Brazilian Journal of Probability and Statistics
2012, Vol. 26, No. 4, 402–422
DOI: 10.1214/11-BJPS169

Stochastic volatility in mean models with heavy-tailed distributions

Carlos A. Abanto-Valle[a], **Helio S. Migon**[a] **and Victor H. Lachos**[b]

[a]*Federal University of Rio de Janeiro*
[b]*Campinas State University*

Abstract. A stochastic volatility in mean (SVM) model using the class of symmetric scale mixtures of normal (SMN) distributions is introduced in this article. The SMN distributions form a class of symmetric thick-tailed distributions that includes the normal one as a special case, providing a robust alternative to estimation in SVM models in the absence of normality. A Bayesian method via Markov-chain Monte Carlo (MCMC) techniques is used to estimate parameters. The deviance information criterion (DIC) and the Bayesian predictive information criteria (BPIC) are calculated to compare the fit of distributions. The method is illustrated by analyzing daily stock return data from the São Paulo Stock, Mercantile & Futures Exchange index (IBOVESPA). According to both model selection criteria as well as out-of-sample forecasting, we found that the SVM model with slash distribution provides a significant improvement in model fit as well as prediction for the IBOVESPA data over the usual normal model.

1 Introduction

In recent years, stochastic volatility (SV) models have been considered as useful tools for modeling time-varying variances, mainly in financial applications where policymakers or stockholders are constantly facing decision problems that usually depend on measures of volatility and risk. An attractive feature of the SV model is its close relationship to financial economic theories (Melino and Turnbull, 1990) and its ability to capture the main empirical properties often observed in daily series of financial returns in a more appropriate way (Carnero et al., 2004).

The relation between expected returns and expected volatility has been extensively examined in recent years. The theory generally predicts a positive relation between expected stock returns and volatility if investors are risk averse. In other words, investors require a larger expected return from a security that is riskier. Empirical studies that attempt to test this important relation, however yield mixed results. French (1987) found a positive and significant relationship and Theodossiou (1995) reported a positive but nonsignificant relationship between stock market volatility and stock returns. Consistent with the asymmetric volatility argument,

Key words and phrases. Feedback effect, Markov chain Monte Carlo, non-Gaussian and nonlinear state space models, scale mixture of normal distributions, stochastic volatility in mean.
Received February 2011; accepted August 2011.

Nelson (1991) and, more recently, Brandt and Kang (2004) reported evidence of a negative and often significant relationship between volatility and returns. Overall, there appears to be stronger evidence of a negative relationship between unexpected returns and innovations to the volatility process, which French (1987) interpreted as indirect evidence of a positive correlation between the expected risk premium and ex ante volatility. This theory, known as feedback volatility, states that bad (good) news decreases (increases) stock prices and increases volatility, therefore determining a further decrease of the price. An alternative explanation for asymmetric volatility where causality runs in the opposite direction is the leverage effect put forward by Black (1976), who asserted that a negative (positive) return shock leads to an increase (decrease) in the firm's financial leverage ratio, which has an upward (downward) effect on the volatility of its stock returns. However, French (1987) and Schwert (1989) argued that leverage alone cannot account for the magnitude of the negative relationship. For example, Campbell and Hentschel (1992) found evidence of both volatility feedback and leverage effects, whereas Bekaert and Wu (2000) presented results suggesting that the volatility feedback effect dominates the leverage effect empirically.

Many empirical studies have shown strong evidence of heavy-tailed conditional mean errors in financial time series; see, for example, Mandelbrot (1963), Liesenfeld and Jung (2000), Chib et al. (2002) and Jacquier et al. (2004). Frequently, the volatility of daily stock returns has been estimated with SV models, but the results have relied on an extensive premodeling of these series to avoid the problem of simultaneous estimation of the mean and variance. Koopman and Uspensky (2002) introduced the SV in mean (SVM) model to deal with this problem and the unobserved volatility is incorporated as an explanatory variable in the mean equation of the returns under the normality assumption of the innovations. In this article we propose to enhance the robustness of the specification of the innovation return in SVM models by introducing SMN distributions. In fact, the flexibility of the SVM with SMN distributions could fit time varying features in the mean of the returns and heavy tails simultaneously. We refer to this generalization as an SVM-SMN class of models. This rich class contains as proper elements the SVM with normal (SVM-N), Student-t (SVM-t), slash (SVM-S) and the contaminated normal (SVM-CN) distributions. The estimation of such intricate models is not straightforward, since volatility now appears in both the mean and the variance equation and hence intensive computational methods are needed for. Inference in the SVM-SMN class of models is performed under a Bayesian paradigm via MCMC methods. An efficient multi-move sampler is developed to simulate the log-volatilities by blocks (Abanto-Valle et al., 2010; Shephard and Pitt, 1997; Watanabe and Omori, 2004).

The remainder of this paper is organized as follows. Section 2 gives a brief introduction about the SMN distributions. Section 3 outlines the general class of the SVM–SMN models as well as the Bayesian estimation procedure using MCMC methods. Section 4 is devoted to application and model comparison among particular members of the SVM–SMN models using the IBOVESPA data set. Finally, some concluding remarks and suggestions for future developments are given in Section 5.

2 Scale mixture of normal distributions

Andrews and Mallows (1974) used the Laplace transform technique to prove that a continuous random variable Y has a SMN distribution if it can be expressed as follows:

$$Y = \mu + \kappa^{1/2}(\lambda)Z, \tag{2.1}$$

where μ is a location parameter, Z is a normal random variable with zero mean and variance σ^2, $\kappa(\lambda)$ is a positive weight function, λ is a mixing positive random variable with density $p(\lambda|\boldsymbol{\nu})$, $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of λ. As in Choy and Chan (2008), we restrict our attention to the case in that $\kappa(\lambda) = 1/\lambda$. Thus, $Y|\lambda \sim \mathcal{N}(\mu, \lambda^{-1}\sigma^2)$ and the marginal p.d.f. of Y with respect to λ is given by

$$f(y|\mu, \sigma^2, \nu) = \int_0^\infty \mathcal{N}(y|\mu, \lambda^{-1}\sigma^2) p(\lambda|\boldsymbol{\nu})\, d\lambda. \tag{2.2}$$

From a suitable choice of the mixing density $p(\cdot|\boldsymbol{\nu})$, a rich class of continuous symmetric distributions can be described by the density given in (2.2) that can readily accommodate thicker tails than the normal process. Note that when $\lambda = 1$ (a degenerate random variable), we retrieve the normal distribution. Apart from the normal model, we explore three different types of heavy-tailed densities based on the choice of the mixing density $p(\cdot|\boldsymbol{\nu})$.

- *The Student-t distribution*, $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$.
  The use of the Student-t distribution as an alternative robust model to the normal distribution has frequently been suggested in the literature (Little, 1988). For the Student-t distribution with location μ, scale σ and degrees of freedom ν, $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$ is equivalent to the following hierarchical form:

$$Y|\mu, \sigma^2, \nu, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{G}(\nu/2, \nu/2), \tag{2.3}$$

  where $\mathcal{G}(\cdot, \cdot)$ denotes the gamma distribution.
- *The Slash distribution*, $Y \sim \mathcal{S}(\mu, \sigma^2, \nu)$, $\nu > 0$.
  This distribution presents heavier tails than those of the normal distribution and it includes the normal case when $\nu \uparrow \infty$. The slash distribution is equivalent to the following hierarchical form:

$$Y|\mu, \sigma^2, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda|\nu \sim \mathcal{B}e(\nu, 1), \tag{2.4}$$

  where $\mathcal{B}e(\cdot, \cdot)$ denotes the beta distribution.

