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

\[ y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{h_t} \epsilon_t, \]  
\[ h_{t+1} = \alpha + \phi h_t + \sigma \eta_t, \]  
\[ \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \]

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 \( h_1 \sim \mathcal{N}(\alpha(\frac{\sigma}{1-\sigma^2}, \frac{1-\rho^2}{1-\sigma^2} \sigma^2)). \) 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. 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 \( \epsilon_t \) in (1a) by a random variable from the GH skew Student’s t-distribution, denoted by \( \omega_t \), which can be written in the form of the normal variance-mean
mixture as

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

(2)

where $\epsilon_t \sim \mathcal{N}(0, 1)$ and $z_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$, where $\mathcal{N}(\cdot, \cdot)$ and $\mathcal{G}(\cdot, \cdot)$ 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:

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

(3a)

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

(3b)

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

(3c)

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

(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

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

(4)

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

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

(5)

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

(6)
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 \( \theta = (\beta_0, \beta_1, \beta_2, \alpha, \phi, \tau^2, \varphi, \nu)' \) be the full parameter vector of the entire class of SVML-GH-ST models, where \( \nu \) is the parameter vector associated with the mixture distribution, \( \tau = \sqrt{1 - \rho^2 \sigma^2} \) and \( \varphi = \rho \sigma \), \( \textbf{h}_{1:T} = (h_1, h_1, \ldots, h_T)' \) be the vector of the log volatilities, \( \textbf{z}_{1:T} = (z_1, \ldots, z_T)' \) be the mixing variables and \( \textbf{y}_{0:T} = (y_0, \ldots, y_T)' \) 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

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

where \( p(y_t, h_{t+1} \mid z_t, h_t, y_{t-1}, \theta) \) is given by equation (4) and \( p(\theta) \) 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 \( (\theta, \textbf{h}_{1:T}, \textbf{z}_{1:T}) \) given \( \textbf{y}_{0:T} \) 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 \( \theta^{(i)} \) and the latent quantities \( \textbf{z}_{1:T}^{(i)} \) and \( \textbf{h}_{1:T}^{(i)} \).

2. Generate \( \theta^{(i)} \) in turn from its full conditional distribution, given \( \textbf{y}_{1:T}, \textbf{h}_{1:T}^{(i-1)} \) and \( \textbf{z}_{1:T}^{(i-1)} \).

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

4. Generate \( \textbf{h}_{1:T} \) by blocks:

   i) For \( l = 1, \ldots, K \), the knot positions are generated as \( k_l, \) the floor of \( [T \times \{(l+u_i)/(K+2)\}] \), where the \( u_i's \) 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 \( y_{k_{l-1}+1:k_{l-1}}, \theta^{(i)}, z_{k_{l-1}+1:k_{l-1}}, h^{(i-1)}_{k_{l-1}} \) and \( h^{(i-1)}_{k_l} \).

iii) For \( l = 1, \ldots, K \), draw \( h^{(i)}_{k_l} \) conditional on \( y_{1:T}, \theta^{(i)}, h^{(i)}_{k_{l-1}} \) 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: \( \beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma_{\bar{\beta}_0}^2), \beta_1 \sim \mathcal{N}(-1,1)(\bar{\beta}_1, \sigma_{\bar{\beta}_1}^2), \beta_2 \sim \mathcal{N}(\bar{\beta}_2, \sigma_{\bar{\beta}_2}^2), \alpha \sim \mathcal{N}(\alpha_0, \tau^2), \varphi \sim \mathcal{N}(\varphi_0, \tau^2/p_0), \phi \sim \mathcal{N}(-1,1)(\phi_0, s_\phi^2), \tau^2 \sim IG(a_\tau/2, S_\tau/2) \) and \( \nu \sim G(a_\nu, b_\nu) \), where \( a_\nu, b_\nu, \alpha_0, \varphi_0, \phi_0, s_\phi^2, a_\tau, S_\tau, p_0 \) 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 \( h_{1:T} \) 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 \( h_{1:T} \) 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 \( h_{1:T} \) in the SVML-SMN class of models. Details on the full conditionals of \( \theta \) and the latent variable \( \lambda_{1:T} \) are given in Appendix A. Some of them are easy to simulate from.

