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, CEP: 21945-970, 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*

Abstract

This paper consider a Bayesian analysis of stochastic volatility models using a class of symmetric normal scale mixtures, which provides an appealing robust alternative to the routine use of the normal distribution in this type of models. Specific distributions examined include the normal, the Student-t, the slash and the variance gamma distribution which are obtained as a sub-class of our proposed class of models. Using a Bayesian paradigm, we explore an efficient Markov chain Monte Carlo (MCMC) algorithm for parameter estimation in this model. Moreover, the mixing parameters obtained as a by-product of the scale mixture representation can be used to identify possible outliers. The methods developed are applied to analyze daily stock returns data on S&P500 index. We conclude that our proposed rich class of normal scale mixture models provides robustification over the traditional normality assumptions often used to model thick-tailed stochastic volatility data.

Key words: stochastic volatility, scale mixture of normal distributions, Markov chain Monte Carlo, non linear state space models.

*Corresponding author. Tel: +55 21 2562-7505 ext. 201
Email address: `cabantovalle@im.ufrj.br` (C. A. Abanto-Valle)

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 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–t model by assuming the flexible class of scale mixtures of normal (SMN) distribution (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 normal (SV–N), the Student-t (SV–t), the slash (SV-S) and variance gamma (SV–VG) distribution. All these distributions have heavier tails than the normal one, and thus can be used for robust inference in these type of models. We refer to this generalization of the SMN class for SV models as SV–SMN distributions. Our work is motivated by the fact that the daily stock returns data on $S\&P500$ index seems to present 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 the Bayesian estimation procedure using MCMC methods. Section 4 is devoted to application and model comparison among particular members of the SV–SMN class using the S&P500 index dataset. Some concluding remarks as well as future developments are deferred to Section 5.

2. SMN distribution

Scale mixtures of normal distribution, which play 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 valued mixing random variable with probability density function (pdf) $h(\lambda|\boldsymbol{\nu})$, independent of $Z \sim \mathcal{N}(0, \sigma^2)$, where $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of λ and $\kappa(.)$ is a 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

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

From a suitable choice of the mixing density $h(.|\boldsymbol{\nu})$, 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 $\kappa(\lambda) = 1$ (a degenerate random variable), we retrieve the normal distribution. Apart from the SV-Normal model, we explore 3 different types of heavy-tailed densities based on the choice of the mixing density $h(.|\boldsymbol{\nu})$. These are as follows.

- *The Student t–distribution,* $Y \sim T(\mu, \sigma^2, \nu)$

  The use of the t-distribution as an alternative robust model to the normal distribution has frequently been suggested in the literature (Little (1988) and 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:

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

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

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

and $\Gamma(a)$ is the gamma function with argument $a > 0$. That is, $Y \sim t_\nu(\mu, \sigma)$ 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, $Y \sim S(\mu, \sigma^2, \nu)$, $\nu > 0$.*

  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

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

  Here the distribution of λ is Beta ($\mathcal{B}e(\nu, 1)$), with density

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

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

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

  The slash distribution has been mainly used in simulation studies because it represents an extreme situation, see for example Andrews et al. (1972), Gross (1973), and Morgenthaler and Tukey (1991).

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

  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

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

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

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

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

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

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

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:

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

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

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

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

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

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: $\lambda_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$, $\lambda_t \sim \mathcal{B}e(\nu, 1)$ 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 as will be seen next.

3.1. Parameter estimation via MCMC

A Bayesian approach to parameter estimation in the SV-SMN class of models defined by equations (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 time T. The Bayesian approach for estimating the SV-SMN class of 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

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

where

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

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

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

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 $p(\mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}, \boldsymbol{\theta} \mid \mathbf{y}_{0:T}, \mathbf{v}_{0:T})$ 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 non linear 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. 2008) in the next section 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 Shepard, 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, $\mathbf{h}_{k_{i-1}+1:k_i-1} = (h_{k_{i-1}+1}, \ldots, h_{k_i-1})'$ for $i = 1, \ldots, K+1$, with $k_0 = 0$ and $k_{K+1} = T$, where $k_i - k_{i-1} \geq 2$ is the size of the $i-$th block. Following Shephard and Pitt (1997) and Omori and Watanabe (2008), the K knots $(k_1, \ldots, k_K)$ are generated randomly using

$$k_i \;=\; \mathrm{int}[T \times \{(i + u_i)/(K + 2)\}], \qquad i = 1, \ldots, K., \tag{17}$$

where the u_i's are independent realizations of the uniform random variable on the interval (0,1) and $\mathrm{int}[x]$ represents the floor of x. We sample the block of disturbances $\boldsymbol{\eta}_{k_{i-1}+1:k_i-1} = (\eta_{k_{i-1}+1}, \ldots, \eta_{k_i-1})$ instead of $\mathbf{h}_{k_{i-1}+1:k_i-1} = (h_{k_{i-1}+1}, \ldots, h_{k_i-1})$, exploring the fact that the innovations η_t are *i.i.d.* with $\mathcal{N}(0, 1)$.

Suppose that $k_{i-1} = t$ and $k_i = t+k+1$ for the $i-$th block, such that $t+k < T$. Then $\boldsymbol{\eta}_{t+1:t+k} = (\eta_{t+1}, \ldots, \eta_{t+k})$ 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

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

where $l(h_r)$ is the log of $f(y_r \mid h_r, \lambda_r)$ given by

$$
l(h_r) \;=\; \text{const} - \frac{h_r}{2} - \frac{1}{2} \lambda_r y_r^2 e^{-h_r}.
$$

