Stochastic volatility with heavy-tails: An approximate Bayesian approach using Hidden Markov Models

Carlos A. Abanto-Valle†, Dipak K. Dey*, Ming-Hui Chen* and Victor H. Lachos*

† Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970, Rio de Janeiro, Brazil

*Department of Statistics, University of Connecticut, U-4120, Storrs, CT 06269, USA

Abstract

In this article, we introduce an approximated Bayesian estimation method for stochastic volatility (SV) models with scale mixtures of normal (SMN) distributions (Abanto-Valle et al., 2010). Bayesian estimation of the parameters of stochastic volatility models in general, and SV models with SMN distributions in particular, is usually regarded as challenging as the likelihood is a high-dimensional multiple integral. Our estimation method is based on the fact that the powerful hidden Markov model (HMM) machinery can be applied in order to evaluate an arbitrarily accurate approximation of the likelihood of an SV model with SMN distributions. Second, we get the posterior mode by using a numerical optimization routine and third, we use importance sampling to sample from the posterior distribution of the parameters using a multivariate normal distribution with mean and variance given by the posterior mode and the inverse of the Hessian matrix evaluated at the posterior mode. However, the HMM approximation leads to simple formula for forecast distributions, for computing appropriately defined residuals, and for decoding, i.e., estimating the volatility of the process. The methods developed are applied to analyze daily stock returns data on S&P500 index. Bayesian model selection criteria results reveal that the SV models based on heavy-tailed SMN distributions provide significant improvement in model fit as well as prediction to the S&P500 index data over the usual normal model.

Keywords: non-Gaussian and nonlinear state-space models, scale mixture of normal distributions, stochastic volatility, Value-at-Risk.

1 Introduction

Stochastic volatility (SV) models were introduced by Taylor (1982; 1986), providing a flexible and parsimonious means of modeling time-varying variances of financial asset returns. These models have significant financial and economic theoretical appeal (Melino and Turnbull, 1990) because of their ability to capture the main empirical properties, i.e., the stylized facts, often observed in daily series of financial returns (Carnero et al., 2004). These properties are crucial not only for describing the return distributions but also for asset allocation, option pricing, forecasting and risk management.

The basic SV model introduced by Taylor (1982) can be written as

$$y_t = e^{\frac{h_t}{2}}\varepsilon_t, \tag{1a}$$

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

where y_t and h_t are the compunded return and the log-volatility on time t, μ, ϕ and σ_η^2 are the mean, the persistence parameter and the volatility of the log-volatility process, respectively, and $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent sequences of independent standard normal random variables. We assume that $|\phi| < 1$, that is the log-volatility is a stationary process. We denote this basic SV model as SV-N.

Despite theoretical advantages, SV models have not been as popular as the GARCH models (Bollerslev, 1986) in practical applications. The main reason is that the likelihood of the SV models is given by a higher-order multiple integral that is analytically intractable. Hence, SV models require much more complex inferential and computational tools. Several estimation methods have been proposed in the literature, for example the quasi-likelihood methods (Harvey et al., 1994), the generalized methods of moments (Andersen and Sorensen, 1996), and simulated maximum likelihood (Danielsson, 1994). Recently Langrock (2011) and Langrock et al. (2012) apply an alternative frequentist estimation method, numerically maximizing an approximation of the likelihood function which is efficiently evaluated using recursive techniques routinely applied for hidden Markov models (HMMs). By the other hand, Bayesian estimation is performed using Markov Chain Monte Carlo (MCMC) methods.

The MCMC approach for SV models is the usual methodology since the seminal work by Jacquier et al. (1994), where Bayesian inference for standard SV models was first developed using 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, approximated 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), Watanabe and Omori (2004) and Omori and Watanabe (2008) suggested the use of random blocks containing some of the components of the log-volatilities in order to reduce the autocorrelation effectively. Meyer et al. (2003) and Skaug and Yu (2008) used the Laplace approximation to compute the likelihood function by integrating out all unknown log-volatilities and applying the Metropolis-Hasting algorithm to sampling the parameters. However, in all of these, the normal distribution was assumed as the basis for parameter inference. Recently, the integrated nested Laplace approximations (INLA) was used inside the R package (R-INLA) to fit SV models (Martino et al., 2010; Ehlers and Zevallos, 2015). However it is restricted to the normal, Student-t and the normal inverse gamma distributions and it is not a trivial task, to generalize inside the R-INLA, to other distributions.

The SV-N model is unable to capture the usual heavy-tailed behavior of conditional distribution of the returns, since they are assumed to be Gaussian. Liesenfeld and Jung (2000), Chib et al. (2002), Jacquier et al. (2004) and Abanto-Valle et al. (2010), amongst others, had provided consistent evidence that leptokurtic distributions, such as the Student's t, the generalized error or the scale mixture of normal (SMN) distributions should be used in order to capture this feature. In this context, the 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.

Abanto-Valle et al. (2010) extended the basic SV model by assuming the flexible class of scale mixtures of normal (SMN) distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993; Chow and Chan, 2008). By using a mixture representation of the SMN distribution they proposed an efficient block sampler to sample the log-volatilities. The SV-SMN class of models contains the normal (SV–N), Student-t (SV–T), slash (SV-S) and variance gamma (SV–VG) distributions. All these distributions have heavier tails than the normal distribution, and thus can be used for robust

inference in these types of models.

In this paper, we apply an alternative Bayesian estimation method to the SV-SMN class of models (Abanto-Valle et al., 2010) without using the mixture representation of the SMN distributions. First, we approximate the likelihood function by integrating out the log-volatilities as suggested by Langrock (2011) and Langrock et al. (2012). Second, we get the maximum a posteriori by using a numerical optimization routine and third, we use importance sampling to sample from the posterior distribution of the parameters using a multivariate normal distribution with mean and variance given by the maximum a posteriori and the inverse of the Hessian matrix evaluated at the maximum a posteriori, respectively.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction to SMN distributions. Section 3 outlines the general class of SV–SMN models as well as the likelihood approximation and evaluation procedure using HMM methods and the Bayesian approach for parameter estimation. Section 4 is devoted to model checking and model selection. In Section 5, we carry out a detailed analysis of the S & P 500 data. Some concluding remarks as well as future developments are deferred to Section 6.

