]]. \quad (26)$$

To calculate the DIC in the context of SV–SMN models, the conditional likelihood $L(\mathbf{y}_{1:T} \mid \alpha, \phi, \sigma^2_\eta, \nu, \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T})$, defined in (14), is used in Eq. (25), where $\boldsymbol{\theta}$ encompasses $(\alpha, \phi, \sigma^2_\eta, \nu)'$, $\boldsymbol{\lambda}_{1:T}$ and $\mathbf{h}_{0:T}$.

As pointed by Stone (2002), Robert and Titterington (2002), Celeux et al. (2006) and Ando (2007), the DIC suffers from some theoretical aspects. First, in the derivation of DIC, Spiegelhalter et al. (2002, p. 604) assumed that the specified parametric family of probability distributions that generate future observations encompasses the true model. This assumption may not always hold true. Secondly, the observed data are used both to construct the posterior distribution and to compute the posterior mean of the expected log-likelihood. Thus, the bias in the estimate of DIC tends to underestimate the true bias considerably. To overcome these theoretical problems in DIC, recently Ando (2007) has proposed the Bayesian predictive information criterion (BPIC) as an improved alternative of the DIC.

3.3.2. Bayesian predictive information criterion

Let us consider $\mathbf{z}_{1:T} = (z_1, z_2, \ldots, z_T)'$ to be a new set of observations generated by the same mechanism as that of the observed data $\mathbf{y}_{1:T}$ drawn from the true model $s(\mathbf{z}_{1:T})$. To evaluate the relative fit of the Bayesian model to the true model $s(\mathbf{z}_{1:T})$, Ando (2007) considered the maximization of the posterior mean of the expected log-likelihood

$$\eta = \int \left[\int \log L(\mathbf{z}_{1:T} \mid \boldsymbol{\theta}) p(\theta \mid \mathbf{y}_{1:T}) s(\mathbf{z}_{1:T}) \mathrm{d}\mathbf{z}_{1:T} \right].$$

It is obvious that η depends on the model fitted, and on the unknown true model $s(\mathbf{z}_{1:T})$. A natural estimator of η is the posterior mean of the log-likelihood,

$$\hat{\eta} = \int \log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) p(\theta \mid \mathbf{y}_{1:T}),$$

where $L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) = \prod_{t=1}^T p(\mathbf{y}_t \mid \boldsymbol{\theta})$. As pointed by Ando (2006, 2007) the quantity $\hat{\eta}$ is generally a positively biased estimator of η, because the same data $\mathbf{y}_{1:T}$ are used both to construct the posterior distribution and to evaluate the posterior mean of the log-likelihood. Therefore, bias correction should be considered, where the bias b is defined as: $b = \int (\hat{\eta} - \eta) s(\mathbf{z}_{1:T}) \mathrm{d}\mathbf{y}_{1:T}$. Ando (2007) evaluated the asymptotic bias as

$$T\hat{b} \approx E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})\}] - \log[L(\mathbf{y}_{1:T} \mid \hat{\boldsymbol{\theta}}) p(\hat{\boldsymbol{\theta}})] + \mathrm{tr}\{J_T^{-1}(\hat{\boldsymbol{\theta}}) I_T(\hat{\boldsymbol{\theta}})\} + 0.5q. \quad (27)$$

Here q is the dimension of $\boldsymbol{\theta}$, $E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[.]$ denotes the expectation with respect to the posterior distribution, $\hat{\boldsymbol{\theta}}$ is the posterior mode, and

$$I_T(\hat{\boldsymbol{\theta}}) = \frac{1}{T} \sum_{t=1}^T \left(\frac{\partial \eta_T(y_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \eta_T(y_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}},$$

$$J_T(\hat{\boldsymbol{\theta}}) = \frac{1}{T} \sum_{t=1}^T \left(\frac{\partial^2 \eta_T(y_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}},$$

with $\eta_T(y_t, \boldsymbol{\theta}) = \log p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}) + \log p(\boldsymbol{\theta})/T$. Thus, correcting the asymptotic bias of the posterior mean of the log-likelihood, the Bayesian predictive information criterion (BPIC; Ando, 2006, 2007) can be written as

$$BPIC = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\}] + 2T\hat{b}. \quad (28)$$

The best model is chosen as the one that has the minimum BPIC. To calculate the BPIC, in the context of SV–SMN models, we use the log-likelihood function $\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\}$ as defined in Eq. (28), where $\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\} = \sum_{t=1}^T \log p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ and $\boldsymbol{\theta} = (\alpha, \phi, \sigma^2_\eta, \nu)'$. Because $p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ does not have closed form, it can be evaluated numerically by using the auxiliary particle filter method (see Kim et al., 1998; Pitt and Shephard, 1999; Chib et al., 2002), which is described next.

3.4. The auxiliary particle filter

In this subsection, we revised the auxiliary particle filtering (APF) method of Pitt and Shephard (1999), which allows us to draw samples from the filtering distribution $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$ by numerical approximation. The method is generically described as follows:

Let us consider $\{(h_{t-1}^{(1)}, w_{t-1}^{(1)}), \ldots, (h_{t-1}^{(N)}, w_{t-1}^{(N)})\} \overset{a}{\sim} p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ where the probability density function, $p(h_{t-1}|\boldsymbol{\theta}, \mathbf{y}_{1:t-1})$, of the continuous random variable, h_{t-1}, is approximated by a discrete variable with random support. It then follows that the one-step ahead predictive distribution $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ can be approximated as:

$$p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) = \int p(h_t \mid h_{t-1}, \boldsymbol{\theta}) p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) \mathrm{d}h_{t-1} \approx \sum_{i=1}^N p(h_t \mid \boldsymbol{\theta}, h_{t-1}^{(i)}) w_{t-1}^{(i)}, \quad (29)$$

where $h_{t-1}^{(i)}$ is a sample from $p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ with weight $w_{t-1}^{(i)}$. The one-step ahead density, $p(y_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$, is then estimated by Monte Carlo averaging of $p(y_t \mid \boldsymbol{\theta}, h_t)$ over the draws of $h_t^{(i)} \sim p(h_t \mid \boldsymbol{\theta}, h_{t-1}^{(i)})$ from Eq. (12b) as follows:

$$p(y_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) = \int p(y_t \mid h_t, \boldsymbol{\theta}) p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) \mathrm{d}h_t \approx \sum_{i=1}^{N} p(y_t \mid \boldsymbol{\theta}, h_t^{(i)}) w_{t-1}^{(i)}. \tag{30}$$

This recursive procedure needs to draw h_t sequentially from the filtered distribution, $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$, which is updated as described in Algorithm 3.2.

Algorithm 3.2.

1. Posterior at $t-1$:
   $\{(h_{t-1}^{(1)}, w_{t-1}^{(1)}), \ldots, (h_{t-1}^{(i)}, w_{t-1}^{(i)}), \ldots, (h_{t-1}^{(N)}, w_{t-1}^{(N)})\} \overset{a}{\sim} p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$
2. For $i = 1, \ldots, N$, calculate $\mu_t^{(i)} = \alpha + \phi h_{t-1}^{(i)}$
3. Sampling (k, h_t):
   For $i = 1, \ldots, N$
   Indicator: k^i such that $P(k^i = k) \propto p(y_t \mid \mu_t^{(k^i)}) w_{t-1}^{(k^i)}$
   Evolution:
   $$h_t^{(i)} \sim \mathcal{N}(\mu_t^{k^i}, \sigma^2)$$
   Weights: compute $w_t^{(i)}$ as follows
   $$w_t^{(i)} \propto \frac{p(y_t \mid \boldsymbol{\theta}, h_t^{(i)})}{p(y_t \mid \boldsymbol{\theta}, \mu_t^{(k^i)})}$$
4. Posterior at t:
   $\{(h_t^{(1)}, w_t^{(1)}), \ldots, (h_t^{(i)}, w_t^{(i)}), \ldots, (h_t^{(N)}, w_t^{(N)})\} \overset{a}{\sim} p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$.