We denote the first and second derivatives of $l(h_r)$ with respect to h_r by l' and l''. Next, we apply a Taylor's series expansion to $\sum_{r=t+1}^{t+k} l(h_r)$ in equation (18) around some preliminary estimate of $\boldsymbol{\eta}_{t+1:t+k}$, denoted by $\hat{\boldsymbol{\eta}}_{t+1:t+k}$. After some simple but tedious algebra, we have the approximate normal density g as follows

$$
\begin{aligned}
&\log f(\boldsymbol{\eta}_{t+1:t+k} | h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k}, \boldsymbol{\theta}) \\
&\quad = \text{const} - \frac{1}{2\sigma_\eta^2} \textstyle\sum_{r=t+1}^{t+k} \eta_r^2 + \frac{1}{2} \sum_{r=t+1}^{t+k-1} l''(\hat{h}_r) \left(\hat{h}_r - \frac{l'(\hat{h}_r)}{l''(\hat{h}_r)} - h_r \right)^2 \\
&\; - \frac{\phi^2 - l''(\hat{h}_{t+k})\sigma_\eta^2}{2\sigma_\eta^2} \left\{ \frac{\sigma_\eta^2}{\phi^2 - l''_F(\hat{h}_{t+k}} \left(l'(\hat{h}_{t+k}) - l''(\hat{h}_{t+k})\hat{h}_{t+k} + \frac{\phi - \alpha_{S_{t+k+1}}}{\sigma_\eta^2} h_{t+k+1} \right) - h_{t+k} \right\}^2 \\
&\quad = \log g,
\end{aligned} \tag{19}
$$

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

From (19), we define auxiliary variables d_r and $\hat{y}_r$ for $r = t+1, \ldots, t+k-1$ as follows:

$$
\begin{aligned}
d_r &= -\frac{1}{l''(\hat{h}_r)}, \\
\hat{y}_r &= \hat{h}_r + d_r l'(\hat{h}_r),
\end{aligned} \tag{20}
$$

For $r = t+k < T$

$$
\begin{aligned}
d_r &= \frac{\sigma_\eta^2}{\phi - \sigma_\eta^2 l''(\hat{h}_{t+k})} \\
\hat{y}_r &= d_r \left[l'(\hat{h}_r) - l''(\hat{h}_r)\hat{h}_r + \frac{(\phi-\alpha)}{\sigma_\eta^2} h_{r+1} \right],
\end{aligned} \tag{21}
$$

and when $r = t+k = T$ we use (20) to define the auxiliary variables.

The resulting normalized density in (19), 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:

$$
\hat{y}_r = h_r + \epsilon_r, \qquad \epsilon_r \sim \mathcal{N}(0, d_r), \tag{22}
$$

$$
h_r = \alpha + \phi h_{r-1} + \sigma_\eta \eta_r, \qquad \eta_r \sim \mathcal{N}(0,1) \tag{23}
$$

Applying the de Jong and Shepard (1995) simulation smoother to this model with the artificial $\hat{\mathbf{y}}_{t+1:t+k}$ 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 (Tierney, 1994; Chib, 1995). In the SV-N case, we use the same procedure with $\lambda_t = 1$ for $t = 1, \ldots, T$.

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 artificial $\hat{\mathbf{y}}_{t+1:t+k}$. Then, we apply the Kalman filter and a disturbance smoother to the linear Gaussian state space model consisting of equations (22) and (23) with the artificial $\hat{\mathbf{y}}_{t+1:t+k}$ to obtain the mean of $\hat{\mathbf{h}}_{t+1:t+k}$ conditional on $\hat{\mathbf{y}}_{t+1:t+k}$ in the linear Gaussian state space model. This is used as the next value of $\hat{\mathbf{h}}_{t+1:t+k}$. In this article, we use five iterations of this procedure to obtain a reasonable sequence of $\hat{\mathbf{h}}_{t+1:t+k}$.

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:

$$\mathrm{DIC} \quad = \quad -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})] + p_D. \tag{24}$$

The second term in (24) 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:

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

In the context of the SV-SMN class of models, $\boldsymbol{\theta}$ encompasses the parameter vector $(\alpha, \phi, \sigma_\eta^2, \nu)'$, $\boldsymbol{\lambda}_{1:T}$ and $\mathbf{h}_{0:T}$. Berg et al. (2004) proposed to use the deviance information criterion (DIC) to compare several specifications of the SV models.

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, Ando (2007) recently proposed the Bayesian predictive information criterion (BPIC) as an improvement over 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

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

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: $b = \int (\hat{\eta} - \eta) s(\mathbf{z}_{1:T}) d\mathbf{y}_{1:T}$. Ando (2007) evaluated the asymptotic bias as

$$T\hat{b} \approx E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})\}] - \log[L(\mathbf{y}_{1:T} \mid \hat{\boldsymbol{\theta}}) p(\hat{\boldsymbol{\theta}})] + \mathrm{tr}\{J_n^{-1}(\hat{\boldsymbol{\theta}}) I_n(\hat{\boldsymbol{\theta}})\} + 0.5q. \tag{26}$$

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

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

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

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

$$BPIC \;=\; -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})] + 2T\hat{b}. \tag{27}$$

The best model is chosen as the minimizer of BPIC. In the context of the SV-SMN class of models, $\boldsymbol{\theta} = (\alpha, \phi, \sigma_\eta^2, \nu)'$ and $\log p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ is evaluated numerically using the auxiliary particle filter method (Kim et al. 1998; Pitt and Shephard 1999; Chib et al. 2002).

4. Empirical Application

This section analyzes the daily closing prices for the S&P500 stock market index[1]. The S&P500 index contains the stocks of 500 Large-Cap corporations, most of which are American, and is used in reference not only to the index but also to the 500 companies that have their common stock included in the index. 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 summarize descriptive statistics for the corrected compounded returns with the time series plot in Figure 4. 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 > three, so that daily

[1]The data set was obtained from the Yahoo finance web site at http://finance.yahoo.com

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

S&P500 returns likely shows a departure from the underlying normality assumption. Thus, we revisit this data with the aim of providing additional inferences by using 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.