- *The contaminated normal distribution*, $Y \sim \mathcal{CN}(\mu, \sigma^2, \boldsymbol{\nu})$, $\boldsymbol{\nu}' = (\delta, \gamma)$.
  Here, λ is a discrete random variable taking one of two states. The probability function of λ, given the parameter vector $\boldsymbol{\nu}' = (\delta, \gamma)$, is denoted by

$$p(\lambda|\boldsymbol{\nu}) = \delta \mathbb{I}_{(\lambda=\gamma)} + (1-\delta)\mathbb{I}_{(\lambda=1)}, \qquad 0 \le \delta < 1, 0 < \gamma < 1, \tag{2.5}$$

where $\mathbb{I}(\cdot)$ denotes an indicator function. It follows then that

$$f(y) = \delta \mathcal{N}(y|\mu, \gamma^{-1}\sigma^2) + (1-\delta)\mathcal{N}(y|\mu, \sigma^2). \tag{2.6}$$

Parameter δ can be interpreted as the proportion of outliers while γ may be interpreted as a scale factor. The contaminated normal distribution reduces to the normal one when $\gamma = 1$.

3 The heavy-tailed stochastic volatility in mean model

The SV in mean model with heavy tails is defined by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{h_t/2}\lambda_t^{-1/2}\varepsilon_t, \tag{3.1a}$$

$$h_t = \alpha + \phi h_{t-1} + \sigma_\eta \eta_t, \tag{3.1b}$$

$$\lambda_t \sim p(\lambda_t|\boldsymbol{\nu}), \varepsilon_t \sim \mathcal{N}(0,1), \eta_t \sim \mathcal{N}(0,1), \tag{3.1c}$$

where y_t and h_t are, respectively, the compounded return and the log-volatility at time t. We assume that $|\phi| < 1$, that is, that the log-volatility process is stationary and that the initial value $h_0 \sim \mathcal{N}(\frac{\alpha}{1-\phi}, \frac{\sigma_\eta^2}{1-\phi^2})$. The innovations ε_t and η_t are assumed to be mutually independent and normally distributed with mean zero and unit variance. In this setup, λ_t is a scale factor, $p(\lambda_t|\boldsymbol{\nu})$ is the mixing density and $\boldsymbol{\nu}$ the parameter that captures the heavy-tailness. The aim of the SVM-SMN class of models is to simultaneously estimate the ex ante relation between returns and volatility and the volatility feedback effect in the presence of outliers. This class of models includes the SVM with Student-t (SVM-t), with slash (SVM-S) and contaminated normal (SVM-CN) distributions as special cases. The first and second models are obtained by choosing the mixing density as $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, $\lambda_t \sim \mathcal{B}e(\nu, 1)$, respectively, where $\mathcal{G}(\cdot,\cdot)$ and $\mathcal{B}e(\cdot,\cdot)$ denote the gamma and beta distributions respectively. In the SVM-CN model λ_t follows a discrete distribution, such that $p(\lambda_t|\boldsymbol{\nu}) = \delta\mathbb{I}_{(\lambda_t=\gamma)} + (1-\delta)\mathbb{I}_{(\lambda_t=1)}$, where $\boldsymbol{\nu}' = (\delta, \gamma)$ and $\mathbb{I}(\cdot)$ is an indicator function. When $\lambda_t = 1$ for all t, we have the SVM with normal distribution of Koopman and Uspensky (2002), and we denote it by SVM-N. Under a Bayesian paradigm, we use MCMC methods to conduct the posterior analysis in the next subsection. Conditional on λ_t, some derivations are common to all members of the SVM-SMN family (see Appendix for details).

3.1 Parameter estimation via MCMC

Let $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \alpha, \phi, \sigma_\eta^2, \boldsymbol{\nu}')'$ be the full parameter vector of the entire class of SVM-SMN models, $\mathbf{h}_{0:T} = (h_0, h_1, \ldots, h_T)'$ be the vector of the log volatilities, $\boldsymbol{\lambda}_{1:T} = (\lambda_1, \ldots, \lambda_T)'$ be the mixing variables and $\mathbf{y}_{0:T} = (y_0, \ldots, y_T)'$ be the information available up to time T, while $\boldsymbol{\nu}$ is the parameter vector associated with the mixture distribution. The Bayesian approach to estimate the parameters in the SVM-SMN models uses the data augmentation principle, which considers $\mathbf{h}_{0:T}$ and $\boldsymbol{\lambda}_{1:T}$ as latent parameters. The joint posterior density of parameters and latent unobservable variables can be written as

$$p(\mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}, \boldsymbol{\theta}|\mathbf{y}_{0:T}) \propto p(\mathbf{y}_{1:T}|y_0, \boldsymbol{\theta}, \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T})p(\mathbf{h}_{0:T}|\boldsymbol{\theta})p(\boldsymbol{\lambda}_{1:T}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (3.2)$$

where $p(\boldsymbol{\theta})$ is the prior distribution. Since $p(\mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}, \boldsymbol{\theta}|\mathbf{y}_{0:T})$ does not have closed form, we first sample the parameters $\boldsymbol{\theta}$, followed by the latent variables $\boldsymbol{\lambda}_{1:T}$ and $\mathbf{h}_{0:T}$ using the Gibbs sampling algorithm (see Algorithm 1 for details of the sampling scheme). Sampling the log-volatilities $\mathbf{h}_{0:T}$ in step 4 of Algorithm 1 is the most difficult task due to the nonlinear setup in the observational equation (3.1a). In order to avoid the higher correlations due to the Markovian structure of the h_t's, in the next subsection we develop a multi-move block sampler to sample $\mathbf{h}_{0:T}$ by blocks (Abanto-Valle et al., 2010; Shephard and Pitt, 1997; Watanabe and Omori, 2004). Details on the full conditionals of $\boldsymbol{\theta}$ and the latent variable $\boldsymbol{\lambda}_{1:T}$ are given in the Appendix.

Algorithm 1.

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

The prior distribution of the parameters in the SVM–SMN class are set as $\beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma_{\beta_0}^2)$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(\bar{\beta}_1, \sigma_{\beta_1}^2)$, $\beta_2 \sim \mathcal{N}(\bar{\beta}_2, \sigma_{\beta_2}^2)$, $\alpha \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2)$, $\phi \sim$

$\mathcal{N}_{(-1,1)}(\bar{\phi}, \sigma_\phi^2)$ and $\sigma_\eta^2 \sim \mathcal{IG}(\frac{T_0}{2}, \frac{M_0}{2})$, where $\mathcal{N}_{(a,b)}(\cdot, \cdot)$ denotes the truncated normal distribution in the interval (a, b). The prior distribution of $\boldsymbol{\nu}$ is model specific (see details in the Appendix).

3.2 A block sampler algorithm

In order to simulate $\mathbf{h}_{0:T} = (h_0, \ldots, h_T)'$ in the SVML-SMN class of models, 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. To sample the vector $\mathbf{h}_{1:T}$, we develop a multi-move block algorithm. In our block algorithm, we divide $\mathbf{h}_{1:T}$ into $K+1$ blocks, $\mathbf{h}_{k_{l-1}+1:k_l-1} = (h_{k_{l-1}+1}, \ldots, h_{k_l-1})'$ for $l = 1, \ldots, K+1$, with $k_0 = 0$ and $k_{K+1} = T$, where $k_l - 1 - k_{l-1} \geq 2$ is the size of the lth block. A suitable selection of K is important to obtain an efficient sampler that can reduce the correlation imposed by the model in the sampling process.