Next, we give some technical details related to the out-of-sample forecasting of aggregated squared returns in SV–SMN models. We refer to the reader to see Tauchen and Pitts (1983) for more details.

3.5. Out-of-sample forecasting of aggregated returns

We have that K-step ahead prediction density can be calculated using the composition method through the following recursive procedure:

$$p(y_{T+K} \mid \mathbf{y}_{1:T}) = \int p(y_{T+K} \mid \lambda_{T+K}, h_{T+K}) p(\lambda_{T+K} \mid \boldsymbol{\theta}) p(h_{T+K} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \mathrm{d}h_{T+K} \mathrm{d}\lambda_{T+K} \mathrm{d}\boldsymbol{\theta},$$

$$p(h_{T+K} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) = \int p(h_{T+K} \mid \boldsymbol{\theta}, h_{T+K-1}) p(h_{T+K-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) \mathrm{d}h_{T+K-1}.$$

Evaluation of the last integrals is straightforward, by using Monte Carlo approximation. To initialize a recursion, we use N draws $\{h_T^{(1)}, \ldots, h_T^{(N)}\}$ and $\{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}\}$ from the MCMC sample. Then given these N draws, sample N draws $\{h_{T+k}^{(1)}, \ldots, h_{T+k}^{(N)}\}$ from $p(h_{T+k} \mid \boldsymbol{\theta}^{(1)}, h_{T+k-1}^{(1)}), \ldots, p(h_{T+k} \mid \boldsymbol{\theta}^{(N)}, h_{T+k-1}^{(N)})$ and $\{\lambda_{T+k}^{(1)}, \ldots, \lambda_{T+k}^{(N)}\}$ from $p(\lambda_{T+k} \mid \boldsymbol{\theta}^{(1)}), \ldots, p(\lambda_{T+k} \mid \boldsymbol{\theta}^{(N)})$, for $k = 1, \ldots, K$, by using Eqs. (12b) and (12c), respectively. Finally, with this N draws $\{h_{T+k}^{(1)}, \ldots, h_{T+k}^{(N)}\}$, sample N draws $\{y_{T+k}^{(1)}, \ldots, y_{T+k}^{(N)}\}$ from $p(y_{T+k} \mid \boldsymbol{\theta}^{(1)}, h_{T+k}^{(1)}), \ldots, p(y_{T+k} \mid \boldsymbol{\theta}^{(N)}, h_{T+k}^{(N)})$, for $k = 1, \ldots, K$. With draws from h_{T+k} and y_{T+k}, the aggregated daily squared return (a common model-free indicator of volatility) can be calculated as $V_K^{(i)} = \sum_{k=1}^{K} y_{T+k}^{2(i)}$ and the aggregated volatility as, $S_K^{(i)} = \sum_{k=1}^{K} \mathrm{e}^{h_{T+k}^{(i)}}$, for $i = 1, \ldots, N$, respectively.

4. Empirical application

This section analyzes the daily closing prices for the S&P500 stock market index. The S&P500 index contains the stocks of 500 Large-Cap corporations. Although a majority of those corporations are US based, it also includes other companies having their common stocks within the index. The data set was obtained from the Yahoo finance web site available to download at http://finance.yahoo.com. The period of analysis is January 5, 1999–September 05, 2008 which yields 2432 observations. Throughout, we will work with the mean corrected returns computed as

$$y_t = 100 \left\{ (\log P_t - \log P_{t-1}) - \frac{1}{T} \sum_{j=1}^{T} (\log P_j - \log P_{j-1}) \right\},$$

where P_t is the closing price on day t.

Table 1 summarizes descriptive statistics for the corrected compounded returns with the time series plot in Fig. 1. For the returns series, the basic statistics viz. the mean, standard deviation, skewness and kurtosis are calculated to be 0.00, 1.13, 0.06 and 5.04, respectively. Note that the kurtosis of the returns is >3, so that daily S&P500 returns likely shows a departure from the underlying normality assumption. Thus, we reanalyze this data with the aim of providing robust inference by using

Table 1
Summary statistics for S&P500 market index series.

	Mean	S.d.	Max	Min	Skewness	Kurtosis
Returns	0.00	1.13	5.58	−6.00	0.05	5.03

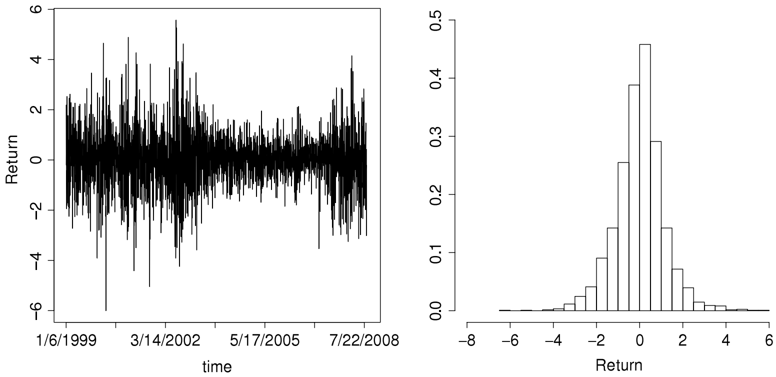


Fig. 1. S&P500 corrected compounded returns with sample period from January 5, 1999 to September 05, 2008. The left panel shows the plot of the raw series and the right panel plots the histogram of returns.

Table 2
Estimation results for the S&P500 daily index returns. The first row: Posterior mean. The second row: Posterior 95% credible interval in parentheses. The third row: Monte Carlo error of the posterior mean. The fourth row: CD statistics.

Parameter	SV-N	SV-t	SV-S	SV–VG
α	−0.0016 (−0.0104, 0.0069) 0.93×10^{-4} −0.11	−0.0040 (−0.0130, 0.0044) 0.90×10^{-4} −0.12	−0.0147 (−0.0270, −0.0043) 1.81×10^{-4} −0.94	−0.0011 (−0.0095, 0.0072) 0.41×10^{-4} 0.51
ϕ	0.9700 (0.9543, 0.9833) 3.04×10^{-4} −1.38	0.9722 (0.9570, 0.9844) 3.03×10^{-4} 0.38	0.9730 (0.9579, 0.9856) 3.11×10^{-4} −1.30	0.9721 (0.9568, 0.9846) 2.99×10^{-4} −0.59
σ^2	0.0447 (0.0293, 0.0652) 5.27×10^{-4} 0.93	0.0411 (0.0273, 0.0599) 5.40×10^{-4} 1.39	0.0404 (0.0254, 0.0594) 5.29×10^{-4} 0.62	0.0402 (0.0270, 0.0607) 4.82×10^{-4} 0.61
ν	– – – –	20.1527 (11.2700, 28.5300) 0.2389 0.69	2.2618 (2.0670, 2.4250) 0.0012 −0.61	17.7880 (9.7930, 30.1460) 0.4535 −0.38

the SMN class of distributions. In our analysis, we compare between the SV-N, SV-t, SV-S and SV–VG distributions from the SMN class of models. All the calculations were performed running stand alone code developed by the authors using an open source C++ library for statistical computation, the Scythe statistical library (Pemstein et al., 2007), which is available for free download at http://scythe.wustl.edu.