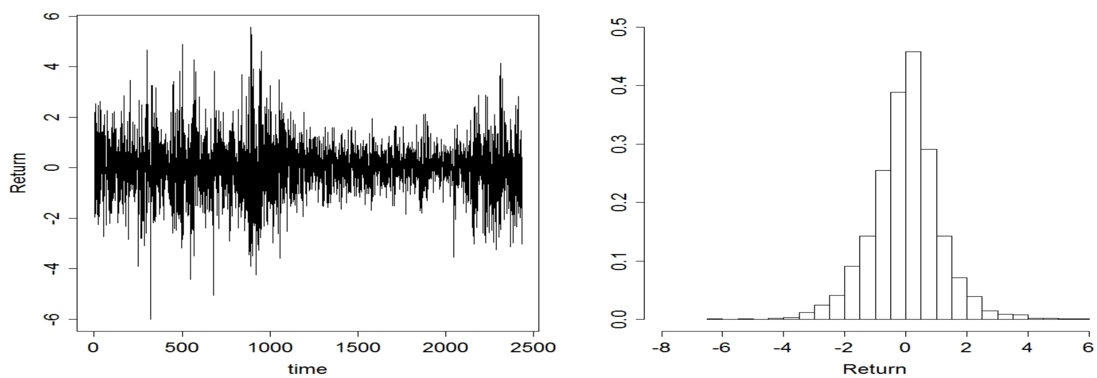

Figure 1: S&P500 corrected compounded returns with sample period from January 5, 1999 to September 05, 2008. Left: raw series. Right: histogram of returns.

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. We set K, the number of blocks as 40 in a such way that each block contained 60 h_t's on average. For all models, we conducted

Table 2: Estimation result for the S&P500 return. 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
	-0.0016	-0.0043	-0.0146	-0.0011
α	(-0.0104,0.0067)	(-0.0132,0.0040)	(-0.0267,-0.0042)	(-0.0095,0.0072)
	0.34×10^{-4}	0.76×10^{-4}	1.86×10^{-4}	0.41×10^{-4}
	-1.09	0.457	-0.98	0.51
	0.9700	0.9725	0.9730	0.9721
ϕ	(0.9542,0.9834)	(0.9575,0.9852)	(0.9579,0.9854)	(0.9568,0.9846)
	3.04×10^{-4}	3.03×10^{-4}	3.17×10^{-4}	2.99×10^{-4}
	-1.94	0.38	-0.72	-0.59
	0.0447	0.0415	0.0406	0.0402
σ^2	(0.0292,0.0652)	(0.0258,0.0590)	(0.0254,0.0598)	(0.0270, 0.0607)
	5.27×10^{-4}	5.40×10^{-4}	5.46×10^{-4}	4.82×10^{-4}
	1.84	-0.27	0.49	0.61
	—	18.2973	2.2618	17.7880
ν	—	(11.2700,28.5300)	(2.0670,2.4250)	(9.7930 ,30.1460)
	—	0.2987	0.0012	0.4535
	—	0.8171	-0.61	-0.38

the MCMC simulation for 60000 iterations. The first 20000 draws were discarded as a burn-in period. Based on the sample of next 40000 samples, 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 40000 draws is stationary is accepted at the 5% level for all the parameters and in all the models considered here. Figures 2, 3, 4 and 5 depicted the sampling results for SV-N, SV-t, SV-S and SV-VG models on the S&P500 return series. We observe a rapid decay of autocorrelations for all the models.

The estimate of the volatility parameters (α, ϕ, σ^2) are consistent with the results

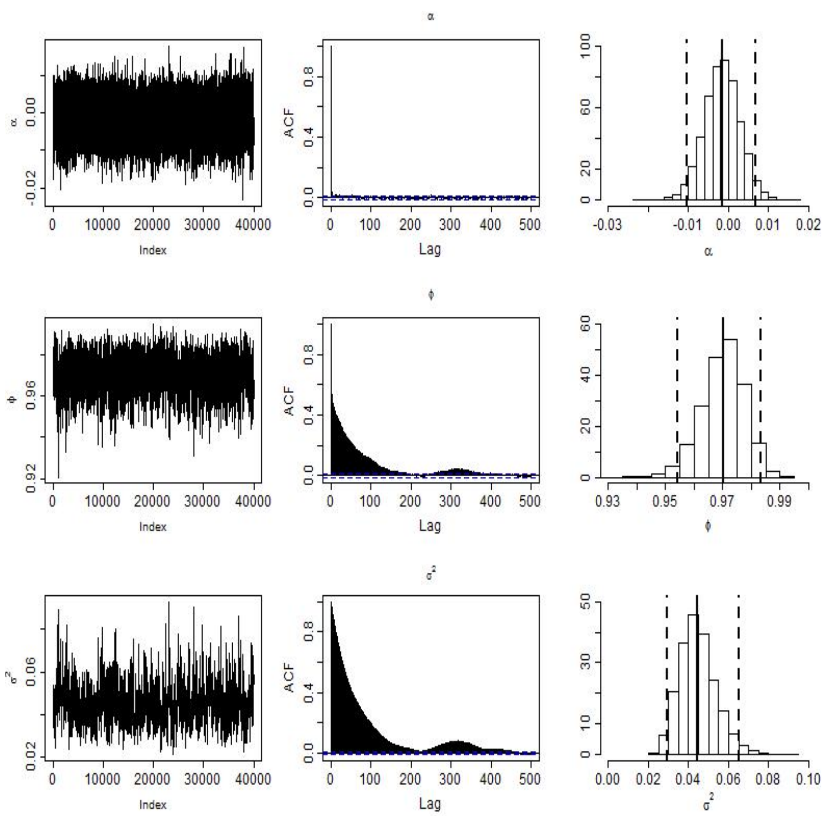

Figure 2: Estimation result for the S&P500 daily index returns (SV-N model). Sample paths (left), sample autocorrelations (middle), posterior histograms (right), the doted line indicate the 2.5% and 97.5% percentiles and the solid line the sample posterior mean.

