Bayesian Analysis of Stochastic volatility in mean model with leverage and asymmetrically heavy-tailed error using Generalized Hyperbolic skew Student's t-distribution

William L. Leão[†], C. A. Abanto-Valle[†] and Ming-Hui Chen[‡]

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

[‡]Department of Statistics, University of Connecticut, U-4120, Storrs, CT 06269, USA

Abstract

A stochastic volatility in mean model with correlated errors using the generalized hyperbolic skew Student-t (GH-ST) distribution provides a robust alternative to the parameter estimation for daily stock returns in the absence of normality. An efficient Markov chain Monte Carlo (MCMC) sampling algorithm is developed for parameter estimation. The Bayesian predictive information criterion (BPIC) is used to assess the fit of the proposed model. The proposed method is applied to an analysis of the daily stock return data from the São Paulo Stock, Mercantile & Futures Exchange index (IBOVESPA). The empirical results reveal that the stochastic volatility in mean model with correlated errors and GH-ST distribution leads to a significant improvement in model fit for the IBOVESPA data over the usual normal model.

Keywords: Feedback and leverage effect, GH skew Student-t distribution, Markov chain Monte Carlo, non-Gaussian and nonlinear state space models, stochastic volatility in mean.

1   Introduction

Stochastic volatility (SV) models were introduced in the financial literature to describe 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 (see, e.g., Liesenfeld and Jung, 2000; Abanto-Valle et al., 2010,

among others) or skewness of the returns. To deal with this problem, Abanto-Valle et al. (2014) propose a new stochastic volatility model based on a generalized skew-Student-t distribution for stock returns, which allows a parsimonious, flexible treatment of the skewness and heavy tails in the conditional distribution of the returns.

However, the volatility of daily stock returns has been estimated with SV models, but the results have relied on an extensive pre-modeling of these series to avoid the problem of simultaneous estimation of the mean and variance. To remedy this problem, Koopman and Uspensky (2002) introduced the SV in mean (SVM) by incorporating the unobserved volatility as an explanatory variable in the mean equation of the returns. They used the simulated maximum likelihood method for parameter estimation and provided an empirical justification that the volatility coefficient in the mean equation is related to the feedback effect, which implies that an increase in the current level of volatility causes agents to increase their forecasts of future volatility and therefore to raise their future required returns. It has also long been recognized in stock markets that there is a negative correlation between today's return and tomorrow's volatility. This phenomenon is called "leverage effect" or "asymmetry". The asymmetric stochastic volatility model is well known to describe these phenomena for stock returns. Markov chain Monte Carlo (MCMC) methods have been used for parameter estimation of SV models with leverage effect. For example, Omori et al. (2007) and Omori and Watanabe (2008) used an efficient mixture sampler and a block sampler for correlated errors, respectively.

In this article, we propose to enhance the robustness of the specification of the innovation returns in SVM models by introducing scale Generalized Hyperbolic skew Student-t distribution with correlated mean and variance errors. The resulting class of models takes into account the asymmetric effect, heavy-tailedness, the feedback, and leverage effects. We refer to this generalization as the SVML-GH-ST model. The flexibility of the SVML-GH-ST model can also capture 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 equations with correlated innovation errors, hence intensive computational methods are needed. Inference in this new SVML-GH-ST model is performed under the Bayesian paradigm via MCMC methods, which permits obtaining the posterior distribution of

parameters via simulation, starting from reasonable prior assumptions on the parameters. We simulate the log-volatilities and the shape parameters by using the block sampler for correlated errors (Omori and Watanabe, 2008) and Metropolis-Hastings algorithms, respectively.

The rest of the paper is structured as follows. Section 2 outlines the SVML-GH-ST model as well as the Bayesian estimation procedure using MCMC methods. Section 3 illustrates our proposed method using simulated data. In Section 4, the proposed class of models is applied to the IBOVESPA daily returns and model comparison is provided among the competing SVML models. Finally, We conclude the paper with some concluding remarks and suggestions for future developments in Section 5.

2 The asymmetric heavy-tailed stochastic volatility in mean model with leverage effect

The basic SV in mean model with leverage effect is defined by

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

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

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

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 ρ measures the correlation between ϵ_t and η_t. When $\rho < 0$, this indicates the so-called leverage effect, a drop in the return followed by an increase in the volatility. Empirical evidence can be found in Ghysels et al. (1996), Harvey and Shephard (1996), Bollerslev and Zhou (2005), Omori et al. (2007) and Nakajima and Omori (2012).

For a joint model of the asymmetric heavy-tailedness, the "feedback" and leverage effects, we replace the normal random variable ϵ_t in (1a) by a random variable from the GH skew Student's t-distribution, denoted by ω_t, which can be written in the form of the normal variance-mean

3

mixture as

$$\omega_t = \mu_\omega + \delta z_t + \sqrt{z_t}\epsilon_t \tag{2}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$ and $z_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$, where $\mathcal{N}(.,.)$ and $\mathcal{G}(.,.)$ indicate the normal and inverse gamma distributions respectively. We assume that $\mu_\omega = -\delta\mu_z$ and $\mu_z = E[z_t] = \nu/(\nu-2)$, such that $E[\omega_t] = 0$

Using the variance-mean mixture representation of the GH skew Student's t-distribution defined by equation (2), the stochastic volatility in mean model with asymmetric heavy-tailedness and leverage effect can be written hierarchically as:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{\frac{h_t}{2}}\{\delta(z_t - \mu_z) + \sqrt{z_t}\epsilon_t\}, \tag{3a}$$

$$h_{t+1} = \alpha + \phi h_t + \sigma_\eta \eta_t, \tag{3b}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}_2\left[\begin{pmatrix} 0 \\ 0, \end{pmatrix}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right], \tag{3c}$$

$$z_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2}). \tag{3d}$$

The model defined for equations (3a)-(3d), will be denoted by SVML-GH-ST. In this setup, equations (3a),(3b) and (3c) with $\delta = 0$ and $z_t = 1 \ \forall t = 1, \ldots, T$, we have the SVM model with leverage effect and normal distribution (SVML-N). Equations (3a)-(3d) with $\delta = 0$ define the SVM model with leverage effect and Student-t distribution (SVML-T)(see Abanto-Valle et al., 2011, for details).