2 Scale mixture of normal distributions

A random variable Y belongs to the SMN family if it can be expressed as

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

where μ is a location parameter, $X \sim \mathcal{N}(0, \sigma^2)$, λ is a positive mixing random variable with cumulative distribution function (*cdf*) $H(\cdot \mid \boldsymbol{\nu})$ and probability density function (*pdf*) $h(\cdot|\boldsymbol{\nu})$, $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of λ, and $\kappa(\cdot)$ is a positive weight function. As in Lange and Sinsheimer (1993) and Choy et al. (2008), we restrict our attention to the case where $\kappa(\lambda) = 1/\lambda$. Given λ, we have $Y|\lambda \sim \mathcal{N}(\mu, \lambda^{-1}\sigma^2)$, and the *pdf* of Y is given by

$$f_{SMN}(y|\mu, \sigma^2, \nu) = \int_{-\infty}^{\infty} \phi(y|\mu, \lambda^{-1}\sigma^2) dH(\lambda|\boldsymbol{\nu}), \tag{3}$$

where $\phi(\cdot \mid \mu, \sigma^2)$ denotes the density of the univariate $\mathcal{N}(\mu, \sigma^2)$ distribution. From equation (3), the *cdf* of the SMN distributions is given by

$$\begin{aligned} F_{SMN}(y|\mu, \sigma^2, \nu) &= \int_{-\infty}^{y} \int_{-\infty}^{\infty} \phi(u|\mu, \lambda^{-1}\sigma^2) dH(\lambda|\boldsymbol{\nu}) du \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{\lambda^{1/2}[y-\mu]}{\sigma}\right) dH(\lambda|\boldsymbol{\nu}), \end{aligned} \tag{4}$$

where $\Phi(\cdot)$ is the *cdf* of the standard normal distribution. The notation $Y \sim \mathcal{SMN}(\mu, \sigma^2, \boldsymbol{\nu}; H)$ will be used when Y has *pdf* (3) and *cdf* (4). As mentioned above, the SMN family constitutes a class of heavy-tailed distributions, including the normal, Student-t, and Slash distributions, which are obtained, respectively, by choosing the mixing variables as: $\lambda = 1$, $\lambda \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, $\lambda \sim \mathcal{B}e(\nu, 1)$, and $\lambda \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$, where $\mathcal{G}(\cdot, \cdot)$, $\mathcal{B}e(\cdot, \cdot)$ and $\mathcal{IG}(., .)$ denote the gamma, beta and the inverse gamma distributions, respectively.

3 The heavy-tailed stochastic volatility model

3.1 Model formulation

In order to capture heavy-tailed features in the marginal distribution of the returns errors, we modify the basic specification of the SV-N model (Taylor, 1982; 1986) replacing the normality assumption of ε_t in equation (1b) by the SMN class of distribution. The resulting model can be expressed as follows:

$$y_t = e^{\frac{h_t}{2}} \varepsilon_t, \tag{5a}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \tag{5b}$$

$$\varepsilon_t \sim SMN(0, 1, \nu), \;\; \eta_t \sim \mathcal{N}(0, 1), \tag{5c}$$

where y_t and h_t are respectively the compounded return and the log-volatility on time t. As before, μ, ϕ and σ_η^2 are the mean, the persistence parameter and the volatility of the log-volatility process, respectively, and ε_t and η_t assumed to be independent. We refer to this generalization as SV-SMN models.

As depicted in Section 2, this class of models include 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 dristibution and thus provide an appealing robust alternative to the usual Gaussian process in SV models

3.2 Likelihood evaluation by iterated numerical integration

To formulate the likelihood in the class of the SV-SMN models, we require the conditional *pdfs* of the random variables y_t, given h_t $(t = 1, \ldots, T)$, and of the random variables h_t, given h_{t-1} $(t = 2, \ldots, T)$. We denote these by $p(y_t \mid h_t)$ and $p(h_t \mid h_{t-1})$, respectively. For any member in the class of SMN distributions, the likelihood of the model defined by equations (5a) , (5b) and (5c) can then be derived as

$$\begin{aligned}
\mathcal{L} &= \int \ldots \int p(y_1, \ldots, y_T, h_1, \ldots, h_T) dh_T \ldots dh_1 \\
&= \int \ldots \int p(y_1, \ldots, y_T \mid h_1, \ldots, h_T) p(h_1, \ldots, h_T) dh_T \ldots dh_1 \\
&= \int \ldots \int p(h_1) p(y_1 \mid h_1) \prod_{t=2}^{T} p(y_t \mid y_{t-1}, h_t) p(h_t \mid h_{t-1}) dh_T \ldots dh_1.
\end{aligned}$$

Hence, the likelihood is a higher-order multiple integral that cannot be evaluated analytically. Through numerical integration, using a simple rectangular rule based on m equidistant intervals, $B_i = (b_{i-1}, b_i)$, $i = 1, \ldots, m$, with midpoints b_i^* and length b, the likelihood can be approximated as follows:

$$\begin{aligned}
\mathcal{L} \approx b^T \sum_{i_1=1}^{m} \ldots \sum_{i_T=1}^{m} & p(h_1 = b_{i_1}^*) p(y_1 \mid h_1 = b_{i_1}^*) \\
& \times \prod_{t=2}^{T} p(h_t = b_{i_t}^* \mid h_{t-1} = b_{i_{t-1}}^*) p(y_t \mid h_t = b_{i_t}^*) = \mathcal{L}_{\text{approx}}\,. \qquad (6)
\end{aligned}$$

This approximation can be made arbitrarily accurate by increasing m, provided that the interval (b_0, b_m) covers the essential range of the log-volatility process. We note that this simple midpoint quadrature is by no means the only way in which the integral can be approximated (cf. Langrock et al., 2012).

3.3 Fast evaluation of the approximate likelihood using HMM techniques

The approximate likelihood, in the form given in (6), can be evaluated numerically, but the evaluation will usually be computationally intractable since it involves m^T summands. However, instead of the brute force summation in (6), an efficient recursive scheme can be used to evaluate the approximate likelihood. To see this, we note that the numerical integration essentially corresponds to a discretization of the state space, i.e., the support of the log-volatility process h_t. Therefore, the approximate likelihood given in (6) can be evaluated using the tools well-established for HMMs, which are the models that have exactly the same dependence structure as the stochastic volatility models, but with a finite and hence discrete state space (cf. Langrock, 2011; Langrock et al., 2012). In the given scenario, the discrete states correspond to the intervals B_i, $i = 1, \ldots, m$, in which the state space has been partitioned. A key property of HMM, which we exploit here, is that the likelihood can be evaluated efficiently using the so-called forward algorithm, a recursive scheme which iteratively traverses forward along the time series, updating the likelihood and the state probabilities in each step (Zucchini et al., 2016). For an HMM, applying the forward algorithm results in a convenient closed-form matrix product expression for the likelihood, and this is exactly what is obtained also for the SVM-SMN class of models:

$$\mathcal{L}_{\text{approx}} = \boldsymbol{\delta}\mathbf{P}(y_1)\boldsymbol{\Gamma}\mathbf{P}(y_2)\boldsymbol{\Gamma}\mathbf{P}(y_3)\cdots\boldsymbol{\Gamma}\mathbf{P}(y_{T-1})\boldsymbol{\Gamma}\mathbf{P}(y_T)\mathbf{1}^\top . \tag{7}$$