2.2 A block sampler algorithm

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

$$
\log f(\eta_{t:t+k-1}|h_t, h_{t+k+1}, y_{t:t+k}, z_{t+1:t+k}) = -\sum_{s=t}^{t+k-1} \frac{\eta_s^2}{2} + \sum_{s=t}^{t+k} l_s - \frac{1}{2\sigma_\eta^2} (h_{t+k+1} - \alpha - \phi h_{t+k})^2 \mathbb{I}(t + k < T),
$$

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 $V_s$, 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 $d_{t+1:t+k} = (d_{t+1}, \ldots, d_{t+k})'$, where $d_s$ and $Q$ 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 $\hat{\eta}_{t:t+k-1}$, we

\footnote{For the last block, we have $y_T | y_{T-1}, h_T \sim N(\beta_0 + \beta_1 y_{T-1} + \beta_2 e^{h_T} + e^{h_T} \delta(z_T - \mu), \delta z_{T} e^{h_T})$.}
have

\[
\log f(\eta_{t:t+k-1}|h_t, h_{t+k+1}, y_{t+1:t+k}, z_{t+1:t+k}) \\
\approx \text{const} - \frac{1}{2} \sum_{r=t+1}^{t+k} \eta_r^2 + \hat{L} + \frac{\partial L}{\partial \eta_{t:t+k-1}} \eta_{t:t+k-1} - \hat{\eta}_{t:t+k-1} (\eta_{t:t+k-1} - \hat{\eta}_{t:t+k-1}) \\
+ \frac{1}{2} (\eta_{t:t+k-1} - \hat{\eta}_{t:t+k-1})^T E(\frac{\partial^2 L}{\partial \eta_{t:t+k-1}^2} \eta_{t:t+k-1}) (\eta_{t:t+k-1} - \hat{\eta}_{t:t+k-1}) \\
= \text{const} - \frac{1}{2} \sum_{r=t+1}^{t+k} \eta_r^2 + \hat{L} + \hat{d}_{t+1:t+k} (h_{t+1:t+k} - \hat{h}_{t+1:t+k}) \\
- \frac{1}{2} (h_{t+1:t+k} - \hat{h}_{t+1:t+k})^T \hat{Q} (h_{t+1:t+k} - \hat{h}_{t+1:t+k}) \\
= \text{const} + \log f^* (\eta_{t:t+k-1}|h_t, h_{t+k+1}, \theta, y_{t+1:t+k}, z_{t+1:t+k}),
\]

(9)

where \( \hat{d}_{t+1:t+k} \), \( \hat{L} \) and \( \hat{Q} \) denote \( d_{t+1:t+k} \), \( L \) and \( Q \) evaluated at \( h_{t+1:t+k} = \hat{h}_{t+1:t+k} \). The expectations are taken with respect to \( y_s \)'s conditional on \( h_s \)'s. We use an information matrix for \( Q \) because we require that \( Q \) is everywhere strictly positive definite. It can be shown that the proposal density \( f^* (\eta_{t:t+k-1}|h_t, h_{t+k+1}, \theta, y_{t+1:t+k}, z_{t+1:t+k}) \) is the posterior density of \( \eta_{t:t+k-1} \) for a linear Gaussian state space model given by equations (10) and (11) below (see Omori and Watanabe, 2008, for details). The mode \( \hat{\eta}_{t:t+k-1} \) can be found by repeating the following algorithm until convergence.

**Algorithm 2**

1. Initialize \( \hat{\eta}_{t:t+k-1} \) and calculate \( \hat{h}_{t+1:t+k} \) 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.

\[
G_s = \hat{M}_s - \hat{N}_s^2 \hat{G}_{s-1}, \quad G_{t+1} = \hat{M}_{t+1}, \\quad J_{t+1} = 0, \quad J_{t+k+1} = 0, \\quad b_s = \hat{d}_s - J_s \hat{K}_{s-1} b_{s-1} - b_{t+1} = \hat{d}_{t+1},
\]

where \( \hat{K}_s = \sqrt{\hat{G}_s} \).
4. Define the auxiliary variables $\hat{y}_s = \hat{\gamma}_s + G_s^{-1} h_s$, 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

$$
\begin{align*}
\hat{y}_s &= c_s + Z_s h_s + H_s \xi_s, \quad s = t + 1, \ldots, t + k, \\
h_{s+1} &= \alpha + \phi h_s + L_s \xi_s, \quad s = t, t + 1, \ldots, t + k,
\end{align*}
$$

