Stochastic volatility in mean models with scale mixtures of normal distributions and correlated errors: A Bayesian approach

C. A. Abanto-Valle[*,a], H. S. Migon[a], V. H. Lachos[b]

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

**Abstract**

The stochastic volatility in mean model with correlated errors using the symmetrical class of scale mixtures of normal distributions is introduced in this article. The scale mixture of normal distributions is an attractive class of symmetric distributions that includes the normal, Student-t, slash and contaminated normal distributions as special cases, providing a robust alternative to estimation in stochastic volatility in mean models in the absence of normality. Using a Bayesian paradigm, an efficient method based on Markov chain Monte Carlo (MCMC) is developed for parameter estimation. The methods developed are applied to analyze daily stock return data from the São Paulo Stock, Mercantile & Futures Exchange index (IBOVESPA). The Bayesian predictive information criteria (BPIC) and the logarithm of the marginal likelihood are used for model selection criteria. The results reveal that the stochastic volatility in mean model with correlated errors and slash distribution provides significant improvement in model fit for the IBOVESPA data over the usual normal model.
*Key words:* Feed-back and leverage effect, Markov chain Monte Carlo, non-Gaussian and nonlinear state space models, scale mixture of normal distributions, stochastic volatility in mean.

## 1. Introduction

The relationships between expected returns and expected volatility have been extensively examined in recent years. Theory generally predicts a positive relation between expected stock returns and volatility if investors are risk averse. That is, equity premium provides more compensation for

[*] Corresponding address: Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970, RJ, Brazil

*Email addresses:* cabantovalle@im.ufrj.br (C. A. Abanto-Valle ), migon@im.ufrj.br (H. S. Migon), hlachos@ime.unicamp.br (V. H. Lachos)

risk during more volatile periods. 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 et al. (1987) found a positive and significant relationship and Baillie and DeGennaro (1990) and Theodossiou and Lee (1995) reported a positive but non significant relationship between stock market volatility and stock returns. Consisting with the asymmetric volatility argument, Nelson (1991), Glosten et al. (1993) 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 et al. (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 stock price reactions to unfavorable events tend to be larger than reactions to favorable ones. This means 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 et al. (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. From an empirical perspective the fundamental difference between the leverage and volatility feedback explanations lies in the causality: the leverage effect explains why a negative return leads to higher subsequent volatility, whereas the volatility feedback effect justifies how an increase in volatility may result in negative returns.

Stochastic volatility (SV) models were introduced in the financial literature for describing time-varying volatilities (Taylor, 1982, 1986). Although the basic SV model offers great flexibility in modeling data with time-varying variances, it can suffer from a lack of robustness in the presence of extreme outlying observations as shown by (Liesenfeld and Jung, 2000; Abanto-Valle et al., 2010, among others). Usually, the volatility of daily stock returns has been estimated with SV models, but the results have relied on a extensive pre-modeling of these series to avoid the problem of simultaneous estimation of the mean and variance. The SV in mean (SVM) model introduced by

Koopman and Uspensky (2002) deals with the simultaneous estimation of the mean and variance. The unobserved volatility is incorporated as an explanatory variable in the mean equation of the returns.

In this article we propose to robustificate the specification of the innovation returns in SVM by introducing scale mixture of normal (SMN) distributions with correlated mean and variance errors. The resulting class of models takes into account the feed-back and leverage effect. We refer to this generalization as SVML-SMN models. Interestingly, this rich class contains as proper elements the SVML with normal (SVML-N), Student-t (SVML-t), slash (SVML-S) and the contaminated normal (SVML-CN) distributions. Indeed, the flexibility of the SVML with SMN distributions could fit time varying features in the mean of the returns and heavy tails simultaneously. The estimation of such intricate models is not straightforward, since volatility now appears in both the mean and the variance equation and hence intensive computational methods are needed for. Inference in this new class of SVML–SMN models is performed under a Bayesian paradigm via MCMC methods, which permits obtaining the posterior distribution of parameters by simulation, starting from reasonable prior assumptions on the parameters. We simulate the log-volatilities and the shape parameters by using the block sampler (Shephard and Pitt, 1997; Omori and Watanabe, 2008; Abanto-Valle et al., 2010) and Metropolis-Hastings algorithms, respectively.

The rest of the paper is structured as follows. Section 2 outlines the general class of the SVML–SMN models as well as the Bayesian estimation procedure using MCMC methods. In Section 3, the proposed class of models is applied to the BOVESPA daily index returns and model comparison is provided among the competing SVML models. Finally, Section 4 contains some concluding remarks and suggestions for future developments.

## 2. The heavy-tailed stochastic volatility in mean model

The SV in mean model with heavy-tails and correlated errors is defined by

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

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

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}_2\left[ \begin{pmatrix} 0 \\ 0, \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \tag{1c}$$

$$\lambda_t \sim p(\lambda_t \mid \boldsymbol{\nu}), \tag{1d}$$

where $y_t$ and $h_t$ are, respectively, the compounded return and the log-volatility at time $t$. We assume that $|\phi| < 1$, i.e., that the log-volatility process is stationary and that the initial value $h_1 \sim \mathcal{N}(\frac{\alpha}{1-\phi}, \frac{(1-\rho^2)\sigma_\eta^2}{1-\phi^2})$. The parameter $\rho$ measures the correlation between $\epsilon_t$ and $\eta_t$. When $\rho < 0$, this indicates the so-called leverage effect, a drop in the return followed by an increase in the volatility. An empirical evidence can be found in Ghysels et al. (1996), Harvey and Shephard (1996) and Bollerslev and Zhou (2005). In this setup, $\lambda_t$ is a scale factor, and $h(\lambda_t \mid \boldsymbol{\nu})$ is the mixing density which capture the heavy-tailness (see Choy and Chan, 2008; Abanto-Valle et al., 2010, for details about the choice of the mixing density). This class of models includes the SVML with Student-t (SVML-t), with slash (SVML-S) and contaminated normal (SVML-CN) distributions as special cases. The first and second models are obtained chosen the mixing density as: $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, $\lambda_t \sim \mathcal{B}e(\nu, 1)$ respectively, where $\mathcal{G}(.,.)$ and $\mathcal{B}e(.,.)$ denote the gamma and beta distributions respectively. In the SVML-CN model $\lambda_t$ follows a discrete distribution. When, $\lambda_t = 1$ for all $t$ and $\rho = 0$, we have the SVM model of Koopman and Uspensky (2002), as a particular case. In the next subsection we use MCMC methods to conduct the posterior analysis under a Bayesian paradigm. Conditionally to $\lambda_t$, some derivations are common to all members of the SVML-SMN family (see Appendix A for details).

Equations (1a)-(1c), can be written in an alternative way as

$$\begin{pmatrix} y_t \\ h_{t+1} \end{pmatrix} \Bigg| \boldsymbol{\theta}, \lambda_t, h_t, y_{t-1} \sim \mathcal{N}\left( \begin{bmatrix} \beta_0 + \beta_1 y_{t-1} + \beta_t e^{h_t} \\ \alpha + \phi h_t \end{bmatrix}, \begin{bmatrix} \lambda_t^{-1} e^{h_t} & \varphi \lambda_t^{-1/2} e^{h_t/2} \\ \varphi \lambda_t^{-1/2} e^{h_t/2} & \tau^2 + \varphi^2 \end{bmatrix} \right). \quad (2)$$

Then, $y_t|\boldsymbol{\theta}, \lambda_t, h_t, h_{t+1}, y_{t-1}$ follows a normal distribution with mean and variance given by

$$\mu_t = \beta_0 + \beta_1 y_{t-1} + \beta_t e^{h_t} + (\varphi/(\varphi^2 + \tau^2))\lambda_t^{-1/2} e^{h_t/2}(h_{t+1} - \alpha - \phi h_t) \quad (3)$$

$$V_t = \lambda_t^{-1} e^{h_t} \tau^2/(\tau^2 + \varphi^2), \quad (4)$$

respectively. The last density will be useful in the derivations of the block sampler given in the next section.

