Methodol Comput Appl Probab (2015) 17:721–738
DOI 10.1007/s11009-013-9389-9

Bayesian Estimation of a Skew-Student-t Stochastic Volatility Model

C. A. Abanto-Valle · V. H. Lachos · Dipak K. Dey

Received: 30 June 2012 / Revised: 17 August 2013 /
Accepted: 24 October 2013 / Published online: 21 November 2013

Abstract In this paper we present a stochastic volatility (SV) model assuming that the return shock has a skew-Student-t distribution. This allows a parsimonious, flexible treatment of skewness and heavy tails in the conditional distribution of returns. An efficient Markov chain Monte Carlo (MCMC) algorithm is developed and used for parameter estimation and forecasting. The MCMC method exploits a skew-normal mixture representation of the error distribution with a gamma distribution as the mixing distribution. The developed methodology is applied to the NASDAQ daily index returns. Bayesian model selection criteria as well as out-of-sample forecasting in a value-at-risk (VaR) study reveal that the SV model based on skew-Student-t distribution provides significant improvement in model fit as well as prediction to the NASDAQ index data over the usual normal model.

Keywords Markov chain Monte Carlo · Non-Gaussian and nonlinear state space models · Skew-Student-t · Stochastic volatility · Value-at-risk

AMS 2000 Subject Classifications 62 F 15 · 60 J 25 · 34 A 34

C. A. Abanto–Valle
Department of Statistics, Federal University of Rio de Janeiro,
CP 68530, Rio de Janeiro, CEP 21945-970, RJ, Brazil
e-mail: cabantovalle@im.ufrj.br

V. H. Lachos
Department of Statistics, Campinas State University,
CP 6065, Campinas, CEP 13083-859, SP, Brazil
e-mail: hlachos@ime.unicamp.br

D. K. Dey (✉)
Department of Statistics, University of Connecticut,
Storrs, CT, USA
e-mail: dipak.dey@uconn.edu

1 Introduction

A large literature in financial econometrics has documented stylized facts which are frequently found in stock and foreign exchange returns: skewness, heavy-tailedness and volatility clustering. These properties are crucial not only for describing the return distributions but also for asset allocation, option pricing, forecasting and risk management.

Stochastic volatility (SV) models were introduced in the financial literature to describe time-varying volatilities (Taylor 1982, 1986). An attractive feature of the SV model is its close relationship to financial economic theories (Melino and Turnbull 1990). However, many empirical studies have pointed out that asset returns data have heavier tails than those of normal distribution (see, e.g., Liesenfeld and Jung 2000; Chib et al. 2002; Jacquier et al. 2004; Abanto-Valle et al. 2010, among others). Then, 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.

The empirical evidence on skewness in the distribution of financial returns is well documented in the literature (Harvey and Sidique 1999, 2000; Mittnik and Paolella 2000; Jondeau and Rockinger 2003; Chen et al. 2012). Hansen (1994) has considered skewness in a GARCH model using skew-Student-t distribution errors allowing for both skewness and heavy tails to co-exist in a time-varying volatility setup. Also, Cappuccio et al. (2004) and Cappuccio et al. (2006) used the skew-GED distribtuion to model skewness and heavy tails in the conditional distribution of returns. More recently, Tsiotas (2012) used the skew-Student-t distribution in SV models, but in his MCMC algorithm, the log-volatilities are drawn using an inefficient single-move algorithm.

In this paper, in order to model simultaneously skewness and heavy-tailedness, we extend the SV model by assuming skew-Student-t (ST) introduced by Branco and Dey (2001) and Azzalini and Capitanio (2003) and hence the SV-ST is defined. Inference in the SV-ST model 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 by using the block sampler algorithm (Shephard and Pitt 1997; Watanabe and Omori 2004; Abanto-Valle et al. 2010, 2011, 2012; Nakajima and Omori 2012). The 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). For the skewness and degrees of freedom parameters of the skew-Student-t distribution, we assume Jeffreys's priors based on Bayes and Branco (2007) and Fonseca et al. (2008). At the model selection stage, we compare in-sample fit using the BPIC criterion (Ando 2007) and out-of-sample forecast in a VaR study.

The remainder of this paper is organized as follows. Section 2 shows a brief review of the skew-normal (Azzalini 1986) and skew-Student-t distribution, including some of its properties. Section 3 describes the SV-ST model through Bayesian estimation procedure using MCMC methods. MCMC output is used to forecast value-at-risk (VaR) thresholds. Section 4 is devoted to application and model comparison among the SV-ST model against the SV-N, SV-T and SV-SN models using the NASDAQ data set. Finally, some concluding remarks as well as future developments are deferred to Section 5.

2 The Univariate Skew-normal and Skew-student-t Distributions

We start by giving an important notation that will be used throughout the paper and present a review of the univariate skew normal (SN) and skew-Student-t (ST) distributions and a study of some related properties of those distributions.

A univariate random variable X is said to follow a skew-normal distribution, $X \sim \mathcal{SN}(\zeta, \omega^2, \lambda)$, with location, scale and asymmetry parameters given by ζ, ω^2 and λ, respectively, if the density of this distribution has the form

$$p\left(x \mid \zeta, \omega^2, \lambda\right) = \frac{2}{\omega}\phi\left(\frac{x-\zeta}{\omega}\right)\Phi\left(\frac{\lambda}{\omega}(x-\zeta)\right), \tag{1}$$

where $\phi(.)$ and $\Phi(.)$ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of the standard normal distribution. When $\lambda = 0$, the density in Eq. (1) becomes $\mathcal{N}\left(\zeta, \sigma^2\right)$ (see, Azzalini 2005, for a comprehensive review). In the next sections, we use the following stochastic representation of the SN distribution (Azzalini 1986; Henze 1986). Let $W \sim \mathcal{N}_{[0,\infty)}(0, 1)$ and $\varepsilon \sim \mathcal{N}(0, 1)$, independently, and let $\delta \in (-1, 1)$, where $\mathcal{N}_{[0,\infty)}(., .)$ and $\mathcal{N}(., .)$ indicate the truncated normal and normal distribution, respectively. The random variable X, defined by

$$X = \zeta + \omega\delta W + \omega\sqrt{1-\delta^2}\varepsilon, \tag{2}$$

follows a univariate skew-normal distributions, that is, $X \sim \mathcal{SN}\left(\zeta, \omega^2, \lambda\right)$, where $\lambda = \delta/\sqrt{1-\delta^2}$.

The kurtosis coefficient of a skew-normal distribution is restricted to the interval [3, 3.8692]. To achieve a higher degree of excess kurtosis, the skew-Student-t distribution has been introduced by Branco and Dey (2001) and Azzalini and Capitanio (2003). A univariate random variable X follows the scalar skew-Student-t distribution, $X \sim \mathcal{ST}(\zeta, \omega^2, \lambda, \nu)$, if it has the following stochastic representation

$$X = \zeta + U^{-1/2}\omega\delta W + U^{-\frac{1}{2}}\omega\left(1-\delta^2\right)^{\frac{1}{2}}\varepsilon, \tag{3}$$

where $W \sim \mathcal{N}_{[0,\infty)}(0, 1)$, $\varepsilon \sim \mathcal{N}(0, 1)$ and $U \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ are independently distributed. The Gamma distribution $\mathcal{G}(a, b)$ is defined with density $p(u \mid a, b) = b^a u^{a-1}\mathrm{e}^{-bu}/\Gamma(a)$. The pdf of X is then given by

$$f(X \mid \zeta, \omega^2, \lambda, \nu) = \frac{2}{\omega}t_\nu\left(\frac{x-\zeta}{\omega}\right)T_{\nu+1}\left(\lambda\omega^{-1}(x-\zeta)\sqrt{\frac{\nu+1}{\nu+\omega^{-2}(x-\zeta)^2}}\right), \tag{4}$$

where $t_\nu(.)$ and $T_\nu(.)$ denote the pdf and cdf of a standard Student-t distribution with ν degrees of freedom. From (3), we have that

$$E(X) = \zeta + \sqrt{\frac{2}{\pi}}k_1\omega\delta, \tag{5}$$

$$V(X) = \omega^2 k_2 - \frac{2}{\pi}k_1^2\omega^2\delta^2, \tag{6}$$

where $\delta = \lambda/\sqrt{1+\lambda^2}$ and $k_m = E(U^{-m/2})$. $E(.)$ and $V(.)$ denote the expected value and variance, respectively. The skew-Student-t nests the traditional symmetric Student's t distribution as a special case when $\lambda = 0$, and the normal distribution as $\nu \to \infty$, and can capture left-tailed or negative skewness when $\lambda < 0$, and positive skewness when $\lambda > 0$.