Here, the $m \times m$-matrix $\boldsymbol{\Gamma} = \left(\gamma_{ij}\right)$ is the analogue to the transition probability matrix in case of an HMM, defined by $\gamma_{ij} = p(h_t = b_j^* \mid h_{t-1} = b_i^*) \cdot b$, which is an approximation of the probability of the log-volatility process changing from some other value in the interval B_i to some value in the interval B_j, this conditional probability is determined by Eq. (5b). The vector $\boldsymbol{\delta}$ is the analogue to the Markov chain initial distribution in case of an HMM, here defined such that δ_i is the density of the $\mathcal{N}(\mu, \frac{\sigma_\eta^2}{1-\phi^2})$-distribution — the stationary distribution of the log-volatility process — multiplied by b. Furthermore, $\mathbf{P}(y_t)$ is an $m \times m$ diagonal matrix with the ith diagonal entry $p(y_t \mid h_t = b_i^*)$, hence the analogue to the matrix comprising the state-dependent probabilities in case of an HMM; this conditional probability is determined by Eq. (5a). Finally, $\mathbf{1}^\top$ is a column vector of ones. Using the matrix product expression given in (7), the computational effort required to evaluate the approximate likelihood is linear in the number of observations, T, and quadratic in the number of

intervals used in the discretization, m.

In practice, this means that the likelihood can typically be calculated in a fraction of a second, even for T in the thousands and say $m = 100$, a value which renders the approximation virtually exact (see the simulation experiments below). Furthermore, the approximation can be made arbitrarily accurate by increasing m (and potentially widening the interval $[b_0, b_m]$).

It should perhaps be noted here that, although we are using the HMM forward algorithm to evaluate the (approximate) likelihood, the specifications of $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$ and $\mathbf{P}(y_t)$ given above do not define exactly an HMM, since in general the row sums of $\boldsymbol{\Gamma}$ will only approximately equal one, and the components of the vector $\boldsymbol{\delta}$ will only approximately sum to one. If desired, this can be remedied by scaling each row of $\boldsymbol{\Gamma}$ and the vector $\boldsymbol{\delta}$ to total 1.

3.4 Bayesian Inference for the heavy-tailed SV class of models

Because we have some constraint in the original parametric space ($\{|\phi| < 1, \sigma_\eta > 0, \nu > 0\}$) of the SV-SMN class of models, we consider the following transformations for the common parameters as follows: $\psi = \log\left(\frac{1+\phi}{1-\phi}\right)$, $\omega = \log(\sigma)$, and $\kappa = \log(\nu)$. Let $\boldsymbol{\theta} = (\mu, \psi, \omega, \kappa)^\top$ and $p(\boldsymbol{\theta})$ be the prior distribution of $\boldsymbol{\theta}$. From equation (7), we then obtain the posterior distribution up to a normalization constant

$$p(\boldsymbol{\theta} \mid \mathbf{y}_T) \propto p(\boldsymbol{\theta}) \mathcal{L}_{\text{approx}}(\boldsymbol{\theta}). \tag{8}$$

Suposse we wish to calculate an expectation $E_{p(\boldsymbol{\theta}|\mathbf{y}_T)}[h(\boldsymbol{\theta})]$, which can be calculated as

$$\begin{aligned} E_{p(\theta|\mathbf{y}_T)}[h(\boldsymbol{\theta})] &= \frac{\int h(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{y}_T) d\boldsymbol{\theta}}{\int p(\boldsymbol{\theta} \mid \mathbf{y}_T) d\boldsymbol{\theta}} \\ &= \frac{\int \frac{h(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_T)}{q(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int \frac{p(\boldsymbol{\theta}|\mathbf{y}_T)}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}} = \frac{E_{q(\boldsymbol{\theta})}\left[h(\boldsymbol{\theta}) w(\boldsymbol{\omega}(\boldsymbol{\theta})\right]}{E_{q(\boldsymbol{\theta})}\left[\boldsymbol{\omega}(\boldsymbol{\theta})\right]}, \end{aligned} \tag{9}$$

where $\boldsymbol{\omega}(\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta}|\mathbf{y}_T)}{q(\boldsymbol{\theta})}$ and now $E_q[.]$ denotes an expected value with respect to $q(\boldsymbol{\theta})$. Therefore a sample of independent draws $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ from $q(\boldsymbol{\theta})$ can be used to estimate $E_{p(\theta|\mathbf{y}_T)}[h(\boldsymbol{\theta})]$ by

$$\bar{h} = \frac{\sum_{i=1}^m \boldsymbol{\omega}(\theta_i) h(\boldsymbol{\theta}_i)}{\sum_{i=1}^m \boldsymbol{\omega}(\theta_i)}. \tag{10}$$

It is shown that using one sample $\theta_i's$ in estimating the ratio in (9) is more efficient than using two samples (one for the numerator and another for denominator) (Chen et al., 2008). It follows from the strong law of large numbers that $\bar{h} \to E_{p(\theta|\mathbf{y}_T)}[h(\boldsymbol{\theta})]$ as $m \to \infty$ almost surely (Geweke, 1989). A variance of $\bar{h}(\boldsymbol{\theta})$ can be consistenly estimated by $\sum_{i=1}^{m} \boldsymbol{\omega}(\boldsymbol{\theta}_i)^2[h(\boldsymbol{\theta}_i) - \bar{h}]^2 / [\sum_{i=1}^{m} \boldsymbol{\omega}(\boldsymbol{\theta}_i)]^2$.