### *2.1. Parameter estimation via MCMC*

Let $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \alpha, \phi, \tau^2, \varphi, \boldsymbol{\nu}')'$ be the full parameter vector of the entire class of SVML-SMN models, where $\boldsymbol{\nu}$ is the parameter vector associated with the mixture distribution, $\tau = \sqrt{1-\rho^2}\sigma_\eta$ and $\varphi = \rho\sigma_\eta$, $\mathbf{h}_{1:T} = (h_1, h_1, \ldots, h_T)'$ 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 $T$. To make Bayesian analysis feasible for parameter estimation in the SVML-SMN class of models, we draw random samples from the posterior distribution of $(\boldsymbol{\theta}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T})$ given $\mathbf{y}_{1:T}$ using MCMC simulation methods. The sampling scheme is described by Algorithm 1.

**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 $\mathbf{h}_{1:T}^{(i)}$.
2. Generate $\boldsymbol{\theta}^{(i)}$ in turn from its full conditional distribution, given $\mathbf{y}_{1:T}$, $\mathbf{h}_{1:T}^{(i-1)}$ and $\boldsymbol{\lambda}_{1:T}^{(i-1)}$.
3. Draw $\boldsymbol{\lambda}_{1:T}^{(i)} \sim p(\boldsymbol{\lambda}_{1:T} \mid \boldsymbol{\theta}^{(i)}, \mathbf{h}_{1:T}^{(i-1)}, \mathbf{y}_{0:T})$.
4. Generate $\mathbf{h}_{0:T}$ by blocks as:
   i) For $l = 1, \ldots, K$, the knot positions are generated as $k_l$, the floor of $[T \times \{(l + u_l)/(K + 2)\}]$, where the $u_l^{\prime}s$ 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 parameters in the SVML–SMN class are set as: $\beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma^2_{\beta_0})$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(\bar{\beta}_1, \sigma^2_{\beta_1})$, $\beta_2 \sim \mathcal{N}(\bar{\beta}_2, \sigma^2_{\beta_2})$, $\alpha \mid \tau^2 \sim \mathcal{N}(\alpha_0, \tau^2/q_0)$, $\varphi \mid \tau^2 \sim \mathcal{N}(\varphi_0, \tau^2/p_0)$, $\phi \sim \mathcal{N}_{(-1,1)}(\phi_0, s^2_\phi)$, $\tau^2 \sim \mathcal{IG}(a_\tau/2, S_\tau/2)$, where $\alpha_0$, $\varphi_0$, $\phi_0$, $s^2_\phi$, $a_\tau$, $S_\tau$, $p_0$ e $q_0$ are known hyper parameters. The prior distribution of $\boldsymbol{\nu}$ is model specific (see details in Appendix A).

As described by Algorithm 1, the Gibbs sampler requires sampling parameters and latent variables from their full conditionals. Sampling the log-volatilities $\mathbf{h}_{1:T}$ in Step 4, due to the nonlinear setup of the observational equation (1a), is the most difficult task. An efficient strategy is to sample from the conditional posterior distribution of $\mathbf{h}_{1:T}$ by dividing it into several blocks and sampling each block given the other blocks. This idea, called the block sampler or multi-move sampler, is developed by Shephard and Pitt (1997) and Watanabe and Omori (2004) in the context of state space modeling. They show that the sampler can produce efficient draws from the target conditional posterior distribution in comparison with a single-move sampler which primitively samples one state, say $h_t$, at a time given the others, $h_s$ $(s \neq t)$. For the SV model with leverage, Omori and Watanabe (2008) develop the associated multi-move sampler and show that it produces efficient samples. In the next subsection, we extend their method for sampling $\mathbf{h}_{1:T}$ in the SVML-SMN class of models. Details on the full conditionals of $\boldsymbol{\theta}$ and the latent variable $\boldsymbol{\lambda}_{1:T}$ are given in Appendix A, some of them are easy to simulate from.

### *2.2. A block sampler algorithm*

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

$$\begin{aligned}
\log f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \mathbf{y}_{t:t+k}, \boldsymbol{\lambda}_{t+1:t+k}) \quad &\doteq \quad -\sum_{s=t}^{t+k-1} \frac{\eta_s^2}{2} + \sum_{s=t}^{t+k} l_s \\
&- \quad \frac{1}{2\sigma_\eta^2}(h_{t+k+1} - \alpha - \phi h_{t+k})^2 \mathbb{I}(t+k < T), \qquad (5)
\end{aligned}$$

where $\mathbb{I}(t+k < T)$ is an indicator variable. Excluding the constant terms $l_s$ denotes the conditional distribution of $y_s$ given $h_s$ and $h_{s+1}$ for $s < T$, which is normal with mean $\mu_s$ and variance $V_s$, which are given by equations (3) and (4) respectively. We define

$$L \quad = \quad \sum_{s=t}^{t+k} l_s - \frac{(h_{t+k+1} - \alpha - \phi h_{t+k})^2}{2\sigma_\eta^2} \mathbb{I}(t+k < T)$$

and $\mathbf{d}_{t+1:t+k} = (d_{t+1}, \dots, d_{t+k})'$, where $d_s$ and $\mathbf{Q}$ are given by equations (B.1) and (B.2) (see Appendix B, for details).

[1]For the last block, we have $y_T \mid y_{T-1}, h_T \sim \mathcal{N}(\beta_0 + \beta_1 y_{T-1} + \beta_2 \mathrm{e}^{h_t}, \lambda_T^{-1} \mathrm{e}^{h_T})$

As $-\frac{1}{2}\sum_{s=t}^{t+k-1}\eta_s^2+L$ in (5) does not have closed form, we use the Metropolis-Hastings acceptance-rejection algorithm (Chib and Greenberg, 1995) to sampling from. To obtain the proposal density, we are going to form an approximated linear state space model that mimics (5), from which sampling is easy. Applying a second order Taylor series expansion to $L$ around the mode $\hat{\boldsymbol{\eta}}_{t:t+k-1}$, we have

$$
\begin{aligned}
&\log f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k}) \\
&\qquad\approx \text{const} - \tfrac{1}{2}\textstyle\sum_{r=t+1}^{t+k}\eta_r^2 + \hat{L} + \left.\frac{\partial L}{\partial \boldsymbol{\eta}'_{t:t+k-1}}\right|_{\boldsymbol{\eta}_{t:t+k-1}=\hat{\boldsymbol{\eta}}_{t:t+k-1}}(\boldsymbol{\eta}_{t:t+k-1}-\hat{\boldsymbol{\eta}}_{t:t+k-1}) \\
&\qquad+\tfrac{1}{2}(\boldsymbol{\eta}_{t:t+k-1}-\hat{\boldsymbol{\eta}}_{t:t+k-1})'E\left.\left(\frac{\partial^2 L}{\partial \boldsymbol{\eta}_{t:t+k-1}\partial \boldsymbol{\eta}'_{t:t+k-1}}\right)\right|_{\boldsymbol{\eta}_{t:t+k-1}=\hat{\boldsymbol{\eta}}_{t:t+k-1}}(\boldsymbol{\eta}_{t:t+k-1}-\hat{\boldsymbol{\eta}}_{t:t+k-1}) \\
&\qquad= \text{const} - \tfrac{1}{2}\textstyle\sum_{r=t+1}^{t+k}\eta_r^2 + \hat{L} + \hat{\mathbf{d}}'_{t+1:t+k}(\mathbf{h}_{t+1:t+k}-\hat{\mathbf{h}}_{t+1:t+k}) \\
&\qquad\qquad-\tfrac{1}{2}(\mathbf{h}_{t+1:t+k}-\hat{\mathbf{h}}_{t+1:t+k})'\hat{\mathbf{Q}}(\mathbf{h}_{t+1:t+k}-\hat{\mathbf{h}}_{t+1:t+k}) \\
&\quad= \text{const} + \log f^*(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \boldsymbol{\theta}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k}),
\end{aligned}
\tag{6}
$$