where $\xi_s \sim \mathcal{N}(0, I_2)$, $c_s = K_s^{-1} J_{s+1} \alpha$, $Z_s = 1 + K_s^{-1} J_{s+1} \phi$, $H_s = K_s^{-1}[1, J_{s+1} \sigma_\eta]$ 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 $\eta_{t:t+k-1}$ ($h_{t:t+k}$) and set $\hat{\eta}_{t:t+k} = 1$ ($\hat{h}_{t:t+k}$) 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 $\hat{y}_{t+1:t+k}$ defined in step 4 of Algorithm 2 enables us to sample $\eta_{t+1:t+k}$ 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 $\hat{h}_{t+1:t+k}$ in Algorithm 2 as follows: the current sample of $\eta_{t:t+k} = 1$ ($h_{t+1:t+k}$) may be taken as an initial value of the $\hat{\eta}_{t:t+k} = 1$ ($\hat{h}_{t+1:t+k}$) in Step 1. Once an initial expansion block $\hat{h}_{t+1:t+k}$ is selected, we can calculate the auxiliary $\hat{y}_{t+1:t+k}$ 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 $\hat{y}_{t+1:t+k}$ yields the mean of $h_{t+1:t+k}$ conditional on $\hat{h}_{t+1:t+k}$ in the linear Gaussian state space model, which is used as the next $\hat{h}_{t+1:t+k}$. By repeating the procedure until the smoothed estimates converge, we obtain the posterior mode of $h_{t+1:t+k}$. 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 $\hat{h}_{t+1:t+k}$ instead of
an optimal one.

Finally, we describe the updating procedure of the knot conditions $h_{kl}$, for $l = 2, \ldots, K$. As
the conditional density $p(h_{kl} | h_{kl-1}, h_{kl+1})$ does not have a closed form, we use the Metropolis-
Hastings algorithm with proposal density $\mathcal{N}(\frac{\alpha(1-\phi)+\phi(h_{kl-1}+h_{kl+1})}{1+\phi^2}, \frac{\sigma^2}{1+\phi^2})$. Let $h^{(i)}_{kl}$ and $h^{(i-1)}_{kl}$ denote the proposal value and the previous iteration value. Thus, the acceptance probability
is given by $\alpha_{MH} = \min\{1, \frac{Q(h^{(i)}_{kl})}{Q(h^{(i-1)}_{kl})}\}$, where $Q(h_{kl})$ is the product of the conditional densities $y_{k_{l-1}} \mid z_{k_{l-1}}, y_{k_{l-2}}, h_{k_{l-1}}, h_{kl} \sim \mathcal{N}(\mu_{k_{l-1}}, V_{k_{l-1}})$ and $y_{k_{l}} \mid z_{k_{l}}, y_{k_{l-1}}, h_{k_{l-1}}, h_{kl} \sim \mathcal{N}(\mu_{k_{l}}, V_{k_{l}})$, with $\mu_s$ 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).
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.


3 Numerical illustration with artificial dataset

In order to assess 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 standalone 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 $\beta_0 = 0.25$, $\beta_1 = 0.03$, $\beta_3 = -0.2$, $\alpha = -0.008$, $\phi = 0.95$, $\sigma^2 = 0.0225$, $\rho = -0.35$ and $\nu = 10$, 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: $\beta_0 \sim \mathcal{N}(0, 100)$, $\beta_1 \sim \mathcal{N}_{[-1,1]}(0.1, 100)$, $\beta_2 \sim \mathcal{N}(-0.1, 100)$, $\alpha \sim \mathcal{N}(0, \tau^2/0.002)$, $\phi \sim \mathcal{N}_{[-1,1]}(0.95, 100)$, $\tau^2 \sim \mathcal{IG}(2.5, 0.025)$, $\varphi \sim \mathcal{N}(-0.3, \tau^2/0.005)$, $\delta \sim \mathcal{N}(0, 1)$ 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 $\beta_1$ and $\phi$, 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 $\beta_1$ and $\phi$ are non-informative.

The number of blocks, $K$, in the block sampler was set equal to 30, so that each block contained 66 $t_s$ 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 $e^{ht}$.