In order to interpret the parameters (λ, ν) in relation to skewness and heavy-tailedness, some skew-Student-t densities are plotted in Fig. 1, considering several combinations of the

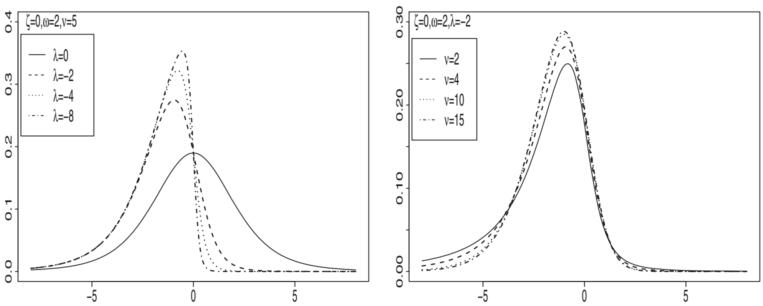

Fig. 1 The skew-Student-t distribution. *Left*: $\zeta = 0, \omega = 2, \nu = 5$ (fixed), $\lambda = 0, -2, -4, -8$. *Right*: $\zeta = 0, \omega = 2, \lambda = -2$ (fixed), $\nu = 2, 4, 10$ and 15

parameter values λ and ν, with ζ and ω held fixed at 0 and 2, respectively . In Fig. 1 (left panel) the densities are drawn using $\lambda = 0, -2, -4, -8$ with ν fixed at 5. As mentioned, $\lambda = 0$ corresponds to a symmetric Student's t-density. We can see that a lower value of λ implies a more negative skewness or left-skewness, as well as, heavier tails. Figure 1 (right panel) shows the densities for ν at 2, 4, 10 and 15 with λ held fixed at -2. We can see that as ν becomes larger, the density becomes less skewed and has lighter tails. Hence, the skewness and heavy-tailedness of the distribution are jointly determined by the combination of values of the parameter λ and ν.

3 The Skew-Student-t Stochastic Volatility Model

3.1 The Model

In order to account for both the excess kurtosis and skewness in stock returns, we introduce the stochastic volatility model with skew-Student-t errors (SV-ST), which is defined as

$$y_t = \mathrm{e}^{\frac{h_t}{2}} \epsilon_t, \tag{7a}$$

$$h_{t+1} = \mu + \varphi(h_t - \mu) + \sigma_\eta \eta_t, \tag{7b}$$

where y_t and h_t are, respectively, the compounded return and the log-volatility at time t. We assume that $|\varphi| < 1$, i.e., the log-volatility process is stationary and that the initial value $h_1 \sim \mathcal{N}\left(\mu, \frac{\sigma_\eta^2}{1-\varphi^2}\right)$, $\epsilon_t \sim \mathcal{ST}(\zeta, \omega^2, \lambda, \nu)$ and $\eta_t \sim \mathcal{N}(0, 1)$ are uncorrelated. We set ζ and ω in such a way that $E(\epsilon_t = 0)$ and $V(\epsilon_t) = 1$, because they imply the martingale return series hypothesis. The SV-ST defined by Eqs. (7a) and (7b) can be written hierarchically using the stochastic representation of the skew-Student-t distribution in (3), as

$$y_t = \left(\zeta + \omega\delta W_t U_t^{-\frac{1}{2}}\right) \mathrm{e}^{\frac{h_t}{2}} + \mathrm{e}^{\frac{h_t}{2}} U_t^{-\frac{1}{2}} \omega (1-\delta^2)^{\frac{1}{2}} \varepsilon_t, \tag{8a}$$

$$h_{t+1} = \mu + \varphi(h_t - \mu) + \sigma_\eta \eta_t, \tag{8b}$$

$$W_t \sim \mathcal{N}_{[0,\infty)}(0, 1), \tag{8c}$$

$$U_t|\nu \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \tag{8d}$$

where ε_t and η_t are mutually independent and normally distributed with zero mean and unit variance, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. In this setup, Eqs. (8a) and (8b), with $\lambda = 0$ (equivalently $\delta = 0$) and $U_t = 1, \forall t = 1, \ldots, T$, define the SV model with normal distribution (SV-N). Equations (8a),(8b) and (8d) with $\lambda = 0$ define the the SV model with Student-t distribution (SV-T). Finally, Eqs. (8a),(8b) and (8c) with $U_t = 1, \forall t = 1, \ldots, T$, results the SV model with skew normal distribution (SV-SN).

3.2 Parameter Estimation via MCMC

Let $\boldsymbol{\theta} = \left(\mu, \varphi, \sigma^2_\eta, \nu, \lambda\right)'$ be the full parameter vector of the entire class of SV-ST model, $\mathbf{h}_{1:T} = (h_1, \ldots, h_T)'$ be the vector of the log volatilities, $\mathbf{U}_{1:T} = (U_1, \ldots, U_T)'$ be the mixing variables, $\mathbf{W}_{1:T} = (W_1, \ldots, W_T)'$ and $\mathbf{y}_{1:T} = (y_1, \ldots, y_T)'$ be the information available up to time T, while ν is the degrees of freedom parameter vector associated with the mixture distribution and λ the skewness parameter. The Bayesian approach to estimate the parameters in the SV-ST model uses the data augmentation principle, which considers $\mathbf{h}_{1:T}$, $\mathbf{W}_{1:T}$ and $\mathbf{U}_{1:T}$ as latent variables. The joint posterior density of parameters and latent unobservable variables can be written as

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{W}_{1:T}, \mathbf{U}_{1:T}, \mathbf{h}_{1:T} \mid \mathbf{y}_{1:T}) \propto\ & p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}, \mathbf{W}_{1:T}, \mathbf{U}_{1:T}, \mathbf{h}_{1:T}) \\ & \times\ p(\mathbf{h}_{1:T} \mid \boldsymbol{\theta}) p(\mathbf{W}_{1:T}) p(\mathbf{U}_{1:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \end{aligned} \tag{9}$$

where $p(\boldsymbol{\theta})$ is the prior distribution. Since the posterior density $p(\boldsymbol{\theta}, \mathbf{W}_{1:T}, \mathbf{U}_{1:T}, \mathbf{h}_{1:T} \mid \mathbf{y}_{1:T})$ does not have closed form, we first sample the parameters $\boldsymbol{\theta}$, followed by the latent variables $\mathbf{W}_{1:T}$, $\mathbf{U}_{1:T}$ and $\mathbf{h}_{1:T}$ using Gibbs sampling. The sampling scheme is described by Algorithm 1. Sampling the log-volatilities $\mathbf{h}_{1:T}$ in step 5 of Algorithm 1 is the most difficult task due to the nonlinear setup in the observational Eq. (8a). In order to avoid the higher correlations due to the Markovian structure of the h_t's, in the next subsection we develop a multi-move block sampler to sample $\mathbf{h}_{0:T}$ by blocks (Shephard and Pitt 1997; Watanabe and Omori 2004; Abanto-Valle et al. 2010, 2011, 2012; Nakajima and Omori 2012). Details on the full conditionals of $\boldsymbol{\theta}$ and the latent variables $\mathbf{U}_{1:T}$ and $\mathbf{W}_{1:T}$ are given in Appendix.