4 Model checking and model selection

4.1 Model checking

The HMM forward algorithm can also be used to obtain forecast distributions for SV-SMN models. For example, it is straightforward to find the cumulative distribution function of the one-step-ahead forecast distribution on day $t-1$, i.e., the conditional distribution of the return on day t, given all previous observations. This is given by

$$F(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_0) \approx \sum_{i=1}^{m} \zeta_i F(y_t \mid y_{t-1}, h_t = b_i^*), \tag{11}$$

where ζ_i is the ith entry of the vector $\boldsymbol{\alpha}_{t-1}/(\boldsymbol{\alpha}_{t-1}\mathbf{1}^\top)$, obtained from the forward probabilities,

$$\boldsymbol{\alpha}_{t-1} = \boldsymbol{\delta}\mathbf{P}(y_1)\boldsymbol{\Gamma}\mathbf{P}(y_2)\boldsymbol{\Gamma}\mathbf{P}(y_3)\cdots\boldsymbol{\Gamma}\mathbf{P}(y_{t-1}),$$

with $\boldsymbol{\delta}$, $\mathbf{P}(y_k)$ and $\boldsymbol{\Gamma}$ defined as above. The corresponding expression for longer forecast horizons is similar (see Chapter 5 of Zucchini et al., 2016, for details). The approximation in (11) usually becomes virtually exact for values of m about 100. A closed-form expression for obtaining state predictions, i.e., volatility predictions, is also available. Furthermore, the forecast distribution given in Eq. (11) can be used in order to perform model checking via residuals (Kim et al., 1998). The one-step-ahead forecast pseudo-residual (or quantile residual) is given by

$$r_t = \Phi^{-1}(F(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_0)), \tag{12}$$

for $t = 1, \ldots, T$. For a correctly specified model, the r_t follows a standard normal distribution (Rosenblatt, 1952; Smith, 1985; Kim et al., 1998; Gerlach et al., 1999; Liesenfeld and Richard, 2003). Thus, forecast pseudo-residuals can be used to identify extreme values, and the suitability of the model can be checked by using, for example, QQ-plots or formal tests for normality.

4.2 Model comparison criteria

Given the wide range of candidate models, it has become increasingly important to be able to discriminate between competing models for a given application. One of the most used methods to compare several competing models for a given data set is derived from the conditional predictive ordinate CPO statistics see (Gelfand et al., 1992; Chen et al., 2000). Let $\mathbf{y}_T = (y_1, y_2, \cdots, y_T)$ be the full data and $\mathbf{y}_{(i)} = (y_1, y_2, \cdots, y_{i-1}, y_{i+1}, \cdots, y_T)$ denote the data with the ith observation deleted. For the ith observation, the CPO is defined as

$$\text{CPO}_i = f(y_i \mid y_{(i)}) = \int_{\boldsymbol{\theta}\in\Theta} f(y_i \mid \boldsymbol{\theta}, y_{(i)})\pi(\boldsymbol{\theta} \mid \mathbf{y}_{(i)})d\boldsymbol{\theta}.$$

The CPO is a cross-validated predictive approach, *i.e.*, it relies on predictive distributions conditioned on the observed data with a single data point deleted. In this paper, we propose a modified version of the CPO called the autoregressive CPO. This CPO, known as prequential CPO, is computed for each new observation using only the data from previous time periods (Dawid, 1984).

Let $\mathbf{y}_t = (y_1, y_2, \cdots, y_t)$ be the data vector at time point t. The conditional predictive ordinate (CPO) based on the forecasting predictive distribution of Y_t given $\mathbf{Y}_{t-1}$ is defined as

$$f(y_t \mid \mathbf{y}_{t-1}) = \left(\int \frac{f(\mathbf{y}_{t-1} \mid \theta)}{f(\mathbf{y}_T \mid \theta)}\pi(\theta \mid \mathbf{y}_T)d\theta\right)^{-1}\left(\int \frac{f(\mathbf{y}_t \mid \theta)}{f(\mathbf{y}_T \mid \theta)}\pi(\theta \mid \mathbf{y}_T)d\theta\right), \tag{13}$$

where θ is a parameter, $f(\mathbf{y}_T \mid \theta)$ is the likelihood function, and $\pi(\theta \mid \mathbf{y})$ is the posterior density function. The CPO given by (13) suggests what values of y_t are likely, given that the model was fitted with observations $y_1, \cdots, y_{t-1}$, and it is possible to see whether the observation supports the model. Since for the proposed model a closed form of the CPO_t is not available, a Monte Carlo estimate of the CPO_t is obtained by using a sample from the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathbf{y}_T)$. Let $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_Q$ be a sample of size Q of $\pi(\boldsymbol{\theta} \mid \mathbf{y}_T)$. A Monte Carlo approximation of the CPO_t is given by the following expression:

$$\widehat{\text{CPO}_t} = \left(\frac{1}{Q}\sum_{q=1}^{Q}\frac{f(\mathbf{y}_{t-1} \mid \boldsymbol{\theta}_q)}{f(\mathbf{y}_T \mid \boldsymbol{\theta}_q)}\right)^{-1}\left(\frac{1}{Q}\sum_{q=1}^{Q}\frac{f(\mathbf{y}_t \mid \boldsymbol{\theta}_q)}{f(\mathbf{y}_T \mid \boldsymbol{\theta}_q)}\right).$$

In addition, a summary statistic of the CPO_t's is the log pseudo marginal likelihood (LPML), defined by LPML$= \sum_{t=1}^{T}\log(\widehat{\text{CPO}_t})$. Models with greater LMPL values have better fit. Another

popular metric of summary statistics for Bayesian model comparison is the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002). This criterion is based on the posterior mean of the deviance. It can be approximated by $\overline{D} = \sum_{q=1}^{Q} D(\boldsymbol{\theta}_q)/Q$, where $D(\boldsymbol{\theta}) = -2\log f(\mathbf{y}_T \mid \boldsymbol{\theta}) = -2\log \mathcal{L}(\boldsymbol{\theta})$. The DIC can be estimated using the Monte Carlo output by $\widehat{\text{DIC}} = \overline{D} + \widehat{p_D} = 2\overline{D} - D(\bar{\boldsymbol{\theta}})$, where $\widehat{p_D}$ is the effective number of parameters an can be evaluated as $\widehat{p_D} = \bar{D} - D(\bar{\boldsymbol{\theta}})$. Given the comparison of two alternative models, the model that best fits a data set is the model with the smallest DIC value. Note that it is important to integrate out all latent variables in the deviance calculation as this yields a more appropriate penalty term $\widehat{p_D}$; (see Kim et al., 2002). For all these criteria, the evaluation of the likelihood function $\mathcal{L}(\boldsymbol{\theta})$ is a key aspect. However, for the SV-SMN class of models it can be evaluated using results given in Subsection 3.2 and 3.4.

Finally, the expected Akaike information criterion (EAIC), the expected Bayesian information criterion (EBIC) and the Log-Predictive Score (LPS), can be estimated by $\widehat{\text{EAIC}} = \overline{D} + 2\sharp(\vartheta)$ and $\widehat{\text{EBIC}} = \overline{D} + \sharp(\vartheta)\log(T)$, where $\sharp(\vartheta)$ is the number of model parameters (see Brooks, 2002), and $LPS = -\frac{1}{T}\sum_{t=1}^{T} \log p(y_t \mid \mathbf{y}_{t-1}, \bar{\boldsymbol{\theta}})$ (Delatola and Griffin, 2011; Abanto-Valle et al., 2014; 2015; Leão et al., 2017).