Equations (3a)-(3d), can be written in an alternative way as

$$\begin{pmatrix} y_t \\ h_{t+1} \end{pmatrix}\Bigg|\boldsymbol{\theta}, z_t, h_t, y_{t-1} \sim \mathcal{N}\left(\begin{bmatrix} \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{\frac{h_t}{2}}\delta(z_t - \mu_z) \\ \alpha + \phi h_t \end{bmatrix}, \begin{bmatrix} z_t e^{h_t} & \varphi\lambda_t^{-1/2}e^{h_t/2} \\ \varphi\sqrt{z_t}e^{h_t/2} & \tau^2 + \varphi^2 \end{bmatrix}\right). \tag{4}$$

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

$$\mu_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{\frac{h_t}{2}}\delta(z_t - \mu_z) + \frac{\varphi}{\varphi^2 + \tau^2}\sqrt{z_t}e^{\frac{h_t}{2}}(h_{t+1} - \alpha - \phi h_t) \tag{5}$$

$$V_t = \frac{\tau^2}{\tau^2 + \varphi^2} z_t e^{h_t}, \tag{6}$$

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 $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \alpha, \phi, \tau^2, \varphi, \nu)'$ be the full parameter vector of the entire class of SVML-GH-ST 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, $\mathbf{z}_{1:T} = (z_1, \ldots, z_T)'$ be the mixing variables and $\mathbf{y}_{0:T} = (y_0, \ldots, y_T)'$ be the information available up to time T. Using the data augmentation principle, we have that the joint posterior density of parameters and latent unobservable variables can be written as

$$p(\boldsymbol{\theta}, \mathbf{h}_{1:T}, \mathbf{z}_{1:T} \mid \mathbf{y}_{0:T}) \quad \propto \quad \left[\prod_{t=1}^{T} p(y_t, h_{t+1} \mid z_t, h_t, y_{t-1}, \boldsymbol{\theta}) p(z_t \mid \boldsymbol{\nu}) \right] p(h_1 \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{7}$$

where $p(y_t, h_{t+1} \mid z_t, h_t, y_{t-1}, \boldsymbol{\theta})$ is given by equation (4) and $p(\boldsymbol{\theta})$ is the prior distribution. To make Bayesian analysis feasible for parameter estimation in the SVML-SMN class of models, we draw random samples from the posterior distribution of $(\boldsymbol{\theta}, \mathbf{h}_{1:T}, \mathbf{z}_{1:T})$ given $\mathbf{y}_{0:T}$ 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 $\mathbf{z}_{1:T}^{(i)}$ 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 $\mathbf{z}_{1:T}^{(i-1)}$.

3. Draw $\mathbf{z}_{1:T}^{(i)} \sim p(\mathbf{z}_{1:T} \mid \boldsymbol{\theta}^{(i)}, \mathbf{h}_{1:T}^{(i-1)}, \mathbf{y}_{0:T})$.

4. Generate $\mathbf{h}_{1:T}$ by blocks as:

   i) For $l = 1, \ldots, K$, the knot positions are generated as k_l, the floor of $[T \times \{(l+u_l)/(K+2)\}]$, where the $u_l's$ are independent realizations of the uniform random variable on the interval (0,1).

ii) For $l = 1, \ldots, K$, generate $h_{k_{l-1}+1:k_l-1}$ jointly conditional on $\mathbf{y}_{k_{l-1}:k_l-1}$, $\boldsymbol{\theta}^{(i)}$, $\mathbf{z}^{(i)}_{k_{l-1}+1:k_l-1}$, $h^{(i-1)}_{k_{l-1}}$ and $h^{(i-1)}_{k_l}$.

iii) For $l = 1, \ldots, K$, draw $h^{(i)}_{k_l}$ conditional on $\mathbf{y}_{1:T}$, $\boldsymbol{\theta}^{(i)}$, $h^{(i)}_{k_l-1}$ and $h^{(i)}_{k_l+1}$.

5. Set $i = i + 1$ and return to 2 until convergence is achieved.

The prior distribution of parameters in the SVML–GH-ST are set as: $\beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma^2_{\beta_0})$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(\bar{\beta}_1, \sigma^2_{\beta_1})$, $\beta_2 \sim \mathcal{N}(\bar{\beta}_2, \sigma^2_{\beta_2})$, $\alpha \mid \tau^2 \sim \mathcal{N}(\alpha_0, \tau^2/q_0)$, $\varphi \mid \tau^2 \sim \mathcal{N}(\varphi_0, \tau^2/p_0)$, $\phi \sim \mathcal{N}_{(-1,1)}(\phi_0, s^2_\phi)$, $\tau^2 \sim \mathcal{I}G(a_\tau/2, S_\tau/2)$ and $\nu \sim \mathcal{G}(a_\nu, b_\nu)$, where a_ν, b_ν, α_0, φ_0, ϕ_0, s^2_ϕ, a_τ, S_τ, p_0 and q_0 are known hyper parameters.

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 (3a), 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 showed that the sampler can produce efficient draws from the target conditional posterior distribution in comparison with a single-move sampler that 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) developed the associated multi-move sampler and showed that it produces efficient samples. In the next subsection, we extend their method to 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 $\mathbf{h}_{1:T} = (h_1, \ldots, h_T)'$ 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, $\mathbf{h}_{k_{l-1}+1:k_l-1} = (h_{k_{l-1}+1}, \ldots, h_{k_l-1})'$ for $l = 1, \ldots, K + 1$, with $k_0 = 1$ and $k_{K+1} = T$, where $k_l - 1 - k_{l-1} \geq 2$ is the size of the $l-$th block. We sample the block

of disturbances $\boldsymbol{\eta}_{k_{l-1}:k_l-2} = (\eta_{k_{l-1}}, \ldots, \eta_{k_l-2})'$ given the end conditions $h_{k_{l-1}}$ and h_{k_l} instead of $\mathbf{h}_{k_{l-1}+1:k_l-1} = (h_{k_{l-1}+1}, \ldots, h_{k_l-1})'$. 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 $\boldsymbol{\eta}_{t:t+k-1} = (\eta_t, \ldots, \eta_{t+k-1})'$ are sampled at once from their full conditional distribution $f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \mathbf{y}_{t:t+k}, \mathbf{z}_{t+1:t+k})[1]$, which without the constant terms is expressed in log scale as

$$
\begin{aligned}
\log f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \mathbf{y}_{t:t+k}, \mathbf{z}_{t+1:t+k}) \quad \dot{=} \quad &- \sum_{s=t}^{t+k-1} \frac{\eta_s^2}{2} + \sum_{s=t}^{t+k} l_s \\
&- \frac{1}{2\sigma_\eta^2} (h_{t+k+1} - \alpha - \phi h_{t+k})^2 \mathbb{I}(t + k < T), \quad (8)
\end{aligned}
$$

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 μ_s and variance V_s, given by equations (5) and (6) respectively. We define

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

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

As $-\frac{1}{2} \sum_{s=t}^{t+k-1} \eta_s^2 + L$ in (8) does not have closed form, we use the Metropolis-Hastings acceptance-rejection algorithm (Chib, 1995) to sample from. To obtain the proposal density, we are going to form an approximated linear state space model that mimics (8), 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