Algorithm 1

1. Set $i = 0$ and get starting values for the parameters $\boldsymbol{\theta}^{(i)}$ and the latent quantities $\mathbf{W}^{(i)}_{1:T}$, $\mathbf{U}^{(i)}_{1:T}$ and $\mathbf{h}^{(i)}_{1:T}$.
2. Generate $\boldsymbol{\theta}^{(i+1)}$ in turn from its full conditional distribution, given $\mathbf{y}_{1:T}$, $\mathbf{h}^{(i)}_{1:T}$,$\mathbf{W}^{(i)}_{1:T}$ and $\mathbf{U}^{(i)}_{1:T}$.
3. Draw $\mathbf{W}^{(i+1)}_{1:T} \sim p(\mathbf{W}_{1:T} \mid \boldsymbol{\theta}^{(i)}, \mathbf{U}^{(i)}_{1:T}, \mathbf{h}^{(i)}_{1:T}, \mathbf{y}_{1:T})$.
4. Draw $\mathbf{U}^{(i+1)}_{1:T} \sim p(\mathbf{U}_{1:T} \mid \boldsymbol{\theta}^{(i+1)}, \mathbf{W}^{(i+1)}_{1:T}, \mathbf{h}^{(i)}_{1:T}, \mathbf{y}_{1:T})$.
5. Generate $\mathbf{h}_{1:T}$ by blocks as:
   i) For $l = 1, \ldots, K$, the knot positions are generated as k_l, the floor of $[T \times \{(l + u_l)/(K + 2)\}]$, where the $u_l's$ are independent realizations of the uniform random variable on the interval (0,1).
   ii) For $l = 1, \ldots, K$, generate the block $h_{k_{l-1}+1:k_l-1}$ jointly conditional on $\mathbf{y}_{k_{l-1}:k_l-1}$, $\boldsymbol{\theta}^{(i+1)}$, $\mathbf{W}^{(i+1)}_{k_{l-1}+1:k_l-1}$, $\mathbf{U}^{(i+1)}_{k_{l-1}+1:k_l-1}$, $h^{(i)}_{k_{l-1}}$ and $h^{(i)}_{k_l}$.
   iii) For $l = 1, \ldots, K$, draw $h^{(i+1)}_{k_l}$ conditional on $\mathbf{y}_{1:T}$, $\boldsymbol{\theta}^{(i)}$, $W^{(i+1)}_{k_l}$, $U^{(i+1)}_{k_l}$, $h^{(i+1)}_{k_l-1}$ and $h^{(i+1)}_{k_l+1}$.
6. Set $i = i + 1$ and return to 2 until convergence is achieved.

In the SV-ST model considered so far, an important modelling assumption is the regularization penalty $p(\nu)$ on the tail thickness. A default Jeffreys' prior was developed by

Fonseca et al. (2008), with a number of desirable properties particularly when learning a fat-tail from a finite dataset. The default Jeffreys's prior for ν takes the form

$$p(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{\frac{1}{2}}, \tag{10}$$

where $\psi'(a) = \frac{d\{\psi(a)\}}{da}$ and $\psi(a) = \frac{d\{\log\Gamma(a)\}}{da}$ are the trigamma and digamma functions, respectively. The interesting feature of this prior is its behavior as ν goes to infinity and it has polynomial tails of the form $p(\nu) \propto \nu^{-4}$. In this case, the tail of the prior decays rather fast for large values of ν and assessing the degree of tail thickness can require prohibitively large samples. To ensure the existence of the variance of U_t in Eq. (8d), we consider $p(\nu)$, such that $\nu > 2$. To the skewness parameter, we assume that $\lambda \sim t_{0.5}(0.0, \frac{\pi^2}{4})$, a Jeffreys' prior suggested by Bayes and Branco (2007), where $t_c(a, b)$ denotes the Student-t distribution with location a, scale b and c degrees of freedom.

3.2.1 Block Sampler

In order to simulate $\mathbf{h}_{1:T} = (h_1, \ldots, h_T)'$ in the SV-ST model, we consider a two-step process: first, we simulate h_1 conditional on $\mathbf{h}_{2:T}$, next $\mathbf{h}_{2:T}$ conditional on h_1. To sample the vector $\mathbf{h}_{2:T}$, we develop a multi-move block algorithm. In our block sampler, we divide it into $K+1$ blocks, $\mathbf{h}_{k_{l-1}+1:k_l-1} = (h_{k_{l-1}+1}, \ldots, h_{k_l-1})'$ for $l = 1, \ldots, K+1$, with $k_0 = 1$ and $k_{K+1} = T$, where $k_l - 1 - k_{l-1} \geq 2$ is the size of the l−th block. We sample the block of disturbances $\boldsymbol{\eta}_{k_{l-1}:k_l-2} = (\eta_{k_{l-1}}, \ldots, \eta_{k_l-2})'$ given the end conditions $h_{k_{l-1}}$ and h_{k_l} instead of $\mathbf{h}_{k_{l-1}+1:k_l-1}$. In order to facilitate the exposition, we omit the dependence on $\boldsymbol{\theta}$, $\mathbf{W}_{t+1:t+k}$ and $\mathbf{U}_{t+1:t+k}$, and suppose that $k_{l-1} = t$ and $k_l = t+k+1$ for the l−th block, such that $t+k < T$. Then $\boldsymbol{\eta}_{t:t+k-1} = (\eta_t, \ldots, \eta_{t+k-1})'$ are sampled at once from their full conditional distribution $f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \mathbf{y}_{t:t+k})$, which without the constant terms is expressed in log scale as

$$\begin{aligned}\log f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}) &= \text{const} - \frac{1}{2}\sum_{r=t}^{t+k-1} \eta_r^2 + \sum_{r=t+1}^{t+k} l(h_r) \\ &\quad - \frac{1}{2\sigma_\eta^2}[h_{t+k+1} - \mu - \varphi(h_{t+k} - \mu)]^2 \mathbb{I}(t+k < T),\end{aligned}$$

where $\mathbb{I}(.)$ is an indicator function. We denote the first and second derivatives of $l(h_r)$ with respect to h_r by l' and l'', where $l(h_r) = \log p(y_r \mid \nu, \lambda, W_r, U_r, h_r)$ is obtained from Eq. (8a). As (11) does not have closed form, we use the Metropolis-Hastings acceptance-rejection algorithm (Tierney 1994; Chib and Greenberg 1995) to sample from. We propose to use the following artificial Gaussian state space model as a proposed density to simulate the block $\boldsymbol{\eta}_{t+1:t+k}$

$$\hat{y}_r = h_r + \xi_r, \qquad \xi_r \sim \mathcal{N}(0, d_r), \quad r = t+1, \ldots, t+k, \tag{11}$$

$$h_{r+1} = \mu + \varphi(h_r - \mu) + \sigma_\eta \eta_r, \qquad \eta_r \sim \mathcal{N}(0, 1), \quad r = t, t+1, \ldots, t+k-1, \tag{12}$$

where the auxiliary variables d_r and $\hat{y}_r$ for $r = t+1, \ldots, t+k-1$ and $t+k = T$ are defined as follows:

$$\begin{aligned} d_r &= -\frac{1}{l''_F(\hat{h}_r)}, \\ \hat{y}_r &= \hat{h}_r + d_r l'(\hat{h}_r). \end{aligned} \tag{13}$$

For $r = t + k < T$, it follows that

$$
\begin{aligned}
d_r &= \frac{\sigma_\eta^2}{\varphi^2 - \sigma_\eta^2 l''_F(\hat{h}_{t+k})}, \\
\hat{y}_r &= d_r \left[l'(\hat{h}_r) - l''_F(\hat{h}_r)\hat{h}_r + \frac{\varphi}{\sigma_\eta^2}[h_{r+1} - \mu(1-\varphi)] \right].
\end{aligned} \tag{14}
$$

We obtain the measurement Eq. (11) by a second-order expansion of l_r around some preliminary estimate of η_r, denoted by $\hat{\eta}_r$, where $\hat{h}_r$ is the estimate of h_r equivalent to $\hat{\eta}_r$, and

$$
l''_F(h_r) = E[l''(h_r)] = -\frac{1}{2} - \frac{\left(\zeta + \omega\delta W_t U_t^{-\frac{1}{2}}\right)^2}{4\omega^2(1-\delta^2)} U_r, \tag{15}
$$

which is everywhere strictly negative. The expectation in (15) is taken with respect to y_r conditional on h_r, W_r, U_r, $\boldsymbol{\theta}$. Since (11)–(12) define a Gaussian state space model, we can apply de Jong and Shephard's simulation smoother (de Jong and Shephard 1995) to perform the sampling. We denote this density by g. Since f is not bounded by g, we use the Metropolis-Hastings acceptance-rejection algorithm to sample from f, as recommended by Chib and Greenberg (1995). In the SV-SN case, we use the same procedure with $U_t = 1$ for $t = 1, \ldots, T$.