where $\hat{\mathbf{d}}_{t+1:t+k}$, $\hat{L}$ and $\hat{\mathbf{Q}}$ denote $\mathbf{d}_{t+1:t+k}$, $L$ and $\mathbf{Q}$ evaluated at $\mathbf{h}_{t+1:t+k} = \hat{\mathbf{h}}_{t+1:t+k}$. The expectations are taken with respect to $y_s$'s conditional on $h_s$'s. We use an information matrix for $\mathbf{Q}$ because we require that $\mathbf{Q}$ is everywhere strictly positive definite. It can be shown that the proposal density $f^*(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \boldsymbol{\theta}, \mathbf{y}_{t+1:t+k}, \boldsymbol{\lambda}_{t+1:t+k})$ is the posterior density of $\boldsymbol{\eta}_{t:t+k-1}$ for a linear Gaussian state space model given by equations (7) and (8) below (see Omori and Watanabe, 2008, for details). The mode $\hat{\boldsymbol{\eta}}_{t:t+k-1}$ can be found by repeating the following algorithm until convergence.

**Algorithm 2.**

1. Initialize $\hat{\boldsymbol{\eta}}_{t:t+k-1}$ and calculate $\hat{\mathbf{h}}_{t+1:t+k}$ using (1b).
2. Evaluate $\hat{d}_s$, $\hat{M}_s$ and $\hat{N}_s$ using equations (B.1), (B.3) and (B.4) respectively.
3. Compute $G_s$, $J_s$ and $b_s$, for $s = t+2, \ldots, t+k$, recursively.

$$
\begin{aligned}
G_s &= \hat{M}_s - \hat{N}_s^2 G_{s-1}^{-1}, \quad G_{t+1} = \hat{M}_{t+1}, \\
J_s &= K_{s-1}^{-1}\hat{N}_s, \quad J_{t+1} = 0, \quad J_{t+k+1} = 0, \\
b_s &= \hat{d}_s - J_s K_{t-1}^{-1} b_{s-1} \quad b_{t+1} = \hat{d}_{t+1},
\end{aligned}
$$

where $K_s = \sqrt{G_s}$.

4. Define the auxiliary variables $\hat{y}_s = \hat{\gamma}_s + G_s^{-1} b_s$, where

$$\hat{\gamma}_s = \hat{h}_s + K_s^{-1} J_{s+1} \hat{h}_{s+1}, \quad s = t+1, \ldots, t+k.$$

5. Consider the linear Gaussian state-space model

$$\hat{y}_s = c_s + Z_s h_s + H_s \xi_s, \quad s = t+1, \ldots, t+k, \tag{7}$$

$$h_{s+1} = \alpha + \phi h_s + L_s \xi_s, \quad s = t, t+1, \ldots, t+k, \tag{8}$$

where $\xi_s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, $c_s = K_s^{-1} J_{s+1} \alpha$, $Z_s = 1 + K_s^{-1} J_{s+1} \phi$, $H_s = K_s^{-1}[1, J_{s+1}\sigma_\eta]$ and $L_s = [0, \sigma_\eta]$.

Apply the Kalman filter and a disturbance smoother (Koopman, 1993) to the linear Gaussian state space model in equations (7) and (8) and obtain the posterior mean of $\boldsymbol{\eta}_{t:t+k-1}$ ($\mathbf{h}_{t:t+k}$) and set $\hat{\boldsymbol{\eta}}_{t:t+k=1}$ ($\hat{\mathbf{h}}_{t:t+k}$) to this value.

6. Return to Step 2 and repeat the procedure until achieving convergence.

Applying the de Jong and Shephard's simulation smoother (de Jong and Shephard, 1995) to the model defined by equations (7) and (8) with the auxiliary variables $\hat{\mathbf{y}}_{t+1:t+k}$ defined in step 4 of Algorithm (2), enables us to sample $\boldsymbol{\eta}_{t+1:t+k}$ from the density $f^*$. Since $f$ is not bounded by $f^*$, we use the Metropolis-Hastings acceptance-rejection algorithm to sample from $f$ as recommended by Chib and Greenberg (1995). In the SVML-N case, we use the same procedure with $\lambda_t = 1$ for $t = 1, \ldots, T$.

In the MCMC sampling procedure, we select the expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ in Algorithm 2 as follows: the current sample of $\boldsymbol{\eta}_{t:t+k=1}$ ($\mathbf{h}_{t+1:t+k}$) may be taken as an initial value of the $\hat{\boldsymbol{\eta}}_{t:t+k=1}$ ($\hat{\mathbf{h}}_{t+1:t+k}$) in Step 1. Once an initial expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ is selected, we can calculate the auxiliary $\hat{\mathbf{y}}_{t+1:t+k}$ variables in Step 4. Then, applying the Kalman filter and a disturbance smoother to the linear Gaussian state space model consisting of equations (7) and (8) with the artificial $\hat{\mathbf{y}}_{t+1:t+k}$ yields the mean of $\mathbf{h}_{t+1:t+k}$ conditional on $\hat{\mathbf{h}}_{t+1:t+k}$ in the linear Gaussian state space model, which is used as the next $\hat{\mathbf{h}}_{t+1:t+k}$. By repeating the procedure until the smoothed estimates converge, we obtain the posterior mode of $\mathbf{h}_{t+1:t+k}$. This is equivalent to the method of scoring to maximize the logarithm of the conditional posterior density. Although, we have just noted that iterating the procedure achieves the mode, this will slow our simulation algorithm if we have to iterate this procedure until full convergence. Instead we suggest to use only five iterations of this procedure to provide reasonably good sequence $\hat{\mathbf{h}}_{t+1:t+k}$ instead of an optimal one.

Finally, we describe the updating procedure of the knot conditions $h_{k_l}$, for $l = 2, \ldots, K$. As the conditional density $p(h_{k_l} \mid h_{k_l-1}, h_{k_l+1})$ does not have a closed form, we use the Metropolis-Hastings algorithm with proposal density $\mathcal{N}(\frac{\alpha(1-\phi)+\phi(h_{k_l-1}+h_{k_l+1})}{1+\phi^2}, \frac{\sigma_\eta^2}{1+\phi^2})$. 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 $\alpha_{MH} = \min\{1, \frac{Q(h_{k_l}^p)}{Q(h_{k_l}^{(i-1)})}\}$, where $Q(h_{k_l})$ is the product of the conditional densities $y_{k_l-1} \mid \lambda_{k_l-1}, y_{k_l-2}, h_{k_l-1}, h_{k_l} \sim \mathcal{N}(\mu_{k_l-1}, V_{k_l-1})$ and $y_{k_l} \mid \lambda_{k_l}, y_{k_l-1}, h_{k_l+1}, h_{k_l} \sim \mathcal{N}(\mu_{k_l}, V_{k_l})$, with $\mu_s$ and $V_s$ are defined by equations (3) and (4) respectively, for $s = k_l - 1$ and $k_l$.

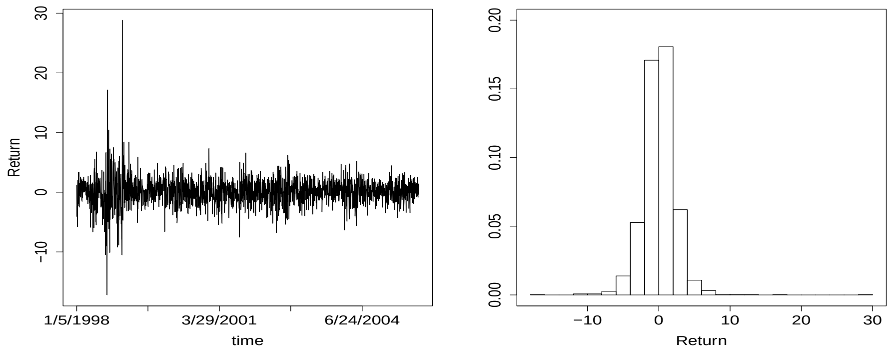


Figure 1: Compounded IBOVESPA returns from January 5, 1998 to September 3, 2005. The left panel shows the plot of the raw series and the right panel the histogram of returns.