---

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

have

$$\log f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \mathbf{z}_{t+1:t+k})$$

$$\approx \text{const} - \tfrac{1}{2}\sum_{r=t+1}^{t+k}\eta_r^2 + \hat{L} + \frac{\partial L}{\partial \boldsymbol{\eta}'_{t:t+k-1}}\bigg|_{\boldsymbol{\eta}_{t:t+k-1}=\hat{\boldsymbol{\eta}}_{t:t+k-1}}(\boldsymbol{\eta}_{t:t+k-1}-\hat{\boldsymbol{\eta}}_{t:t+k-1})$$

$$+\tfrac{1}{2}(\boldsymbol{\eta}_{t:t+k-1}-\hat{\boldsymbol{\eta}}_{t:t+k-1})'E\big(\frac{\partial^2 L}{\partial\boldsymbol{\eta}_{t:t+k-1}\partial\boldsymbol{\eta}'_{t:t+k-1}}\big)\bigg|_{\boldsymbol{\eta}_{t:t+k-1}=\hat{\boldsymbol{\eta}}_{t:t+k-1}}(\boldsymbol{\eta}_{t:t+k-1}-\hat{\boldsymbol{\eta}}_{t:t+k-1})$$

$$= \text{const} - \tfrac{1}{2}\sum_{r=t+1}^{t+k}\eta_r^2 + \hat{L} + \hat{\mathbf{d}}'_{t+1:t+k}(\mathbf{h}_{t+1:t+k}-\hat{\mathbf{h}}_{t+1:t+k})$$

$$-\tfrac{1}{2}(\mathbf{h}_{t+1:t+k}-\hat{\mathbf{h}}_{t+1:t+k})'\hat{\mathbf{Q}}(\mathbf{h}_{t+1:t+k}-\hat{\mathbf{h}}_{t+1:t+k})$$

$$= \text{const} + \log f^*(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \boldsymbol{\theta}, \mathbf{y}_{t+1:t+k}, \mathbf{z}_{t+1:t+k}), \qquad (9)$$

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 $f^*(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \boldsymbol{\theta}, \mathbf{y}_{t+1:t+k}, \mathbf{z}_{t+1:t+k})$ is the posterior density of $\boldsymbol{\eta}_{t:t+k-1}$ for a linear Gaussian state space model given by equations (10) and (11) 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 (3b).

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.

$$G_s = \hat{M}_s - \hat{N}_s^2 G_{s-1}^{-1}, \quad G_{t+1} = \hat{M}_{t+1},$$

$$J_s = K_{s-1}^{-1}\hat{N}_s, \quad J_{t+1} = 0, \quad J_{t+k+1} = 0,$$

$$b_s = \hat{d}_s - J_s K_{t-1}^{-1} b_{s-1} \quad b_{t+1} = \hat{d}_{t+1},$$

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

$$\hat{y}_s \;\; = \;\; c_s + Z_s h_s + H_s\xi_s, \quad s = t+1,\ldots,t+k, \tag{10}$$

$$h_{s+1} \;\; = \;\; \alpha + \phi h_s + L_s\xi_s, \quad s = t,t+1,\ldots,t+k, \tag{11}$$

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 (10) and (11) 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 simulation smoother (de Jong and Shephard, 1995) to the model defined by equations (10) and (11) 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 (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 (10) and (11) 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 a 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 $y_{k_l-1} \mid z_{k_l-1}, y_{k_l-2}, h_{k_l-1}, h_{k_l} \sim \mathcal{N}(\mu_{k_l-1}, V_{k_l-1})$ and $y_{k_l} \mid z_{k_l}, y_{k_l-1}, h_{k_l+1}, h_{k_l} \sim \mathcal{N}(\mu_{k_l}, V_{k_l})$, with μ_s and V_s are defined by equations (5) and (6) respectively, for $s = k_l - 1$ and k_l.
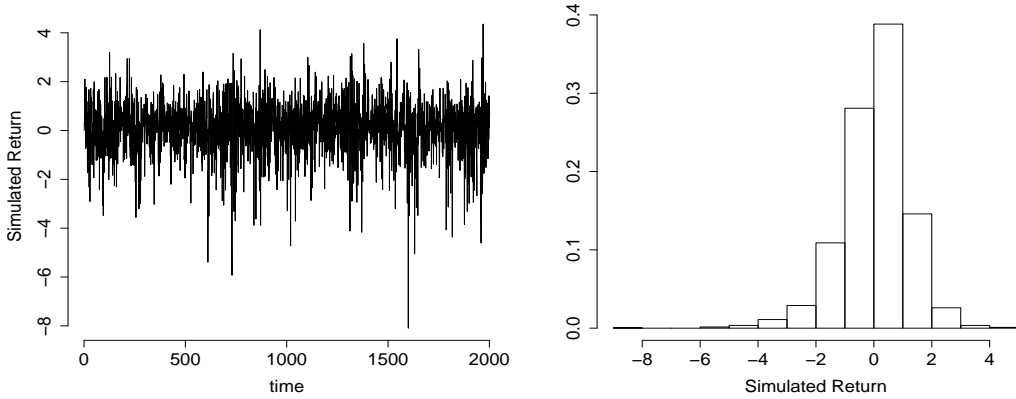

Figure 1: Simulated dataset from the SVML-GH-ST: Time series of returns (left) and the histogram (right).
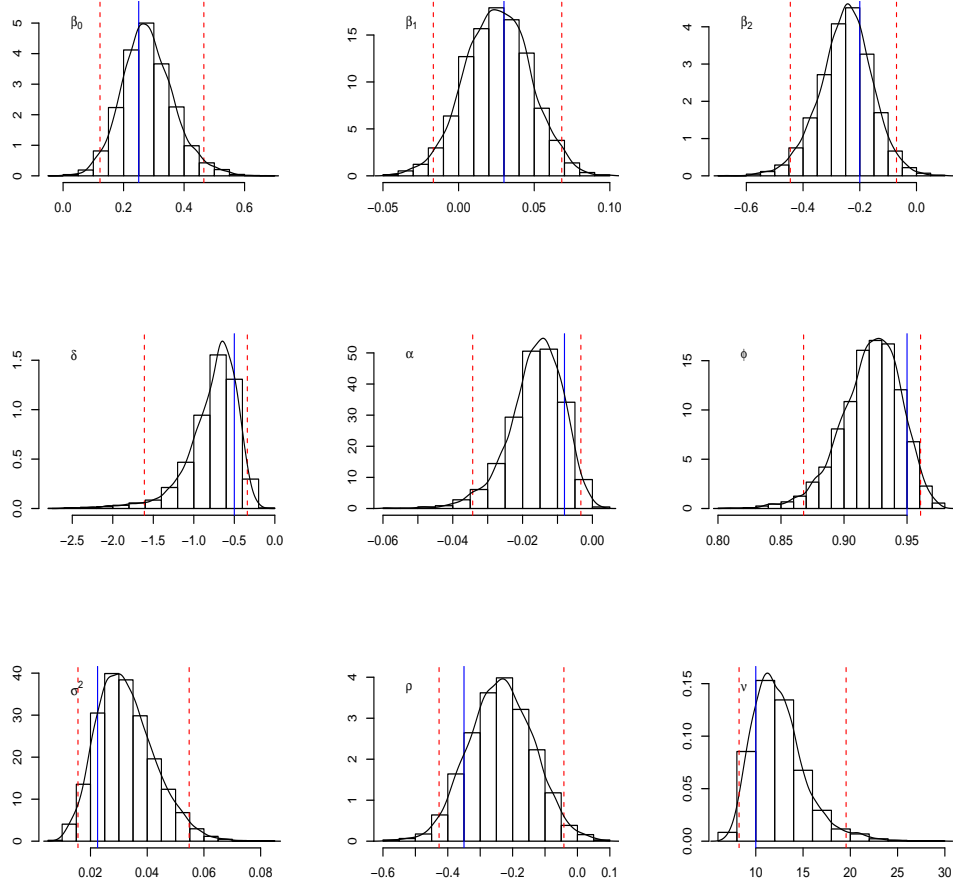
10