The procedure to select the expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ is described in the Algorithm 2.

Algorithm 2

1. Initialize $\hat{\mathbf{h}}_{t+1:t+k}$.
2. Evaluate recursively $l'(\hat{h}_r)$ and $l''_F(\hat{h}_r)$ for $r = t+1, \ldots, t+k$.
3. Conditional on the current values of the vector of parameters $\boldsymbol{\theta}$, $\mathbf{U}_{t+1:t+k}$, $\mathbf{W}_{t+1:t+k}$, h_t and h_{t+k+1}, define the auxiliary variables $\hat{y}_r$ and d_r using Eqs. (13) or (14) for $r = t+1, \ldots, t+k$.
4. Consider the linear Gaussian state-space model in (11) and (12). Apply the Kalman filter and a disturbance smoother (Koopman 1993) and obtain the posterior mean of $\boldsymbol{\eta}_{t:t+k}$ ($\mathbf{h}_{t:t+k}$) and set $\hat{\boldsymbol{\eta}}_{t:t+k}$ ($\hat{\mathbf{h}}_{t:t+k}$) to this value.
5. Return to step 2 and repeat the procedure until achieving convergence.

Finally, we describe the updating procedure for h_1 and the knot conditions h_{k_l}, for $l = 1, \ldots, K$. First, we simulate h_1 from $p(h_1 \mid h_2, \boldsymbol{\theta}, \mathbf{y}_{1:T})$ by using the Metropolis-Hasting (MH) algorithm with the normal density, $\mathcal{N}\left(\mu + \varphi[h_2 - \mu], \sigma_\eta^2\right)$, as a proposal. Then, the acceptance probability is given by $\alpha_{MH} = \min\left\{1, \frac{Q(h_1^p)}{Q\left(h_1^{(i-1)}\right)}\right\}$, where $Q(h_1)$ is the conditional density of $y_1 \mid \boldsymbol{\theta}, W_1, U_1, h_1$. Let h_1^p and $h_1^{(i-1)}$ denote the proposal and the previous iteration values. As the density $p\left(h_{k_l} \mid h_{k_l-1}, h_{k_l+1}\right)$ does not have a closed form, we use the MH algorithm with proposal $\mathcal{N}\left(\frac{\mu(1-\varphi)^2 + \varphi(h_{k_l-1} + h_{k_l+1})}{1+\varphi^2}, \frac{\sigma_\eta^2}{1+\varphi^2}\right)$. As before, $h_{k_l}^p$ and $h_{k_l}^{(i-1)}$ denote the proposal and the previous iteration values, respectively. Thus, the acceptance probability is given by $\alpha_{MH} = \min\left\{1, \frac{Q\left(h_{k_l}^{p}\right)}{Q\left(h_{k_l}^{(i-1)}\right)}\right\}$, where $Q(h_{k_l})$ is the conditional density of $y_{k_l} \mid \boldsymbol{\theta}, W_{k_l}, U_{k_l}, h_{k_l}$.

3.3 Forecasting Returns, Volatility and Value-at-Risk

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

$$p(y_{T+K} \mid \mathbf{y}_{1:T}) = \int \Big[p(y_{T+K} \mid U_{T+K}, W_{T+K}, h_{T+K}) p(W_{T+K} \mid \boldsymbol{\theta}) p(U_{T+K} \mid \boldsymbol{\theta}) \times p(h_{T+K} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \Big] dh_{T+K} dW_{T+K} dU_{T+K} d\boldsymbol{\theta},$$

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

Numerical evaluation of the last integrals is straightforward. To initialize the recursion, we use $h_T^{(i)}$ and $\boldsymbol{\theta}^{(i)}$, for $i = 1, \ldots, N$, from the MCMC output. Given these N draws, sample $h_{T+k}^{(i)}$ from $p\left(h_{T+k} \mid \boldsymbol{\theta}^{(i)}, h_{T+k-1}^{(i)}\right)$, $W_{T+k}^{(i)}$ from $p\left(W_{T+k} \mid \boldsymbol{\theta}^{(i)}\right)$ and $U_{T+k}^{(i)}$ from $p\left(U_{T+k} \mid \boldsymbol{\theta}^{(i)}\right)$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$, by using Eqs. (8b), (8c) and (8d), respectively. Finally, using Eq. (8a), we sample $y_{T+k}^{(i)}$ from $p\left(y_{T+k} \mid \boldsymbol{\theta}^{(i)}, W_{T+k}^{(i)}, U_{T+k}^{(i)}, h_{T+k}^{(i)}\right)$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$.

In order to emphasize applications in risk management, we compute the Value-at-Risk (VaR) to measure the risk of an investment position. VaR summarizes the expected maximum loss over a target horizon within a given confidence level α. Specifically, we can define the VaR over a one-step horizon, with probability α, as

$$\alpha = P(y_{T+1} < -\mathrm{VaR}_{T+1}). \tag{16}$$

Then, the quantile VaR is given by

$$\mathrm{VaR}_{T+1}^{(i)} = -\left[D^{-1}\left(\boldsymbol{\theta}^{(i)}\right) \mathrm{e}^{\frac{1}{2} h_{T+1}^{(i)}} \right], \tag{17}$$

where D^{-1} is inverse CDF for the distribution D, so this is the standardised quantile of D. Then, the final one-step-ahead VaR is the Monte Carlo posterior mean estimate, given by:

$$\mathrm{VaR}_{T+1} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{VaR}_{T+1}^{(i)}. \tag{18}$$

A standard approach to test the accuracy of VaR forecasts is to assess the violation rate, which is estimates as $\hat{\alpha} = x/m$, where x is defined by

$$x = \sum_{t=T+1}^{T+m} I(y_t < -(\mathrm{VaR})_t) \tag{19}$$

and is the number of violations during the time interval of length m.

In order to examine the accuracy of VaR forecasts, we adopt the unconditional coverage test introduced in Kupiec (1995). This is a likelihood ratio test with χ_1^2-distributed test statistic

$$LRuc = 2\{\log[\hat{\alpha}^x(1-\hat{\alpha})^{m-x}] - \log[\alpha^x(1-\alpha)^{m-x}]\}. \tag{20}$$

The null hypothesis is that the achieved violation rate is equal to the predetermined nominal probability α. For more details see Kupiec (1995).

Following Christoffersen (1998) and Christoffersen and Pelletier (2004), the test of independence tests for the clustering of VaR exceptions under the hypothesis of an independently distributed failure process against the alternative hypothesis of first order Markov failure process. The independence test statistic is

$$LRind = 2\left\{\log\left[(1-\hat{\pi}_{01})^{m-x-T_{01}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{x-T_{11}}\hat{\pi}_{11}^{T_{11}}\right] - \log[\hat{\alpha}^x(1-\hat{\alpha})^{m-x}]\right\}, \tag{21}$$

where $\hat{\pi}_{01} = T_{01}/(m-x)$, $\hat{\pi}_{11} = T_{11}/x$ and T_{ij} denotes the number of observations with a j following an i. $LRind \sim \chi_1^2$. As a practical matter, if the sample at hand has $T_{11} = 0$, which can easily happen in small samples and with small coverage rates, then we calculate

$$LRind = 2\left\{\log\left[(1-\hat{\pi}_{01})^{m-x-T_{01}}\hat{\pi}_{01}^{T_{01}}\right] - \log[\hat{\alpha}^x(1-\hat{\alpha})^{m-x}]\right\}. \tag{22}$$

Finally, the correct conditional coverage jointly tests for independence and correct coverage, with the test statistics as:

$$LRcc = LRind + LRuc \tag{23}$$

The magnitude of violating returns, i.e., the expected loss given a violation are also important quantities to be calculated (and not only their rate). Thus, measures of loss magnitude are also considered here. The absolute deviation (AD) of violating returns, considered by McAleer and da Veiga (2008), is defined by

$$AD_t = \mid y_t - (-(\mathrm{VaR})_t) \mid, \tag{24}$$

which is defined only when y_t is a violation. The maximum and the mean AD are calculated here to compare the competing VaR models: models with lower maximum (mean) ADs are preferred.