## 3. Empirical Application

This section analyzes the daily closing prices of the IBOVESPA. The IBOVESPA 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, $y_t = 100(\log P_t - \log P_{t-1})$, where $P_t$ is the closing price on day $t$.

The corrected compounded IBOVESPA returns are plotted in Figure 1 as a time series plot and also 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 raise (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 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 SVML-N, SVML-t, SVML-S and SVML-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 2.1. We set the prior distributions of the common parameters as: $\beta_0 \sim \mathcal{N}(0, 100)$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(0.1, 100)$, $\beta_2 \sim \mathcal{N}(-0.1, 100)$, $\phi \sim \mathcal{N}_{(-1,1)}(0.95, 100)$, $\tau^2 \sim \mathcal{IG}(2.5, 0.025)$, $\alpha \mid \tau^2 \sim \mathcal{N}(0, \tau^2/0.002)$ and $\varphi \mid \tau^2 \sim \mathcal{N}(-0.3, \tau^2/0.005)$. The prior distributions on the shape parameters were chosen as: $\nu \sim \mathcal{G}(12, 0.8)$ and $\nu \sim \mathcal{G}(2, 0.25)$ for the SVML-t and the SVML-S models, respectively. For the SVML-CN, we set $\delta \sim \mathcal{B}e(2, 2)$ and $\gamma \sim \mathcal{B}e(2, 4)$. 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 $p(\lambda_t \mid \nu)$. 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., 2007), 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 SVML-N, SVML-t and the SVML-S models, we conducted the MCMC simulation for 50000 iterations. However, for the SVML-CN model, we used 210000 iterations. In both cases, the first 10000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain, only every $10th$ (SVML-N, SVML-t and SVML-S models) and $100th$ (SVML-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). Table 1 summarizes the results. According to the CD values, the null hypothesis that the sequence of 4000 (2000) draws is stationary was accepted at the 5% level for all the parameters in all the models considered here. The inefficiency factor is defined by $1 + \sum_{s=1}^{\infty} \rho_s$ where $\rho_s$ is the sample autocorrelation at lag $s$. It measures how well the MCMC chain mixes (see, e.g, Chib, 1972). It is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. When the inefficiency factor is equal to $m$, we need to draw MCMC samples $m$ times as many as uncorrelated samples. From Table 1, we found a well mixing of the MCMC chain produced by our algorithm.

From Table 1, the posterior mean and 95% interval of $\phi$ in the SVML-S is higher than those of the other three models. However, for all the models, we found that the posterior means of $\phi$ are above 0.93, showing higher persistence. We found that the persistence of the SVML-t and the SVML-S are slightly different from that the SVML-N and SVML-CN models. The posterior mean of $\sigma_\eta^2$ is smaller in the SVML-S than those of the SVML-N, SVML-t and the SVML-CN models, indicating that the volatility of the SVML-S is less variable than those of the other three models. We also found that the posterior mean of $\sigma_\eta^2$ of the SVML-t and the SVML-CN model are smaller than the SVML-N case.

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 $\beta_0$ is always positive and statistically significant for the SVML-N, SVML-t and SVML-S. In the SVML-CN model the 95% interval of $\beta_0$ contains zero. We found that the posterior mean of $\beta_1$ 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 $\beta_2$ 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. Although the posterior credibility interval of $\beta_2$ barely contains zero for all the models, its posterior distribution is primarily located in the negative range as shown in Table 2. 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 higher expected returns.

As expected for all the models considered here, the posterior means of $\rho$, the correlation coefficient between shocks to return at time $t$ and shocks to volatility at time $t+1$, is always negative

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

| Parameter | SVML-N | SVML-t | SVML-S | SVML-CN |
|---|---|---|---|---|
| | 0.1409 | 0.1801 | 0.2293 | 0.1421 |
| $\beta_0$ | (0.0031,0.2820) | (0.0269,0.3388) | (0.0899,0.3736) | (-0.0003,0.2779) |
| | 0.61 | -1.67 | -0.26 | 0.70 |
| | 1.33 | 2.34 | 3.43 | 1.33 |
| | 0.0299 | 0.0242 | 0.0230 | 0.0295 |
| $\beta_1$ | (-0.0562,0.0239) | (-0.0219,0.0682) | (-0.0142,0.0593) | (-0.0137,0.0753) |
| | -0.18 | -0.18 | -0.76 | -0.07 |
| | 1.49 | 1.60 | 1.44 | 0.88 |
| | -0.0179 | -0.0343 | -0.0776 | -0.0192 |
| $\beta_2$ | (-0.0559,0.0193) | (-0.0894,0.0160) | (-0.1606,0.0063) | (-0.0571,0.0198) |
| | 0.69 | 0.66 | 0.30 | 0.79 |
| | 1.27 | 3.02 | 4.88 | -0.92 |
| | 0.0713 | 0.0407 | 0.0185 | 0.0722 |
| $\alpha$ | (0.0271, 0.1196) | (0.0147,0.0758) | (0.0342,0.1175) | (0.0002,0.0059) |
| | 1.47 | -1.49 | -1.00 | -1.32 |
| | 23.55 | 28.31 | 6.22 | 15.59 |
| | 0.9368 | 0.9579 | 0.9677 | 0.9376 |
| $\phi$ | (0.8940,0.9765) | (0.9184,0.9855) | (0.9745,0.9947) | (0.8929,0.9761) |
| | -1.33 | 1.47 | 0.34 | -0.96 |
| | 25.29 | 36.05 | 11.40 | 2.38 |
| | 0.0708 | 0.0426 | 0.0279 | 0.0701 |
| $\sigma_\eta^2$ | (0.0250,0.1214) | (0.0146,0.0818) | (0.0163,0.0430) | (0.0262,0.1211) |
| | 1.31 | -1.46 | -0.65 | 1.41 |
| | 28.74 | 44.37 | 15.46 | 2.86 |
| | -0.3112 | -0.3445 | -0.3217 | -0.3081 |
| $\rho$ | (-0.4677,-0.1774) | (-0.5319,-0.1779) | (-0.4559,-0.1889) | (-0.4601,-0.1715 ) |
| | 1.36 | -1.75 | -0.06 | 1.90 |
| | 11.13 | 21.96 | 6.49 | 1.93 |
| | – | 10.9988 | 1.8787 | – |
| $\nu$ | – | (6.9690,16.9087) | (1.5257,2.3398) | – |
| | – | 1.32 | -0.48 | – |
| | – | 26.25 | 15.20 | – |
| | – | – | – | 0.6342 |
| $\delta$ | – | – | – | (0.3849,0.8643) |
| | – | – | – | -0.82 |
| | – | – | – | 3.62 |
| | – | – | – | 0.9730 |
| $\gamma$ | – | – | – | (0.8928,0.9982) |
| | – | – | – | -1.03 |
| | – | – | – | 2.75 |

Table 2: IBOVESPA return series: estimated $P(\beta_2 < 0)$

| | SVML-N | SVML-t. | SVML-S | SVML-CN |
|---|---|---|---|---|
| $P(\beta_2 < 0)$ | 0.8295 | 0.9055 | 0.9634 | 0.8320 |

and the 95% posterior credibility intervals do not contain zero. This result indicates the parameter is statistically significant. Hence, we may conclude that there is a strong and significant "leverage effect" for the IBOVESPA returns data set.

The magnitude of the tail fatness is measured by the shape parameter $\nu$ in the SVML-t and SVML-S models. In the SVML-CN case it is measured by $\delta$. The posterior means of $\nu$ are almost 11 and 1.9 in the SVML-t and SVML-S models respectively. In the SVML-CN the posterior mean of $\delta$ is 0.63, with $\gamma$ as a scale factor, has a posterior mean of 0.97. These results seem to indicate that the measurement error of the stock returns are better explained by heavy-tailed distributions.