Figure 2: Simulated dataset. Histograms and estimated densities from the MCMC output for the SVML-GH-ST. The solid line indicates the true value and the dotted line the 95% credible interval

11

3 Numerical illustration with artificial dataset

In order to asses the performance of the MCMC algorithms described in the previous section, we present results based on a simulated dataset. All the calculations were performed running stand alone code developed by the authors using the Scythe statistical library (Pemstein et al., 2011), which is available for free download at http://scythe.wustl.edu. We simulated a dataset of 2000 observations of the SVM-L-GH-SST distribution using $\beta_0 = 0.25$, $\beta_1 = 0.03$, $\beta_3 = -0.2$, $\alpha = -0.008$, $\phi = 0.95$, $\sigma^2 = 0.0225$, $\rho = -0.35$ e $\nu = 10$, which correspond to typical values found in daily series of returns. Figure 1 shows the raw data and the histograms of the simulated dataset.

We set the prior distributions as: $\beta_0 \sim \mathcal{N}(0, 100)$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(0.1, 100)$, $\beta_2 \sim \mathcal{N}(-0.1, 100)$, $\alpha | \tau^2 \sim \mathcal{N}(0, \tau^2/0.002)$, $\phi | \tau^2 \sim \mathcal{N}_{(-1,1)}(0.95, 100)$, $\tau^2 \sim \mathcal{IG}(2.5, 0.025)$, $\varphi | \tau^2 \sim \mathcal{N}(-0.3, \tau^2/0.005)$, $\delta \sim \mathcal{N}(0, 1)$ and $\nu \sim \mathcal{G}(12, 0.5)$, where $\mathcal{N}(., .)$, $\mathcal{N}_{(a,b)}(., .)$, $\mathcal{G}(., .)$ and $\mathcal{IG}(., .)$ indicate the normal, the truncated normal, the gamma and the inverse gamma distributions respectively. The priors's means for β_1 and ϕ, 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 β_1 and ϕ are non-informative.

The number of blocks, K, in the block sampler was set equal to 30, so that each block contained 66 $h'_t s$ on average. We conducted the MCMC simulation for 50000 iterations. The first 10000 draws were discarded as a burn-in period, and then the next 40000 were recorded. In order to reduce the autocorrelation between successive values of the simulated chain, only every $10th$ values of the chain were stored. With the resulting 4000 we calculated the posterior means, the 95% intervals and the convergence diagnostic (CD) statistics proposed by Geweke (1992) for all the parameters.

The proposed algorithm is evaluated in terms of how well it estimates the true parameter values. From Table 1 and Figure 2, it can be seen that the estimated results for the parameters appear quite reasonable, because all the 95% credibility intervals include true values. According
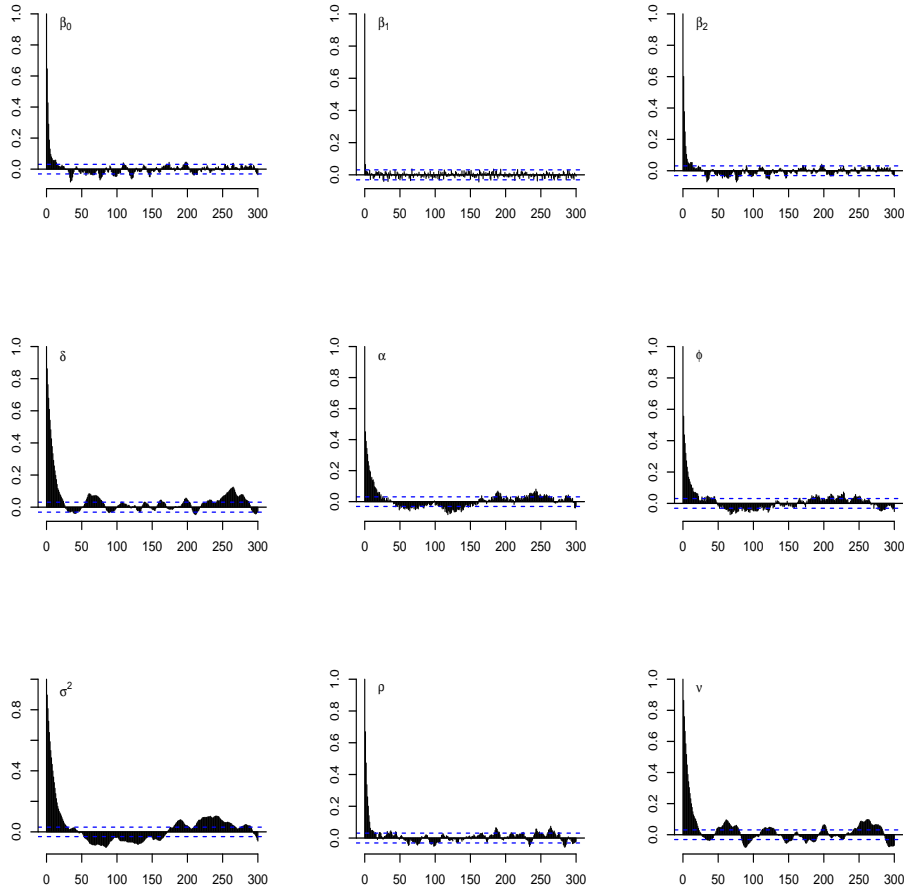
Figure 3: SVML-GH-ST, simulated dataset. Autocorrelation function (acf) for the parameters obtained from the MCMC output.
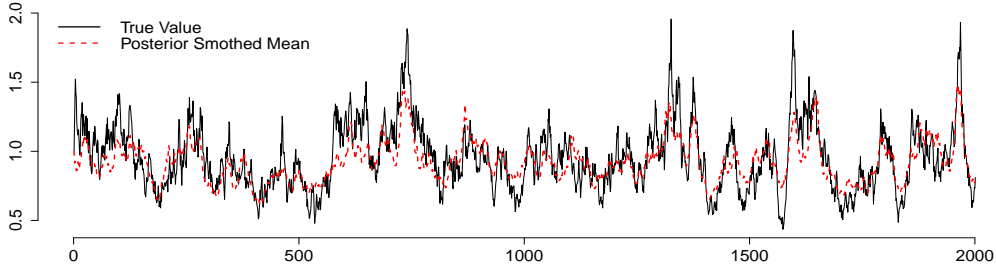
Figure 4: SVML-GH-ST, simulated dataset. True values(solid line) and posterior smoothed mean (dotted line) of $\mathrm{e}^{\frac{h_t}{2}}$.

to the CD values, the null hypothesis that the sequence of 4000 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 ρ_s is the sample autocorrelation at lag s. It measures how well the MCMC chain mixes (see, e.g, Kim et al., 1998). It is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. When the inefficiency factor is equal to m, we need to draw MCMC samples m times as many as the number of uncorrelated samples. From Table 1, we found that our algorithm produces a good mixing of the MCMC chain. This fact is confirmed by Figure 3, where the the autocorrelation function (acf) of the parameters shows a faster decay.