In all cases, we simulated the h_t's in a multi-move fashion with stochastic knots based on the method described in Section 3.1. We set the prior distributions of the common parameters as: $\alpha \sim \mathcal{N}(0.0, 100.0)$, $\phi \sim \mathcal{N}_{(-1,1)}(0.95, 100.0)$, $\sigma_\eta^2 \sim \mathcal{IG}(2.5, 0.025)$. The prior distributions on the shape parameters were chosen as: $\nu \sim \mathcal{G}(12.0, 0.8)$, $\nu \sim \mathcal{G}(0.2, 0.05)$ and $\nu \sim \mathcal{G}(2.0, 0.25)$ for the SV-t model, the SV-S model and the SV–VG model, respectively. The initial values of the parameters are randomly generated from the prior distributions. We set all the log-volatilities, h_t, to be zero. Finally the initial $\boldsymbol{\lambda}_{1:T}$ are generated from the prior $p(\lambda_t \mid \nu)$.

We set K, the number of blocks to be 40 in such a way that each block contained 60 h_t's on average. For all the models, we conducted the MCMC simulation for 60 000 iterations. The first 20 000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain,only every 10th values of the chain are stored. With the resulting 4000 values, we calculated the posterior means, the 95% credible intervals, the Monte Carlo error of the posterior means and the convergence diagnostic (CD) statistics (Geweke, 1992). Table 2 summarizes these results. According to the CD values, the null hypothesis that the sequence of 4000 draws is stationary is accepted at 5% level for all the parameters and in all the models considered here. Fig. 2 shows the sampling results for the SV-S model on the S&P500 return series. As expected, we observe a rapid decay of autocorrelations for all the parameters.

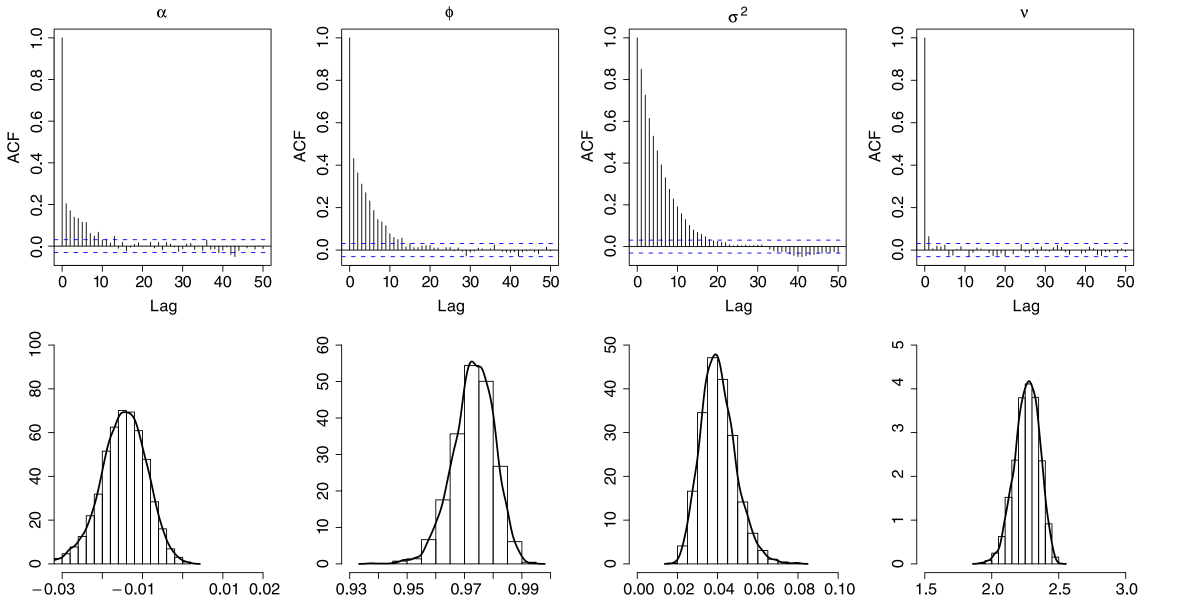


Fig. 2. Estimation results for the S&P500 daily index returns (SV-S model). The top row shows plots of sample autocorrelations and the bottom row shows posterior histograms and overlaid density estimates.

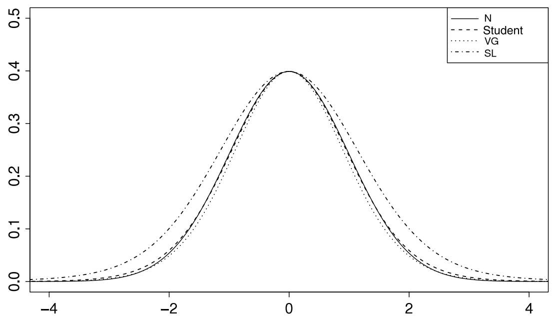


Fig. 3. S&P500 index. Density curves of the univariate normal, student-t, slash and variance gamma using the estimated tail-fatness parameter from the respective SV model.

The estimate of the volatility parameters (α, ϕ, σ^2) are consistent with the results found in the previous literature (e.g. Chib et al., 2002; Omori et al., 2007). The posterior mean of ϕ is close to one, which indicates a well-known high persistence of volatility asset returns. The posterior mean of ϕ for the SV-N model is lower than the other models and the estimates of σ^2 for the SV-t, SV-S and SV–VG models are slightly lower than the SV-N model. Thus, the models allowing heavy-tail errors seem to explain the excess of returns as a realization of the disturbance ϵ_t, which decreases the variance of the volatility process.

The magnitude of the tail fatness is measured by the shape parameter ν in the SV-t, SV-S and SV–VG models. The posterior mean of ν in the SV-t model is 20.1527, which is in accordance with the literature (Nakajima and Omori, 2008). In the SV-S model, the posterior mean of ν is 2.2618, and in the SV–VG model the posterior mean of ν is 17.7880. These results seem to indicate that the measurement error of the stock returns are better explained by heavy-tailed distributions.

To illustrate the tail behavior, we plot the normal $(\mathcal{N}(0, 1))$ density, student-$t(\mathcal{T}(0, 1, \nu))$ density with ν degrees of freedom, the slash $(\mathcal{S}(0, 1, \nu))$ density with shape parameter ν and the variance gamma $(\mathcal{VG}(0, 1, \nu))$ density with shape parameter ν. We set ν as the posterior mean of the respective SV model (see Table 2). Fig. 3 depicts the four density curves (the student-t, slash and variance gamma have been rescaled to be comparable, see Wang and Genton (2006)). The density of the variance gamma emphasizes on the sharpness around the mean rather than the tail fatness, so that the student-t and slash distributions have fatter tails than the standard normal and variance gamma distributions. Note that the slash distribution has fatter tail than the other distributions that we have considered. Therefore, the SV-S and SV-t models attribute a relatively larger proportion of extreme return values to ε_t instead of η_t than those of SV-N and SV–VG models, making the volatility of the SV-S and SV-t models less variable.