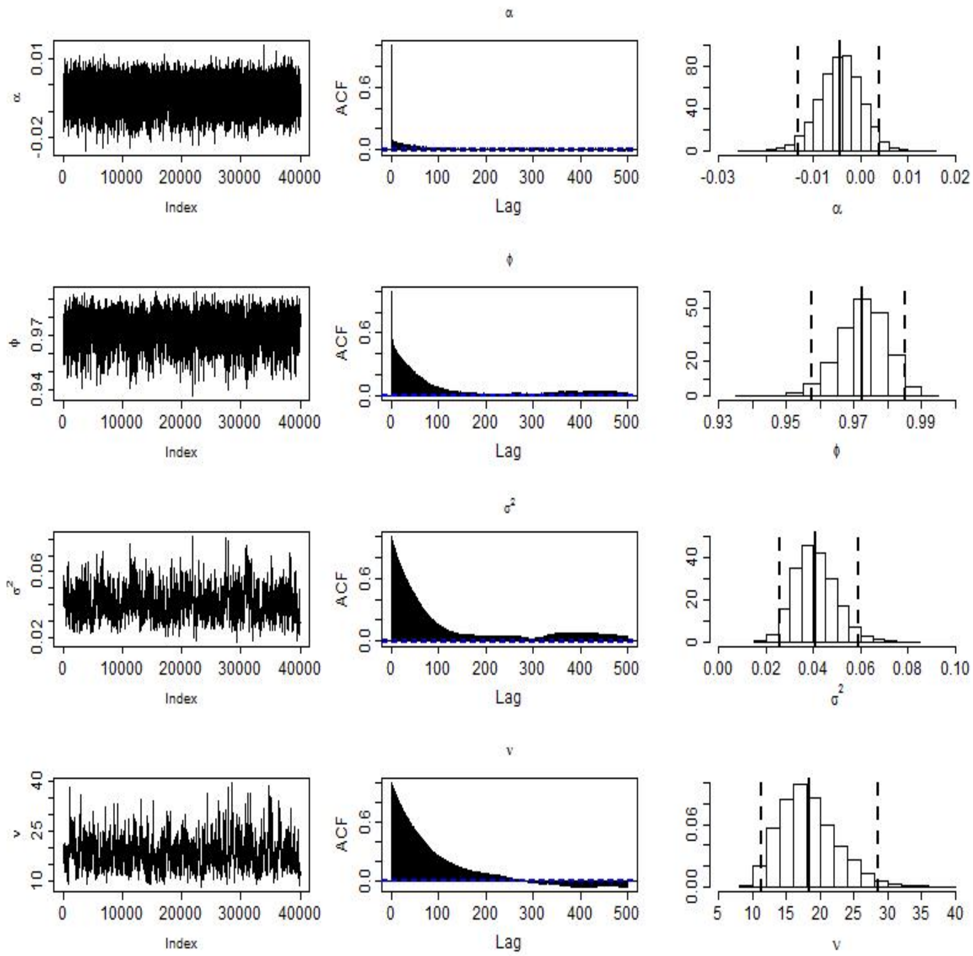

Figure 3: Estimation result for the S&P500 daily index returns (SV-t model). Sample paths (left), sample autocorrelations (middle), posterior histograms (right), the doted line indicate the 2.5% and 97.5% percentiles and the solid line the sample posterior mean.

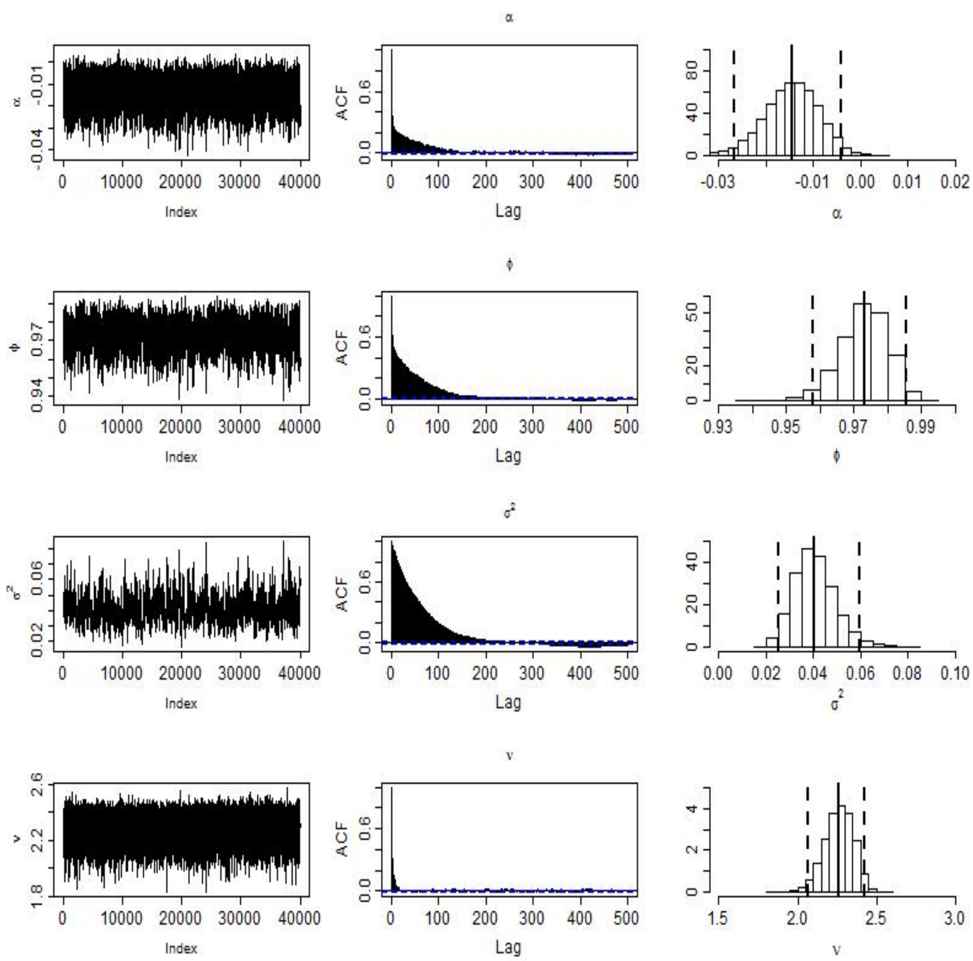

Figure 4: Estimation result for the S&P500 daily index returns (SV-S model). Sample paths (left), sample autocorrelations (middle), posterior histograms (right), the doted line indicate the 2.5% and 97.5% percentiles and the solid line the sample posterior mean.