5 Empirical Application

This section analyzes the daily closing prices for the S&P500 stock market index. The S&P500 index contains the stocks of 500 Large-Cap corporations. Although a majority of those corporations are US based, it also includes other companies having their common stocks within the index. The data set was obtained from the Yahoo finance web site available to download at http://finance.yahoo.com. The period of analysis is January 8, 1990 - December 31, 2012, which yields 5793 observations. Throughout, we will work with the compounded returns computed as: $y_t = 100 \times (\log P_t - \log P_{t-1})$, where P_t is the closing price on day t.

Table 1 summarizes descriptive statistics for the compounded returns with the time series plot in Figure 1. For the returns series, the basic statistics viz. the mean, standard deviation, skewness and kurtosis are calculated to be 0.02, 1.17, -0.23 and 11.47, respectively. Note that the kurtosis of the returns is > 3, so that daily S &P 500 returns likely shows a departure from the underlying

Table 1: Summary statistics for S &P 500 return series

	mean	s.d.	max	min	skewness	kurtosis
Returns	0.02	1.17	10.95	-9.047	-0.23	11.47

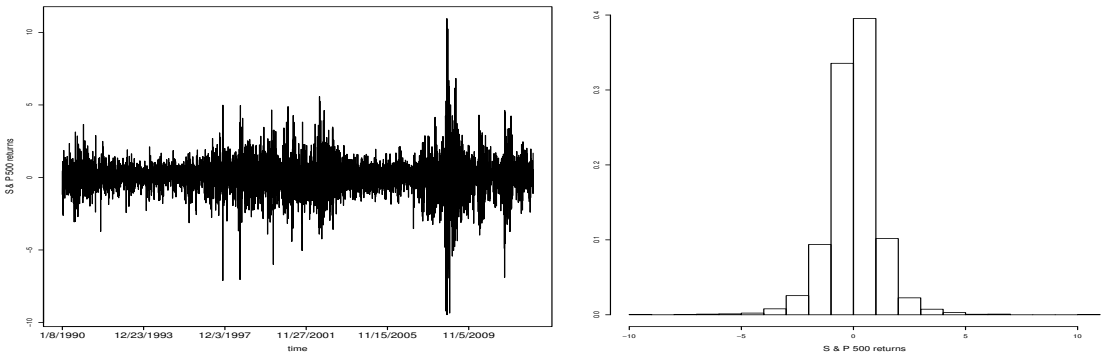

Figure 1: S&P 500 compounded returns with sample period from January 8, 1990 to December 31, 2012. The left panel shows the plot of the raw series and the right panel plots the histogram of returns.

normality assumption. Thus, we reanalyze this data with the aim of providing robust inference 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.

We set the prior distributions of the common parameters of the SV-N, SV-T, SV-S and SV-VG models as: $\mu \sim \mathcal{N}(0, 100)$, $\psi \sim \mathcal{N}(4.5, 100)$, $\omega \sim \mathcal{N}(-1.5, 100)$. The prior distributions on the transformed shape parameters were chosen as: $\kappa \sim \mathcal{N}(2.5, 100)$ (SV-T), $\kappa \sim \mathcal{N}(0.7, 100)$ (SV-S) and $\kappa \sim \mathcal{N}(2.5, 100)$ (SV-VG), respectively. All these priors are slightly flat.

Now, we apply the methodology described in Section 3 to fit the SV-N, SV-T, SV-S and SV-VG models. All the calculations were performed using stand-alone code developed by the authors using the Rcpp interface inside R. First, we approximate the likelihood using $b_m = -b_0 = 3$ and $m = 50, 100, 150, 200$. Second, we obtain numerically the posterior mode using the *optim* routine in the R package. Table 2 reports the results for each model fitted here. It is important to observe that for $m = 100, 150, 200$, the results are almost the same. In the SV-S and the SV-VG the computing time increases because we do not have a R function to evaluate the density of the slash

and variance gamma distribution and we do it using numerical integration. Finally, we apply the importance sampling algorithm to draw a random sample from the posterior distribution of the parameters using a multivariate normal distribution with mean and variance given by the posterior mode the inverse of the Hessian matrix evaluated at the posterior mode, using $b_m = -b_0 = 3$ and $m = 100$, respectively. For each model, we draw a sample of size 500.

Based on the sample of size 500 obtained by importance sampling, we calculate estimates of the posterior means and variances of μ, ϕ, σ_η^2, ν as described in Section 3.4 equations (9)-(10). Table 3 reports the posterior mean, 95% credible interval and variance of the estimators. For all models considered, we find that the posterior means of ϕ were very close to 0.99, indicating a high persistence of the log-volatility process. The persistence in the log-volatility process underlying the SV-N model was smaller than that obtained under the other three models. The posterior mean of σ_η under the SV-T, SV-S and SV-VG were smaller than the one under the SV-N, indicating that the volatility processes of the SV-T, SV-S and SV-VG were less variable than those in the SV-N case.

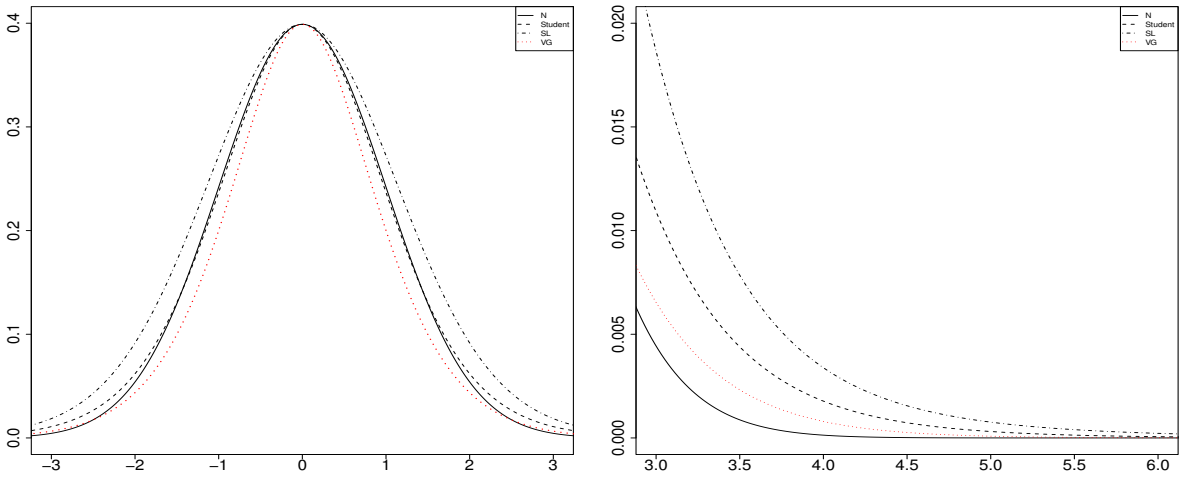

Figure 2: Density curves of the univariate normal, Student-t, slash and variance gamma using the estimated tail-fatness parameter from the respective SV model.