4 Empirical Application

This section analyzes the daily closing prices of the NASDAQ Composite index. The NASDAQ Composite is a stock market index of the common stocks and similar securities listed on the NASDAQ stock market, meaning that it has over 3000 components. It is highly followed in the U.S. as an indicator of the performance of stocks of technology and growth companies. Since both U.S. and non-U.S. companies are listed on the NASDAQ stock market, the index is not exclusively a U.S. 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 July 5, 1988 - July 3, 2003, which yields 3784 observations. Throughout, we work with the compounded return expressed as a percentage, $y_t = 100 \times (\log P_t - \log P_{t-1})$, where P_t is the closing price on day t. The compounded NASDAQ index returns are plotted in Fig. 2 as a time series plot and also as a histogram. We clearly identify the period of elevate volatility around of the turn of the Millennium associated with the collapse of the Tech

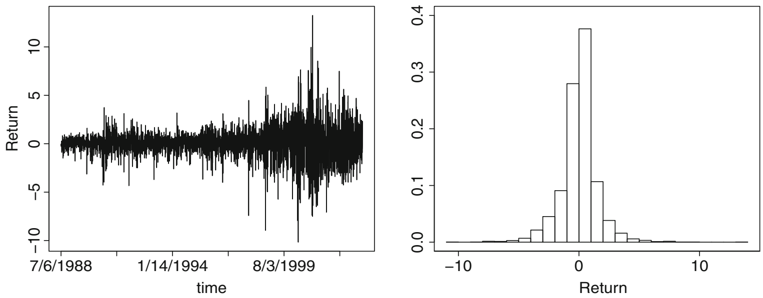

Fig. 2 Compounded NASDAQ returns from July 6, 1988 to July 3, 2003. The *left panel* shows the plot of the raw series and the *right panel* the histogram of returns

bubble. We are particularly interested in understanding the importance of excess of kurtosis and skewness in the NASDAQ index return and we avoid confounding our results by this highly influential outlier. There are some interesting pattern we observe in this history of NASDAQ returns. The average return is 0.04 percent daily with a daily standard deviation of 1.54. Returns exhibit negative skewness of -0.02 and kurtosis of 9.13. Note also that the returns have a large range (minimum, -10.16 and maximum, 13.25). We use the the Jarque-Bera (JB) statistics to test the normality assumption of the returns. In light of the JB statistics (5923.6), the null hypothesis of normality is rejected (p-value of 0.00) due to negative skewness and excess of kurtosis.

Now, we analyze the NASDAQ index returns with the aim of providing robust inference. In our analysis, we fit and compare the SV-N, SV-T, SV-SN and SV-ST models. In all cases, we simulated the h_t's in a multi-move fashion with stochastic knots based on the method described in Section 3.2. We fix the number of blocks K to be 95 in such a way that each block contained 40 h_t's on average. We set the prior distribution of the common parameters as: $\varphi \sim \mathcal{N}_{(-1,1)}(0.95, 100)$, $\sigma^2 \sim \mathcal{IG}(2.5, 0.025)$, $\mu \sim \mathcal{N}(0, 100)$. For the parameter φ the priors' mean and variance are 0.0032 and 0.3328, respectively. This prior setup is equivalent to the uniform distribution on interval (-1, 1), which gives zero mean and variance of 0.3333. We assume that $\lambda \sim t_{0.5}(0.0, \frac{\pi^2}{4})$, a Jeffreys' prior suggested by Bayes and Branco (2007), where $t_c(a, b)$ denotes the Student-t distribution with location a, scale b and c degrees of freedom. Finally, for ν, we assume the prior given by Eq. (10). All the calculations were performed running stand-alone code developed by us 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.

For all models, we conducted the MCMC simulation for 50000 iterations. In all cases, the first 10000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain, only every 20*th* values of the chain were stored. With the resulting 2000 values, we calculated the posterior means, the 95 % credible intervals and the convergence diagnostic (CD) statistics (Geweke 1992). If the sequence of the recorded MCMC output is stationary, it converges in distribution to the standard normal. According to the CD the null hypothesis that the sequence of 2000 draws is stationary was accepted at the 5 % level, CD $\in (-1.96, 1.96)$, for all the parameters in all the models considered here. In Table 1 we report, the inefficiency factor, which is defined by $1 + \sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s. It measures how well the MCMC chain mixes (see, e.g., Kim et al. 1998). It is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws.

Table 1 Estimation results for the NASDAQ returns. First row: Posterior mean. Second row: Posterior 95 % credible interval in parentheses. Third row: CD statistics

Parameter	SV-N	SV-T	SV-SN	SV-ST
	0.0623	0.0457	0.0417	0.0314
μ	(−0.9523, 0.9764)	(−1.3032, 1.2315)	(−1.1524, 1.0409)	(−1.5076, 1.3985)
	0.70	0.94	−0.91	−0.66
	1.12	0.98	1.13	0.76
	0.9944	0.9963	0.9954	0.9967
φ	(0.9897,0.9984)	(0.9926,0.9994)	(0.9745,0.9947)	(0.9932,0.9995)
	1.12	−0.43	0.11	−1.58
	2.89	3.28	1.74	2.76
	0.0172	0.0118	0.0143	0.0107
σ_η^2	(0.0115,0.0246)	(0.0075,0.0171)	(0.0097,0.0204)	(0.0070,0.0155)
	−0.13	0.39	0.81	0.87
	7.77	13.06	6.33	10.67
	–	–	−1.3908	−1.1528
λ	–	–	(−1.6280, −1.1470)	(−1.4020, −0.8820)
	–	–	0.23	−1.65
	–	–	1.49	7.80
	–	19.3369	–	19.6797
ν	–	(11.3700,35.3600)	–	(11.4000,36.5000)
	–	−1.78	–	1.72
	–	19.22	–	30.61
	–	0.0564	–	0.0556
$\frac{1}{\nu}$	–	(0.0283,0.0879)	–	(0.0274,0.0877)
	–	1.51	–	−1.36
	–	18.23	–	30.49

When the inefficiency factor is equal to m, we need to draw MCMC samples m times as many as the number of uncorrelated samples. Examining the inefficiency coeficients, we found that our algorithm produces a good mixing of the MCMC chain.

From Table 1, consistent with the existing evidence of great persistence in the log-volatility process, we found that the posterior means of φ and 95 % posterior credible intervals very close to the unity. Being the posterior mean of φ of the SV-ST model slightly higher than those of the other three models. The posterior mean of σ_η^2 is smaller in the SV-ST than those of the SV-N, SV-T and the SV-SN models, indicating that the log-volatility process of the SV-ST is less variable than those of the other ones.

In the SV-T and SV-ST models, the magnitude of the tail-fatness is measured by the degrees of freedom, ν, parameter. We found that the posterior mean of ν are 19.34 and 19.68, respectively, which indicates tail-fatness. In Table 1, we report the posterior mean of $1/\nu$, for both models, which, in both cases, are over 3.5 standard deviation from zero. Since the SV-N and SV-SN models are nested in the limit when $1/\nu$ approaches to zero. This provides strong evidence of heavy-tailness of conditional distributions of the returns.

Regarding the skewness parameter, λ, in the SV-SN and SV-ST models, we found that the posterior means are -1.3908 and -1.1528, respectively. In both models, the 95 % credible interval does not contain zero, that is the negativity of λ is credible. This supports the strong evidence of skewnesses in the NASDAQ data set.

The magnitudes of the mixing parameter U_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 SV-T and SV-ST models can accommodate an outlier by inflating the variance component for that observation in the conditional distribution with smaller U_t value. This fact is shown in Fig. 3 where we depicted the posterior mean of the mixing variable U_t for the SV-T (top panel) and SV-ST (bottom panel) models, respectively.