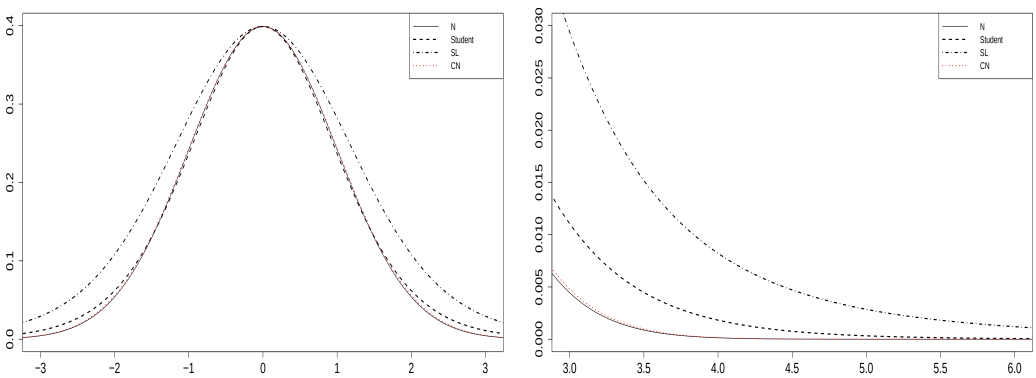


Figure 2: Density curves of the univariate normal, Student-t, slash and contaminated normal distributions using the estimated tail-fatness parameter from the respective SVM model.

The reason why the estimated volatility of the SVML-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 $\nu$ degrees of freedom, the slash ($\mathcal{S}(0,1,\nu)$) density with shape parameter $\nu$ and the contaminated normal ($\mathcal{CN}(0,1,\delta,\gamma)$). We set $\nu$, $\delta$ and $\gamma$ as the posterior mean of the respective SVML 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 SVML-SMN class of models considered here attributes a relatively larger proportion of extreme return values to $\epsilon_t$ instead of $\eta_t$ than the SVML-N model, making the volatility of the SVML-t, SVML-S and SVML-CN 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 smoothed mean of $e^{\frac{h_t}{2}}$. The posterior smoothed mean of $e^{\frac{h_t}{2}}$ of the SVML-t, SVML-S and SVML-CN models show smoother movements than that from the SVML-N model (solid line). Extreme returns, such a 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 $e^{\frac{h_t}{2}}$ by $\lambda_t^{-1}e^{\frac{h_t}{2}}$. This can have a substantial impact, for instance, in the valuation of derivative instruments and several strategic or tactical asset allocation topics.

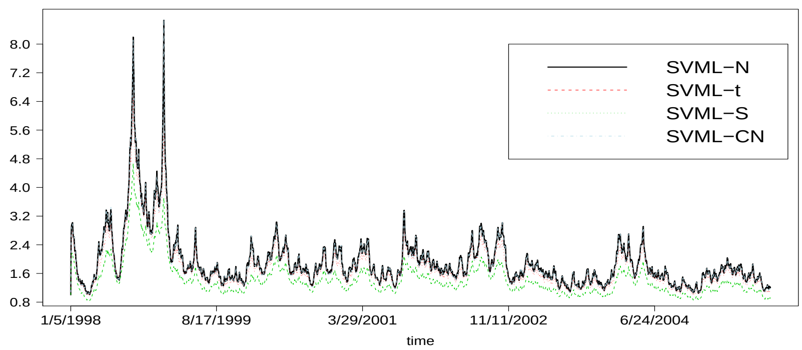


Figure 3: IBOVESPA data set. Posterior smoothed mean of $e^{\frac{h_t}{2}}$

To assess the goodness of the estimated models, we calculate the Bayesian predictive information criteria, BPIC (Ando, 2006, 2007) and the logarithm marginal likelihood, log −ML (Chib, 1995; Chib and Jeliazov, 2001). The BPIC criteria is defined as

$$BPIC \quad = \quad -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\}] + 2T\hat{b}, \tag{9}$$

where, $\hat{b}$ is given by

$$\hat{b} \quad \approx \quad \frac{1}{T}\left\{E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})\}] - \log[p(\mathbf{y}_{1:T} \mid \hat{\boldsymbol{\theta}})p(\hat{\boldsymbol{\theta}})] + \mathrm{tr}\{J_T^{-1}(\hat{\boldsymbol{\theta}})I_T(\hat{\boldsymbol{\theta}})\} + 0.5q\right\}. \tag{10}$$

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

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

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

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

The marginal likelihood is defined as the integral of the likelihood with respect to the prior density of the parameter. Following Chib (1995), we estimate the logarithm of the marginal likelihood $m(y)$

$$\log m(y) \quad = \quad \log p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) - \log p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}). \tag{11}$$

The equality holds for any values of $\boldsymbol{\theta}$; but we use the posterior mode of $\hat{\boldsymbol{\theta}}$ to obtain a stable estimate of $m(y)$. In the SVML-SMN, the log-likelihood function, $\log p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})$, is estimated using the auxiliary particle filter (see, e.g. Pitt and Shephard, 1999; Omori et al., 2007) with 10000 particles. The number of iterations for the reduced runs is set to 5000. Table 3 shows the BPIC and the logarithm of the estimated marginal likelihoods. The BPIC criterion indicates 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 logarithm of the estimated marginal likelihood also selects the SVML-S model as the best.

The robustness aspects of the SVML-SMN models can be studied through the influence of outliers on the posterior distribution of the parameters. We consider only the SVML-t and the SVML-S

Table 3: IBOVESPA return data set. DIC: deviance information criterion, BPIC: Bayesian predictive information criterion.

| | BPIC | | log −ML | |
|---|---|---|---|---|
| Model | BPIC | Ranking | log Chib | Ranking |
| SVML-N | 8074.6 | 4 | -4069.9 | 4 |
| SVML-t | 8076.6 | 2 | -4044.6 | 3 |
| SVML-S | 8074.1 | 1 | -3766.4 | 1 |
| SVM-CN | 8075.9 | 3 | -3871.0 | 2 |

models for illustrative purposes. We study the influence of three contaminated observations on the posterior estimates of mean and 95% credible interval of model parameters. The observations in $t = 1861, 1870$ and 1887, which corresponds 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 Figures 4 we plot the posterior mean and 95% credible interval of $\phi$, $\sigma_\eta^2$ and $\rho$, respectively, for the SVML-N, the SVML-t and the SVML-S models. Clearly, the SVML-S and the SVML-t models are less affected by variations of $k$ than the SVML-N model, meaning substantial robustness of the estimates over the usual normal process in the presence of outlying observations.