We sample the block of disturbances $\boldsymbol{\eta}_{k_{l-1}+1:k_l-1} = (\eta_{k_{l-1}+1}, \ldots, \eta_{k_l-1})'$ given the end conditions $h_{k_{l-1}}$ and h_{k_l} instead of $\mathbf{h}_{k_{l-1}+1:k_l-1}$, exploiting the fact that the innovations η_t are i.i.d. with $\mathcal{N}(0, 1)$ distribution. To facilitate the exposition, suppose that $k_{l-1} = t$ and $k_l = t + k + 1$ for the lth block. Then $\boldsymbol{\eta}_{t+1:t+k}$ are sampled at once from their full conditional distribution, which omitting the dependence on $\mathbf{y}_{t:t+k}, \boldsymbol{\lambda}_{t+1:t+k}$, in order to facilitate the exposition is denoted by $f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \boldsymbol{\theta})$ and expressed in the log scale as

$$\begin{aligned} \log f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \boldsymbol{\theta}) \\ = \text{const} - \frac{1}{2}\sum_{r=t+1}^{t+k} \eta_r^2 + \sum_{r=t+1}^{t+k} l(h_r) \\ - \frac{1}{2\sigma_\eta^2}(h_{t+k+1} - \alpha - \phi h_{t+k})^2 \mathbb{I}(t+k < T), \end{aligned} \tag{3.3}$$

where $\mathbb{I}(\cdot)$ denotes an indicator function. We denote the first and second derivatives of $l(h_r)$ with respect to h_r by l' and l'', where $l(h_r) = \log p(y_r|y_{r-1}, \beta_0, \beta_1, \beta_2, \lambda_r, h_r)$ is obtained from equation (3.1a). As (3.3) does not have closed form, we use the Metropolis–Hastings acceptance-rejection algorithm (Tierney, 1994; Chib and Greenberg, 1995) to sample from. We propose to use the following artificial Gaussian state space model as a proposed density to simulate the block $\boldsymbol{\eta}_{t+1:t+k}$:

$$\hat{y}_r = h_r + \varepsilon_r, \qquad \varepsilon_r \sim \mathcal{N}(0, d_r), \qquad r = t+1, \ldots, t+k, \tag{3.4}$$

$$h_r = \alpha + \phi h_{r-1} + \sigma_\eta \eta_r, \qquad \eta_r \sim \mathcal{N}(0, 1), \tag{3.5}$$

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

$$\begin{aligned} d_r &= -\frac{1}{l_F''(\hat{h}_r)}, \\ \hat{y}_r &= \hat{h}_r + d_r l'(\hat{h}_r). \end{aligned} \tag{3.6}$$

For $r = t + k < T$,

$$\begin{aligned} d_r &= \frac{\sigma_\eta^2}{\phi^2 - \sigma_\eta^2 l''_F(\hat{h}_{t+k})}, \\ \hat{y}_r &= d_r\left[l'(\hat{h}_r) - l''_F(\hat{h}_r)\hat{h}_r + \frac{\phi}{\sigma_\eta^2}(h_{r+1} - \alpha)\right]. \end{aligned} \tag{3.7}$$

We obtain the observational equation (3.4) by a second-order expansion of $l(h_r)$ around some preliminary estimate of η_r, denoted by $\hat{\eta}_r$, where $\hat{h}_r$ is the estimate of h_r equivalent to $\hat{\eta}_r$. Since $l''(h_r)$ is

$$\begin{aligned} l''(h_r) = &-\tfrac{1}{2}\lambda_r(y_r - \beta_0 - \beta_1 y_{r-1} - \beta_2 e^{h_t})^2 e^{-h_r} \\ &- \beta_2\lambda_r(y_r - \beta_0 - \beta_1 y_{r-1} - \beta_2 e^{h_r}) - \beta_2^2\lambda_r e^{h_r}, \end{aligned}$$

which can be positive for some values of h_r, we define $l''_F(h_r)$ as

$$l''_F(h_r) = E[l''(h_r)] = -\tfrac{1}{2} - \beta_2^2\lambda_r e^{h_r}, \tag{3.8}$$

which is strictly negative everywhere. The expectation (3.8) is taken with respect to y_r conditional on h_r and λ_r, $\beta_0, \beta_1, \beta_2$ and y_{r-1}. Since (3.4)–(3.5) define a Gaussian state space model, we can apply the simulation smoother (de Jong and Shephard, 1995) to perform the sampling. Let g be the density based on the model given by equations (3.4)–(3.5). Since f is not bounded by g, we use the Metropolis–Hastings acceptance-rejection algorithm to sample from f, as recommended by Chib and Greenberg (1995). In the SVM-N case, we use the same procedure with $\lambda_t = 1$ for $t = 1, \dots, T$.

We select the expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ as follows. Once an initial expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ is selected, we can calculate the auxiliary $\hat{\mathbf{y}}_{t+1:t+k}$ by using equations (3.6) and (3.7). In the MCMC implementation, the previous sample of $\mathbf{h}_{t+1:t+k}$ can be taken as an initial value of $\hat{\mathbf{h}}_{t+1:t+k}$. Then, applying the Kalman filter and a disturbance smoother to the linear Gaussian state space model consisting of equations, (3.4) and (3.5), with the artificial $\hat{\mathbf{y}}_{t+1:t+k}$ yields the mean of $\mathbf{h}_{t+1:t+k}$ conditional on $\hat{\mathbf{h}}_{t+1:t+k}$ in the linear Gaussian state space model, which is used as the next $\hat{\mathbf{h}}_{t+1:t+k}$. By repeating the procedure until the smoothed estimates converge, we obtain the posterior mode of $\mathbf{h}_{t+1:t+k}$. This is equivalent to the method of scoring to maximize the logarithm of the conditional posterior density. Although we have just noted that iterating the procedure achieves the mode, this will slow our simulation algorithm if we have to iterate this procedure until full convergence. Instead we suggest using only five iterations of this procedure to provide reasonably good sequence $\hat{\mathbf{h}}_{t+1:t+k}$ instead of an optimal one. The procedure is summarized in Algorithm 2.

Algorithm 2.

1. Initialize $\hat{\mathbf{h}}_{t+1:t+k}$.
2. Evaluate recursively $l'(\hat{h}_r)$ and $l''_F(\hat{h}_r)$ for $r = t+1, \dots, t+k$.
3. Conditional on the current values of the vector of parameters $\boldsymbol{\theta}$, $\boldsymbol{\lambda}_{t+1:t+k}$, h_t and h_{t+k+1}, define the auxiliary variables $\hat{y}_r$ and d_r using equations (3.6) or (3.7) for $r = t+1, \dots, t+k$.
4. Consider the linear Gaussian state–space model in (3.4) and (3.5). Apply the Kalman filter and a disturbance smoother (Koopman, 1993) and obtain the posterior mean of $\boldsymbol{\eta}_{t:t+k}$ $(\mathbf{h}_{t:t+k})$ and set $\hat{\boldsymbol{\eta}}_{t:t+k}$ $(\hat{\mathbf{h}}_{t:t+k})$ to this value.
5. Return to step 2 and repeat the procedure until achieving convergence.

Finally, we describe the updating procedure for h_0 and the knot conditions h_{k_l}, for $l = 1, \dots, K$. We simulate $h_0|h_1, \boldsymbol{\theta} \sim \mathcal{N}(\alpha + \phi h_1, \sigma_\eta^2)$. As the density $p(h_{k_l}|h_{k_l-1}, h_{k_l+1})$ does not have a closed form, we use the Metropolis–Hastings algorithm with proposal $\mathcal{N}(\frac{\alpha(1-\phi)+\phi(h_{k_l-1}+h_{k_l+1})}{1+\phi^2}, \frac{\sigma_\eta^2}{1+\phi^2})$. Let $h_{k_l}^p$ and $h_{k_l}^{(i-1)}$ denote the proposal value and the previous iteration value. Thus, the acceptance probability is given by $\alpha_{\mathrm{MH}} = \min\{1, \frac{Q(h_{k_l}^p)}{Q(h_{k_l}^{(i-1)})}\}$, where $Q(h_{k_l})$ is the conditional density of $y_{k_l}|\lambda_{k_l}, y_{k_l-1}, h_{k_l}$.