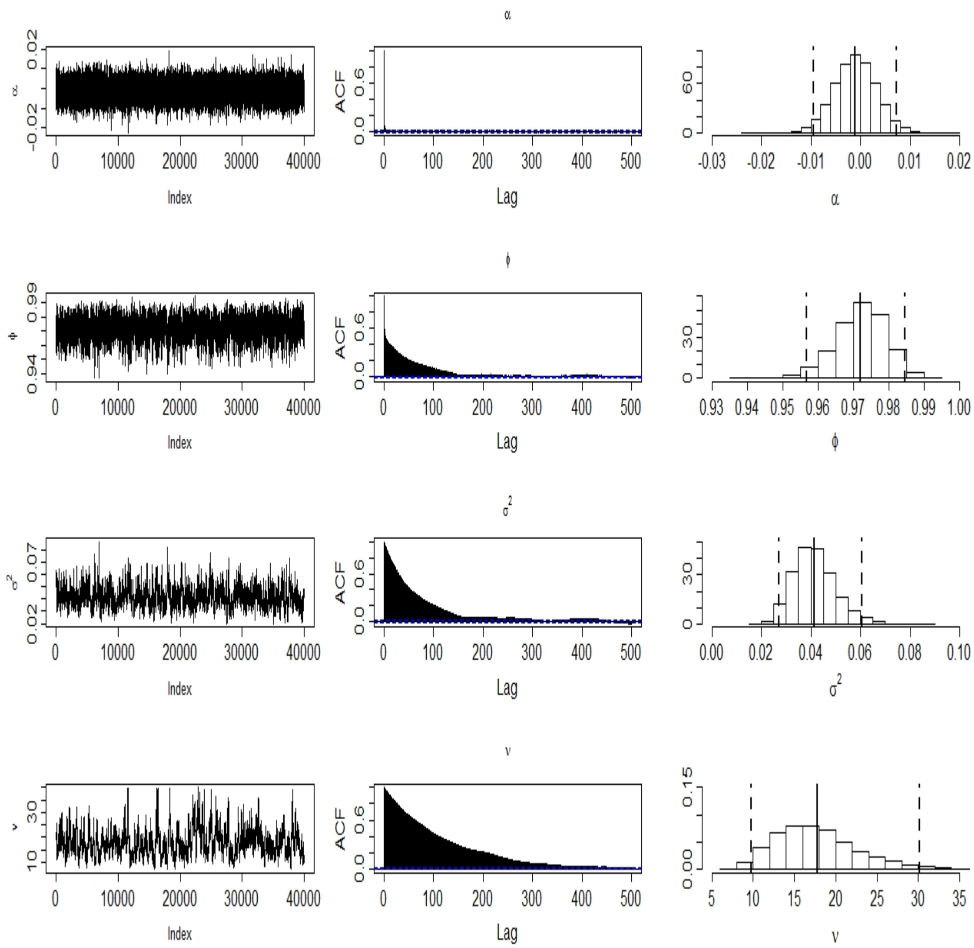

Figure 5: Estimation result for the S&P500 daily index returns (SV-VG model). Sample paths (left), sample autocorrelations (middle), posterior histograms (right), the doted line indicate the 2.5% and 97.5% percentiles and the solid line the sample posterior mean.

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 18.2973, 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.

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 SV-SMN class of models can accommodate an outlier by inflating the variance component for that observation in the conditional normal distribution with smaller λ_t value. This fact is shown in Figure 6 where we depicted the posterior mean of the mixing variable λ_t for the SV-t (top panel), the SV-S (middle panel) and the SV-VG (bottom panel) model.

In Figure 7, we show the graph of e^{h_t} estimated by the SV-N versus the e^{h_t} and $\lambda_t^{-1}\mathrm{e}^{h_t}$ estimated by the SV-t (top panel), SV-S (middle panel) and SV-VG (bottom panel). It can seen from Figure 7 that the SV-N, SV-t and SV-VG models produce similar estimates to e^{h_t}. However, Figure 7 (middle panel) indicates that the volatility process estimated by the SV-S model is different from the other compet-

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

	DIC		BPIC	
Model	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

ing SV models. This can have a substantial impact, for instance, in the valuation of derivative instruments and several strategic or tactical asset allocation topics. It is clear that the SV-S model accommodate in a different way, possible outliers by inflating the variance e^{h_t} by $\lambda_t^{-1}e^{h_t}$. For example, in this model the observations labeled as A, B and C corresponding to April 14, 2000, July 24, 2002 and July 29, 2002 respectively have their fitted values of e^{h_t} smaller than the corresponding $\lambda_t^{-1}e^{h_t}$.

We use the deviance information criterion (DIC) and the Bayesian predictive information criterion (BPIC) to compare between 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 model fit than the basic SV-N model, with the SV-S model relatively better among all the models, suggesting that the SP&500 data demonstrate sufficient departure from underlying normality assumptions. The DIC also selects the SV-S model as the best model.