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 $\rho_s$ 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.
Table 1: Simulated dataset: summary results for the SVML-GH-ST model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% CI</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0.2500</td>
<td>0.2810</td>
<td>(0.1220, 0.4650)</td>
<td>6.26</td>
<td>-0.12</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.0300</td>
<td>0.0260</td>
<td>(-0.0170, 0.0680)</td>
<td>1.28</td>
<td>0.95</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.2000</td>
<td>-0.2500</td>
<td>(-0.4450, -0.0700)</td>
<td>5.38</td>
<td>0.01</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.0080</td>
<td>-0.0160</td>
<td>(-0.0340, -0.0030)</td>
<td>10.36</td>
<td>-0.95</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9500</td>
<td>0.9210</td>
<td>(0.8680, 0.9610)</td>
<td>10.67</td>
<td>-1.01</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0225</td>
<td>0.0330</td>
<td>(0.0160, 0.0550)</td>
<td>21.83</td>
<td>0.96</td>
</tr>
<tr>
<td>$\delta$</td>
<td>-0.5000</td>
<td>-0.7680</td>
<td>(-1.6100, -0.3400)</td>
<td>21.31</td>
<td>-0.32</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.3500</td>
<td>-0.2350</td>
<td>(-0.4270, -0.0420)</td>
<td>7.02</td>
<td>0.52</td>
</tr>
<tr>
<td>$\nu$</td>
<td>10.0000</td>
<td>12.4430</td>
<td>(8.2330, 19.5520)</td>
<td>20.02</td>
<td>0.15</td>
</tr>
</tbody>
</table>

In Figure 4, the smoothed mean calculated from the MCMC output (dotted line) and true values (solid line) of $e^{\frac{h_t}{T}}$ 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.

<table>
<thead>
<tr>
<th>Média</th>
<th>D.p.</th>
<th>Mínimo</th>
<th>Máximo</th>
<th>Assimetria</th>
<th>Curtose</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>2.34</td>
<td>-17.21</td>
<td>28.83</td>
<td>0.83</td>
<td>19.18</td>
</tr>
</tbody>
</table>

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: $\beta_0 \sim \mathcal{N}(0, 100)$, $\beta_1 \sim \mathcal{N}(-1,1)(0.1, 100)$, $\beta_2 \sim \mathcal{N}(-0.1, 100)$, $\phi \sim \mathcal{N}(-1,1)(0.95,100)$, $\tau^2 \sim \mathcal{IG}(2.5, 0.025)$, $\alpha | \tau^2 \sim \mathcal{N}(0, \tau^2/0.002)$ and $\varphi | \tau^2 \sim \mathcal{N}(-0.3, \tau^2/0.005)$. 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 \( z_{1:T} \) 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 \( h_t \)'s 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 \( \phi \). For all the models, the posterior means of \( \phi \) 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 \( \sigma_n^2 \) 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 \( \beta_0 \) is always positive and statistically significant for each one of the models fitted. 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 might be not significant. The \( \beta_2 \) parameter, which measures both the \textit{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 4. This

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SVML-N</th>
<th>SVML-T</th>
<th>SVML-GH-ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0.1409</td>
<td>0.1801</td>
<td>0.1510</td>
</tr>
<tr>
<td></td>
<td>(0.0031, 0.2820)</td>
<td>(0.0269, 0.3388)</td>
<td>(0.0180, 0.2920)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.18</td>
<td>-0.18</td>
<td>-1.13</td>
</tr>
<tr>
<td></td>
<td>(-0.0562, 0.0239)</td>
<td>(-0.0219, 0.0682)</td>
<td>(-0.0220, 0.0690)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.49</td>
<td>1.60</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>(-0.0559, 0.0193)</td>
<td>(-0.0894, 0.0160)</td>
<td>(-0.0700, 0.0140)</td>
</tr>
<tr>
<td>$\sigma^2_\eta$</td>
<td>0.0713</td>
<td>0.0407</td>
<td>0.049</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9368</td>
<td>0.9579</td>
<td>0.9510</td>
</tr>
<tr>
<td></td>
<td>(0.8940, 0.9765)</td>
<td>(0.9184, 0.9855)</td>
<td>(0.9150, 0.9800)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-1.33</td>
<td>1.47</td>
<td>-1.68</td>
</tr>
<tr>
<td></td>
<td>(-0.4677, 0.1774)</td>
<td>(-0.5319, 0.1779)</td>
<td>(-0.4490, 0.2060)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>2.355</td>
<td>28.31</td>
<td>20.05</td>
</tr>
<tr>
<td></td>
<td>(0.0271, 0.1196)</td>
<td>(0.0147, 0.0758)</td>
<td>(0.0210, 0.0850)</td>
</tr>
<tr>
<td>$\nu^2$</td>
<td>0.0708</td>
<td>0.0426</td>
<td>0.0550</td>
</tr>
<tr>
<td></td>
<td>(0.0250, 0.1214)</td>
<td>(0.0146, 0.0818)</td>
<td>(0.0240, 0.0940)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>- 10.9988</td>
<td>18.7150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.9690, 16.9087)</td>
<td>(10.3000, 29.5600)</td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>1.32</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26.25</td>
<td>23.75</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>- 1.2200</td>
<td>- 0.3810</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.2200, 0.1600)</td>
<td>(-1.79)</td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>18</td>
<td>10.22</td>
<td></td>
</tr>
</tbody>
</table>
Table 4: IBOVESPA dataset: $P(\beta_2 < 0)$ estimated from the MCMC output.