The magnitudes of the mixing parameter λ_t are associated with extremeness of the corresponding observations. In the Bayesian paradigm, the posterior mean of the mixing parameter can be used to identify a possible outlier (see, for instance Rosa et al. (2003)). The heavy-tailed SV–SMN models can accommodate an outlier by inflating the variance component for

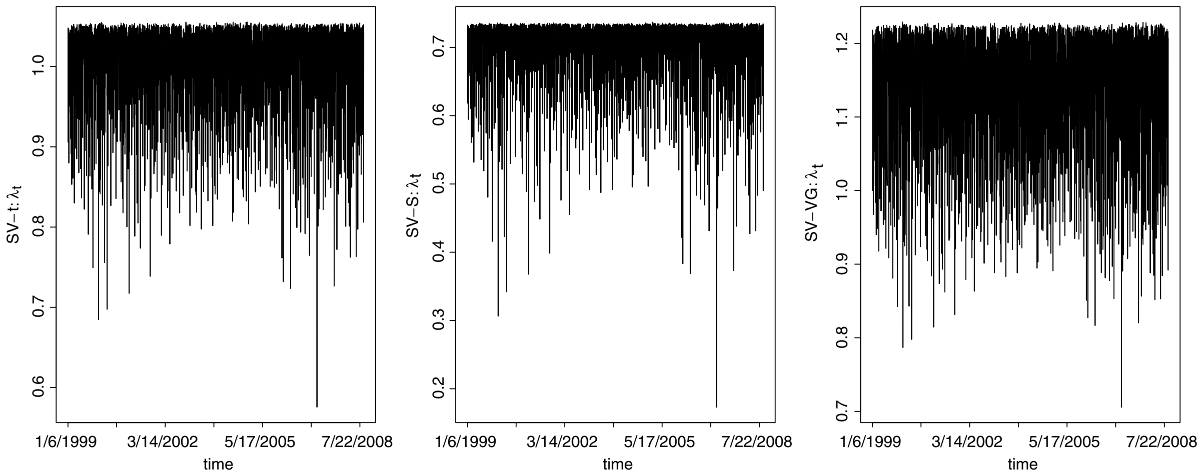


Fig. 4. Comparison of the estimated mixing variables λ_t for the SP&500 index data.

Table 3
SP&500 return data set. DIC: deviance information criterion, BPIC: Bayesian predictive information criterion.

Model	DIC		BPIC	
	Value	Ranking	BPIC	Ranking
SV-N	6889.6	3	7603.1	4
SV-*t*	6888.1	2	6957.4	2
SV-S	6878.4	1	6951.4	1
SV–VG	6906.8	4	7406.5	3

that observation in the conditional normal distribution with smaller λ_t value. This fact is shown in Fig. 4 where we depicted the posterior mean of the mixing variable λ_t for the SV-t (left panel), SV-S (middle panel) and the SV–VG (right panel) models.

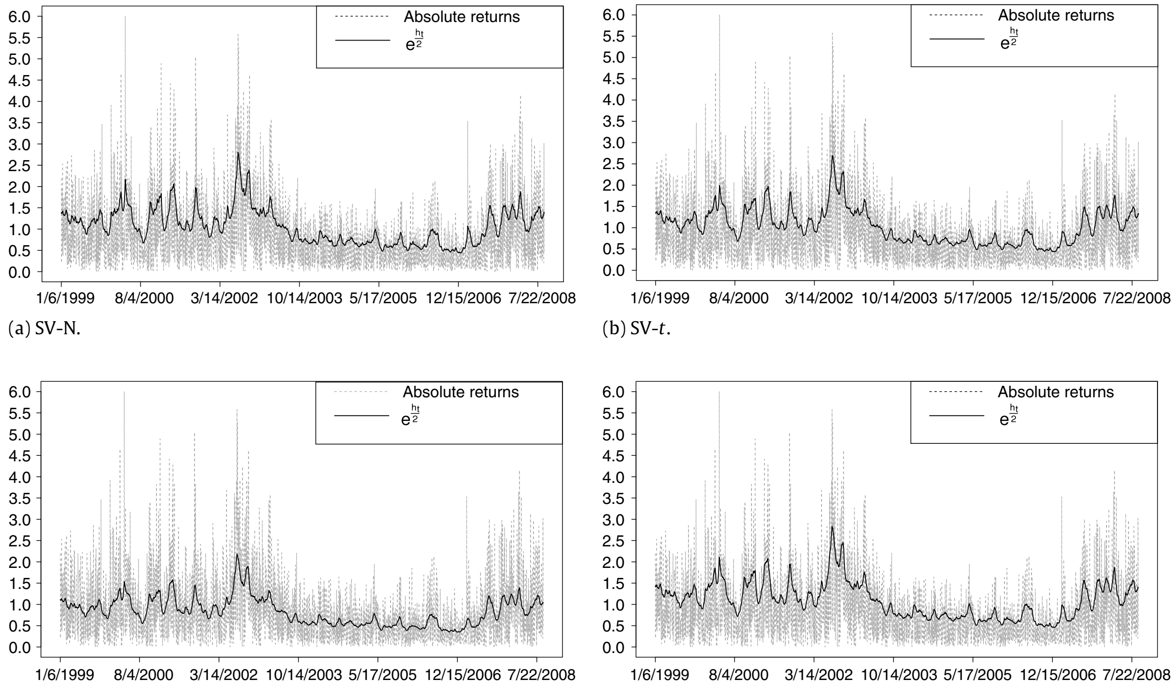
In Fig. 5a–d, we plot the smoothed mean of $e^{\frac{h_t}{2}}$ jointly with the absolute returns for the SV-N, SV-t, SV-S and SV–VG models. It can be seen from Fig. 5a, b and d that the SV-N, SV-t and SV–VG models produce similar estimates to $e^{\frac{h_t}{2}}$. However, the SV-S model in Fig. 5c exhibits smoother movements than the other competing SV models. Clearly, extreme returns make a clear difference. The models with heavy tails accommodate possible outliers in a somewhat different way by inflating the variance $e^{\frac{h_t}{2}}$ by $\lambda_t^{-\frac{1}{2}} e^{\frac{h_t}{2}}$. This can have a substantial impact, for instance, in the valuation of derivative instruments and several strategic or tactical asset allocation topics.

Next, we use the deviance information criterion (DIC) and the Bayesian predictive information criterion (BPIC) to compare between all the competing models. In both cases, the best model has the smallest DIC (BPIC). From Table 3, the BPIC criterion indicates that the SV–SMN models with heavy tails present better fit than the basic SV-N model, with the SV-S model relatively better among all the considered models, suggesting that the SP&500 data demonstrate sufficient departure from underlying normality assumptions. As expected, the DIC also selects the SV-S model as the best model.

Forecasting asset price volatility has become an important area in empirical finance research, because volatility plays a significant role in asset-pricing models, portfolio management and trading strategies. Using the particle filter algorithm (see Section 3.4), we have calculated the predictive distribution of $p(h_t \mid \mathbf{y}_{1:t-1}, \hat{\boldsymbol{\theta}})$, for $t = 1, \ldots, T$, where $\hat{\boldsymbol{\theta}}$ is the mode of the posterior distribution (see Section 3.4 for details). Fig. 6 depicted the mean of $\{e^{\frac{h_t}{2}} \mid \mathbf{y}_{1:t-1}, \hat{\boldsymbol{\theta}}\}$ with the absolute returns to the SP&500 index. Note that the SV-S and SV-t models exhibit smoother movements than those from the SV-N and SV–VG models. Once again, difference in extreme returns is clearly manifested once the associated volatility values jump up more under the SV-N and SV–VG models than the SV-S and SV-t models.