The robustness of the SV-SMN class models can be study through the influence of outliers on the posterior distribution of the parameters. We consider only the SV-S model for illustration as it is the best chosen model using model selection measures. We study the influence of three contaminated observations on the posterior

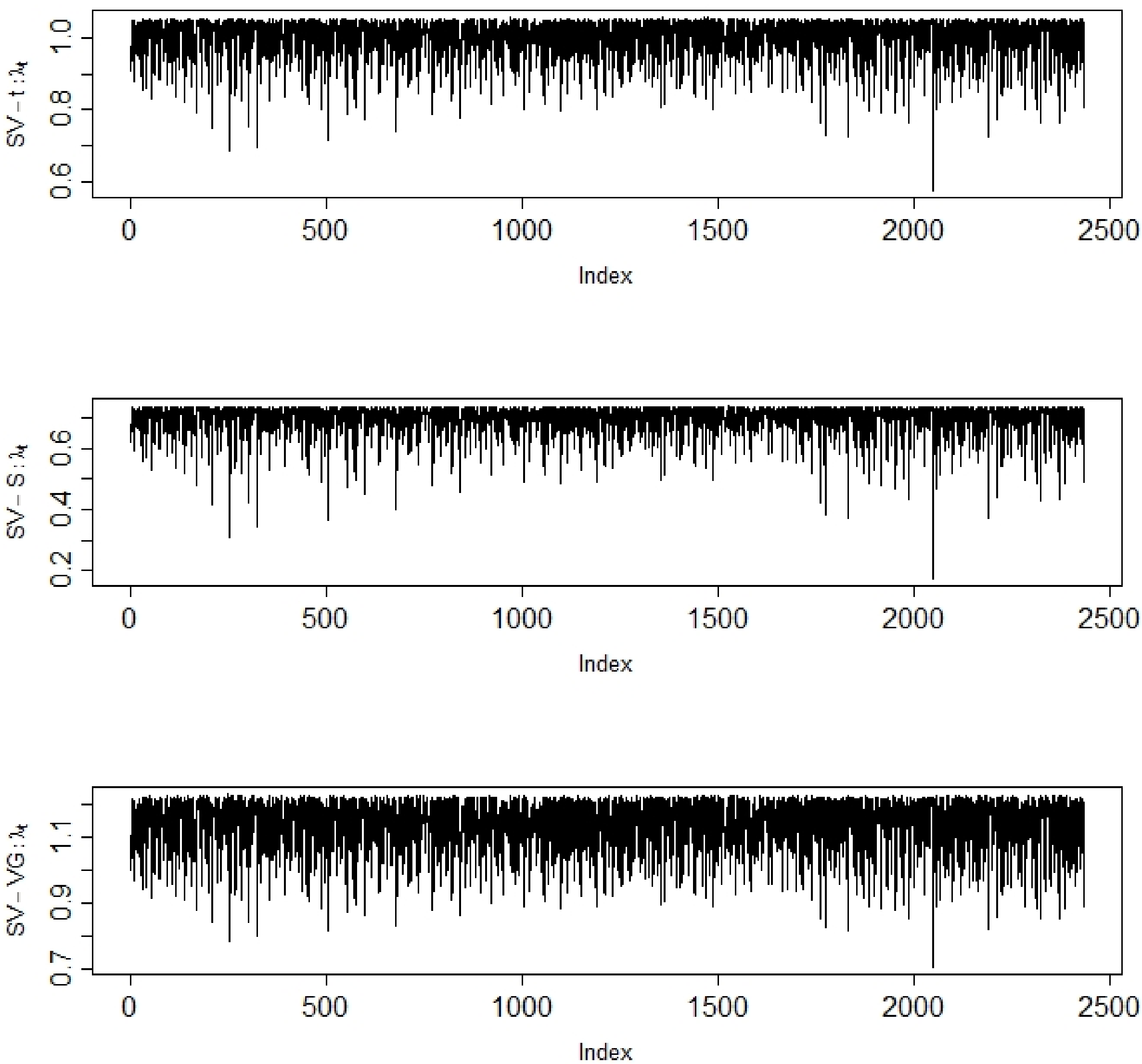

Figure 6: Comparison of the estimated mixing variables λ_t for the SP&500 index

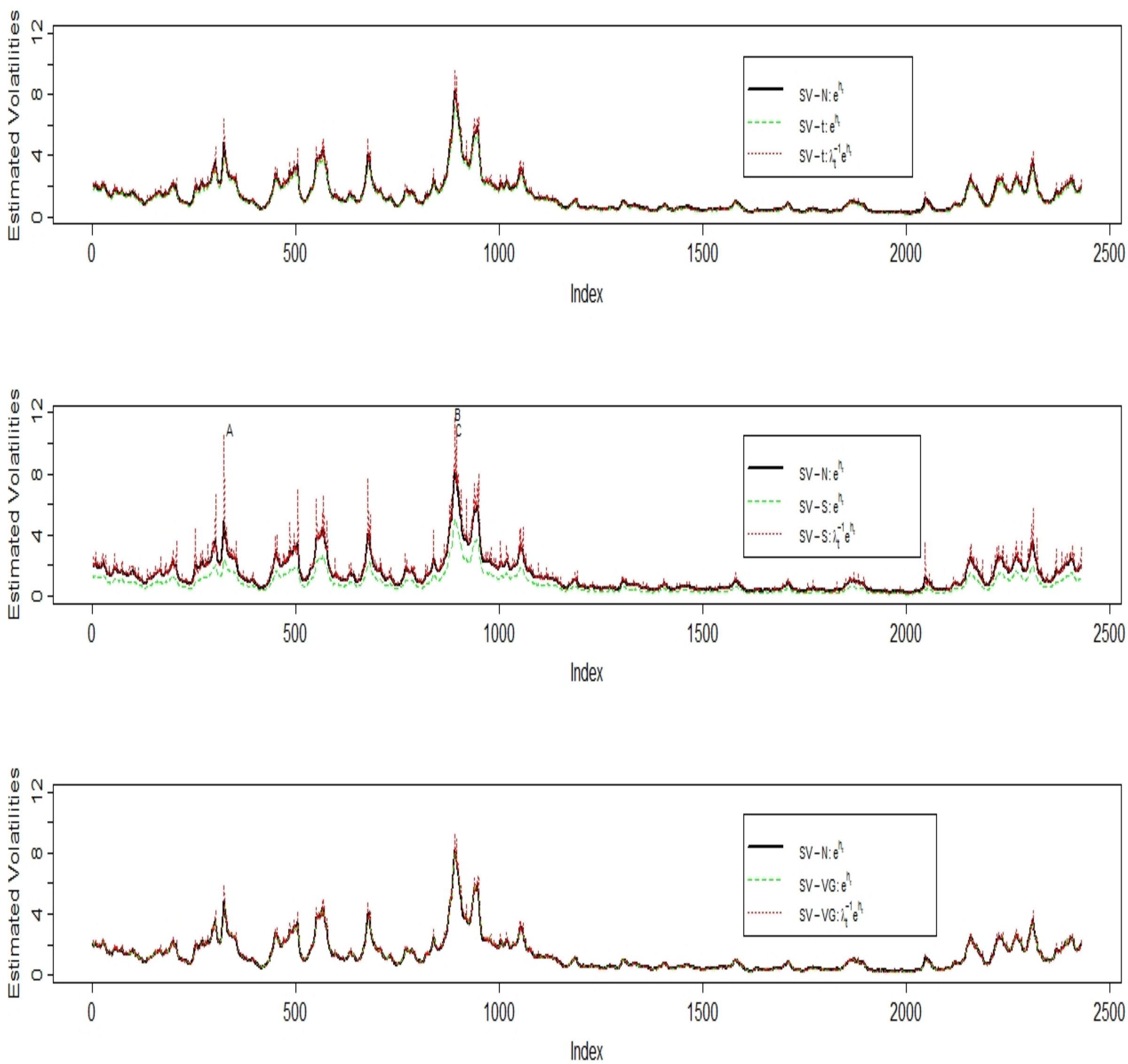

Figure 7: Comparison of the estimated volatilities for SP&500 index

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 Figures 8 and 9, we plot the posterior mean and 95% credible interval of ϕ and σ^2_η, respectively, for the SV-N and SV-S models. Clearly, the SV-S model is less affected by variations of k than the SV-N model signifying substantial robustness over the normal model in presence of outlying observations.