14

Table 1: Simulated dataset: summary results for tge SVML-GH-ST model.

Parmater	True value	Posterior mean	95% CI	IF	CD
β_0	0.2500	0.2810	$(0.1220, 0.4650)$	6.26	-0.12
β_1	0.0300	0.0260	$(-0.0170, 0.0680)$	1.28	0.95
β_2	-0.2000	-0.2500	$(-0.4450, -0.0700)$	5.38	0.01
α	-0.0080	-0.0160	$(-0.0340, -0.0030)$	10.36	-0.95
ϕ	0.9500	0.9210	$(0.8680, 0.9610)$	10.67	-1.01
σ^2	0.0225	0.0330	$(0.0160, 0.0550)$	21.83	0.96
δ	-0.5000	-0.7680	$(-1.6100, -0.3400)$	21.31	-0.32
ρ	0.3500	-0.2350	$(-0.4270, -0.0420)$	7.02	0.52
ν	10.0000	12.4430	$(8.2330, 19.5520)$	20.02	0.15

In Figure 4, the smoothed mean calculated from the MCMC output (dotted line) and true values (solid line) of $e^{\frac{h_t}{2}}$ are shown. They show that the estimated values follow the behavior of the true volatilities.

4 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 dataset 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 compounded IBOVESPA returns are plotted in Figure 5 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 5, the returns are slightly skew (0.83) with heavy tails. From Table

Table 2: Summary statistics for the IBOVESPA returns.

Média	D.p.	Minimo	Máximo	Assimetria	Curtose
0.06	2.34	−17.21	28.83	0.83	19.18

2, note also that the returns have a large range (minimum, -17.21 and maximum, 28.83). Some extreme observations, explained by turbulence 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.
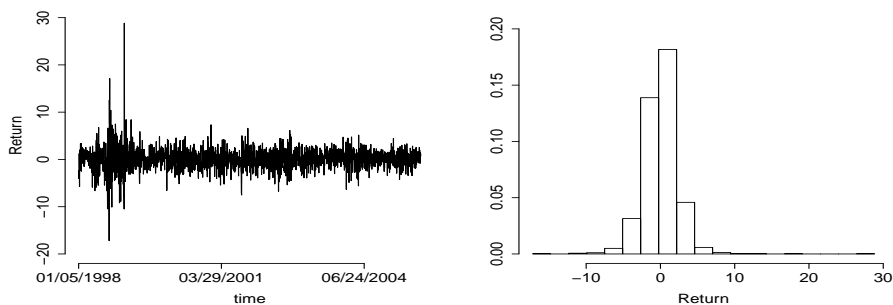

Figure 5: Compounded IBOVESPA returns from January 5, 1998 to September 3, 2005. The left panel shows the plot of the raw series and the right panel the histogram of returns.

We fitted the SVML-N, SVML-T and SVML-GH-ST. 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 parameter was chosen as:

$\nu \sim \mathcal{G}(12, 0.8)$ for the SVML-T and the SVML-GH-ST models, respectively. For the SVML-GH-ST, we set $\delta \sim \mathcal{N}(0, 100)$. 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 $\mathbf{z}_{1:T}$ were generated from the prior $p(z_t \mid \nu)$.

For the block sampler algorithm, we set the number of blocks K to be 30 in such a way that each block contained 66 h'_ts on average. For the SVML-N, SVML-T and the SVML-GH-ST models, we conducted the MCMC simulation for 50000 iterations. In all the cases, the first 10000 draws were discarded as a burn-in period. As before, in order to reduce the autocorrelation between successive values of the simulated chain, only every $10th$ values of the chain were stored. With the resulting 4000 values, we calculated the posterior means, the 95% credible intervals and the convergence diagnostic (CD) statistics (Geweke, 1992). Table 3 summarizes the results. According to the CD values, the null hypothesis that the sequence of 4000 draws is stationary was accepted at the 5% level for all the parameters in all the models considered here. From Table 3 and Figure 7, we found that our algorithm produces a good mixing of the MCMC chain.

Table 3 shows that the posterior mean and 95% interval of ϕ. For all the models, the posterior means of ϕ are above 0.93, showing higher persistence, as expected. We found that the persistence of the SVML-T and the SVML-GH-ST are slightly different from that the SVML-N. The posterior mean of σ_η^2 is smaller in the SVML-T and SVML-GH-ST than the SVML-N model, indicating that the volatilities of the SVML-T and SVML-GH-ST are less variable than the equivalent SVML-N model.

The posterior means together with the posterior 95% intervals of the three parameters, which govern the mean process for each of the three models, are reported in Table 3. In all cases the posterior mean of β_0 is always positive and statistically significant for each one of the models fitted. 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 might 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. Although the posterior credibility interval of β_2 barely contains zero for all the models, its posterior distribution is primarily located in the negative range, as shown in Table 4. This

17

Table 3: 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.

Parâmetro	SVML-N	SVML-T	SVML-GH-ST
β_0	0.1409	0.1801	0.1510
	(0.0031,0.2820)	(0.0269,0.3388)	(0.0180,0.2920)
	0.61	-1.67	-1 38
	1.33	2.34	2.43
β_1	0.0299	0.0242	0.023
	(-0.0562,0.0239)	(-0.0219,0.0682)	(-0.0220,0.0690)
	-0.18	-0.18	-1.13
	1.49	1.60	1.91
β_2	-0.0179	-0.0343	-0.026
	(-0.0559,0.0193)	(-0.0894,0.0160)	(-0.0700,0.0140)
	0.69	0.66	1.21
	1.27	3.02	2.26
α	0.0713	0.0407	0.049
	(0.0271, 0.1196)	(0.0147,0.0758)	(0.0210,0.0850)
	1.47	-1.49	1.65
	23.55	28.31	20.05
ϕ	0.9368	0.9579	0.9510
	(0.8940,0.9765)	(0.9184,0.9855)	(0.9150,0.9800)
	-1.33	1.47	-1.68
	25.29	36.05	25.00
σ_η^2	0.0708	0.0426	0.0550
	(0.0250,0.1214)	(0.0146,0.0818)	(0.0240,0.0940)
	1.31	-1.46	1.74
	28.74	44.37	28.43
ρ	-0.3112	-0.3445	-0.3500
	(-0.4677,-0.1774)	(-0.5319,-0.1779)	(-0.4490,-0.2060)
	1.36	-1.75	-0.83
	11.13	21.96	9.62
ν	–	10.9988	18.7150
	–	(6.9690,16.9087)	(10.3000,29.5600)
	–	1.32	0.16
	–	26.25	23.75
δ	–	–	-0.3810
	–	–	(-1.2200,0.1600)
	–	–	-1.79
	–	18 –	10.22

Table 4: IBOVESPA dataset: $P(\beta_2 < 0)$ estimated from the MCMC output.