## 4. Conclusions

This article presented a Bayesian implementation of a robust alternative to estimation in the stochastic volatility in mean model with correlated errors, as an extension of the model proposed by Koopman and Uspensky (2002), via MCMC methods. The SVML 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 scale mixtures of skew-normal distributions. We studied three specific sub-classes, 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 SVML-SMN model. We illustrated our methods through an empirical application of the IBOVESPA return series, which showed that the SVML-S model provides better fit than the SVML-N model in terms of parameter estimates, interpretation and robustness aspects. The $\beta_2$ estimate, which measures both the *ex ante* relationship between returns and volatility and

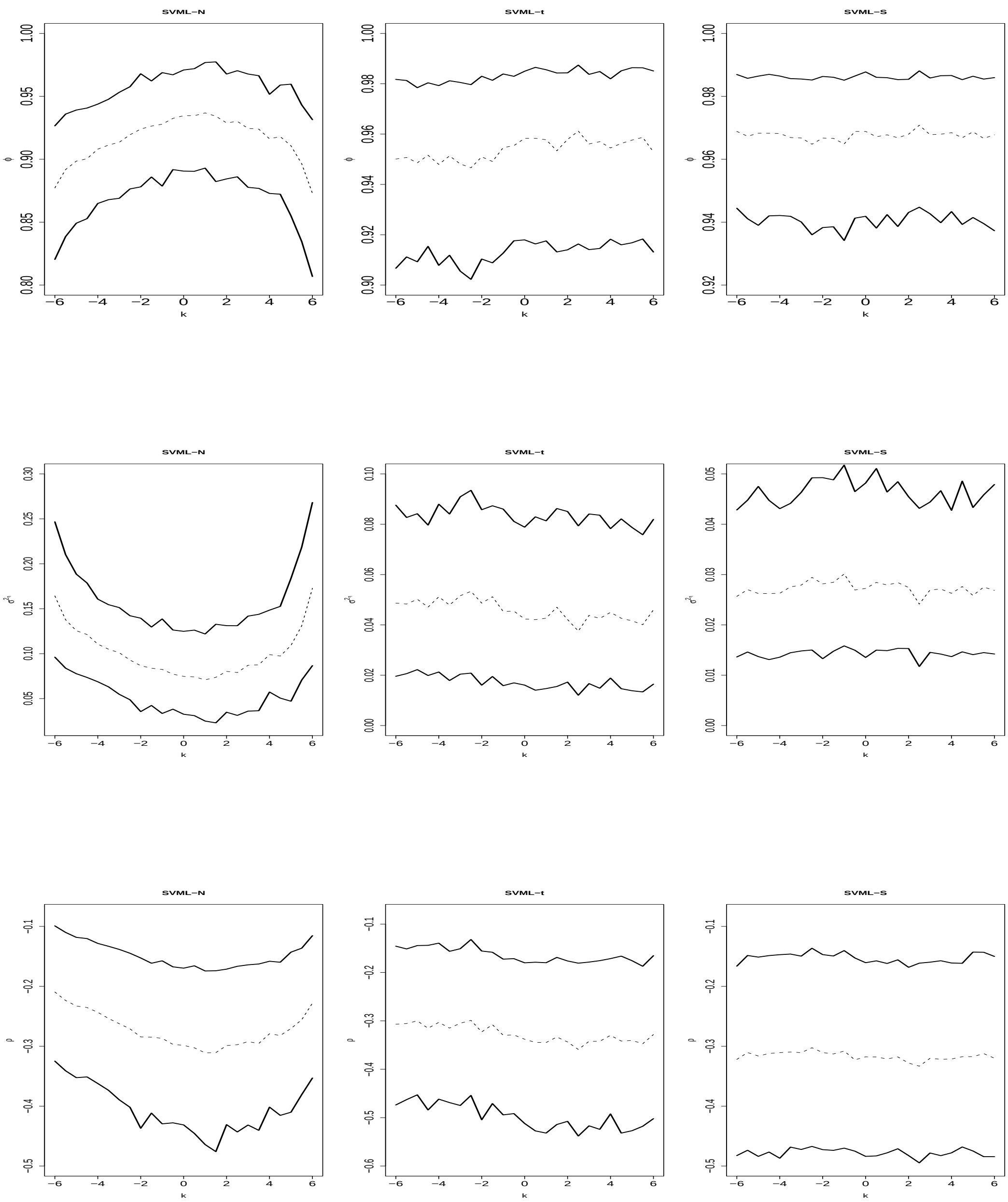


Figure 4: Posterior mean (dashed line) and 95% credible interval (solid line) of fitting the SVML-N, SVML-t and SVML-S models to the IBOVESPA data set. Top: $\phi$, middle: $\sigma^2$ and bottom: $\rho$. The observations which corresponds to July 5, 2005, July 28, 2005 and August 22, 2005, respectively, are contaminated by $ky_t$, where $k$ varies from -6 and 6 with increments of 0.5 units.

the volatility feedback effect, was found to be negative. The results are line with those of French et al. (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 $\phi$ combined with larger negative values for the in-mean parameter. By the other hand, as the posterior mean and 95% posterior credibility interval contains only negative values, we can conclude that there is a strong and significant "leverage effect" for the IVOBESPA returns data set.

Our SVML-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 set-up, 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.

**Acknowledgements**

The authors would like to thank the Executive 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 Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq-Brazil). The second author was supported by CNPq, FAPERJ and CAPES. The third author acknowledges partial financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and CNPq-Brazil.

**Appendix A: 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 SVML-SMN class of models.

*Full conditional distribution of $\beta_0$, $\beta_1$ and $\beta_2$*

Let $m_t$ and $V_t$ be defined by

$$m_t = \begin{cases} \lambda_t^{-1} e^{\frac{h_t}{2}} \frac{\varphi}{\tau^2+\varphi^2}(h_{t+1} - \alpha - \phi h_t), & t < T, \\ 0, & t = T, \end{cases} \qquad V_t = \begin{cases} \lambda_t^{-1} e^{h_t} \frac{\varphi}{\tau^2+\varphi^2}, & t < T, \\ \lambda_t^{-1} e^{h_t}, & t = T, \end{cases}$$

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

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

$$\beta_1 \mid \mathbf{y}_{0:T}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T}, \beta_0, \beta_1 \sim \mathcal{N}(\frac{b_{\beta_1}}{a_{\beta_1}}, \frac{1}{a_{\beta_1}})\mathbb{I}_{|\beta_2|<1}, \tag{A.2}$$

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

where $a_{\beta_0} = \sum_{t=1}^T \frac{1}{V_t} + \frac{1}{\sigma^2_{\beta_0}}$, $b_{\beta_0} = \sum_{t=1}^T \frac{w_t}{V_t} + \frac{\bar{\beta}_0}{\sigma^2_{\beta_0}}$, $a_{\beta_1} = \sum_{t=1}^T \frac{y^2_{t-1}}{V_t} + \frac{1}{\sigma^2_{\beta_1}}$, $b_{\beta_1} = \sum_{t=1}^T \frac{z_t y_{t-1}}{V_t} + \frac{\bar{\beta}_1}{\sigma^2_{\beta_1}}$, $a_{\beta_2} = \sum_{t=1}^T \frac{e^{2h_t}}{V_t} + \frac{1}{\sigma^2_{\beta_2}}$, $b_{\beta_2} = \sum_{t=1}^T \frac{r_t e^{h_t}}{V_t} + \frac{\bar{\beta}_2}{\sigma^2_{\beta_2}}$, $w_t = y_t - \beta_1 y_{t-1} - \beta_2 e^{h_t} - m_t$, $z_t = y_t - \beta_0 - \beta_2 e^{h_t} - m_t$, $r_t = y_t - \beta_0 - \beta_1 y_{t-1} - m_t$ and $\mathbb{I}_{|\beta_2|<1}$, an indicator variable.

*Full conditional distribution of $\alpha$, $\phi$, $\varphi$ and $\tau^2$*

We assume the following prior's distributions: $\alpha \mid \tau^2 \sim \mathcal{N}(\alpha_0, \tau^2/q_0)$, $\varphi \mid \tau^2 \sim \mathcal{N}(\varphi_0, \tau^2/p_0)$, $\phi \sim \mathcal{N}_{(-1,1)}(\phi_0, s^2_\phi)$, $\tau^2 \sim \mathcal{GI}(a_\tau/2, S_\tau/2)$, where $\alpha_0$, $\varphi_0$, $\phi_0$, $s^2_\phi$, $a_\tau$, $S_\tau$, $p_0$ e $q_0$ are known hyper parameters.