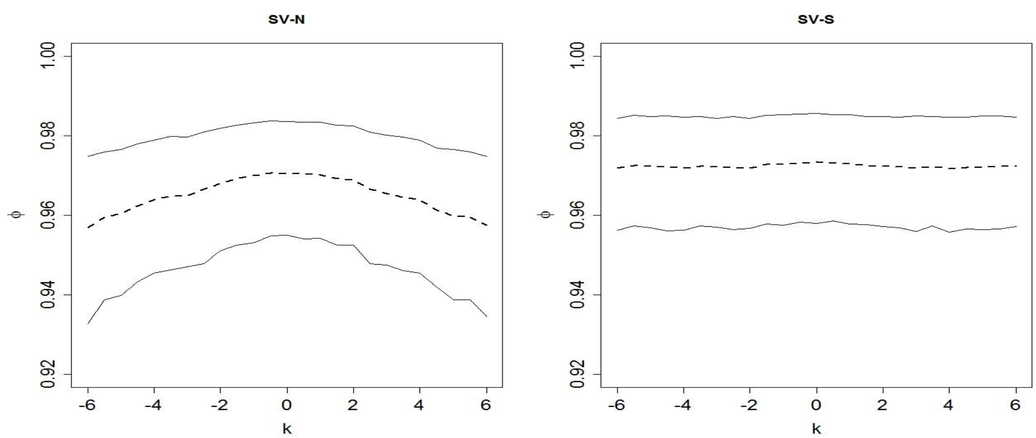

Figure 8: Posterior mean (dashed line) and 95% credible interval (solid line) for ϕ of fitting the SV-N and SV-S models for the SP&500 index

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

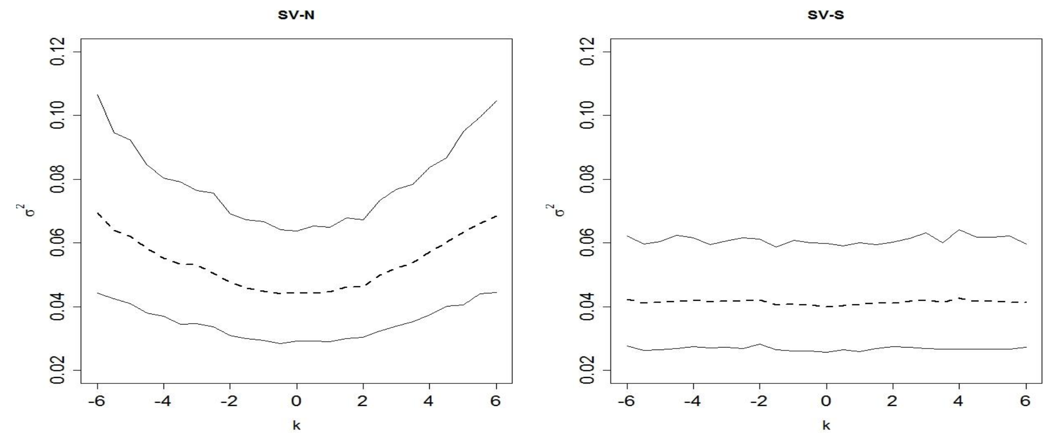

Figure 9: Posterior mean (dashed line) and 95% credible interval (solid line) for σ^2 of fitting the SV-N and SV-S models for the SP&500 index

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, 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. An empirical application is given using the SP&500 index return series, which show that the SV-S model provide better model fitting than the SV-N model in terms of parameter estimates, interpretation and robustness.