4 Empirical application

This section analyzes the daily closing prices of the IBOVESPA, which is an index of about 50 stocks that are traded on the São Paulo Stock, Mercantile & Futures Exchange. The index is composed of a theoretical portfolio with the stocks that accounted for 80% of the volume traded in the last 12 months and that were traded on at least 80% of the trading days. It is revised quarterly, to keep it representative of the volume traded. On average, the components of the IBOVESPA represent 70% of all the stock value traded. The data set was obtained from the Yahoo finance web site, available to download at "http://finance.yahoo.com." The period of analysis is January 5, 1998–October 3, 2005, which yields 1917 observations. Throughout, we work with the compounded return expressed as a percentage, $y_t = 100 \times (\log P_t - \log P_{t-1})$, where P_t is the closing price on day t.

The compounded IBOVESPA returns are plotted in Figure 1 as a time series plot and also as a histogram. The mean and standard deviation of returns are 0.06 and 2.34, respectively. As can be easily seen in Figure 1, the returns are slight skew (0.83) with heavy tails. Note also that the returns have a large range (minimum, -17.21 and maximum, 28.83). Some extreme observations, explained by turbulences in financial markets that occurred by August 1998 and January 1999 (the Russian and Brazilian exchange rate crises, respectively), contribute to the large kurtosis (19.18) of the IBOVESPA returns. As a result, the IBOVESPA returns

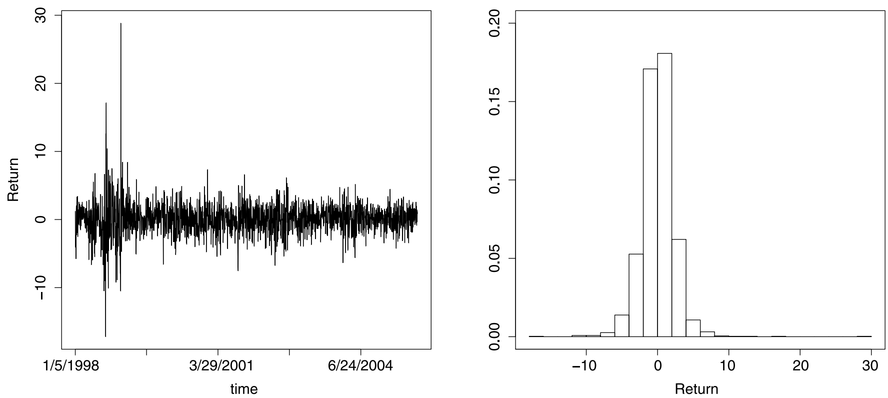

Figure 1 *Compounded IBOVESPA returns from January* 5, 1998 *to September* 3, 2005. *The left panel shows the plot of the raw series and the right panel the histogram of returns.*

likely depart from the underlying 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 the SVM-N, SVM-t, SVM-S and SVM-CN models.

In all cases, we simulated the h_t's in a multi-move fashion with stochastic knots based on the method described in Section 3.1. We set the prior distributions of the common parameters as $\beta_0 \sim \mathcal{N}(0, 100)$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(0.1, 100)$, $\beta_2 \sim \mathcal{N}(-0.1, 100)$, $\alpha \sim \mathcal{N}(0.0, 100)$, $\phi \sim \mathcal{N}_{(-1,1)}(0.95, 100)$ and $\sigma_\eta^2 \sim \mathcal{IG}(2.5, 0.025)$. The prior distributions on the shape parameters were chosen as $\nu \sim \mathcal{G}(12.0, 0.8)$ and $\nu \sim \mathcal{G}(2.0, 0.25)$ for the SVM-t model and the SVM-S model, respectively. For the SVM-CN, we set $\delta \sim \mathcal{B}e(2, 2)$ and $\gamma \sim \mathcal{B}e(2, 4)$. The priors' means for ϕ and β_1, are, respectively, 0.0032 and 0.0003, and their variances, 0.3328 and 0.3329. In both cases, the priors are equivalent to the uniform distribution on interval $(-1, 1)$, which gives zero mean and variance of 0.3333. Thus, it is clear that the priors considered for ϕ and β_1 are noninformative.

The initial values of the parameters were randomly generated from the prior distributions. We set all the log-volatilities, h_t, to be zero. Finally, the initial $\boldsymbol{\lambda}_{1:T}$ were generated from the prior $p(\lambda_t|\nu)$. All the calculations were performed running stand-alone code developed by us using an open source C++ library for statistical computation, the Scythe statistical library (Pemstein et al., 2011), which is available for free download at http://scythe.wustl.edu.

For the block sampler algorithm, we set the number of blocks K to be 60 in such a way that each block contained 32 h_t's on average. For the SVM-N, SVM-t and the SVM-S models, we conducted the MCMC simulation for 50,000 iterations. However, for the SVM-CN model, we used 210,000 iterations. In both cases, the first 10,000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain, only every 10th (SVM-N,

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

Parameter	SVM-N	SVM-t	SVM-S	SVM-CN
β_0	0.2491	0.3004	0.3205	0.2798
	(0.1050, 0.3976)	(0.1419, 0.4627)	(0.1589, 0.4889)	(0.0783, 0.4824)
	−0.37	0.29	0.12	−0.92
β_1	0.0313	0.0289	0.0298	0.0396
	(−0.0122, 0.0763)	(−0.0162, 0.0746)	(−0.0148, 0.0750)	(−0.0051, 0.0833)
	−0.12	−0.34	−0.70	1.80
β_2	−0.0402	−0.0616	−0.0959	−0.0612
	(−0.0772, −0.0046)	(−0.1086, −0.0158)	(−0.1701, −0.0297)	(−0.1245, −0.0024)
	1.12	−0.14	0.08	0.79
α	0.0235	0.0047	0.0116	0.0025
	(0.0093, 0.0408)	(0.0056, 0.0321)	(0.0032, 0.0225)	(0.0002, 0.0059)
	−1.03	0.12	1.56	−1.32
ϕ	0.9814	0.9851	0.9858	0.9977
	(0.9686, 0.9919)	(0.9735, 0.9944)	(0.9745, 0.9947)	(0.9950, 0.9996)
	1.05	0.02	−1.57	1.06
σ_η^2	0.0173	0.0122	0.0109	0.0008
	(0.0102, 0.0272)	(0.0070, 0.0198)	(0.0061, 0.0182)	(0.0006, 0.0012)
	−0.91	0.60	1.72	−0.71
ν	–	16.2892	2.4657	–
	–	(10.7400, 24.0800)	(2.0880, 2.7380)	–
	–	0.22	−0.55	–
δ	–	–	–	0.1188
	–	–	–	(0.0277, 0.3321)
	–	–	–	1.64
γ	–	–	–	0.2952
	–	–	–	(0.1488, 0.4371)
	–	–	–	0.10

SVM-t and SVM-S models) and 100th (SVM-CN model) values of the chain were stored. With the resulting 4000 (2000) values, we calculated the posterior means, the 95% credible intervals and the convergence diagnostic (CD) statistics (Geweke, 1992). If the sequence of the recorded MCMC output is stationary, it converges in distribution to the standard normal. According to the CD, the null hypothesis that the sequence of 4000 (2000) draws is stationary was accepted at the 5% level [CD $\in (-1.96, 1.96)$] for all the parameters in all the models considered here. Table 1 summarizes the results.