To assess the goodness of the estimated models, we calculate the Bayesian predictive information criteria, BPIC (Ando 2006, 2007) and the log-predictive score, LPS (Good 1952; Gneiting and Raftery 2007; Delatola and Griffin 2011). The BPIC criterion is defined as

$$BPIC = -2E_{\boldsymbol{\theta}} \mid \mathbf{y}_{1:T}[\log\{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\}] + 2T\hat{b}, \tag{25}$$

where $\hat{b}$ is given by

$$\hat{b} \approx \frac{1}{T}\left\{E_{\boldsymbol{\theta}}|\mathbf{y}_{1:T}[\log\{p(\mathbf{y}_{1:T}|\boldsymbol{\theta})p(\boldsymbol{\theta})\}] - \log[p(\mathbf{y}_{1:T}|\hat{\boldsymbol{\theta}})p(\hat{\boldsymbol{\theta}})] + \mathrm{tr}\{J_T^{-1}(\hat{\boldsymbol{\theta}})I_T(\hat{\boldsymbol{\theta}})\} + 0.5q\right\}. \tag{26}$$

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

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

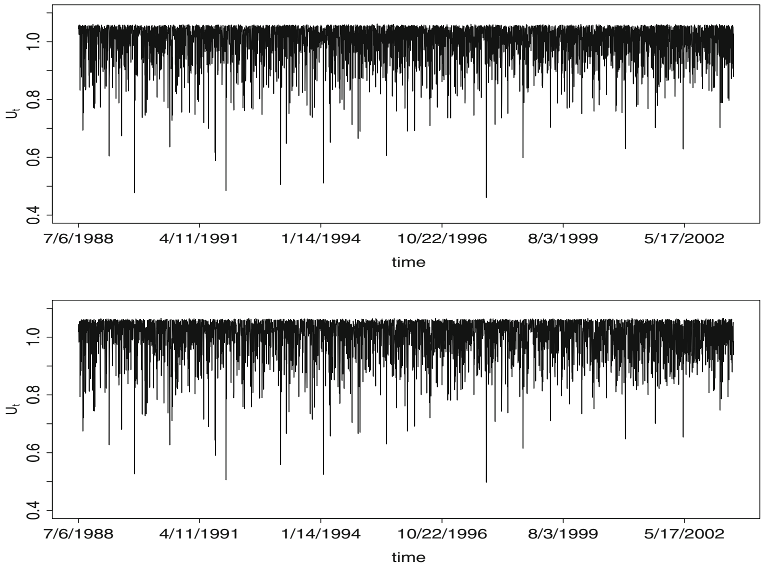

Fig. 3 NASDAQ data set: posterior smoothed mean of mixture variable U_t for the SV-T (*top panel*) and SV-ST (*bottom panel*) models

with $\eta_T(y_t, \boldsymbol{\theta}) = \log p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}) + \log p(\boldsymbol{\theta})/T$.

The LPS provide summary measures for the evaluation of probabilistic forecast, by assigning a numerical score based on the predictive distribution and on the event or value that materializes. The average log predictive score for one-step ahead predictions is given by

$$\text{LPS} = -\frac{1}{T}\sum_{t=1}^{T} \log p\left(y_t|\mathbf{y}_{1:t-1}, \hat{\boldsymbol{\theta}}\right), \tag{27}$$

where $\hat{\boldsymbol{\theta}}$ is an estimate of the model parameters and $p(y_t|\mathbf{y}_{1:t-1}, \hat{\boldsymbol{\theta}})$ the one-step ahead predictive density.

In the SV-N, SV-T, SV-SN, and SV-ST models, the log-likelihood function, $\log p(\mathbf{y}_{1:T}|\boldsymbol{\theta})$, is estimated using the auxiliary particle filter (see, e.g., Pitt and Shephard 1999) with 10000 particles. From Table 2, the BPIC and the LPS indicate the SV-ST model is the best model among all the models considered here, suggesting that the NASDAQ index return data demonstrate sufficient departure from underlying normality assumptions and asymmetry.

In Fig. 4 , we plot the smoothed mean of $e^{\frac{h_t}{2}}$ obtained from the MCMC output for the SV-N (solid line) and the SV-ST (dotted line). From a practical point of view, we are mainly interested in whether we find a significant difference between the two series. Therefore, in the bottom panel of Figure 4, we plot the smoothed mean of the difference of $e^{\frac{h_t}{2}}$ obtained from the SV-N and SV-T models. Some extreme returns make the differences clear. This can have a substantial impact, for instance, in the valuation of derivative instruments and several strategic or tactical asset allocation topics.

In order to examine the performance of VaR forecast for the competing models, we use the data from July 7, 2003 to October 10, 2004 as the validation period, giving $m = 321$ trading days. In the moving window approach, we use the first T observations in the period July 6, 1988 - July 3, 2003 to estimate the model and to forecast the $(T + 1)$th observation; the sample is then rolled forward by one observation, so that the second to the $(T + 1)$th observations are used to forecast the $(T + 2)$th observation. This process is repeated until the end of the sample, i.e., the $(T+m)$th observation. We thus obtain 321 volatility forecasts and VaR estimates with confidence levels of 5 %. The competing models were: RiskMetrics, SV-N, SV-T, SV-N and SV-ST. The results of 321 one-step-ahead forecasts are presented in Table 3, along with the violation rates, p-values of the unconditional coverage test (LR_{uc}), the independent test (LR_{ind}), the correct coverage test (LR_{cc}), the maximum ADs and mean ADs. In fact, we expect that the violation rates are close to the nominal probability $\alpha = 0.05$. Clearly, skewed errors are highly important at $\alpha = 0.05$ and a SV-SN or SV-ST specification seems best under that choice. The results suggest that at the 5 % quantile of the distribution the shape of the error distribution, especially whether it is skewed, is very

Table 2 Nasdaq return data set. BPIC: Bayesian predictive information criterion, LPS: average Log-predictive score

Model	BPIC	Ranking	LPS	Ranking
SV-N	15628.4	3	2.0650	3
SV-T	11662.6	2	1.5383	2
SV-SN	16012.4	4	2.1292	4
SV-ST	11575.3	1	1.5360	1

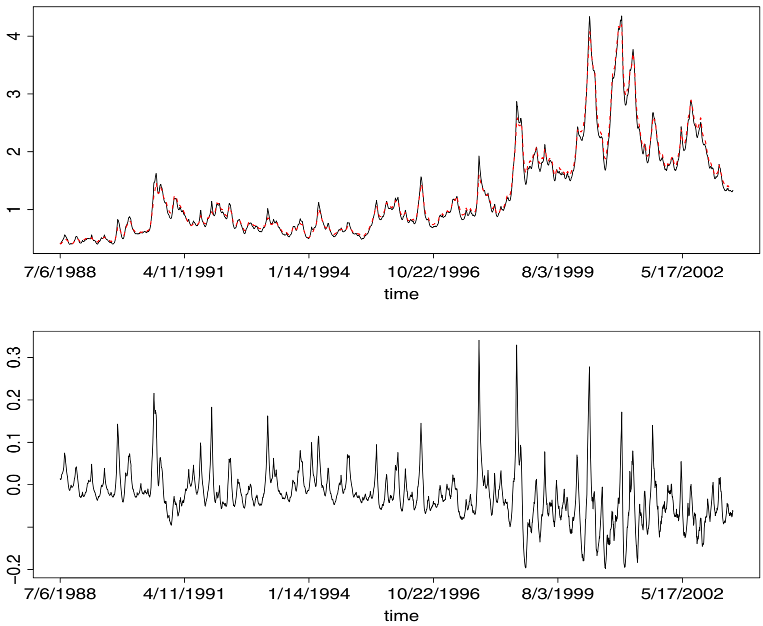

Fig. 4 NASDAQ data set. *Top*: Posterior smoothed mean of $e^{\frac{h_t}{2}}$. SV-N (*solid line*) and SV-ST (*dotted line*). *Bottom*: Posterior smoothed mean of the difference of $e^{\frac{h_t}{2}}$ in both models

important. According to the unconditional coverage test we accept the null hypothesis that the achieved violation rate is equal to 5 % for all the models. Acording to the independence test, we accept the null hypothesis of independently distributed failures for the SV-SN and SV-ST models, because their $p-$ values are 0.155 and 0.159 respectively. Finally, using the correct conditional coverage test jointly test for independence and correc coverage, we accept this hypothesis for the SV-N and SV-T, with the corresponding $p-$values equals to 0.352 and 0.420, respectively. According to the violation rate the SV-ST gives the best performance. Considering the maximum ADs, the best model to VaR forecast is the SV-ST. As the period under investigation is a quite period, the VaR forecast results at the 1 % level are almost the same for all the competing models. So, they are not reported here.