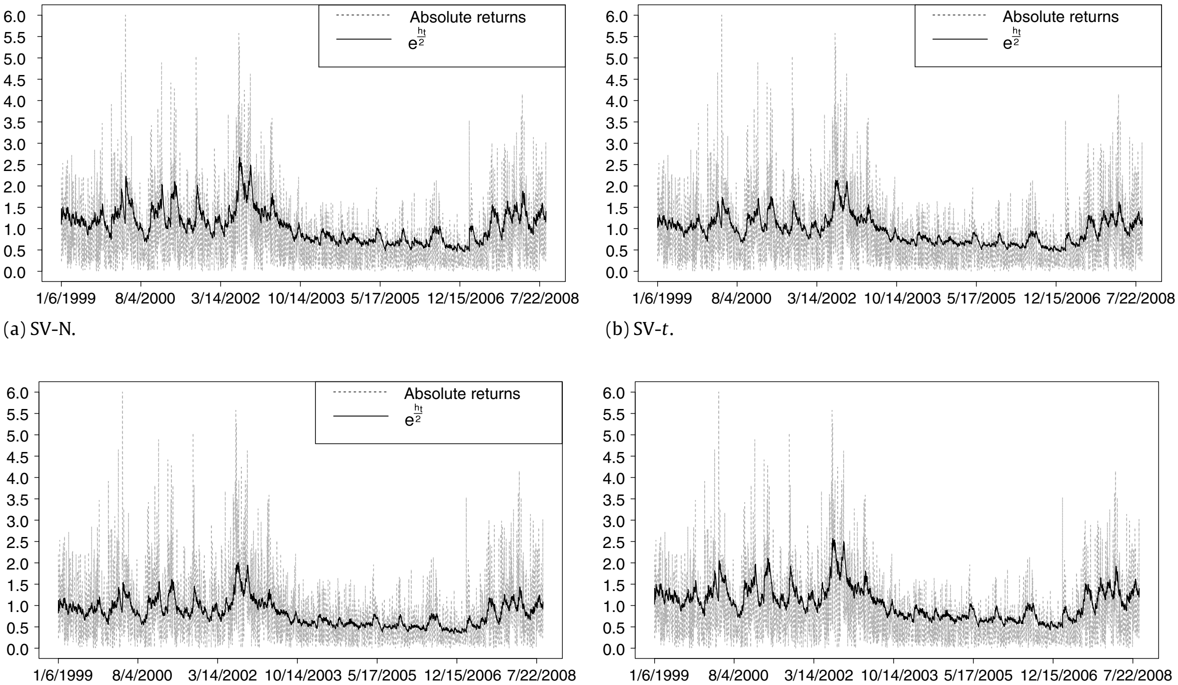
We evaluate the SV–SMN models by using the out-of-sample forecasting of the squared returns aggregated over certain period of time. Based on the 2432 observations of returns used previously, we calculate the forecast over the following $1, 2, \ldots, 10$ days as described in Section 3.5.

Fig. 7 plots the posterior means and 95% posterior credibility interval of the aggregated squared returns together with the realized values. The 95% posterior interval of the aggregated volatility, e^{h_t}, are also plotted. For all models (a)-(d), the 95% intervals of the aggregated squared returns are much wider than those for the aggregated volatility. The 95% posterior credibility interval of the aggregated squared returns for the SV-S model include the realized values for days from 1 to 7. The SV-t model shows similar forecasts except the day 6. The SV–VG only include realized values of the aggregated squared returns for days from 1 to 5. The SV-N shows the worst behavior, it includes only the realized values for days 1, 4 and 5.



(a) SV-N. (b) SV-t. (c) SV-S. (d) SV–VG.

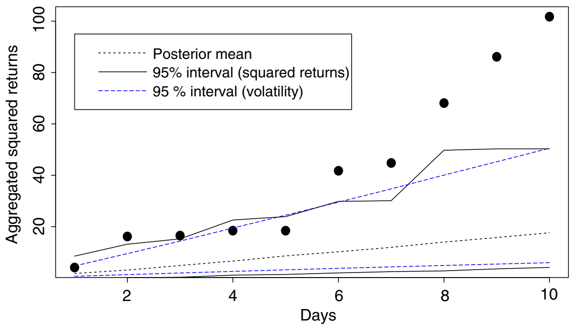
Fig. 5. Posterior smoothed mean (solid line) of $e^{\frac{h_t}{2}}$ for (a) SV-N, (b) SV-t, (c) SV-S and (d) SV–VG models. The dashed line indicates the absolute returns of the S&P500 index data.



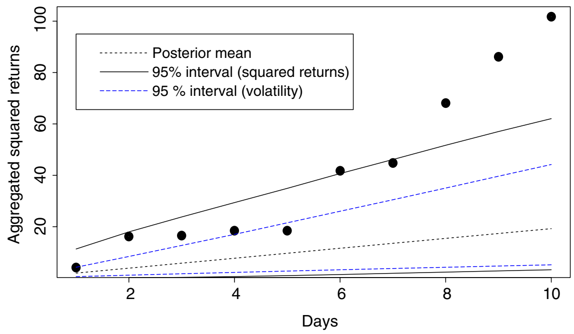
(a) SV-N. (b) SV-t. (c) SV-S. (d) SV–VG.

Fig. 6. Posterior mean (solid line) of $e^{\frac{h_t}{2}} \mid \hat{\boldsymbol{\theta}}, \mathbf{y}_{1:t-1}$ for (a) SV-N, (b) SV-t, (c) SV-S and (d) SV–VG models. The dashed line indicates the absolute returns of the S&P500 index data.

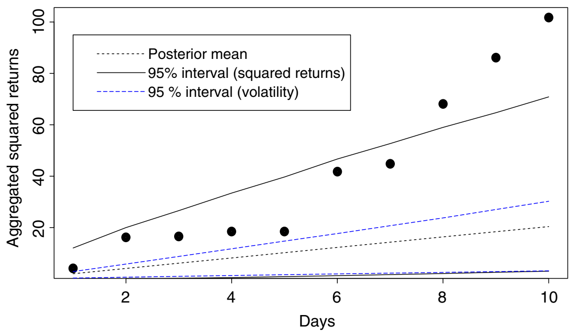
The robustness aspects of the SV–SMN models can be studied through the influence of outliers on the posterior distribution of the parameters. We consider only the SV-t and the SV-S models for illustrative purposes. We study the



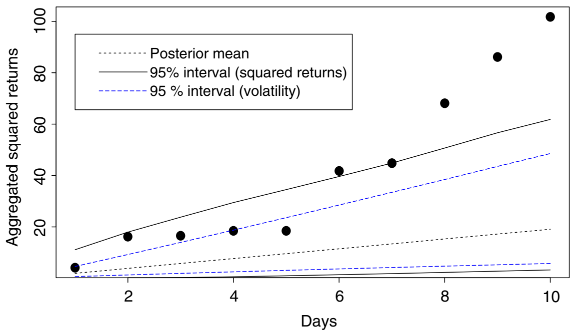
(a) SV-N.



(b) SV-t.



(c) SV-S.



(d) SV–VG.

Fig. 7. Out-of-sample forecast of the aggregated squared returns for (a) SV-N, (b) SV-t, (c) SV-S and (d) SV–VG models.