<table>
<thead>
<tr>
<th></th>
<th>SVML-N</th>
<th>SVML-T</th>
<th>SVML-GH-ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\beta_2 &lt; 0)$</td>
<td>0.8295</td>
<td>0.9055</td>
<td>0.8948</td>
</tr>
</tbody>
</table>

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 $\rho$, 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 $\delta$ 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 $\nu$ in the SVML-T and SVML-GH-ST models. The posterior means of $\nu$ are almost 11 and 19 in the SVML-T and SVML-GH-ST respectively. This difference can explained by the $\delta$, 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 $e^{h_t}$. The posterior smoothed mean of $e^{h_t}$ 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 $e^{h_t}$ by $\frac{1}{4}e^{\frac{h_t}{2}}$. 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

$$BPIC = -2E_{\theta|y_{1:T}}[\log(p(y_{1:T} | \theta))] + 2T\hat{b},$$  

(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 $b$ is given by

$$b \approx \frac{1}{T} \left\{ E_{\theta|y_{1:T}}[\log\{p(y_{1:T} | \theta)p(\theta)\}] - \log[p(y_{1:T} | \hat{\theta})p(\hat{\theta})] + \text{tr}\{J_T^{-1}(\hat{\theta})I_T(\hat{\theta})\} + 0.5q \right\} \quad (13)$$

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

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

$$J_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial \eta_T(y_t, \theta)}{\partial \theta} \right) \bigg|_{\theta = \hat{\theta}},$$

with $\eta_T(y_t, \theta) = \log p(y_t | y_{1:t-1}, \theta) + \log p(\theta)/T$.

In the SVML class of models, the log-likelihood function, $\log p(y_{1:T} | \theta)$, 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>
<thead>
<tr>
<th>Modelo</th>
<th>BPIC</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVML-N</td>
<td>8084.6</td>
<td>3</td>
</tr>
<tr>
<td>SVML-T</td>
<td>8081.8</td>
<td>2</td>
</tr>
<tr>
<td>SVML-GH-ST</td>
<td>8075.4</td>
<td>1</td>
</tr>
</tbody>
</table>

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 $\beta_2$ estimate, which measures both the \textit{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 $\phi$ 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 IBOVESPA 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 $\nu$ and $\delta$ 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.

\textbf{Appendix A: The Full conditionals}

In this appendix, we describe the full conditional distributions for the parameters and the mixing latent variables $z_{1:T}$ of the SVML-GH-ST model.
Full conditional distribution of $\beta_0$, $\beta_1$ and $\beta_2$

Let $m_t$ and $V_t$ be defined by

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

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

$$
\begin{align*}
    \beta_0 \mid y_{0:T}, h_{1:T}, \lambda_{1:T}, \beta_1, \beta_2 &\sim \mathcal{N}\left(\frac{b_{\beta_0}}{a_{\beta_0}}, \frac{1}{a_{\beta_0}}\right), \\
    \beta_1 \mid y_{0:T}, h_{1:T}, \lambda_{1:T}, \beta_0, \beta_1 &\sim \mathcal{N}\left(\frac{b_{\beta_1}}{a_{\beta_1}}, \frac{1}{a_{\beta_1}}\right) \mathbb{I}_{|\beta_2|<1}, \\
    \beta_2 \mid y_{0:T}, h_{1:T}, \lambda_{1:T}, \beta_0, \beta_1 &\sim \mathcal{N}\left(\frac{b_{\beta_2}}{a_{\beta_2}}, \frac{1}{a_{\beta_2}}\right),
\end{align*}
$$