Table 3 Nasdaq return data set. Violation rate in 321 one-step-ahead forecast, *P*-values of the uncondtional coverage test, and the rank of the value-at-risk (VaR) at the 5 % level

	Violation Rate (%)	LR_{uc} $p-$value	LR_{ind} $p-value$	LR_{cc} $p-$value	AD of violation Maximum	Mean
RiskMetriks	0.059	0.262	0.005	0.010	1.053	0.399
SV-N	0.062	0.329	0.004	0.011	1.014	0.409
SV-T	0.059	0.462	0.005	0.014	1.027	0.399
SV-SN	0.056	0.624	0.155	0.352	0.917	0.355
SV-ST	0.049	0.989	0.159	0.420	0.887	0.352

5 Conclusions

In this article, we presented a Bayesian implementation of the stochastic volatility model with skew-Student-t (SV-ST) errors as an alternative to the normal (symmetric) assumption in the conditional distribution of the returns. The SV-ST model allows a parsimonious yet flexible treatment of both skewness and tail thickness. Under a Bayesian perspective, we developed a fast and efficient MCMC sampling procedure to estimate all the parameters and latent quantities in our proposed SV-ST model. We use objective priors for the degrees of freedom and the skewness parameters, ν and λ, based on Fonseca et al. (2008) and Bayes and Branco (2007), respectively. As a by product of the MCMC algorithm, we were able to produce an estimate of the latent information process which can be used in financial modeling. The use of mixing variable, $\mathbf{U}_{1:T}$ not only simplifies the full conditional distributions required for the Gibbs sampling algorithm, but also provides a mean for outlier diagnostics. We illustrated our methods through an empirical application of the NASDAQ return series, which showed that the SV-ST model provides better fit than the SV-N, SV-T and SV-SN models in terms of parameter estimates, interpretation and robustness aspects. On the other hand, since the posterior mean and 95 % posterior credibility interval of the parameter λ contains only negative values, we can conclude that there is a strong evidence of skewness in the NASDAQ data set.

This paper assesses the possibility of general Bayesian forecasting for carrying out 1-day ahead VaR forecasting across a range of competing parametric time-varying models, viz, the RiskMetrics, SV-N, SV-T, SV-SN and SV-ST models. For the NASDAQ data, the SV-ST ranked best, followed by the SV-SN model. In light of the results, the SV-ST is able to capture the skewness and execess of kurtosis we observe in practice.

Acknowledgments The research of Carlos A. Abanto-Valle was supported by CNPq grant 303421/2012-6. V. H. Lachos acknowledges financial support from FAPESP and CNPq.

Appendix A: The Full Conditionals

In this appendix, we describe the full conditional distributions for the parameters and the mixing latent variables $\mathbf{U}_{1:T}$ and $\mathbf{W}_{1:T}$ for the SV-ST model model.

Full conditional distribution of μ, φ and σ_η^2

The prior distributions of the common parameters are set as: $\mu \sim N\left(\bar{\mu}, \sigma_\mu^2\right)$, $\varphi \sim \mathcal{N}_{(-1,1)}\left(\bar{\varphi}, \sigma_\varphi^2\right)$, $\sigma_\eta^2 \sim \mathcal{IG}\left(\frac{T_0}{2}, \frac{M_0}{2}\right)$. We have the following full conditional for μ:

$$\mu \mid \mathbf{h}_{1:T}, \varphi, \sigma_\eta^2 \sim \mathcal{N}\left(\frac{b_\mu}{a_\mu}, \frac{1}{a_\mu}\right), \tag{A.1}$$

where $a_\mu = \frac{1}{\sigma_\alpha^2} + \frac{(T-1)(1-\varphi)^2}{\sigma_\eta^2} + \frac{(1-\varphi)^2}{\sigma_\eta^2}$ and $b_\mu = \frac{\bar{\mu}}{\sigma_\mu^2} + \frac{(1-\varphi^2)}{\sigma_\eta^2} h_1 + \frac{\sum_{t=1}^{T-1}(h_{t+1}-\varphi h_t)(1-\varphi)}{\sigma_\eta^2}$. In a similar way, the conditional posterior of φ is given by

$$p\left(\varphi \mid \mathbf{h}_{1:T}, \mu, \sigma_\eta^2\right) \propto Q(\varphi) \exp\left\{-\frac{a_\varphi}{2}\left(\psi - \frac{b_\varphi}{a_\varphi}\right)^2\right\} \mathbb{I}_{|\varphi|<1}, \tag{A.2}$$

where $Q_{\varphi} = \sqrt{1-\varphi^2}\exp\left\{-\frac{1}{2\sigma_\eta^2}\left[\left(1-\varphi^2\right)(h_1-\mu)^2\right\}\right.$, $a_{\varphi} = \frac{\sum_{t=1}^{T-1}(h_t-\mu)^2}{\sigma_\eta^2} + \frac{1}{\sigma_\psi^2}$, $b_{\varphi} = \frac{\sum_{t=1}^{T-1}(h_t-\mu)(h_{t+1}-\mu)}{\sigma_\eta^2} + \frac{\bar{\varphi}}{\sigma_\varphi^2}$ and $\mathbb{I}_{|\varphi|<1}$ is an indicator variable. As $p\left(\varphi \mid \mathbf{h}_{0:T}, \alpha, \sigma_\eta^2\right)$ in (A.2) does not have closed form, we sample from it by using the Metropolis-Hastings algorithm with truncated $\mathcal{N}_{(-1,1)}\left(\frac{b_\psi}{a_\psi}, \frac{1}{a_\psi}\right)$ as the proposal density.

Finally, the full conditional of σ_η^2 is $\mathcal{IG}\left(\frac{T_1}{2}, \frac{M_1}{2}\right)$, where $T_1 = T_0 + T$ and $M_1 = M_0 + \left[\left(1-\psi^2\right)(h_1-\mu)^2\right] + \sum_{t=1}^{T-1}\left[h_{t+1}-\mu-\psi\left(h_t-\mu\right)\right]^2$.

Full conditional of ν, λ, U_t and W_t

We, set ζ and ω in such a way that $E(y_t \mid h_t) = 0$ and $V(y_t \mid h_t) = \mathrm{e}^{h_t}$. So, we have $\zeta = -\sqrt{\frac{2}{\pi}}k_1\delta\omega$ and $\omega^2 = \left[k_2 - \frac{2}{\pi}k_1^2\delta^2\right]^{-1}$, where $k_1 = \sqrt{\frac{\nu}{2}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$, $k_2 = \frac{\nu}{\nu-2}$ and $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. Then the full the conditionals of ν and λ follows:

$$
\begin{aligned}
p(\nu \mid .) \propto & \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{\frac{1}{2}} \\
& \times \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}} \mathrm{e}^{-\frac{\nu}{2}\sum_{t=1}^{T}(U_t - \log U_t)}\left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} && \text{(A.3)} \\
& \times \left(\frac{1}{\omega}\right)^{T} \mathrm{e}^{-\frac{1}{2\omega^2(1-\delta^2)}\sum_{t=1}^{T} U_t \mathrm{e}^{-h_t}\left(y_t - \zeta - \omega\delta W_t U_t^{-\frac{1}{2}}\mathrm{e}^{\frac{h_t}{2}}\right)^2} \mathbb{I}_{\nu>2}, && \text{(A.3)}
\end{aligned}
$$

$$
p(\lambda \mid .) \propto \left(1 + \frac{2\lambda}{\frac{\pi^2}{4}}\right)^{-\frac{3}{4}}\left(\frac{1}{1-\delta^2}\right)^{\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\omega^2\left(1-\delta^2\right)}\sum_{t=1}^{T} U_t \mathrm{e}^{-h_t}\left(y_t-\zeta-\omega\delta W_t U_t^{-\frac{1}{2}}\mathrm{e}^{\frac{h_t}{2}}\right)^2}. \tag{A.4}
$$

Since the above full conditional distributions are not in any known closed form, we must simulate ν and λ using the Metropolis-Hastings algorithm. The proposal density used are $\mathcal{N}_{(\nu>2)}\left(\mu_\nu, \tau_\nu^2\right)$ and $\mathcal{N}\left(\mu_\lambda, \tau_\lambda^2\right)$, with $\mu_\upsilon = x - \frac{q'(x)}{q''(x)}$ and $\tau_\upsilon^2 = \max\left\{0.001, (-q''(x))^{-1}\right\}$ for $\upsilon = \nu$ or λ, where x is the value of the previous iteration, $q(.)$ is the logarithm of the conditional posterior density, and $q'(.)$ and $q''(.)$ are the first and second derivatives respectively.