The posterior means of ν are 11.2189, 2.8899 and 7.7817 for the SV-T, SV-S and SV-VG models,

Table 2: Posterior mode of the parameters obtained when fitting the SV-N, SV-T, SV-S and SV-VG models to the S & P500 returns (using $m = 50, 100, 150, 200$ and $b_{max} = -b_{min} = 3$) and time in minutes to get the posterior mode.

SVM-N					
m	μ	ψ	ω	κ	time
50	-0.1993	5.2685	-2.0206	–	0.09
100	-0.2018	5.2672	-2.0195	–	0.19
150	-0.2021	5.2673	-2.0196	–	0.38
200	-0.2021	5.2673	-2.0196	–	0.64
SV-T					
m	μ	ψ	ω	κ	time
50	-0.3743	5.6124	-2.2085	2.3942	0.65
100	-0.3757	5.6073	-2.2071	2.3947	1.29
150	-0.3757	5.6073	-2.2071	2.3947	2.07
200	-0.3757	5.6073	-2.2071	2.3947	3.23
SV-S					
m	μ	ψ	ω	κ	time
50	-0.6116	5.5272	-2.1593	1.0477	4.57
100	-0.6110	5.5271	-2.1592	1.4095	8.05
150	-0.6110	5.5271	-2.1592	1.4095	12.34
200	-0.6110	5.5271	-2.1592	1.4095	16.77
SV-VG					
m	μ	ψ	ω	κ	time
50	-0.1699	5.6711	-2.2427	1.9977	45.53
100	-0.1699	5.6711	-2.2427	1.9977	98.45
150	-0.1699	5.6711	-2.2427	1.9977	158.21
200	-0.1699	5.6711	-2.2427	1.9977	197.35

Table 3: Estimation results for the S&P500 daily index returns. The first row: Posterior mean. The second row: Posterior 95% credible interval in parentheses. The third row: standard error of the posterior mean.

Parameter	SV-N	SV-T	SV-S	SV-VG
	-0.2122	-0.3891	-0.6069	-0.1333
μ	(-0.5260,0.1263)	(-0.7551,0.0287)	(-0.9963,-0.2667)	(-0.5328,0.2554)
	0.0089	0.0177	0.0131	0.0237
	0.9906	0.9933	0.9927	0.9939
ϕ	(0.9839,0.99936)	(0.9877,0.9956)	(0.9871,0.9952)	(0.9882,0.9958)
	1.12×10^{-4}	2.57×10^{-4}	1.15×10^{-4}	1.12×10^{-4}
	0.0173	0.0121	0.0133	0.0114
σ_η^2	(0.0126,0.0241)	(0.0086, 0.0175)	(0.0094,0.0187)	(0.0079, 0.0158)
	1.39×10^{-4}	8.79×10^{-4}	5.29×10^{-4}	1.31×10^{-4}
	—	11.2189	2.8899	7.7817
ν	—	(7.8815,14.6406)	(2.3395,3.4719)	(5.3534,10.0791)
	—	0.2844	0.0012	0.06762

respectively. These results seem to indicate that the measurement errors of the stock returns are better explained by heavy-tailed distributions. The reason why the estimated volatility of the SV-SMN models is more persistent and less variable can be understood by comparing the densities of these distributions. To illustrate the tail behavior, we plot the normal ($\mathcal{N}(0,1)$) density, Student-t ($\mathcal{T}(0,1,\nu)$) density with ν degrees of freedom, the slash ($\mathcal{S}(0,1,\nu)$) density with shape parameter ν and the variance gamma with shape parameter ($\mathcal{VG}(0,1,\nu)$). We set ν as the posterior mean of the respective SV model (see Table 3 for details). Figure 2 depicts the four density curves (the Student-t, slash and variance gamma distributions have been rescaled to be comparable; see Wang and Genton, 2006, for details). All the distributions have heavier tails than that of the normal distribution. Note that the slash distribution has a fatter tail than the other distributions that we have considered (see Figure 2 right panel). Therefore, the SV-T, SV-S and SV-VG models attribute a relatively larger proportion of extreme return values ϵ_t instead of η_t than the SV-N model, making the volatility of the SV-T, SV-S and SV-VG models less variable. It also increases the persistence of these models' volatility. This interpretation is confirmed by comparing the volatility estimates.

In Figure 3, we plot the decoded volatilities $e^{\frac{h_t}{2}}$. We find smoother trajectories under the models SV-T, SV-S and SV-VG than for the SV-N model. Extreme returns, such as during the sub-prime crisis, make the differences clear. This can have a substantial impact, for instance, on the valuation of derivative instruments and several strategic or tactical asset allocation topics.

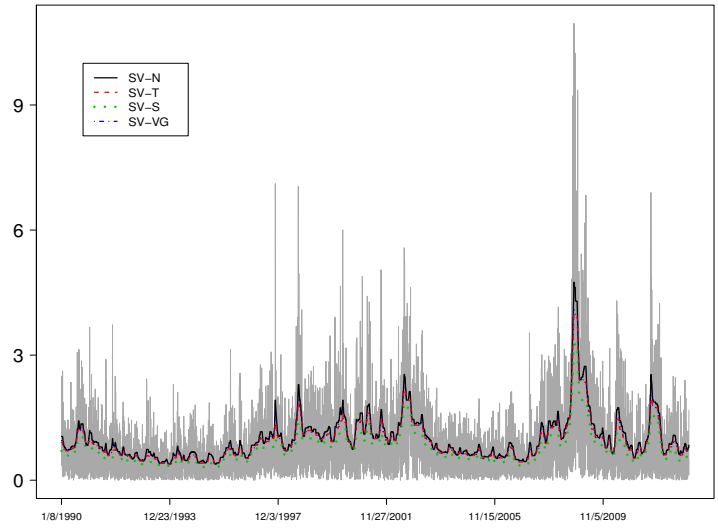

Figure 3: Decoded $e^{\frac{h_t}{2}}$ using the viterbi algorithm. The solid line (SV-N), dotted red line (SV-T), dotted green line (SV-S) and the dootted blue line (SV-VG). The grey line indicates the absolute returns.

Next, we compare the models in terms of their relative in-sample fit using the DIC, LPS, EAIC, EBIC and LPML criteria. Models with greater LMPL values have better fit and models with smallest EAIC, EBIC and LPS have the best fit. In table 4, we report the results. All the criteria considered here indicate that the SV-VG model has the best fit.