After some simples but tedious algebra, we have

$$\alpha \mid . \sim \mathcal{N}(\frac{B_\alpha}{A_\alpha}, \frac{\tau^2}{A_\alpha}), \tag{A.4}$$

$$\varphi \mid . \sim \mathcal{N}(\frac{B_\varphi}{A_\varphi}, \frac{\tau^2}{A_\varphi}), \tag{A.5}$$

where $A_\alpha = q_0 + \frac{1+\phi}{1-\phi} + T - 1$, $B_\alpha = \alpha_0 q_0 + (1+\phi)h_1 + \sum_{t=1}^{T-1} k_t$, $k_t = h_{t+1} - \phi h_t - \varphi(y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})\lambda_t^{\frac{1}{2}} e^{-\frac{h_t}{2}}$, $A_\varphi = p_0 + \sum_{t=1}^{T-1}(y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})^2 \lambda_t e^{-h_t}$, $B_\varphi = \varphi_0 p_0 + \sum_{t=1}^{T-1} c_t (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})\lambda_t^{\frac{1}{2}} e^{-\frac{h_t}{2}}$ and , $c_t = h_{t+1} - \alpha - \phi h_t$. In a similar way, the conditional posterior of $\phi$ is given by

$$p(\phi \mid .) \propto Q(\phi) \exp\Big\{ -\frac{A_\phi}{2}\Big(\phi - \frac{B_\phi}{A_\phi}\Big)^2 \Big\}, \tag{A.6}$$

where

$$Q(\phi) = \sqrt{1-\phi^2}\exp\{-\frac{1-\phi^2}{2\tau^2}(h_1 - \frac{\alpha}{1-\phi})^2\},$$

$l_t = h_{t+1} - \alpha - \varphi(y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})\lambda_t^{\frac{1}{2}} e^{-\frac{h_t}{2}}$, $A_\phi = \frac{1}{s_\phi^2} + \sum_{t=1}^{T-1}\frac{h_t^2}{\tau^2}$, $B_\phi = \frac{\phi_0}{s_\phi^2} + \sum_{t=1}^{T-1}\frac{l_t h_t}{\tau^2}$ and $\mathbb{I}_{|\phi|<1}$ is an indicator variable. As $p(\phi \mid \mathbf{h}_{1:T}, \alpha, \sigma_\eta^2)$ in (A.6) 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. The conditional posterior of $\tau^2$ is $\mathcal{IG}(\frac{T_1}{2}, \frac{M_1}{2})$, where $T_1 = a_\tau + T + 1$ and $M_1 = (1-\phi^2)(h_1 - \frac{\alpha}{1-\phi})^2 + \sum_{t=1}^{T-1}(c_t - \varphi\lambda_t^{\frac{1}{2}} e^{-\frac{h_t}{2}}(y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t})^2 + p_0(\varphi-\varphi_0)^2 + q_0(\alpha-\alpha_0)^2 + S_\tau$. Once $\tau^2$ and $\varphi$ are sampled, respectively, from their conditional posteriors, we can calculate $\rho$ and $\sigma^2$ through $\sigma_\eta^2 = \tau^2 + \varphi^2$ and $\rho = \varphi/\sigma_\eta$.

*Full conditional of $\lambda_t$ and $\nu$*

- **SV-t case**

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

$$p(\lambda_t \mid y_t, y_{t-1}, h_t, h_{t+1}, \beta_0, \beta_1, \beta_2, \varphi, \tau^2, \nu) \propto \lambda_t^{\frac{\nu+1}{2}-1} e^{-\frac{\lambda_t}{2}[(\frac{\tau^2+\varphi^2}{\tau^2})u_t^2 e^{-h_t}+\nu]} Q(\lambda_t), \quad t < T, \tag{A.7}$$

where $Q_t(\lambda_t) = e^{[\lambda_t^{\frac{1}{2}}\frac{\varphi}{\tau^2}u_t e^{-\frac{h_t}{2}}(h_{t+1}-\alpha-\phi h_t)]}$ and $u_t = y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t}$. From (A.7), the full conditional of $\lambda_t$ does not have closed form. We simulate from it using the Metropolis-Hastings algorithm with proposal density $\mathcal{G}(\frac{\nu+1}{2}, \frac{1}{2}[(\frac{\tau^2+\varphi^2}{\tau^2})u_t^2 e^{-h_t} + \nu])$ . A proposed $\lambda_t^*$ is accepted com probability $a = \min\{1, \frac{Q(\lambda_t^*)}{Q(\lambda_t^{(i-1)})}\}$. For $t = T$, the full conditional of $\lambda_T$ is $\mathcal{G}(\frac{\nu+1}{2}, \frac{1}{2}[u_T^2 e^{-h_T} + \nu])$.

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

$$p(\nu \mid \boldsymbol{\lambda}_{1:T}) \propto \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} e^{-\frac{\nu}{2}[\sum_{t=1}^{T}(\lambda_t - \log\lambda_t)+2b_\nu]}}{[\Gamma(\frac{\nu}{2})]^T}\mathbb{I}_{2<\nu\leq 40}. \tag{A.8}$$

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

$$\log\Gamma(\nu) = \frac{\log(2\pi)}{2} + \frac{2\nu-1}{2}\log(\nu) - \nu + \frac{\theta}{12\nu}, \qquad 0 < \theta < 1. \tag{A.9}$$

Taking the second derivative of $\ell(\nu)$ from (A.8) and using (A.9), we have that

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

because in practical applications $T \geq 2$.

• **SV-S case**

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

$$p(\lambda_t \mid y_t, y_{t-1}, h_t, h_{t+1}, \beta_0, \beta_1, \beta_2, \varphi, \tau^2, \nu) \propto \lambda_t^{\nu+\frac{1}{2}-1} \mathrm{e}^{-\frac{\lambda_t}{2}[(\frac{\tau^2+\varphi^2}{\tau^2})u_t^2 \mathrm{e}^{-h_t}]} Q(\lambda_t)\mathbb{I}_{0<\lambda_t<1}, \quad t < T, \tag{A.10}$$

where $Q_t(\lambda_t) = \mathrm{e}^{[\lambda_t^{\frac{1}{2}} \frac{\varphi}{\tau^2} u_t \mathrm{e}^{-\frac{h_t}{2}}(h_{t+1}-\alpha-\phi h_t)]}$ and $u_t = y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 \mathrm{e}^{h_t}$. From (A.10), the full conditional of $\lambda_t$ does not have closed form. We simulate from it using the Metropolis-Hastings algorithm with proposal density that is $\lambda_t \sim \mathcal{G}_{(0<\lambda_t<1)}(\nu + \frac{1}{2}, \frac{1}{2}[(\frac{\tau^2+\varphi^2}{\tau^2})u_t^2 \mathrm{e}^{-h_t}])$, the right truncated gamma distribution. For $t = T$, the full conditional of $\lambda_T$ is $\mathcal{G}(\nu + \frac{1}{2}, \frac{u_T^2 \mathrm{e}^{-h_T}}{2})$.

Assuming that a prior distribution of $\nu \sim \mathcal{G}(a_\nu, b_\nu)$, the full conditional distribution of $\nu$ is given by

$$p(\nu \mid \mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}) \propto \nu^{T+a_\nu-1} \mathrm{e}^{-\nu(b_\nu - \sum_{t=1}^{T} \log \lambda_t)} \mathbb{I}_{\nu>1}. \tag{A.11}$$

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

• **SVM-CN case**

Here $\lambda_t$ is a discrete random variable and $\boldsymbol{\nu} = (\delta, \gamma)'$. To sample from $\lambda_t$, 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

$$\begin{aligned} p(S_t \mid \delta, \gamma, \varphi, \tau^2, \beta_0, \beta_1, \beta_2, h_t, y_t, y_{t-1}) &\propto \delta^{S_t}(1-\delta)^{1-S_t}\gamma^{\frac{S_t}{2}} \\ &\times \mathrm{e}^{-\frac{1}{2}[(\frac{\tau^2+\varphi^2}{\tau^2})\mathrm{e}^{-h_t}(\gamma S_t + 1 - S_t)u_t^2]} \\ &\times \mathrm{e}^{[\gamma^{\frac{S_t}{2}} \frac{\varphi}{\tau^2} u_t (h_{t+1}-\alpha-\phi h_t)\mathrm{e}^{-\frac{h_t}{2}}]} \quad t < T. \end{aligned} \tag{A.12}$$

That is, $S_t \mid \delta, \gamma, \varphi, \tau^2, \beta_0, \beta_1, \beta_2, h_t, y_t, y_{t-1}$ has a Bernoulli distribution, where $u_t = y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 \mathrm{e}^{h_t}$. For $t = T$, we omit the last term in (A.12).