From Table 1, the posterior mean and 95% interval of ϕ in the SVM-CN are higher than those of the other three models. However, for all the models, we found that the posterior means of ϕ are above 0.9814, showing higher persistence. We found that the persistence of the SVM-t and the SVM-S are slightly different from the SVM-N model. The posterior mean of σ_η^2 is smaller in the SVM-CN than those of the SVM-N, SVM-t and the SVM-S models, indicating that the volatility of the SVM-CN is less variable than those of the other three models. We also found that the posterior mean of σ_η^2 of the SVM-t and the SVM-S model are smaller than the SVM-N case, too.

The posterior means together with the posterior 95% intervals of the three parameters, which govern the mean process for each of the four models, are reported in Table 1. We observed that in all the cases the posterior mean of β_0 is always positive and statistically significant, because the 95% interval does not contain zero. We found that the posterior mean of β_1 is positive and similar to the first-order autocorrelation (not reported here). Since the 95% posterior interval contains zero, this coefficient could be not significant. The β_2 parameter, which measures both the ex ante relationship between returns and volatility and the volatility feedback effect, has a negative posterior mean for all the models. We found β_2 is statistically significant because in all cases the 95% posterior credibility interval does not contain zero. This result confirms previous results in the literature and indicates that when investors expect higher persistent levels of volatility in the future they require compensation for this in the form of higher expected returns. The magnitude of the tail fatness is measured by the shape parameter ν in the SVM-t and SVM-S models. In the SVM-CN case it is measured by δ. The posterior means of ν are, respectively, 16.3 and 2.5 in the SVM-t and SVM-S models. In the SVM-CN the parameter δ can be interpreted as the proportion of outliers present in the data set, and its posterior mean is equal to 0.12. The parameter γ is a scale factor, and it is estimated as 0.29. These results suggest that the measurement error of the stock returns are better explained by heavy-tailed distributions.

The reason why the estimated volatility of the SVM-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 contaminated normal ($\mathcal{CN}(0, 1, \delta, \gamma)$). We set ν, δ and γ as the posterior mean of the respective SVM model (see Table 1 for details). Figure 2 depicts the four density curves (the Student-t, slash and contaminated normal distributions have been rescaled to be comparable. See Wang and Genton, 2006). All the distributions have fatter tails than 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 SVM-SMN class of models considered here attributes a relatively larger proportion of extreme return values to ε_t instead of η_t than the SVM-N model, making the volatility of the SVM-t, SVM-S and SVM-CN models less variable. It also increases the persistence of these models' volatility.

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,

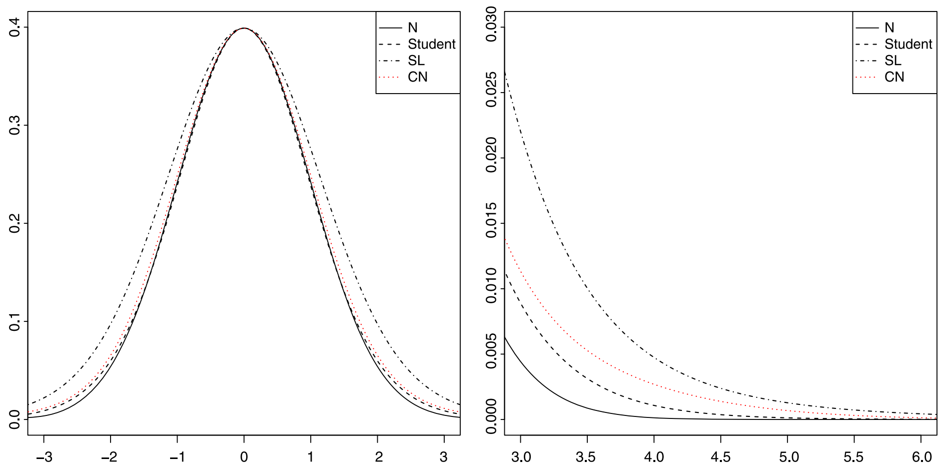

Figure 2 *Density curves of the univariate normal, Student-t, slash, variance gamma and contaminated normal distributions using the estimated tail-fatness parameter from the respective SVM model.*

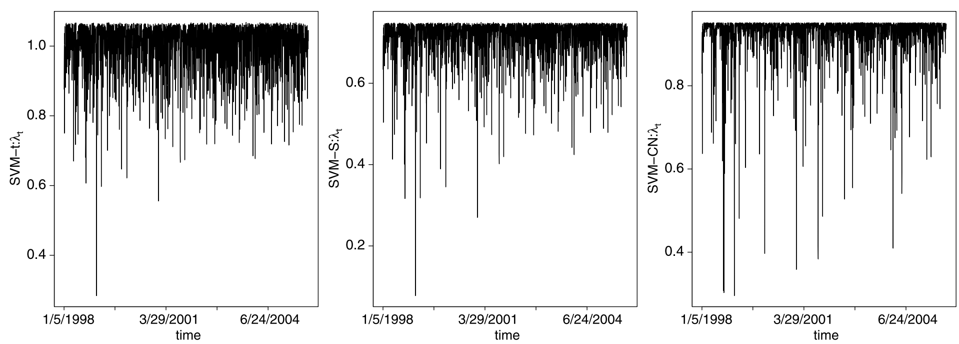

Figure 3 *IBOVESPA data set: posterior smoothed mean of mixture variable* λ_t *for the SVM-t (left panel), SVM-S (middle panel) and SVM-CN (right panel) models.*

Rosa et al., 2003). The heavy-tailed SV-SMN 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 3 where we depicted the posterior mean of the mixing variable λ_t for the SVM-t (left panel), SVM-S (middle panel) and SVM-CN (right panel) models, respectively.

In Figure 4, we plot the smoothed mean of $e^{h_t/2}$. The posterior smoothed mean of $e^{h_t/2}$ of the SVM-t, SVM-S and SVM-CN models show smoother movements than that from the SVM-N model (solid line). Extreme returns, such as during the Brazilian exchange rate crises in January 1999, make the differences clear. The models with heavy tails accommodate possible outliers in a somewhat different way by inflating the variance $e^{h_t/2}$ by $\lambda_t^{-1}e^{h_t/2}$. This can have a substantial im-

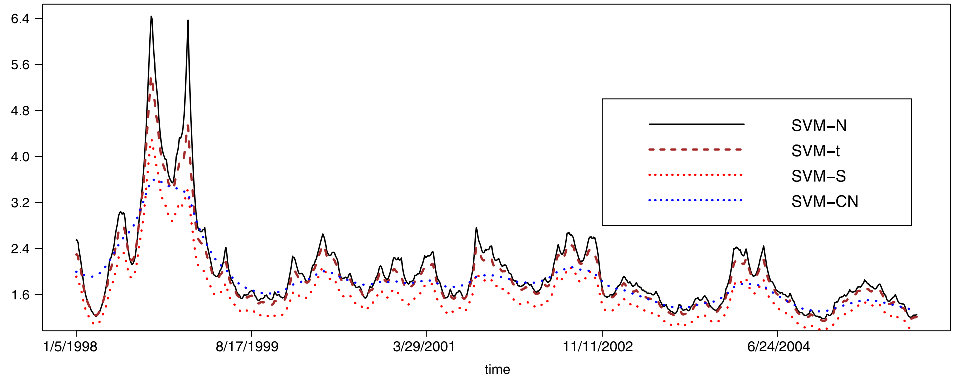

Figure 4 *IBOVESPA data set. Posterior smoothed mean of* $e^{h_t/2}$. *SVM-N (solid line), SVM-T (dotted line), SVM-S (pointed line) and SVM-CN (pointed-dotted line).*

pact, for instance, in the valuation of derivative instruments and several strategic or tactical asset allocation topics.