We additionally performed an out-of-sample analysis of the forecast performance for the models covered in Table 4. For the observation period January 8, 1990 until May 19, 2017, each return series was divided into a calibration and a validation sample:

- Calibration sample: from January 8, 1998 until December 31, 2012.

Table 4: SP&500 return data set. DIC: deviance information criterion, LPS: log-predictive score, EAIC: expected AIC, EBIC: expected BIC. LPML:log-pseudo marginal likelihood.

	DIC		LPS		EAIC		EBIC		LPML	
Model	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking
SV-N	15705.3	4	1.3550	4	15708.4	4	15728.3	4	-7861.92	4
SV-T	15663.2	2	1.3512	2	15667.3	2	15694.0	2	-7843.99	2
SV-S	15672.4	3	1.3520	3	15676.5	3	15703.2	3	-7844.91	3
SV-VG	15656.0	1	1.3506	1	15660.1	1	15686.8	1	-7836.57	1

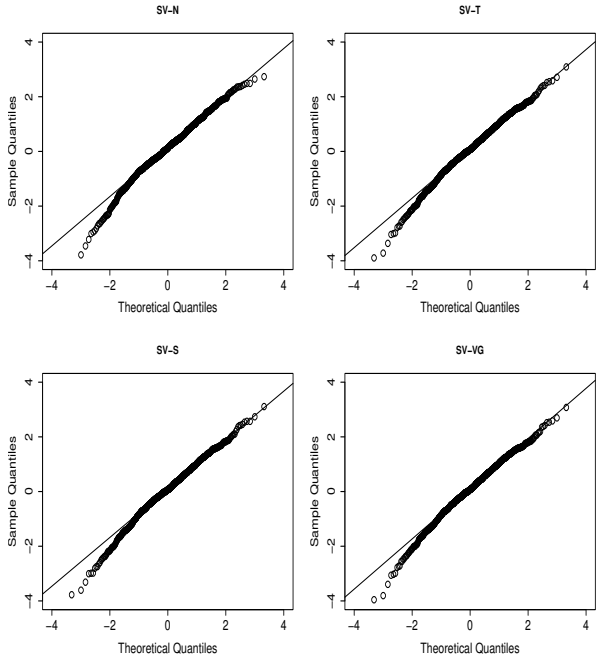

Figure 4: QQ-plot of the forecast pseudo-residuals: SV-N (top-left), SV-T (top-right), SV-S (bottom-left) and SV-VG (bottom-right) for the IBVSP returns.

- Validation sample: from January 2, 2013 until May 19, 2017.

As a first step, the SV-N, SV-T, SV-S and SV-VG models were fitted to the calibration sample of each series. This was done by using the HMM method with $m = 100$, a value that is large enough to ensure that any anomalies that may occur could not be attributed to inaccuracies in the approximation of the likelihood. Then, for each one of the observations in the validation sample, the (one-step-ahead forecast) pseudo-residual was computed according to equation (12).

The QQ-plot for the S & P 500 returns under the SV-N, SV-T, SV-S and SV-VG models are shown in Figure 4. The QQ-plot indicates a lack of fit in the left tail and in the right tail for the SV-N. For the SV-T, SV-S and SV-VG models, the QQ-plot indicates a lack of fit in the left tail. The indicated mis-specification could be caused by the presence of correlation between the perturbation terms defined by equations (5a) and (5b).

The QQ-plot discussed above is useful for assessing the relative fit of a model, but, for the purpose of assessing the risk associated with a share or index, it is the extreme left tail of the forecast distribution that is of particular interest. It determines the value-at-risk (VaR), defined as the maximum possible loss of a portfolio (over a specified period) at a given confidence level. For example, the one-day 1% VaR is the 0.01-quantile of the one-day-ahead forecast distribution. Whenever the return falls below that quantile, an *exception* is said to have occurred. If the model used for forecasting is correct, then, using a $100\alpha\%$VaR, the number of exceptions, X, in n days follows a Binomial(n, α) distribution. This distributional result makes it possible to implement back testing, where the adequacy of the time series model is assessed through a comparison of the observed number of exceptions and the corresponding theoretical distribution. A standard approach to test the accuracy of VaR forecasts is to assess the violation rate, which is estimated as $\hat{\alpha} = X/n$. 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^2_1-distributed test statistic

$$LRuc \quad = \quad 2\{\log[\hat{\alpha}^x(1-\hat{\alpha})^{n-x}] - \log[\alpha^x(1-\alpha)^{n-x}]\}. \tag{14}$$

The null hypothesis is that the achieved violation rate is equal to the predetermined nominal probability α. See Kupiec (1995) for more details. According to the unconditional coverage test, we accept the null hypothesis that the achieved violation rate is equal to 5% for all the returns

Table 5: Violation rate (VR) as a percentage in n one-day-ahead forecast, P-values of the unconditional coverage test at the 1% and 5% levels.

			0.01		0.05	
Return	Model	n	VR (%)	P-value	VR (%)	P-value
S & P 500	SV-N	1102	0.0172	0.0285	0.0435	0.3161
	SV-T	1102	0.0154	0.0935	0.0481	0.7701
	SV-S	1102	0.0163	0.0528	0.0508	0.9012
	SV-VG	1102	0.0154	0.0935	0.0472	0.6654

under all the models considered here. We reject that the achieved violation is 1% only for S & P 500 under the SV-N model.

6 Conclusions

In this article, we presented an easy-to-implement Bayesian estimation approach for the SV-SMN class of models. The commonly made Gaussian assumption of the mean innovation was replaced by univariate heavy-tailed processes, known as scale mixtures of normal distributions. While we focused on practical and computational aspects of fitting these models to real data, there may of course also be of interest in deriving theoretical properties of the estimators.

We illustrated our methods through an empirical application of the S&P500 index return series, which shows that the SV-VG model provides better model fitting than the SV-N model in terms of parameter estimates, interpretation, robustness aspects and out-of-sample forecast.

Our SV-SMN class of models showed considerable flexibility to accommodate outliers, however their robustness aspects could be seriously affected by the presence of skewness and heavy-tailedness simultaneously. To remedy this problem, the scale mixtures of skew-normal distributions can be used, or alternatively, the conditional distribution of the returns could be modeled nonparametrically (Langrock et al., 2015).

Finally, another important stylized fact often attributed to financial time series, the so-called *leverage effect*, is not explicitly incorporated in the class of models presented in this paper. Models

with leverage effect involve a (negative) correlation between the innovations in the returns and subsequent innovations in the log-volatility process. While relatively easy to accomplish when both innovations are Gaussian — in which case the joint distribution of the innovations can simply be taken to be a bivariate normal — it is not quite as straightforward to formulate corresponding models where the distribution of the innovations in the returns is from the general class of SMN distributions. We believe that the way forward to constructing corresponding models is via the use of copulas, which can be used to couple arbitrary marginal densities, in particular SMN distributions and normal distributions, respectively. This strategy was first proposed by Smith (2007), but has since not been pursued further, and in particular not within classes of models as flexible as the one discussed in the present paper. The copula-based extension of SV models, which is beyond the scope of the present paper, is currently under investigation.