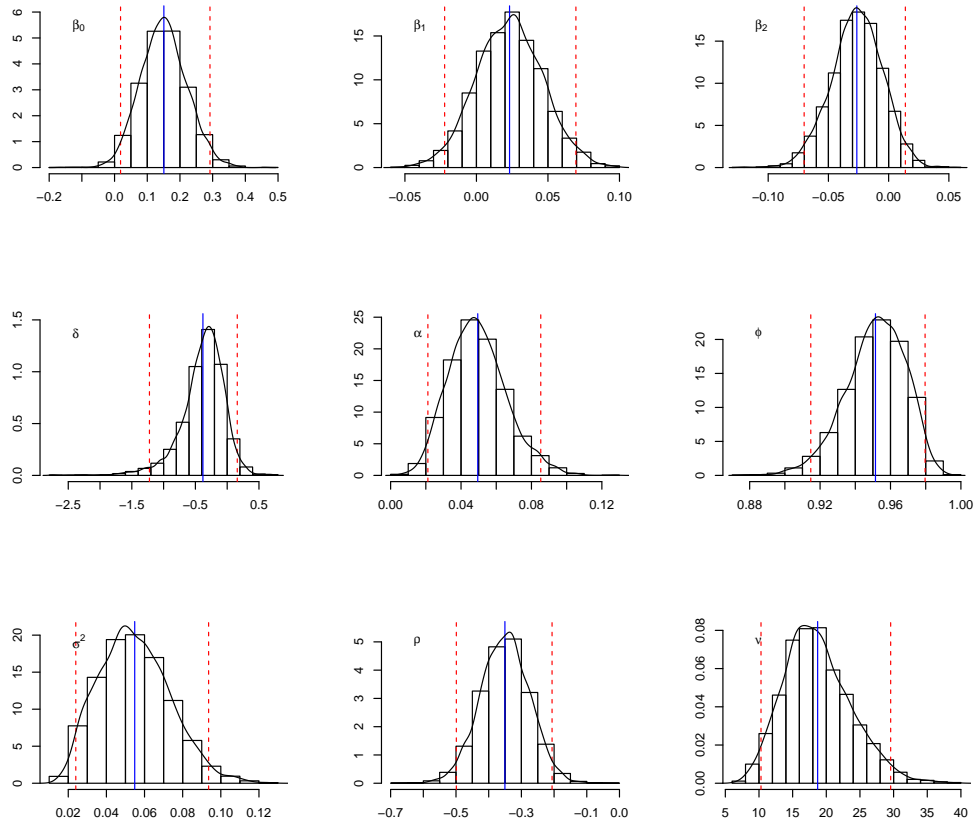|	SVML-N	SVML-T.	SVML-GH-ST
$P(\beta_2 < 0)$	0.8295	0.9055	0.8948

Figure 6: IBOVESPA dataset. Histograms and estimated densities from the MCMC output for the SVML-GH-ST. The solid line indicates the posterior mean and the dotted line the 95% credible interval
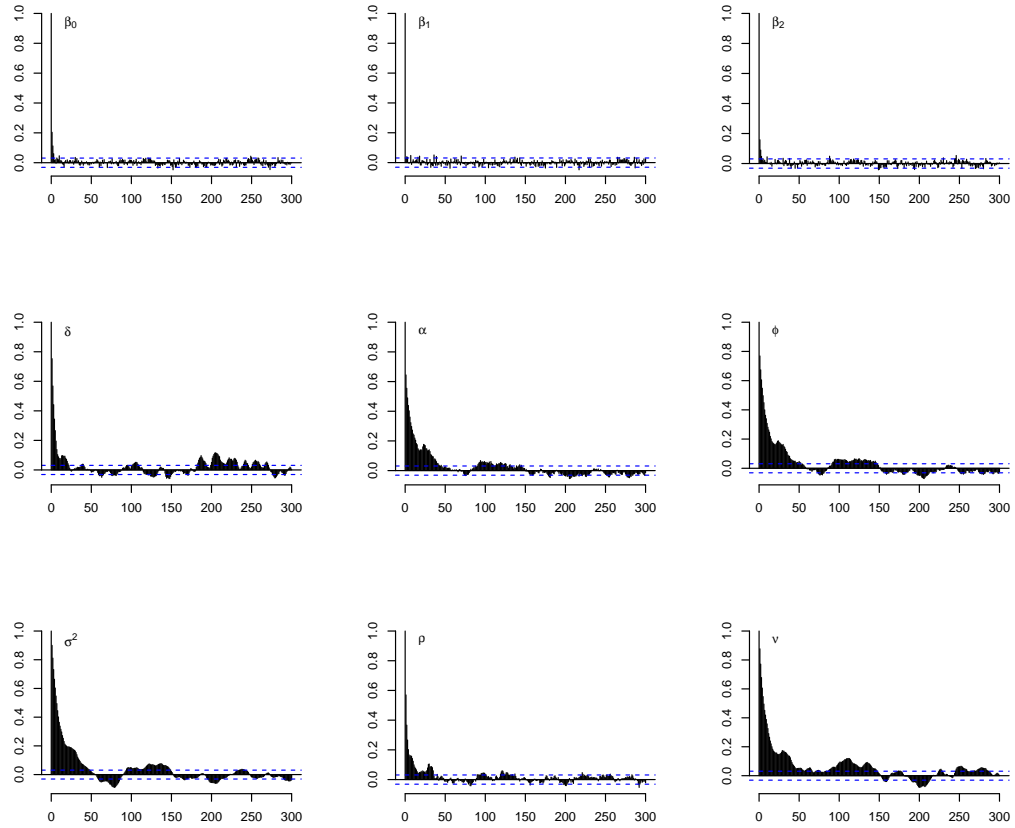
Figure 7: IBOVESPA dataset. Autocorrelation function (acf) for the parameters obtained from the MCMC output.

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 ρ, the correlation coefficient between shocks to return at time t and shocks to volatility at time $t + 1$, are always negative and the 95% posterior credibility intervals do not contain zero. This result indicates the parameter is statistically significant. Hence, we may conclude there is a strong and significant "leverage effect" for the IBOVESPA returns dataset.

We found that the posterior mean of δ is -0.381 which indicates the returns are slightly asymmetric. We found that the 95% credible interval contains zero, but from Figure 6, its posterior distribution is concentrated in the negative range.

The magnitude of the tail fatness is measured by the shape parameter ν in the SVML-T and SVML-GH-ST models. The posterior means of ν are almost 11 and 19 in the SVML-T and SVML-GH-ST respectively. This difference can explained by the δ, the extra asymmetry parameter which is considered in the specification of the SVML-GH-ST model. These results seem to indicate that the measurement errors of the stock returns are better explained by heavy-tailed distributions.