where $a_{\beta_0} = \sum_{t=1}^{T} \frac{1}{V_t} + \frac{1}{\sigma^2_{\beta_0}}$, $b_{\beta_0} = \sum_{t=1}^{T} \frac{y_{t-1}^2}{V_t} + \frac{\beta_0}{\sigma^2_{\beta_0}}$, $a_{\beta_1} = \sum_{t=1}^{T} \frac{y_{t-1}^2}{V_t} + \frac{1}{\sigma^2_{\beta_1}}$, $b_{\beta_1} = \sum_{t=1}^{T} \frac{y_{t-1}^2}{V_t} + \frac{\beta_1}{\sigma^2_{\beta_1}}$, $a_{\beta_2} = \sum_{t=1}^{T} \frac{1}{V_t} + \frac{1}{\sigma^2_{\beta_2}}$, $b_{\beta_2} = \sum_{t=1}^{T} \frac{y_{t-1}^2}{V_t} + \frac{\beta_2}{\sigma^2_{\beta_2}}$,

$$
\begin{align*}
    w_t &= y_t - \beta_1 y_{t-1} - \beta_2 e^{h_t} - e^{h_t} \delta(z_t - \mu_2) - m_t, \\
    r_t &= y_t - \beta_0 - \beta_1 y_{t-1} - e^{h_t} \delta(z_t - \mu_2) - m_t \text{ and } \mathbb{I}_{|\beta_2|<1},
\end{align*}
$$

an indicator variable.

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

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

After some simple but tedious algebra, we have

$$
\begin{align*}
    \alpha \mid &\sim \mathcal{N}\left(\frac{B_{\alpha}}{A_{\alpha}}, \frac{\tau^2}{A_{\alpha}}\right), \\
    \varphi \mid &\sim \mathcal{N}\left(\frac{B_{\varphi}}{A_{\varphi}}, \frac{\tau^2}{A_{\varphi}}\right), \\
    \delta \mid &\sim \mathcal{N}\left(\frac{B_{\delta}}{A_{\delta}}, \frac{1}{A_{\delta}}\right),
\end{align*}
$$

where $A_{\alpha} = q_0 + \frac{1}{\tau^2} + T - 1$, $B_{\alpha} = \alpha_0 q_0 + (1 + \phi) h_1 + \sum_{t=1}^{T} k_t$, $k_t = h_{t+1} - \phi h_t - \varphi g_t z_t^{-\frac{1}{2}} e^{\frac{-h_t}{2}}$, $A_{\varphi} = p_0 + \sum_{t=1}^{T} g_t^2 z_t^{-1} e^{-h_t}$, $B_{\varphi} = \varphi_0 p_0 + \sum_{t=1}^{T} c_t g_t z_t^{-\frac{1}{2}} e^{\frac{-h_t}{2}}$, $A_{\delta} = \frac{1}{\tau^2} \sum_{t=1}^{T} \frac{1}{\sqrt{a_t}} (z_t - p_0 + \sum_{t=1}^{T} c_t g_t z_t^{-\frac{1}{2}} e^{\frac{-h_t}{2}}), B_{\delta} = \varphi_0 p_0 + \sum_{t=1}^{T} c_t g_t z_t^{-\frac{1}{2}} e^{\frac{-h_t}{2}}.$
\[ p(\phi \mid \cdot) \propto Q(\phi) \exp \left\{ \frac{-A_\phi}{2} \left( \phi - \frac{B_\phi}{A_\phi} \right)^2 \right\}, \quad (A.7) \]

where

\[ Q(\phi) = \sqrt{1 - \phi^2} \exp \left\{ -\frac{1 - \phi^2}{2\tau^2}(h_1 - \frac{\alpha}{1 - \phi})^2 \right\}, \]

\[ l_t = h_{t+1} - \alpha - \varphi(y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t}) \lambda_t e^{-\frac{h_t}{T}}, \quad A_\phi = \frac{1}{\sigma^2_\phi} + \sum_{t=1}^{T-1} \frac{h_t^2}{\tau^2}, \quad B_\phi = \frac{\phi_0}{\sigma^2_\phi} + \sum_{t=1}^{T-1} \frac{h_t y_t}{\tau^2} \]