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 $\delta$ is given by

$$p(\delta \mid \gamma, \mathbf{S}_{1:T}) \propto \delta^{\delta_0-1}(1-\delta)^{\delta_1-1}\prod_{t=1}^{T}\delta^{S_t}(1-\delta)^{1-S_t}, \tag{A.13}$$

which is $\delta \mid \gamma, \mathbf{S}_{1:T} \sim \mathcal{B}e(\delta_0^*, \delta_1^*)$, where $\delta_0^* = \delta_0 + \sum_{t=1}^T S_t$ and $\delta_1^* = \delta_1 + T - \sum_{t=1}^T S_t$. The full conditional of $\gamma$ is given by

$$\begin{aligned} p(\gamma \mid \beta_0, \beta_1, \beta_2, \mathbf{S}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{0:T}) &\propto (1-\gamma)^{\gamma_1 - 1} \gamma^{\gamma_0 + \sum_{t=1}^T \frac{S_t}{2} - 1} \mathrm{e}^{-\frac{\gamma}{2} \sum_{t=1}^T (\frac{\tau^2 + \varphi^2}{\tau^2}) \mathrm{e}^{-h_t} S_t u_t^2} \\ &\times \mathrm{e}^{\sum_{t=1}^{T-1} [\gamma^{\frac{S_t}{2}} \frac{\varphi}{\tau^2} u_t (h_{t+1} - \alpha - \phi h_t) \mathrm{e}^{-\frac{h_t}{2}}]}, \end{aligned} \tag{A.14}$$

As (A.14) does not have closed form, we can sample from it by using the Metropolis-Hastings algorithm. The the right truncated gamma distribution $\mathcal{G}_{0<\gamma<1}(\gamma_0 + \sum_{t=1}^T \frac{S_t}{2}, \frac{1}{2} \sum_{t=1}^T (\frac{\tau^2+\varphi^2}{\tau^2}) \mathrm{e}^{-h_t} S_t u_t^2)$ can be used as a proposal density. A proposed $\gamma^*$ is accepted with probability $a_{MH\gamma} = \min\{1, \frac{Q(\gamma^*)}{Q(\gamma^{(i-1)})}\}$, where $\mathrm{Q}(\gamma) = (1-\gamma)^{\gamma_1 - 1} \mathrm{e}^{\sum_{t=1}^{T-1} [\gamma^{\frac{S_t}{2}} \frac{\varphi}{\tau^2} u_t (h_{t+1} - \alpha - \phi h_t) \mathrm{e}^{-\frac{h_t}{2}}]}$ and $\gamma^{(i-1)}$ denotes the previous iteration value.

## Appendix B: Some derivations of the Block sampler

First, we define

$$\begin{aligned} d_s &= \frac{\partial L}{\partial h_s} = -\frac{1}{2} + \frac{(y_s - \mu_s)^2}{2V_s} + \frac{(y_s - \mu_s)}{V_s} \frac{\partial \mu_s}{\partial h_s} + \frac{(y_{s-1} - \mu_{s-1})}{V_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_s} \\ &- \phi \frac{(h_{s+1} - \alpha - \phi h_s)}{\sigma_\eta^2} \mathbb{I}(t+k<T), \qquad s = t+1, \ldots, t+k, \end{aligned} \tag{B.1}$$

and

$$\mathbf{Q} = \begin{pmatrix} M_{t+1} & N_{t+2} & 0 & \ldots & 0 \\ N_{t+2} & M_{t+2} & N_{t+3} & \ldots & 0 \\ 0 & N_{t+3} & M_{t+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & N_{t+k} \\ 0 & \ldots & 0 & N_{t+k} & M_{t+k} \end{pmatrix} \tag{B.2}$$

where

$$\begin{aligned} M_s &= -E\left[\frac{\partial^2 L}{\partial h_s^2}\right] = \frac{1}{2} + \frac{1}{V_s}\left(\frac{\partial \mu_s}{\partial h_s}\right)^2 + \frac{1}{V_{s-1}}\left(\frac{\partial \mu_{s-1}}{\partial h_s}\right)^2 \\ &+ \frac{\phi^2}{\sigma_\eta^2} \mathbb{I}(t+k<T), \qquad s = t+1, \ldots, t+k, \end{aligned} \tag{B.3}$$

$$N_s = -E\left[\frac{\partial^2 L}{\partial h_s \partial h_{s-1}}\right] = \frac{1}{V_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_s}, \quad s = 2, \ldots, T, \tag{B.4}$$

with $N_{t+1} = 0$. Next, we define

$$\frac{\partial \mu_s}{\partial h_s} = \begin{cases} \beta_2 \mathrm{e}^{h_s} + \frac{\rho}{\sigma_\eta} \lambda_s^{-\frac{1}{2}} \mathrm{e}^{\frac{h_s}{2}} \left[ \frac{(h_{s+1} - \alpha - \phi h_s)}{2} - \phi \right], & s = 1, \ldots, T-1, \\ \beta_2 \mathrm{e}^{h_s}, & s = T, \end{cases} \tag{B.5}$$

$$\frac{\partial \mu_{s-1}}{\partial h_s} = \begin{cases} 0, & s = 1, \\ \frac{\rho}{\sigma_\eta} \lambda_{s-1}^{-\frac{1}{2}} \mathrm{e}^{\frac{h_{s-1}}{2}}, & s = 2, \ldots, T. \end{cases} \tag{B.6}$$

## References


Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H., 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.

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

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

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

Baillie, R. T., DeGennaro, R. P., 1990. Stock returns and volatility. Journal of Financial and Quantitative Analysis 25, 203–214.

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

Black, F., 1976. Studies of stock price volatility changes. Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section, 177–181.

Bollerslev, T., Zhou, H., 2005. Volatility puzzles: a simple framework for gauging return-volatility regressions. Journal of Econometrics 131, 123–150.

Brandt, M. W., Kang, Q., 2004. Generalized autoregressive conditional heteroskedasticy. Journal of Financial Economics 72, 217–257.

Campbell, J. Y., Hentschel, L., 1992. No news is good news, an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Chib, S., 1972. Markov chain Monte Carlo methods: computation and inference. North-Holland, Amsterdam.

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

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The American Statistician 49, 327–335.

Chib, S., Jeliazov, I., 2001. Marginal likelihood from the Metropolis-Hastings output. Journal of the American Statistical Association 96, 270–281.

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

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

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

French, K. R., Schert, W. G., Stambugh, R. F., 1987. Expected stock return 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: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), Bayesian Statistics. Vol. 4. Oxford University Press, Oxford, U.K., pp. 169–193.

Ghysels, E., Harvey, A. C., Renault, E., 1996. Stochastic volatility. In: Maddala, G., Rao, C. R. (Eds.), Handbook of Statistics. Vol. 14. North-Holland, Amsterdam, pp. 119–191.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess returns on stocks. Journal of Finance 48, 1791–1801.

Harvey, A. C., Shephard, N., 1996. The estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

Koopman, S., 1993. Disturbance smoothers for State Space models. Biometrika 80, 117–126.

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

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

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

Nelson, D. B., 1991. Leverage, heavy-tails and correlated jumps in stochastic volatility models. Econometrica 59, 347–370.

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

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

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

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

Salazar, E., Migon, H. S., Ferreira, M. A. R., 2009. Objective Bayesian analysis for exponential power regression models. Technical report, Federal University of Rio de Janeiro, Departament of Statistics.

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

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

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

Taylor, S., 1986. Modeling Financial Time Series. Wiley, Chichester.

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

Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Annal of Statistics 21, 1701–1762.

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

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