To assess the goodness of the estimated models, we calculate the deviance information criteria, DIC (Spiegelhalter et al., 2002), and the Bayesian predictive information criteria, BPIC (Ando, 2006, 2007). The DIC is defined as

$$\text{DIC} = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log p(\mathbf{y}_{1:T}|\boldsymbol{\theta})] + p_D. \tag{4.1}$$

The second term in (4.1) measures the complexity of the model by the effective number of parameters, p_D, defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$p_D = 2\big[\log p(\mathbf{y}_{1:T}|\bar{\boldsymbol{\theta}}) - E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log p(\mathbf{y}_{1:T}|\boldsymbol{\theta})]\big]. \tag{4.2}$$

To calculate the DIC in the context of SVM-SMN models, we use the conditional likelihood $p(\mathbf{y}_{1:T}|\alpha, \phi, \sigma_\eta^2, \boldsymbol{\nu}', \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T})$, in this case $\boldsymbol{\theta}$ encompasses $(\alpha, \phi, \sigma_\eta^2, \boldsymbol{\nu}')'$, $\boldsymbol{\lambda}_{1:T}$ and $\mathbf{h}_{0:T}$.

The BPIC criterion is defined as

$$\text{BPIC} = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T}|\boldsymbol{\theta})\}] + 2T\hat{b}, \tag{4.3}$$

where $\hat{b}$ is given by

$$\begin{aligned}\hat{b} \approx \frac{1}{T}\{&E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T}|\boldsymbol{\theta})p(\boldsymbol{\theta})\}] \\ &- \log[p(\mathbf{y}_{1:T}|\hat{\boldsymbol{\theta}})p(\hat{\boldsymbol{\theta}})] + \text{tr}\{J_T^{-1}(\hat{\boldsymbol{\theta}})I_T(\hat{\boldsymbol{\theta}})\} + 0.5q\}.\end{aligned} \tag{4.4}$$

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

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

Table 2 *IBOVESPA return data set. DIC: deviance information criterion, BPIC: Bayesian predictive information criterion*

Model	DIC		BPIC	
	Value	Ranking	BPIC	Ranking
SVM-N	8055.53	3	8229.62	4
SVM-t	8054.90	2	8165.19	2
SVM-S	8038.64	1	7960.33	1
SVM-CN	8076.36	4	8222.54	3

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

with $\eta_T(y_t, \boldsymbol{\theta}) = \log p(y_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta}) + \log p(\boldsymbol{\theta})/T$. In the SVM-SMN class of models, the log-likelihood function, $\log p(\mathbf{y}_{1:T}|\boldsymbol{\theta})$, is estimated using the auxiliary particle filter (see, e.g., Pitt and Shephard, 1999) with 10,000 particles.

Next, we use the deviance information criterion (DIC) and the Bayesian predictive information criterion (BPIC) to compare all the competing models. In both cases, the best model has the smallest DIC (BPIC). From Table 2, the BPIC criterion indicates that the SVM-SMN models with heavy tails present a better fit than the basic SVM-N model, with the SVM-S model relatively better among all the considered models, suggesting that the IBOVESPA return data demonstrate sufficient departure from underlying normality assumptions. As expected, the DIC also selects the SVM-S model as the best.

We evaluate the SVM-SMN models by using the out-of-sample forecasting of the squared returns aggregated over a certain period of time. Based on the 1917 observations of returns used previously, we calculate the forecast over the following $1, 2, \ldots, 10$ days as described by Abanto-Valle et al. (2010). Figure 5 plots the posterior means and 95% posterior credibility interval of the aggregated squared returns together with the observed values. The 95% posterior intervals of the aggregated volatility, e^{h_t}, are also plotted. For all models, the 95% intervals of the aggregated squared returns are much wider than those for the aggregated volatility. The 95% posterior credibility interval of the aggregated squared returns for the SVM-S model does not include the observed values for days from 3, 4 and 10. The SVM-t model shows different forecasts, and days 3, 4, 5, 7 and 10 are outside the 95% credibility intervals. The SVM-CN model includes all the observed values of the aggregated squared returns for days from 1 to 10. The SVM-N model shows the worst behavior: it includes only the observed values for day 1.

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

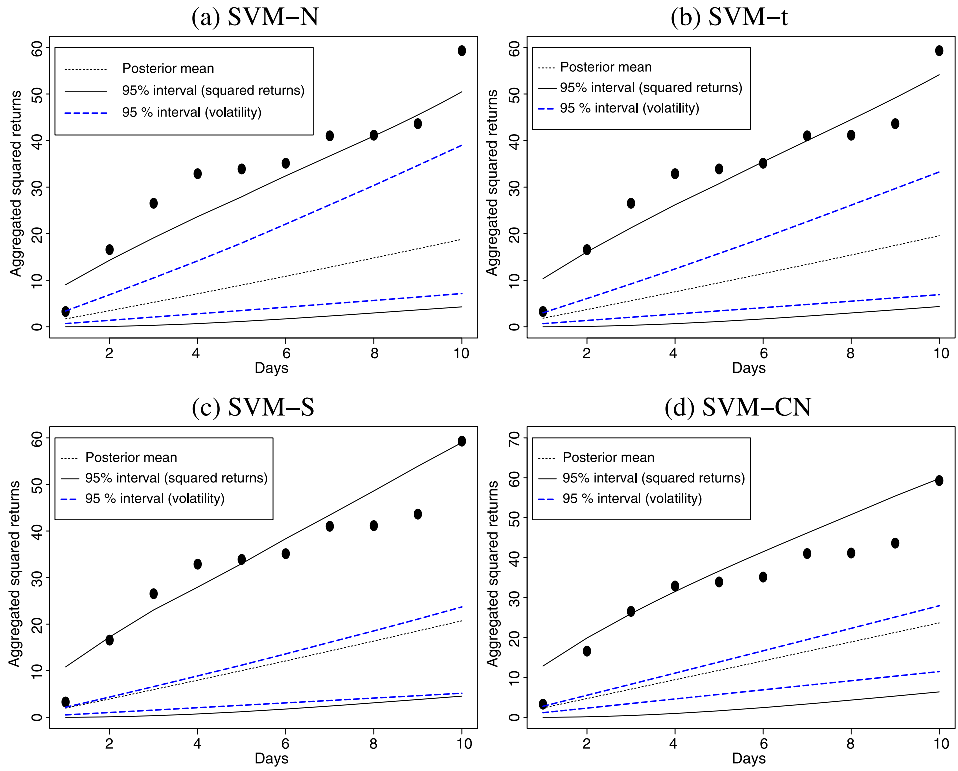

Figure 5 *IBOVESPA data set. Out-of-sample forecast of the aggregated squared returns for the* (a) *SVM-N*, (b) *SVM-t*, (c) *SVM-S and* (d) *SVM-CN models.*

mean and 95% credible interval of model parameters. The observations in $t = 1861$, 1870 and 1887, which correspond to July 5, 2005, July 28, 2005 and August 22, 2005, respectively, are contaminated by ky_t, where k varies from -6 and 6 with increments of 0.5 units. In Figure 6, we plot the posterior mean and 95% credible interval of ϕ and σ_η^2, respectively, for the SVM-N, the SVM-t and the SVM-S models. Clearly, the SVM-S and the SVM-t models are less affected by variations of k than the SVM-N model, meaning substantial robustness of the estimates over the usual normal process in the presence of outlying observations.