and \( I_{|\phi| < 1} \) is an indicator variable. As \( p(\phi \mid h_1: T, \alpha, \sigma^2_\eta) \) in (A.7) does not have closed form, we sample from it by using the Metropolis-Hastings algorithm with truncated \( \mathcal{N}(\alpha, 1) \) as the proposal density. The conditional posterior of \( \tau^2 \) is \( IG\left(\frac{T_1}{2}, \frac{M_1}{2}\right) \), where \( T_1 = a_\tau + T + 2 \) and \( M_1 = (1 - \phi^2)(h_1 - \frac{\alpha}{1 - \phi})^2 + \sum_{t=1}^{T-1} (c_t - \varphi z_t \frac{1}{\sigma^2_\eta} e^{-\frac{h_t}{T}} y_t)^2 + p_0(\phi - \varphi_0)^2 + q_0(\alpha - \alpha_0)^2 + S_\tau \). Once \( \tau^2 \) and \( \varphi \) are sampled, respectively, from their conditional posteriors, we can calculate \( \rho \) and \( \sigma^2_\eta \) through \( \sigma^2_\eta = \tau^2 + \varphi^2 \) and \( \rho = \varphi/\sigma_\eta \).

**Full conditional of \( z_t \) and \( \nu \)**

The full conditional of \( z_t \) is

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

where the values of \( \lambda, \vartheta \) and \( \gamma \) are the parameters of a distribution \( GIG(\lambda, \vartheta, \gamma) \) whose values are given by

\[ \lambda = -\frac{\nu + 1}{2} \]

\[ \gamma^2 = \delta^2 \varphi^2 + \frac{\tau^2}{\tau^2} \]

\[ \vartheta^2 = \frac{\varphi^2 + \frac{\tau^2}{\tau^2} e^{-h_t} \left( y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} + \mu z e^{h_t/2} \delta \right)^2 + \nu}{2} \]
We sample $z_t$ by the Metropolis-Hastings algorithm. We use $\text{GIG}(\lambda, \vartheta, \gamma)$ as the proposal distribution 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

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

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

$$p(\nu \mid z_{1:T}) \propto \frac{2^\nu}{\Gamma(\frac{\nu}{2})} \exp\left\{ -\sum_{t=1}^{T} \frac{1}{2 V_t} \left[ y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{ht} - e^{\frac{ht}{2}} \delta (z_t - \frac{\nu}{\nu - 2}) - m_t \right]^2 \right\}
- \frac{\nu}{2} \sum_{t=1}^{T} \left[ \frac{1}{z_t} + \log z_t \right] \nu^{\nu-1} \exp\{-b_\nu \nu\} \mathbb{I}_{4<\nu<40}.$$

We sample $\nu$ by the Metropolis-Hastings acceptance-rejection algorithm (Tierney, 1994; Chib, 1995). Let $\nu^*$ denote the mode (or approximate mode) of $p(\nu \mid z_{1:T})$, and let $\ell(\nu) = \log p(\nu \mid z_{1:T})$. 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

$$d_s = \frac{\partial L}{\partial h_s} = -\frac{1}{2} + \frac{(y_s - \mu_s)^2}{2 V_s} + \frac{(y_s - \mu_s)}{V_s} \frac{\partial \mu_s}{\partial h_s} + \frac{(y_{s-1} - \mu_{s-1})}{V_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_s} - \phi \frac{1}{\sigma_h^2} \mathbb{I}(t+k < T), \quad s = t+1, \ldots, t+k,$$

and

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

\[
M_s = -E \left[ \frac{\partial^2 L}{\partial h_s^2} \right] = \frac{1}{2} + \frac{1}{V_s} \left( \frac{\partial \mu_s}{\partial h_s} \right)^2 + \frac{1}{V_{s-1}} \left( \frac{\partial \mu_{s-1}}{\partial h_s} \right)^2
+ \frac{\phi^2}{\sigma^2} \mathbb{I}(t + k < T), \quad s = t + 1, \ldots, t + k, \quad (B.3)
\]

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

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

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

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

**Acknowledgements**

The first author gratefully acknowledges financial support from the Fundação de Amparo à Pesquisa do Estado de Rio de Janeiro (FAPERJ). C. A. Abanto-Valle is deeply indebted to CNPq-Brazil and FAPERJ.

**References**