Now, we compare the volatilities estimates. In Figure 8, we plot the smoothed mean of $\mathrm{e}^{\frac{h_t}{2}}$. The posterior smoothed mean of $\mathrm{e}^{\frac{h_t}{2}}$ of the SVML-T, SVML-GH-ST 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 $\mathrm{e}^{\frac{h_t}{2}}$ by $z_t^{\frac{1}{2}}\mathrm{e}^{\frac{h_t}{2}}$. This can have a substantial impact, for instance, on the valuation of derivative instruments and several strategic or tactical asset allocation topics.

To assess the goodness of the estimated models, we calculate the Bayesian predictive information criteria, BPIC (Ando, 2006; 2007). The BPIC criterion is defined as

$$BPIC \;\; = \;\; -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\}] + 2T\hat{b}, \tag{12}$$
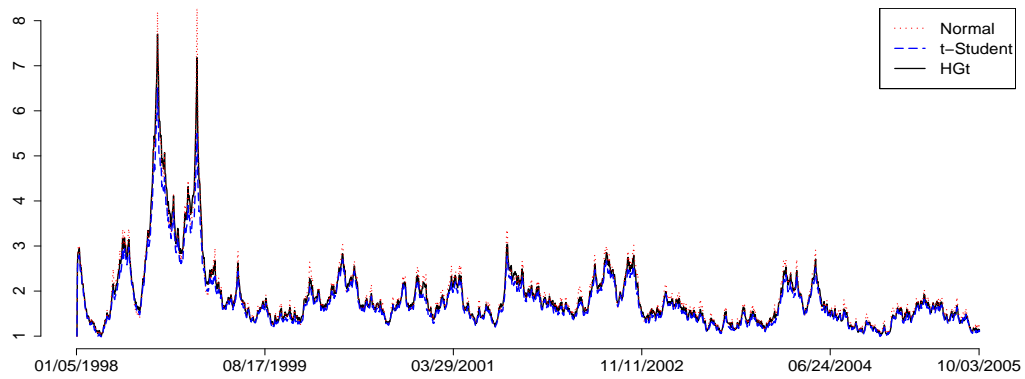
Figure 8: IBOVESPA dataset. Posterior smoothed mean (dotted line) of $e^{\frac{h_t}{2}}$, SVML-GH-ST (solid line), SVML-T (dotted line) , SVML-N (tiny line).

where $\hat{b}$ is given by

$$\hat{b} \approx \frac{1}{T}\left\{E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})\}] - \log[p(\mathbf{y}_{1:T} \mid \hat{\boldsymbol{\theta}})p(\hat{\boldsymbol{\theta}})] + \mathrm{tr}\{J_T^{-1}(\hat{\boldsymbol{\theta}})I_T(\hat{\boldsymbol{\theta}})\} + 0.5q\right\} (13)$$

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

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

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

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

In the SVML class of models, 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. Table 5 shows the BPIC. According with the BPIC criterion the SVML-GH-ST model is relatively better among all the considered models, suggesting that the IBOVESPA return data demonstrate sufficient departure from underlying normality assumptions.

Table 5: IBOVESPA dataset. Bayesian predictive information criteria (BPIC)

Modelo	BPIC	Ranking
SVML-N	8084.6	3
SVML-T	8081.8	2
SVML-GH-ST	8075.4	1

5   Conclusions

This article presented a Bayesian implementation of a robust alternative for estimation in the stochastic volatility in mean model with correlated errors, as an extension of the model proposed

by Koopman and Uspensky (2002) and Abanto-Valle et al. (2011), 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 the variance-mean mixture of the normal distribution. 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-GH-ST model. We illustrated our methods through an empirical application of the IBOVESPA return series, which showed that the SVML-GH-ST model provides better fit than the SVML-N and SVML-T model in terms of parameter estimates, interpretation and robustness aspects. 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 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 ϕ combined with larger negative values for the in-mean parameter. On the other hand, since 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 dataset.

Our SVML-GH-ST models showed considerable flexibility to accommodate outliers, but their robustness aspects could be seriously affected by the prior of the ν and δ parameters. In this set-up, for example, it would be possible to study different objective priors for form parameter in the GH-ST distributions in the same spirit of the works of Fonseca et al. (2008) and Salazar et al. (2009). Nevertheless, a deeper investigation of this modification is beyond the scope of the present paper, but provides stimulating topics for further research.

Appendix A: The Full conditionals

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

Full conditional distribution of β_0, β_1 and β_2

Let m_t and V_t be defined by

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

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

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

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

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

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

Full conditional distribution of α, ϕ, φ, δ and τ^2

We assume the following prior distributions: $\alpha \mid \tau^2 \sim \mathcal{N}(\alpha_0, \tau^2/q_0)$, $\varphi \mid \tau^2 \sim \mathcal{N}(\varphi_0, \tau^2/p_0)$, $\phi \sim \mathcal{N}_{(-1,1)}(\phi_0, s^2_\phi)$, $\delta \sim \mathcal{N}(\delta_0, s^2_\delta)$, $\tau^2 \sim \mathcal{GI}(a_\tau/2, S_\tau/2)$, where α_0, φ_0, ϕ_0, s^2_ϕ, δ_0, s^2_δ, a_τ, S_τ, p_0 and q_0 are known hyper parameters.

After some simple but tedious algebra, we have

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

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

$$
\delta \mid . \sim \mathcal{N}\left(\frac{B_\delta}{A_\delta}, \frac{1}{A_\delta}\right), \tag{A.6}
$$

where $A_\alpha = q_0 + \frac{1+\phi}{1-\phi} + T - 1$, $B_\alpha = \alpha_0 q_0 + (1+\phi)h_1 + \sum_{t=1}^{T-1}k_t$, $k_t = h_{t+1} - \phi h_t - \varphi g_t z_t^{-\frac{1}{2}}\mathrm{e}^{-\frac{h_t}{2}}$, $A_\varphi = p_0 + \sum_{t=1}^{T-1}g_t^2 z_t^{-1}\mathrm{e}^{-h_t}$, $B_\varphi = \varphi_0 p_0 + \sum_{t=1}^{T-1}c_t g_t z_t^{-\frac{1}{2}}\mathrm{e}^{-\frac{h_t}{2}}$, $A_\delta = -\frac{\varphi}{\tau^2}\sum_{t=1}^{T-1}\frac{1}{\sqrt{z_t}}(z_t -$