5 Conclusions

A Bayesian implementation of a robust alternative to estimation in the stochastic volatility in mean model (Koopman and Uspensky, 2002) via MCMC methods was presented in this article. The SVM enabled us to investigate the dynamic relationship between returns and their time-varying volatility. The Gaussian assumption of the mean innovation was replaced by univariate thick-tailed processes, known as

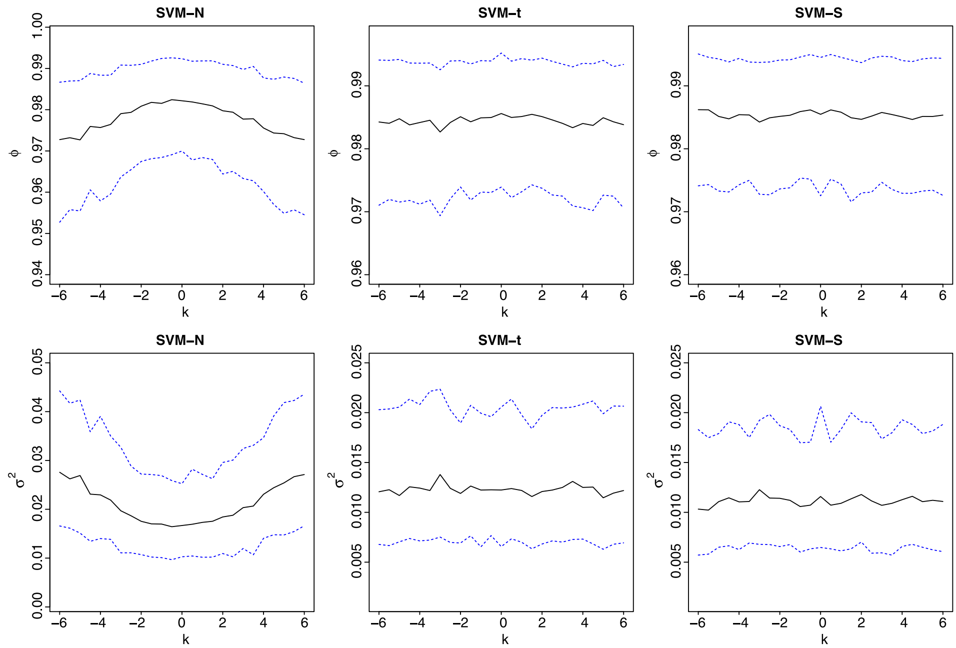

Figure 6 *Posterior mean (solid line) and 95% credible interval (dashed line) for fitting the SVM-N, SVM-t and SVM-S models to the IBOVESPA data set. top*: ϕ, *bottom*: σ_η^2. *The observations, which correspond to July* 5, 2005, *July* 28, 2005 *and August* 22, 2005, *respectively, are contaminated by* ky_t, *where k varies from* -6 *and* 6 *with increments of* 0.5 *units.*

scale mixtures of normal distributions. We studied three specific subclasses, viz. the Student-t, slash and the contaminated normal distributions, and compared 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 SVM-SMN model. We illustrated our methods through an empirical application of the IBOVESPA return series, which showed that the SVM-S model provides better fit than the SVM-N model in terms of parameter estimates, interpretation, robustness aspects and out-of-sample forecast of the aggregated squared returns. The β_2 estimate, which measures both the ex ante relationship between returns and volatility and the volatility feedback effect, was found to be negative. The results are in line with those of French (1987), who found a similar relationship between unexpected volatility dynamics and returns, and confirm the hypothesis that investors require higher expected returns when unanticipated increases in future volatility are highly persistent. This is consistent with our findings of higher values of ϕ combined with larger negative values for the in-mean parameter.

Our SVM-SMN models showed considerable flexibility to accommodate outliers, however, their robustness aspects could be seriously affected by the prior of the $\boldsymbol{\nu}$ parameter and the presence of skewness. In this setup, two natural extensions are still possible. The first would be to study different objective priors for form parameter in the Student-t and slash models in the same spirit of the works of Fonseca et al. (2008) and Salazar et al. (2009). The second would be to incorporate skewness and heavy-tailedness simultaneously using scale mixtures of skew-normal (SMSN) distributions, as proposed in Lachos et al. (2010). Nevertheless, a deeper investigation of these modifications is beyond the scope of the present paper, but provides stimulating topics for further research.

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 SVM-SMN class of models.

Full conditional distribution of $\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$

For parameters β_0, β_1 and β_2, we set the prior distributions as $\beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma^2_{\beta_0})$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(\bar{\beta}_1, \sigma^2_{\beta_1})$, $\beta_2 \sim \mathcal{N}(\bar{\beta}_2, \sigma^2_{\beta_2})$. The full conditionals are given by

$$\beta_0|\mathbf{y}_{0:T}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T}, \beta_1, \beta_2 \sim \mathcal{N}\left(\frac{b_{\beta_0}}{a_{\beta_0}}, \frac{1}{a_{\beta_0}}\right), \tag{A.1}$$

$$\beta_1|\mathbf{y}_{0:T}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T}, \beta_0, \beta_1 \sim \mathcal{N}_{(-1,1)}\left(\frac{b_{\beta_1}}{a_{\beta_1}}, \frac{1}{a_{\beta_1}}\right), \tag{A.2}$$

$$\beta_2|\mathbf{y}_{0:T}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T}, \beta_0, \beta_1 \sim \mathcal{N}\left(\frac{b_{\beta_2}}{a_{\beta_2}}, \frac{1}{a_{\beta_2}}\right), \tag{A.3}$$

where $a_{\beta_0} = \sum_{t=1}^T \lambda_t e^{-h_t} + \frac{1}{\sigma^2_{\beta_0}}$, $b_{\beta_0} = \sum_{t=1}^T \lambda_t e^{-h_t}(y_t - \beta_1 y_{t-1} - \beta_2 e^{h_t}) + \frac{\bar{\beta}_0}{\sigma^2_{\beta_0}}$, $a_{\beta_1} = \sum_{t=1}^T \lambda_t e^{-h_t} y^2_{t-1} + \frac{1}{\sigma^2_{\beta_1}}$, $b_{\beta_1} = \sum_{t=1}^T \lambda_t e^{-h_t}(y_t - \beta_0 - \beta_2 e^{h_t}) y_{t-1} + \frac{\bar{\beta}_1}{\sigma^2_{\beta_1}}$, $a_{\beta_2} = \sum_{t=1}^T \lambda_t e^{h_t} + \frac{1}{\sigma^2_{\beta_0}}$, $b_{\beta_2} = \sum_{t=1}^T \lambda_t (y_t - \beta_0 - \beta_1 y_{t-1}) + \frac{\bar{\beta}_2}{\sigma^2_{\beta_2}}$.

Full conditional distribution of $\boldsymbol{\alpha}$, $\boldsymbol{\phi}$ and $\boldsymbol{\sigma^2_\eta}$

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

$$\alpha|\mathbf{h}_{0:T}, \phi, \sigma^2_\eta \sim \mathcal{N}\left(\frac{b_\alpha}{a_\alpha}, \frac{1}{a_\alpha}\right) \tag{A.4}$$

$a_\alpha = \frac{1}{\sigma_\alpha^2} + \frac{T}{\sigma_\eta^2} + \frac{1+\phi}{\sigma_\eta^2(1-\phi)}$ and $b_\alpha = \frac{\bar{\alpha}}{\sigma_\alpha^2} + \frac{(1+\phi)}{\sigma_\eta^2} h_0 + \frac{\sum_{t=1}^{T}(h_t - \phi h_{t-1})}{\sigma_\eta^2}$. In a similar way, the conditional posterior of ϕ is given by

$$p(\phi|\mathbf{h}_{0:T}, \alpha, \sigma_\eta^2) \propto Q(\phi) \exp\left\{-\frac{a_\phi}{2\sigma_\eta^2}\left(\phi - \frac{b_\phi}{a_\phi}\right)^2\right\}\mathbb{I}_{|\phi|<1}, \tag{A.5}$$

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 $p(\phi|\mathbf{h}_{0:T}, \alpha, \sigma_\eta^2)$ in (A.5) does not have closed form, we sample from it by 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.