Acknowledgements

The research of Carlos A. Abanto-Valle was partially supported by the CNPq-Brazil. Dr. Chen's research was partially supported by US NIH grants #GM70335 and #P01CA142538.

References

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V., and Enriquez, I. (2010), "Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions," *Computational Statistics and Data Analysis*, 54, 2883–2898.

Abanto-Valle, C. A., Lachos, V. H., and Dey, D. K. (2015), "Bayesian Estimation of a Skew-Student-t Stochastic Volatility Model," *Methodology and Computing in Applied Probability*, 17(3), 721–738.

Abanto-Valle, C. A., Wang, C., Wang, X., Wang, F.-X., and Chen, M.-H. (2014), "Bayesian inference for stochastic volatility models using the generalized skew-t distribution with applications to the Shenzhen Stock Exchange returnsl," *Statistics and its Interface*, 7, 487–502.

Andersen, T., and Sorensen, B. (1996), "GMM estimation of a stochastic volatility model: A Monte Carlo study," *Journal of Business and Economic Statistics*, 13, 329–352.

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

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

Brooks, S. (2002), "Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde," *J. R. Stat. Soc. Series B*, 64, 616–618.

Carnero, M. A., Peña, D., and Ruiz, E. (2004), "Persistence and Kurtosis in GARCH and Stochastic volatility models," *Journal of Financial Econometrics*, 2, 319–342.

Chen, M.-H., Huan, L., Ibrahim, J., and Kim, S. (2008), "Bayesian Variable Selection and Computation for Generalized Linear Models with Conjugate Priors," *Bayesian Analysis*, 3, 585–614.

Chen, M.-H., Shao, Q., and Ibrahim, J. (2000), *Monte Carlo Methods in Bayesian Computation*, New York: Springer-Verlag.

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

Chow, S. T. B., and Chan, J. S. K. (2008), "Scale mixtures distributions in statistical modelling," *Australian & New Zeland Journal of Statistics*, 50, 135–146.

Choy, S., Wan, W. Y., and Chan, C. (2008), "Bayesian Student-t Stochastic Volatility Models via Scale Mixtures," *Advances in Econometrics*, 23, 595–618.

Danielsson, J. (1994), "Stochastic volatility in asset prices: Estimation with simulated maximum likelihood," *Journal of Econometrics*, 64, 375–400.

Dawid, A. P. (1984), "Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach," *Journal of Royal Statistics Society, Series B*, 64, 278–292.

Delatola, E.-I., and Griffin, J. E. (2011), "Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility," *Bayesian Analysis*, 6, 901–926.

Ehlers, R., and Zevallos, M. (2015), "Bayesian Estimation and Prediction of Stochastic Volatility Models via INLA," *Communications in Statistics—Simulation and Computation*, 44, 683–693.

Gelfand, A., Dey, D., and Chang, H. (1992), "Model determination using predictive distributions with implementation via sampling-based methods," in *Bayesian Statistics 4*, eds. J. O. B. J. M. Bernardo, and M. H. DeGroot, pp. 147–167.

Gerlach, R., Carter, C., and Kohn, R. (1999), "Diagnostics for time series analysis," *Journal of Time Series Analysis*, 20, 309–330.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994), "Multivariate stochastic volatility models," *Review of Economical Studies*, 61, 3–30.

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

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

Kim, S., Chen, M., and Dey, D. (2002), "Flexible generalized t-link models for binary response data," *Biometrika*, 95, 93–106.

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

Kupiec, P. H. (1995), "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives*, 3, 73–84.

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

Langrock, R. (2011), "Some applications of nonlinear and non-Gaussian state-space modelling by means of hidden Markov models," *Journal of Applied Statistics*, 38, 2955–2970.

Langrock, R., MacDonald, I. L., and Zucchini, W. (2012), "Some nonstandard stochastic volatility models and their estimation usingstructured hidden Markov models," *Journal of Empirical Finance*, 19, 147–161.

Langrock, R., Michelot, T., Sohn, A., and Kneib, T. (2015), "Semiparametric stochastic volatility modelling using penalized splines," *Computational Statistics*, 19, 517–537.

Leão, W. L., Abanto-Valle, C. A., and Chen, M.-H. (2017), "Bayesian analysis of stochastic volatility-in-mean model with leverage and asymmetrically heavy-tailed error using generalized hyperbolic skew Student's t-distribution," *Statistics and its Interface*, 10, 529–541.

Liesenfeld, R., and Jung, R. C. (2000), "Stochastic volatility models: Conditional normality versus heavy-tailed distrutions," *Journal of Applied Econometics*, 15, 137–160.

Liesenfeld, R., and Richard, J.-F. (2003), "Univariate and multivariate stochastic volatility models: estimation and diagnostics," *Journal of Empirical Finance*, 10, 505–531.

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

Martino, S., Aas, K., Lindqvist, O., Neef, L., and Rue, H. (2010), "Estimating stochastic volatility models using integrated nested laplace approximations," *The European Journal of Finance*, 17, 487–503.

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

Meyer, R., Fournier, D. A., and Berg, A. (2003), "Stochastic volatility: Bayesian computation using automatic differentiation and the extended Kalman filter," *Econometrics Journal*, 8, 323–361.

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

Rosenblatt, M. (1952), "Remarks on a multivariate transformation," *Annals of Mathematical Statistics Volume 23*, 23, 470–472.

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

Skaug, H. J., and Yu, J. (2008), Automated likelihood based inference for stochastic volatility models,, Working paper, Singapur Management University.

Smith, D. R. (2007), A Stochastic Volatility Model with Fat Tails, Skewness and Leverage Effects,, Technical report.
URL: *http://dx.doi.org/10.2139/ssrn.1078625*

Smith, J. Q. (1985), "Diagnostic checks of non-standard time series models," *Journal of Forecasting*, 4, 283–291.

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

Taylor, S. (1982), "Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75," in *Time Series Analysis: Theory and Practice, Vol 1*, ed. O. Anderson, pp. 203–226.

Taylor, S. (1986), *Modeling Financial Time Series*, Chichester: Wiley.

Wang, J., and Genton, M. (2006), "The multivariate skew-slash distribution," *Journal of Statistical Planning and Inference*, 136, 209–220.

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

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction Using R, 2nd Edition*, Boca Raton, FL: Chapman & Hall.