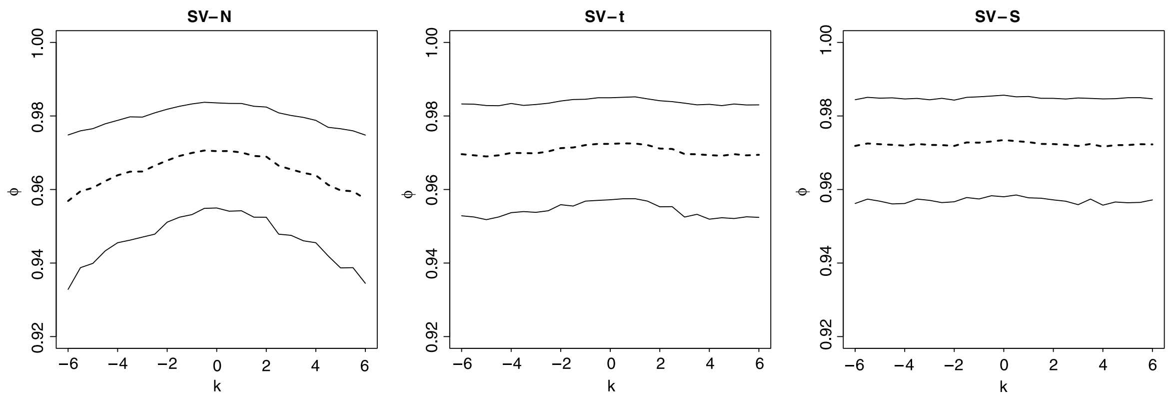


Fig. 8. Posterior mean (dashed line) and 95% credible interval (solid line) for ϕ of fitting the SV-N, SV-t and SV-S models for the S&P500 index data.

influence of three contaminated observations on the posterior estimates of mean and 95% credible interval of model parameters. The observations in $t = 1566$, 1582 and 1599, which corresponds to March 5, 2005, April 20, 2005 and May 16, 2005, respectively, are contaminated by ky_t, where k varies from -6 and 6 with increments of 0.5 units. In Figs. 8 and 9, we plot the posterior mean and 95% credible interval of ϕ and σ_η^2, respectively, for the SV-N, the SV-t and the SV-S models. Clearly, the SV-S and the SV-t models are less affected by variations of k than the SV-N model, meaning substantial robustness of the estimates over the usual normal process in the presence of outlying observations.

5. Conclusions

This article discusses a Bayesian implementation of some robust alternatives to stochastic volatility models via MCMC methods. The Gaussian assumption of the mean innovation was replaced by univariate thick-tailed processes, known as scale mixtures of normal distributions. We study three specific sub-classes, viz. the Student-t, the slash and the variance gamma distributions and compare parameter estimates and model fit with the default normal model. Under a Bayesian perspective, we constructed an algorithm based on Markov Chain Monte Carlo (MCMC) simulation methods to estimate all the parameters and latent quantities in our proposed SV–SMN class of models. As a by-product of the MCMC algorithm,

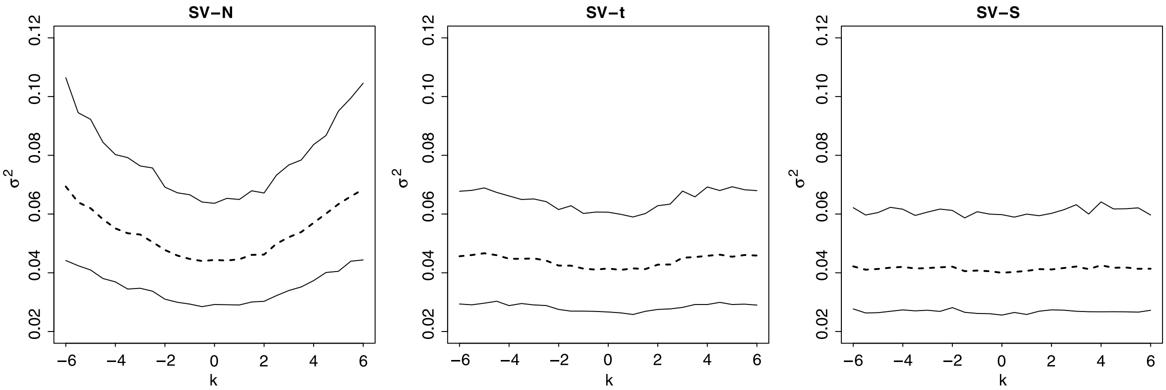


Fig. 9. Posterior mean (dashed line) and 95% credible interval (solid line) for σ^2 of fitting the SV-N, SV-t and SV-S models for the S&P500 index.

we were able to produce an estimate of the latent information process which can be used in financial modeling. The use of mixing variable, $\boldsymbol{\lambda}_{1:T}$ for normal scale mixture distributions not only simplifies the full conditional distributions required for the Gibbs sampling algorithm, but also provides a mean for outlier diagnostics. We illustrate our methods through an empirical application of the S&P500 index return series, which shows that the SV-S model provide better model fitting than the SV-N model in terms of parameter estimates, interpretation, robustness aspects and out-of-sample forecast of the aggregated squared returns.

In future, we plan to extend our research in several directions with the aim of exploring the robustness aspect of the parameter estimates. For instance, in this paper the estimated volatility of financial asset return changes does not accommodate sudden structural breaks. Recently, the SV model with jumps (Barndorff-Nielsen and Shephard, 2001; Chib et al., 2002) and the regime switching models (So et al., 1998; Shibata and Watanabe, 2005; Abanto-Valle et al., 2009) have received considerable attention. The volatility of daily stock index returns has been estimated with SV models but usually results have relied on extensive pre-modeling of these series, thus avoiding the problem of simultaneous estimation of the mean and variance. The SV in mean (SVM) (Koopman and Uspensky, 2002) model deals with this problem and incorporates the unobserved volatility as explanatory variable in the mean equation of the returns. Indeed, the flexibility of the SVM with SMN distributions could fit time-varying features in the mean of the returns and heavy tails simultaneously. The estimation of such intricate models is not straightforward since volatility now appears in both the mean and the variance equation. This requires modifications of the multi-move algorithm to sampling the log-volatilities. We plan to explore our methods along those lines. Furthermore, our SV–SMN models have shown considerable flexibility to accommodate outliers, however its robustness aspects could be seriously affected by the presence of skewness. Lachos et al. (2009) have recently proposed a remedy to incorporate skewness and heavy-tailedness simultaneously using scale mixtures of skew-normal (SMSN) distributions. We conjecture that the methodology presented in this paper can be undertaken under univariate and multivariate setting of SMSN distributions and should yield satisfactory results in situations where data exhibit non-normal behavior, although at the expense of additional complexity in its implementation. Nevertheless, a deeper investigation of those modifications is beyond the scope of the present paper, but provides stimulating topics for further research.

Acknowledgments

The authors would like to thank the Editor and two anonymous referees for their constructive comments which substantially improved the quality of this paper. The first author acknowledges financial support from the Fundação de Amparo à Pesquisa do Estado de Rio de Janeiro (FAPERJ) grants E-26/171.092/2006. The research of D. Bandyopadhyay was supported by grants P20 RR017696-06 from the United States National Institutes of Health. The research of V.H Lachos was supported in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Appendix. The full conditionals

In this appendix, we describe the full conditional distributions for the parameters and the mixing latent variables $\boldsymbol{\lambda}_{1:T}$ of the SV–SMN class of models.