In future, we plan to extend our research into other directions of exploring robustness. In this paper, our estimated volatility of financial asset return changes does not accommodate sudden structural changes. 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., 2008) have received considerable attention. We plan to explore our model considering robustness along those lines. Furthermore, our SV-SMN models has shown great flexibility to accommodate outliers, however its robustness aspects could be seriously affected by presence of skewness. Lachos et al. (2008) 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 certain situations, 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 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.

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:

$$p(\alpha \mid \mathbf{h}_{0:T}, \phi, \sigma_\eta^2) \quad \propto \exp\{-\tfrac{a_\alpha}{2}(\alpha - \tfrac{b_\alpha}{a_\alpha})^2\}, \tag{28}$$

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

$$p(\phi \mid \mathbf{h}_{0:T}, \alpha, \sigma_\eta^2) \quad \propto Q(\phi) \exp\{-\tfrac{a_\phi}{2\sigma_\eta^2}(\phi - \tfrac{b_\phi}{a_\phi})^2\} \mathbb{I}_{|\phi|<1} \tag{29}$$

where $Q_\phi = \sqrt{1-\phi^2} \exp\{-\frac{1}{2\sigma_\eta^2}[(1-\phi^2)(h_0 - \frac{\alpha}{1-\phi})^2\}$, $a_\phi = \sum_{t=1}^T h_{t-1}^2 + \frac{\sigma_\eta^2}{\sigma_\phi^2}$, $b_\phi = \sum_{t=1}^T h_{t-1}(h_t - \alpha) + \bar{\phi}\frac{\sigma_\eta^2}{\sigma_\phi^2}$ and $\mathbb{I}_{|\phi|<1}$ is an indicator variable. As $p(\phi \mid \mathbf{h}_{0:T}, \alpha, \sigma_\eta^2)$ in (29) 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 $M_1 = M_0 + [(1-\phi^2)(h_0 - \frac{\alpha}{1-\phi})^2] + \sum_{t=1}^T (h_t - \alpha - \phi h_{t-1})^2$.

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

$$p(\lambda_t \mid y_t, h_t, \nu) \quad \propto \lambda_t^{\frac{\nu+1}{2}-1} \mathrm{e}^{-\frac{\lambda_t}{2}(y_t^2 \mathrm{e}^{-h_t} + \nu)}, \tag{30}$$

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

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

$$p(\nu \mid \boldsymbol{\lambda}_{1:T}) \quad \propto \quad \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} \mathrm{e}^{-\frac{\nu}{2}\sum_{t=1}^T [(\lambda_t - \log \lambda_t) + 2b_\nu]}}{\Gamma(\frac{\nu}{2})} \mathbb{I}_{2<\nu\leq 40}. \tag{31}$$

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 proof the concavity of $\ell(\nu)$, we use the result of Abramowitz and Stegun (1970), in which the $\log\Gamma(\nu)$ could be approximated as

$$\log\Gamma(\nu) \;=\; \frac{\log(2\pi)}{2} + \frac{2\nu-1}{2}\log(\nu) - \nu + \frac{\theta}{12\nu}, \qquad 0<\theta<1. \tag{32}$$

Taking the second derivative of $\ell(\nu)$ from (36) and using (32), we have that

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

- **SV-S case**

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

$$p(\lambda_t \mid y_t, h_t, \nu) \;\propto \lambda_t^{\nu+\frac{1}{2}-1}\mathrm{e}^{-\frac{\lambda_t}{2}y_t^2\mathrm{e}^{-h_t}}\mathbb{I}_{0<\lambda_t<1}, \tag{33}$$

that is $\lambda_t \sim \mathcal{G}_{(0<\lambda_t<1)}(\nu+\frac{1}{2}, \frac{1}{2}y_t^2e^{-h_t})$,i.e., 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

$$p(\nu \mid \mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}) \;\propto \nu^{T+a_\nu-1}\mathrm{e}^{-\nu(b_\nu-\sum_{t=1}^{T}\log\lambda_t)}\mathbb{I}_{\nu>1}. \tag{34}$$

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

$$p(\lambda_t \mid y_t, h_t, \nu) \;\propto \lambda_t^{-\frac{\nu}{2}+\frac{1}{2}-1}\mathrm{e}^{-\frac{1}{2}(\lambda_t y_t^2\mathrm{e}^{-h_t}+\frac{\nu}{\lambda_t})}, \tag{35}$$

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

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

$$p(\nu \mid \mathbf{y}_{1:T}, \mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}) \quad \propto \quad \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} \mathrm{e}^{-\frac{\nu}{2}\sum_{t=1}^{T}[(\frac{1}{\lambda_t}+\log \lambda_t)+2b_\nu]}}{\Gamma(\frac{\nu}{2})} \mathbb{I}_{0<\nu\leq 40} \tag{36}$$

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 $\mathcal{N}_{(0,40)}(\mu_\nu, \sigma_\nu^2)$.

References

Abanto-Valle, C. A., Migon, H. S., Lopes, H. F., 2008. Bayesian modeling of financial returns: A relationship between volatility and trading volume. Technical report 214, Federal University of Rio de Janeiro, Departament of Statistics.

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 leverge 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 heteroskedasticy. 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 Zeland 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 Zeland Journal of Statistics 50, 135–146.

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

Fama, E., 1965. Portfolio analysis in a stable paretian market. Managament 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., David, 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 forthcoming.

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

Lachos, V. H., Ghosh, P., Arellano-Valle, R. B., 2008. Likelihood based inference for skew–normal/independent linear mixed models. Statistica Sinica (to appear).

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. J. Comput. Graph. Stat 2, 175–198.

Liesenfeld, R., Jung, R. C., 2000. Stochastic volatility models: Conditional normality versus heavy-tailed distrutions. Journal of Applied Econometics 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.

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 relationshis 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.

Watanabe, T., Omori, Y., 2004. A multi-move sampler for estimate non-Gaussian time series model: Comments on Shepard and Pitt (1997). Bimetrika 91, 246–248.

Yu, J., 2005. On leverage in stochastic volatility model. Journal of Econometrics 127, 165–178.