As $U_t \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, the conditional posterior of U_t is given by

$$
p(U_t \mid h_t, W_t, \nu, \lambda) \propto Q(U_t) U_t^{\frac{\nu+1}{2}-1} \mathrm{e}^{-\frac{U_t}{2}\left[\nu + \frac{\mathrm{e}^{-h_t}\left(y_t - \zeta \mathrm{e}^{\frac{h_t}{2}}\right)^2}{\omega^2\left(1-\delta^2\right)}\right]}, \tag{A.5}
$$

where $Q(U_t) = \mathrm{e}^{\frac{U_t^{\frac{1}{2}}\delta W_t \mathrm{e}^{-\frac{h_t}{2}}\left(y_t - \zeta\mathrm{e}^{\frac{h_t}{2}}\right)}{\omega\left(1-\delta^2\right)}}$. As $p(U_t \mid h_t, W_t, \nu, \lambda)$ in (A.5) does not have closed form, we sample from it by using the Metropolis-Hastings algorithm with

$\mathcal{G}\left(\frac{\nu+1}{2}, \frac{1}{2}\left[\nu + \frac{e^{-h_t}\left(y_t - \zeta e^{\frac{h_t}{2}}\right)^2}{\omega^2(1-\delta^2)}\right]\right)$ as the proposal density. Finally, from Eqs. (8a) and (8c), we have the full conditional of W_t is the $\mathcal{N}_{[0,\infty)}\left(\frac{\delta U_t^{\frac{1}{2}} e^{-\frac{h_t}{2}}\left[y_t - \zeta e^{\frac{h_t}{2}}\right]}{\omega}, \frac{1}{1-\delta^2}\right)$.

References

Abanto-Valle CA, Bandyopadhyay D, Lachos VH, Enriquez I (2010) Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. Comput Stat Data Anal 54:2883–2898

Abanto-Valle CA, Migon HS, Lachos VH (2011) Stochastic volatility in mean models with scale mixtures of normal distributions and correlated errors: a bayesian approach. J Stat Plan Infer 141:1875–1887

Abanto-Valle CA, Migon HS, Lachos VH (2012) Stochastic volatility in mean models with heavy-tailed distributions. Braz J Probab Stat 26:402–422

Ando T (2006) Bayesian inference for nonlinear and non-gaussian stochastic volatility model wit leverge effect. J Jpn Stat Soc 36:173–197

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

Azzalini A (1986) Further results on a class of distributions which include the normal ones. Statistica 46:199–208

Azzalini A (2005) The skew-normal distribution and related multivariate families. Scand J Stat 32:159–188

Azzalini A, Capitanio A (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. J R Stat Soc: Ser B 65:367–389

Bayes CL, Branco MD (2007) Bayesian inference for the skewness parameter of the scalar skew-normal distribution. Braz J Probab Stat 21:141–163

Branco MD, Dey DK (2001) A general class of multivariate skew-elliptical distributions. J Multivar Anal 79:99–113

Cappuccio N, Lubian D, Raggi D (2004) Mcmc bayesian estimation of a skew-GED stochastic volatility model. Stud Nonlinear Dyn Econ 8:article 6

Cappuccio N, Lubian D, Raggi D (2006) Investigating asymmetry in U.S. stock market indexes: evidence from a stochastic volatility model. Appl Financ Econ 16:479–490

Carter CK, Kohn R (1994) On Gibbs sampling for state space models. Biometrika 81:541–553

Chen Q, Gerlach R, Lub Z (2012) Bayesian value-at-risk and expected shortfall forecasting via the asymmetric laplace distribution. Comput Stat Data Anal 56:3498–3516

Chib S, Greenberg E (1995) Understanding the Metropolis-Hastings algorithm. Am Stat 49:327–335

Chib S, Nardari F, Shepard N (2002) Markov Chain Monte Carlo methods for stochastic volatility models. J Econ 108:281–316

Christoffersen P (1998) Evaluating interval forecasts. Int Econ Rev 39:841–862

Christoffersen P, Pelletier D (2004) Backtesting value-at-risk: a duration-based approach. J Financ Econ 2:84–108

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

Delatola EI, Griffin JE (2011) Bayesian nonparametric modelling of the return distribution with stochastic volatility. Bayesian Anal 6:901–926

Fonseca TCO, Ferreira MAR, Migon HS (2008) Objective Bayesian analysis for the student-t regression model. Biometrika 95:325–333

Frühwirth-Schnater S (1994) Data augmentation and dynamic linear models. J Time Ser Anal 15:183–202

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics, vol 4. Oxford University Press, Oxford, pp 169–193

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction and estimation. J Am Stat Assoc 6:901–926

Good IJ (1952) Rational decisions. J R Stat Soc, Ser B 14:107–114

Hansen BE (1994) Autoregressive conditional density estimation. J Financ 35:705–730

Harvey C, Sidique A (1999) Autoregressive conditional skewness. J Financ Quant Anal 34:465–487

Harvey C, Sidique A (2000) Conditional skewness in asset pricing tests. J Financ 55:1263–1296

Henze N (1986) A probabilistic representation of the skew-normal distribution. Scand J Stat 13:271–275
Jacquier E, Polson N, Rossi P (2004) Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. J Econ 122:185–212
Jondeau E, Rockinger M (2003) Conditional volatility, skewness and kurtosis: existence, persistence and co-movements. J Econ Dyn Control 27:1699–1737
Kim S, Shepard N, Chib S (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. Rev Econ Stud 65:361–393
Koopman S (1993) Disturbance smoothers for State Space models. Biometrika 80:117–126
Kupiec PH (1995) Techniques for verifying the accuracy of risk measurement models. J Deriv 3:73–84
Liesenfeld R, Jung RC (2000) Stochastic volatility models: conditional normality versus heavy-tailed distrutions. J Appl Econ 15:137–160
McAleer M, da Veiga B (2008) Single-index and portfolio models for forecasting value-at-risk thresholds. J Forecast 27:217–235
Melino A, Turnbull SM (1990) Pricing foreign options with stochastic volatility. J Econ 45:239–265
Mittnik S, Paolella MS (2000) Conditional density and value-at-risk prediction of asian currency exchange rates. J Forecast 19:313–333
Nakajima J, Omori Y (2012) Stochastic volatility model with leverage and asymmetrically heavy-tailed error using gh skew student's t-distribution. Comput Stat Data Anal 56:3690–3704
Pemstein D, Quinn KV, Martin AD (2007) The Scythe statistical library: an open source C++ library for statistical computation. J Stat Softw VV(II):1–29
Pitt M, Shephard N (1999) Filtering via simulation: auxiliary particle filter. J Am Stat Assoc 94:590–599
Rosa GJM, Padovani CR, Gianola D (2003) Robust linear mixed models with Normal/Independent distributions and bayesian MCMC implementation. Biom J 45:573–590
Shephard N, Pitt M (1997) Likelihood analysis of non-Gaussian measurements time series. Biometrika 84:653–667
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. North-Holland, Amsterdam, pp 203–226
Taylor S (1986) Modeling financial time series. Wiley, Chichester
Tierney L (1994) Markov chains for exploring posterior distributions (with discussion). Ann Stat 21:1701–1762
Tsiotas G (2012) On generalised asymmetric stochastic volatility models. Comput Stat Data Anal 56:151–172
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