A.1. Full conditional distribution of α, ϕ and σ_η^2

The prior distributions of the common parameters are set as: $\alpha \sim N(\bar{\alpha}, \sigma_\alpha^2)$, $\phi \sim \mathcal{N}_{(-1,1)}(\bar{\phi}, \sigma_\phi^2)$, $\sigma_\eta^2 \sim \mathcal{IG}(\frac{T_0}{2}, \frac{M_0}{2})$. Together with (15), we have the following full conditional for α:

$$p(\alpha \mid \mathbf{h}_{0:T}, \phi, \sigma_\eta^2) \propto \exp\left\{-\frac{a_\alpha}{2}\left(\alpha - \frac{b_\alpha}{a_\alpha}\right)^2\right\}, \tag{31}$$

which is the normal distribution with mean $\frac{b_\alpha}{a_\alpha}$ and variance $\frac{1}{a_\alpha}$, where $a_\alpha = \frac{1}{\sigma_\alpha^2} + \frac{T}{\sigma_\eta^2} + \frac{1+\phi}{\sigma_\eta^2(1-\phi)}$ and $b_\alpha = \frac{\bar{\alpha}}{\sigma_\alpha^2} + \frac{(1+\phi)}{\sigma_\eta^2}h_0 + \frac{\sum_{t=1}^{T}(h_t-\phi h_{t-1})}{\sigma_\eta^2}$. Similarly, by using (15), we have that the conditional posterior of ϕ is given by

$$p(\phi \mid \mathbf{h}_{0:T}, \alpha, \sigma_\eta^2) \propto Q(\phi)\exp\left\{-\frac{a_\phi}{2\sigma_\eta^2}\left(\phi - \frac{b_\phi}{a_\phi}\right)^2\right\}\mathbb{I}_{|\phi|<1} \tag{32}$$

where $Q_\phi = \sqrt{1-\phi^2}\exp\{-\frac{1-\phi^2}{2\sigma_\eta^2}(h_0 - \frac{\alpha}{1-\phi})^2\}$, $a_\phi = \sum_{t=1}^{T} h_{t-1}^2 + \frac{\sigma_\eta^2}{\sigma_\phi^2}$, $b_\phi = \sum_{t=1}^{T} h_{t-1}(h_t - \alpha) + \bar{\phi}\frac{\sigma_\eta^2}{\sigma_\phi^2}$ and $\mathbb{I}_{|\phi|<1}$ is an indicator variable. As $p(\phi \mid \mathbf{h}_{0:T}, \alpha, \sigma_\eta^2)$ in (32) does not have closed form, we sample from using the Metropolis–Hastings algorithm with truncated $\mathcal{N}_{(-1,1)}(\frac{b_\phi}{a_\phi}, \frac{\sigma_\eta^2}{a_\phi})$ as the proposal density.

From (15), the conditional posterior of σ_η^2 is $\mathcal{IG}(\frac{T_1}{2}, \frac{M_1}{2})$, where $T_1 = T_0 + T + 1$ and $M_1 = M_0 + [(1-\phi^2)(h_0 - \frac{\alpha}{1-\phi})^2] + \sum_{t=1}^{T}(h_t - \alpha - \phi h_{t-1})^2$.

A.2. Full conditional of λ_t and ν

- *SV-t case*

  As $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, the full conditional of λ_t is given by

$$p(\lambda_t \mid y_t, h_t, \nu) \propto \lambda_t^{\frac{\nu+1}{2}-1} e^{-\frac{\lambda_t}{2}(y_t^2 e^{-h_t}+\nu)}, \tag{33}$$

which is the gamma distribution, $\mathcal{G}(\frac{\nu+1}{2}, \frac{y_t^2 e^{-h_t}+\nu}{2})$.

We assume the prior distribution of ν as $\mathcal{G}(a_\nu, b_\nu)\mathbb{I}_{2<\nu\leq 40}$. Then, the full conditional of ν is

$$p(\nu \mid \boldsymbol{\lambda}_{1:T}) \propto \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} e^{-\frac{\nu}{2}\sum_{t=1}^{T}[(\lambda_t - \log\lambda_t)+2b_\nu]}}{[\Gamma(\frac{\nu}{2})]^T}\mathbb{I}_{2<\nu\leq 40}. \tag{34}$$

We sample ν by the Metropolis–Hastings acceptance–rejection algorithm (Tierney, 1994; Chib, 1995). Let ν^* denote the mode (or approximate mode) of $p(\nu \mid \boldsymbol{\lambda}_{1:T})$, and let $\ell(\nu) = \log p(\nu \mid \boldsymbol{\lambda}_{1:T})$. As $\ell(\nu)$ is concave, we use the proposal density $\mathcal{N}_{(2,40)}(\mu_\nu, \sigma_\nu^2)$, where $\mu_\nu = \nu^* - \ell'(\nu^*)/\ell''(\nu^*)$ and $\sigma_\nu^2 = -1/\ell''(\nu^*)$. $\ell'(\nu^*)$ and $\ell''(\nu^*)$ are the first and second derivatives of $\ell(\nu)$ evaluated at $\nu = \nu^*$. To prove the concavity of $\ell(\nu)$, we use the result of Abramowitz and Stegun (1970), in which the $\log\Gamma(\nu)$ could be approximated as

$$\log\Gamma(\nu) = \frac{\log(2\pi)}{2} + \frac{2\nu-1}{2}\log(\nu) - \nu + \frac{\theta}{12\nu}, \quad 0 < \theta < 1. \tag{35}$$

Taking the second derivative of $\ell(\nu)$ from (34) and using (35), we have that

$$\ell''(\nu) = -\frac{T\theta}{3\nu^3} - \frac{(T + 2a_\nu - 2)}{2\nu^2} < 0.$$

- *SV-S case*

  Using the fact that $\lambda_t \sim \mathcal{B}e(\nu, 1)$, we have the full conditional of λ_t given as

$$p(\lambda_t \mid y_t, h_t, \nu) \propto \lambda_t^{\nu+\frac{1}{2}-1} e^{-\frac{\lambda_t}{2}y_t^2 e^{-h_t}}\mathbb{I}_{0<\lambda_t<1}, \tag{36}$$

that is $\lambda_t \sim \mathcal{G}_{(0<\lambda_t<1)}(\nu + \frac{1}{2}, \frac{1}{2}y_t^2 e^{-h_t})$, the right truncated gamma distribution.
Assuming that a prior distribution of $\nu \sim \mathcal{G}(a_\nu, b_\nu)$, the full conditional distribution of ν is given by

$$p(\nu \mid \mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}) \propto \nu^{T+a_\nu-1} e^{-\nu(b_\nu - \sum_{t=1}^{T}\log\lambda_t)}\mathbb{I}_{\nu>1}. \tag{37}$$

Then, the full conditional of ν is $\mathcal{G}_{\nu>1}(T + a_\nu, b_\nu - \sum_{t=1}^{T}\log\lambda_t)$, i.e. the left truncated gamma distribution. We simulate from the right and left truncated gamma distributions using the algorithm proposed by Philippe (1997).

- *SV–VG case*

  As $\lambda_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$, the full conditional of λ_t is given by

$$p(\lambda_t \mid y_t, h_t, \nu) \propto \lambda_t^{-\frac{\nu}{2}+\frac{1}{2}-1} e^{-\frac{1}{2}(\lambda_t y_t^2 e^{-h_t} + \frac{\nu}{\lambda_t})}, \tag{38}$$

which is the generalized inverse gaussian distribution, $\mathcal{GIG}(-\frac{\nu}{2}+\frac{1}{2}, y_t^2 e^{-h_t}, \nu)$.