Finally, the full conditional 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 $\boldsymbol{\lambda_t}$ and $\boldsymbol{\nu}$

SV-t case. As $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, then $\lambda_t|y_t, y_{t-1}, h_t, \boldsymbol{\theta} \sim \mathcal{G}(\frac{\nu+1}{2}, ([y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t}]^2 e^{-h_t} + \nu)/2)$. We assume the prior distribution of ν as $\mathcal{G}(a_\nu, b_\nu)\mathbb{I}_{2<\nu\leq 40}$. Then, the full conditional of ν is

$$p(\nu|\boldsymbol{\lambda}_{1:T}) \propto \frac{[\nu/2]^{T\nu/2}\nu^{a_\nu - 1}e^{-(\nu/2)[\sum_{t=1}^{T}(\lambda_t - \log\lambda_t) + 2b_\nu]}}{[\Gamma(\nu/2)]^T}\mathbb{I}_{2<\nu\leq 40}. \tag{A.6}$$

As (A.6) does not have closed form, we sample ν by using the Metropolis–Hastings acceptance-rejection algorithm (Tierney, 1994; Chib and Greenberg, 1995). Let ν^* denote the mode (or approximate mode) of $p(\nu|\boldsymbol{\lambda}_{1:T})$, and let $\ell(\nu) = \log p(\nu|\boldsymbol{\lambda}_{1:T})$. 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 $\nu = \nu^*$ (see, Abanto-Valle et al., 2010, for details to prove the concavity of $\ell(\nu)$).

SV-S case. Using the fact that $\lambda_t \sim \mathcal{B}e(\nu, 1)$, then

$$\lambda_t|y_t, y_{t-1}, h_t, \boldsymbol{\theta} \sim \mathcal{G}_{(0<\lambda_t<1)}\left(\nu + \frac{1}{2}, \frac{1}{2}[y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t}]^2 e^{-h_t}\right),$$

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 $\mathcal{G}_{\nu>1}(T + a_\nu, b_\nu - \sum_{t=1}^{T}\log\lambda_t)$, that is, the left truncated gamma distribution.

SVM-CN case. Here λ_t is a discrete random variable and $\boldsymbol{\nu} = (\delta, \gamma)'$. To sample from λ_t, we introduce an auxiliary variable, S_t, such that $P(S_t = 1) = \delta$ and $\lambda_t = \gamma S_t + 1 - S_t$. The full conditional of S_t is given by

$$\begin{aligned} & p(S_t|\delta, \gamma, \beta_0, \beta_1, \beta_2, h_t, y_t, y_{t-1}) \\ & \quad \propto \delta^{S_t}(1-\delta)^{1-S_t}\gamma^{S_t/2} \\ & \qquad \times e^{-(1/2)[e^{-h_t}(\gamma S_t + 1 - S_t)(y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})^2]}. \end{aligned} \tag{A.7}$$

That is, $S_t|\delta, \gamma, \beta_0, \beta_1, \beta_2, h_t$ has a Bernoulli distribution. We assume that $\delta \sim \mathcal{B}e(\delta_0, \delta_1)$ and $\gamma \sim \mathcal{B}e(\gamma_0, \gamma_1)$. Then, the full conditional of $\delta|\gamma, \mathbf{S}_{1:T} \sim \mathcal{B}e(\delta_0^*, \delta_1^*)$, where $\delta_0^* = \delta_0 + \sum_{t=1}^T S_t$ and $\delta_1^* = \delta_1 + T - \sum_{t=1}^T S_t$. Thus, the full conditional of γ is given by

$$\begin{aligned} & p(\gamma|\beta_0, \beta_1, \beta_2, \mathbf{S}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{0:T}) \\ & \quad \propto (1-\gamma)^{\gamma_1 - 1}\gamma^{\gamma_0 - 1 + \sum_{t=1}^T S_t/2} \\ & \qquad \times e^{-(\gamma/2)\sum_{t=1}^T e^{-h_t} S_t (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})^2}. \end{aligned} \tag{A.8}$$

As (A.8) does not have closed form, we can sample from it by using the Metropolis–Hastings algorithm. Then the right truncated gamma distribution $\mathcal{G}_{0<\gamma<1}(\gamma_0 + \sum_{t=1}^T \frac{S_t}{2}, -\frac{1}{2}\sum_{t=1}^T e^{-h_t} S_t (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})^2)$ can be used as a proposal density.

Acknowledgments

We would like to thank the Editor and an anonymous referee for their useful comments, which improved the quality of this paper. The research of Carlos A. Abanto-Valle was supported by the CNPq. Helio S. Migon was supported by CNPq and CAPES/FAPERJ-PRONEX. V. H. Lachos acknowledges partial financial support from FAPESP and CNPq.

References

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H. and Enriquez, I. (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics & Data Analysis* **54**, 2883–2898. MR2727721

Ando, T. (2006). Bayesian inference for nonlinear and non-Gaussian stochastic volatility model with leverge effect. *Journal of Japan Statistical Society* **36**, 173–197. MR2338878

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika* **94**, 443–458. MR2380571

Andrews, D. F. and Mallows, S. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Ser. B* **36**, 99–102. MR0359122

Bekaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies* **13**, 1–42.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section* 177–181. Alexandria, VA: American Statistical Association.

Brandt, M. W. and Kang, Q. (2004). On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. *Journal of Financial Economics* **72**, 217–257.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* **31**, 281–318.

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.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *The American Statistician* **49**, 327–335.

Chib, S., Nardari, F. and Shepard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* **108**, 281–316. MR1894758

Choy, S. T. B. and Chan, J. S. K. (2008). Scale mixtures distributions in statistical modelling. *Australian & New Zealand Journal of Statistics* **50**, 135–146. MR2516871

de Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika* **82**, 339–350. MR1354233

Fonseca, T. C. O., Ferreira, M. A. R. and Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika* **95**, 325–333. MR2521587

French, K. R., Schert, W. G. and Stambugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics* **19**, 3–29.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 169–193. Oxford, UK: Clarendon Press. MR1380276

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. MR2083256

Koopman, S. J. (1993). Disturbance smoothers for state space models. *Biometrika* **80**, 117–126. MR1225219

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

Lachos, V. H., Ghosh, P. and Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal/independent linear mixed models. *Statistica Sinica* **20**, 303–322. MR2640696

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

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

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* **36**, 394–419.

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

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**, 347–370. MR1097532

Pemstein, D., Quinn, K. V. and Martin, A. D. (2011). The Scythe statistical library: An open source C++ library for statistical computation. *Journal of Statistical Software* **42**, 1–26.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association* **94**, 590–599. MR1702328

Rosa, G. J. M., Padovani, C. R. and Gianola, D. (2003). Robust linear mixed models with Normal/Independent distributions and bayesian MCMC implementation. *Biometrical Journal* **45**, 573–590. MR1998137

Salazar, E., Migon, H. S. and Ferreira, M. A. R. (2009). Objective Bayesian analysis for exponential power regression models. Technical report, Federal Univ. Rio de Janeiro, Dept. Statistics.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance* **44**, 1115–1153.

Shephard, N. and Pitt, M. (1997). Likelihood analysis of non-Gaussian measurements time series. *Biometrika* **84**, 653–667. MR1603940

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, Ser. B* **64**, 621–622. MR1979380

Theodossiou, P. and Lee, U. (1995). Relationship between volatility and expected returns across international stock markets. *Journal of Business Finance and Accounting* **22**, 289–300.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics* **22**, 1701–1762. MR1329166

Wang, J. and Genton, M. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference* **136**, 209–220. MR2207182

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. MR2050475

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

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