$\mu_z)\,(h_{t+1} - \alpha - \phi h_t)$

$\left(\frac{\varphi^2+\tau^2}{\tau^2}\right)\sum_{t=1}^{T-1}\frac{1}{z_t e^{h_t/2}}(z_t-\mu_z)\left(y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 e^{h_t}\right)$, $B_\delta = \left(\frac{\varphi^2+\tau^2}{\tau^2}\right)\sum_{t=1}^{T-1}\left(\frac{1}{z_t}(z_t-\mu_z)^2\right)+$ $\frac{1}{z_n}(z_n-\mu_z)^2 + \frac{1}{s_\delta^2}$, $c_t = h_{t+1}-\alpha-\phi h_t$ and $g_t = y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} - e^{\frac{h_t}{2}}\delta(z_t-\mu_z)$. In a similar way, the conditional posterior of ϕ is given by

$$p(\phi\mid .) \;\propto\; Q(\phi)\exp\left\{-\frac{A_\phi}{2}\left(\phi-\frac{B_\phi}{A_\phi}\right)^2\right\}, \tag{A.7}$$

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.7) 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 τ^2 is $\mathcal{IG}(\frac{T_1}{2},\frac{M_1}{2})$, where $T_1 = a_\tau + T + 2$ and $M_1 = (1-\phi^2)(h_1-\frac{\alpha}{1-\phi})^2 + \sum_{t=1}^{T-1}(c_t-\varphi z_t^{-\frac{1}{2}}e^{-\frac{h_t}{2}}g_t)^2 + p_0(\varphi-\varphi_0)^2 + q_0(\alpha-\alpha_0)^2 + S_\tau$. Once τ^2 and φ are sampled, respectively, from their conditional posteriors, we can calculate ρ and σ_η^2 through $\sigma_\eta^2 = \tau^2 + \varphi^2$ and $\rho = \varphi/\sigma_\eta$.

Full conditional of z_t and ν

The full conditional of z_t is

$$p(z_t\mid .) \propto Q(z_t)\left(\frac{\gamma}{\vartheta}\right)^\lambda \frac{z^{\lambda-1}}{2K_\lambda(\gamma,\vartheta)}\exp\left\{-\frac{1}{2}\left(\vartheta^2 z^{-1}+\gamma^2 z\right)\right\}$$

where the values of λ, ϑ and γ are the parameters of a distribution $\mathrm{GIG}(\lambda,\vartheta,\gamma)$ whose values are given by

$$\lambda = -\frac{\nu+1}{2}$$
$$\gamma^2 = \delta^2\frac{\varphi^2+\tau^2}{\tau^2}$$
$$\vartheta^2 = \frac{\varphi^2+\tau^2}{\tau^2}e^{-h_t}\left(y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 e^{h_t}+\mu_z e^{h_t/2}\delta\right)^2+\frac{\nu}{2}$$

We sample z_t by the Metropolis-Hastings algorithm. We use $\text{GIG}(\lambda, \vartheta, \gamma)$ as the proposal distrisbution such that z_t^* and $z_t^{(i-1)}$ are the proposal value and previous iteration value respectively. Thus, the acceptance probability is given by $\alpha_{MH} = \min\{1, \frac{Q(z_t^*)}{Q(z_t^{(i-1)})}\}$ where

$$Q(z_t) = \exp\left\{ \frac{\varphi}{\tau^2}\left[z_t^{-1/2} e^{-h_t/2}(h_{t+1} - \alpha - \phi h_t)\left(y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} + \mu_z e^{h_t/2}\delta\right) - \right.\right.$$

$$\left.\left. - z_t^{1/2}\delta(h_{t+1} - \alpha - \phi h_t)\right]\mathbb{I}_{\{t<n\}} \right\}$$

We assume the prior distribution of ν as $\mathcal{G}(a_\nu, b_\nu)\mathbb{I}_{4<\nu\geq 40}$. Then, the full conditional of ν is

$$p(\nu \mid \mathbf{z}_{1:T}) \propto \frac{\frac{\nu}{2}^{\frac{T\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)^T} \exp\left\{ -\sum_{t=1}^{T}\frac{1}{2V_t}\left[y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} - e^{\frac{h_t}{2}}\delta\left(z_t - \frac{\nu}{\nu-2}\right) - m_t\right]^2 \right.$$

$$\left. -\frac{\nu}{2}\sum_{t=1}^{T}\left[\frac{1}{z_t} + \log z_t\right] \right\} \nu^{a_\nu-1}\exp\{-b_\nu\nu\}\mathbb{I}_{4<\nu\geq 40}. \tag{A.8}$$

We sample ν by the Metropolis-Hastings acceptance-rejection algorithm (Tierney, 1994; Chib, 1995). Let ν^* denote the mode (or approximate mode) of $p(\nu \mid \mathbf{z}_{1:T})$, and let $\ell(\nu) = \log p(\nu \mid \mathbf{z}_{1:T})$. We use the proposal density $\mathcal{N}_{(4,40)}(\mu_\nu, \sigma_\nu^2)$, 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 $\nu = \nu^*$.

Appendix B: Some derivations of the block sampler

First, we define

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

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

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

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

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

$$
\frac{\partial \mu_s}{\partial h_s} =
\begin{cases}
\beta_2 e^{h_s} + \frac{1}{2}e^{\frac{h_s}{2}}\delta(z_s - \mu_z) + \frac{\varphi}{\varphi^2+\tau^2}\sqrt{z_s}\,e^{\frac{h_s}{2}}\left[\frac{(h_{s+1}-\alpha-\phi h_s)}{2} - \phi\right], & s = 1,\ldots,T-1, \\
\beta_2 e^{h_s} + \frac{1}{2}e^{\frac{h_s}{2}}\delta(z_s - \mu_z), & s = T,
\end{cases}
\tag{B.5}
$$

$$
\frac{\partial \mu_{s-1}}{\partial h_s} =
\begin{cases}
0, & s = 1, \\
\frac{\varphi}{\varphi^2+\tau^2}\sqrt{z_{s-1}}\,e^{\frac{h_{s-1}}{2}}, & s = 2,\ldots,T.
\end{cases}
\tag{B.6}
$$

Acknowledgements

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.

Abanto-Valle, C. A., Migon, H. S., and Lachos, V. H. (2011), "Stochastic volatility in meanmodels with scale mixtures of normal distributions and correlated errors: A Bayesian approach," *Journal of Statistical Planning and Inference*, 54, 1875–1887.

Abanto-Valle, C. A., Wang, C., Wang, X., Wang, F.-X., and Chen, M.-H. (2014), "Bayesian Inference for Stochastic Volatility Models Using the Generalized Skew-t Distribution with

Applications to the Shenzhen Stock Exchange Return," *Statistical and its Interface (forth-coming)*, .

Ando, T. (2006), "Bayesian Inference for nonlinear and non-gaussian stochastic volatility model wit leverage 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.

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

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

de Jong, P., and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339–350.

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.

French, K. R., Schert, W. G., and 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 *Bayesian Statistics. Vol. 4.*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 169–193.

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

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

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

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

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

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

Nakajima, J., and Omori, Y. (2012), "Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution," *Computational Statistics & Data Analysis*, 56, 3690–3704.

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

Omori, Y., and 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., 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.

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

Shephard, N., and 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 *Time Series Analysis: Theory and Practice, Vol 1*, ed. O. Anderson, Amsterdam: North-Holland, pp. 203–226.

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

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

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.