We assume the prior distribution of ν as $\mathcal{G}(a_\nu, b_\nu)\mathbb{I}_{0<\nu\leq 40}$. Then, the full conditional of ν is

$$p(\nu \mid \mathbf{y}_{1:T}, \mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}) \propto \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} \mathrm{e}^{-\frac{\nu}{2}\sum_{t=1}^{T}[(\frac{1}{\lambda_t}+\log\lambda_t)+2b_\nu]}}{[\Gamma(\frac{\nu}{2})]^T}\mathbb{I}_{0<\nu\leq 40} \tag{39}$$

which is log-concave. Thus, we sample ν by the Metropolis–Hastings acceptance–rejection algorithm as in the case of the SV-t model with proposal density $\mathcal{N}_{(0,40)}(\mu_\nu, \sigma^2_\nu)$.

References

Abanto-Valle, C.A., Migon, H.S., Lopes, H.F., 2009. Bayesian modeling of financial returns: A relationship between volatility and trading volume. Applied Stochastic Modeling in Business and Industry, in press (doi:10.1002/asmb.789).

Abramowitz, M., Stegun, N., 1970. Handbook of Mathematical Functions. Dover Publications, Inc, New York.

Ando, T., 2006. Bayesian inference for nonlinear and non-gaussian stochastic volatility model wit leverage effect. Journal of Japan Statistical Society 36, 173–197.

Ando, T., 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika 94, 443–458.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J., 1972. Robust Estimates of Location: Survey and Advances. Princeton University Press, Princeton, NJ.

Andrews, D.F., Mallows, S.L., 1974. Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B 36, 99–102.

Barndorff-Nielsen, O., Shephard, N., 2001. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253–280.

Berg, A., Meyer, R., Yu, J., 2004. Deviance Information criterion for comparing stochastic volatility models. Journal of Business and Economic Statistics 22, 107–120.

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

Carnero, M.A., Peña, D., Ruiz, E., 2004. Persistence and kurtosis in GARCH and stochastic volatility models. Journal of Financial Econometrics 2, 319–342.

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.

Celeux, G., Forbes, F., Robert, C.P., Titterington, D.M., 2006. Deviance information criteria for missing data models. Bayesian Analysis 1, 651–674.

Chen, C.W.S., Liu, F.C., So, M.K.P., 2008. Heavy-tailed-distributed threshold stochastic volatility models in financial time series. Australian & New Zealand Journal of Statistics 50, 29–51.

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

Chib, S., Nardari, F., Shepard, N., 2002. Markov Chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108, 281–316.

Chow, S.T.B., Chan, J.S.K., 2008. Scale mixtures distributions in statistical modelling. Australian & New Zealand Journal of Statistics 50, 135–146.

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

Fama, E., 1965. Portfolio analysis in a stable paretian market. Management Science 11, 404–419.

Fernández, C., Steel, M.F.J., 2000. Bayesian regression analysis with scale mixtures of normals. Econometric Theory 16, 80–101.

Frühwirth-Schnater, S., 1994. Data augmentation and dynamic linear models. Journal of Time Series Analysis 15, 183–202.

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

Gross, A.M., 1973. A Monte Carlo swindle for estimators of location. Journal of the Royal Statistical Society, Series C. Applied Statistics 22, 347–353.

Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–418.

Jacquier, E., Polson, N., Rossi, P., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122, 185–212.

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

Koopman, S.J., Uspensky, E.H., 2002. The stochastic volatility in mean model: Empirical evidence from international stock markets. Journal of Applied Econometrics 17, 667–689.

Lachos, V.H., Ghosh, P., Arellano-Valle, R.B., 2009. Likelihood based inference for skew-normal/independent linear mixed models. Statistica Sinica (in press).

Lange, K.L., Little, R., Taylor, J., 1989. Robust statistical modeling using t distribution. Journal of the American Statistical Association 84, 881–896.

Lange, K.L., Sinsheimer, J.S., 1993. Normal/independent distributions and their applications in robust regression. Journal of Computational and Graphical Statistics 2, 175–198.

Liesenfeld, R., Jung, R.C., 2000. Stochastic volatility models: Conditional normality versus heavy-tailed distributions. Journal of Applied Econometrics 15, 137–160.

Little, R., 1988. Robust estimation of the mean and covariance matrix from data with missing values. Applied Statistics 37, 23–38.

Madan, D., Seneta, E., 1990. The variance gamma (v.g) model for share market return. Journal Business 63, 511–524.

Mahieu, R., Schotman, P.C., 1998. Am empirical application of stochastic volatility models. Journal of Applied Econometrics 13, 333–360.

Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of Business 36, 314–419.

Melino, A., Turnbull, S.M., 1990. Pricing foreign options with stochastic volatility. Journal of Econometrics 45, 239–265.

Morgenthaler, S., Tukey, J., 1991. Configural Polysampling: A Route to Practical Robustness. Wiley, New York.

Nakajima, J., Omori, Y., 2008. Leverage, heavy-tails and correlated jumps in stochastic volatility models. Computational Statistics & Data Analysis. doi:10.1016/j.csda.2008.03.015.

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

Omori, Y., Watanabe, T., 2008. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. Computational Statistics & Data Analysis 52, 2892–2910.

Pemstein, D., Quinn, K.V., Martin, A.D., 2007. The scythe statistical library: An open source C++ library for statistical computation. Journal of Statistical Software V, 1–29.

Philippe, A., 1997. Simulation of right and left truncated gamma distributions by mixtures. Statistics and Computing 7, 173–181.

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

Robert, C.P., Titterington, D.M., 2002. Discussion on "Bayesian measures of model complexity and fit". Biometrical Journal 64, 573–590.

Rosa, G.J.M., Padovani, C.R., Gianola, D., 2003. Robust linear mixed models with Normal/Independent distributions and bayesian MCMC implementation. Biometrical Journal 45, 573–590.

Shephard, N., Pitt, M., 1997. Likelihood analysis of non-Gaussian measurements time series. Biometrika 84, 653–667.

Shibata, M., Watanabe, T., 2005. Bayesian analysis of a Markov switching stochastic volatility model. Journal of the Japan Statistical Society 35, 205–219.

So, M., Lam, K., Li, W., 1998. A stochastic volatility model with Markov Switching. Journal of Business and Economic Statistics 15, 183–202.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64, 621–622.

Stone, M., 2002. Discussion on "Bayesian measures of model complexity and fit". Journal of the Royal Statistical Society, Series B 64, 621.

Tauchen, G.E., Pitts, M., 1983. The price variability-volume relationships in speculative markets. Econometrica 51, 485–506.
Taylor, S., 1982. Financial returns modelled by the product of two stochastic processes—a study of the daily sugar prices 1961–75. In: Anderson, O. (Ed.), Time Series Analysis: Theory and Practice, vol. 1. pp. 203–226.
Taylor, S., 1986. Modeling Financial Time Series. Wiley, Chichester.
Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Annal of Statistics 21, 1701–1762.
Wang, J., Genton, M., 2006. The multivariate skew-slash distribution. Journal of Statistical Planning and Inference 136, 209–220.
Watanabe, T., Omori, Y., 2004. A multi-move sampler for estimate non-Gaussian time series model: Comments on Shepard and Pitt (1997). Biometrika 91, 246–248.
Yu, J., 2005. On leverage in stochastic volatility model. Journal of Econometrics 127, 165–178.