



## Threshold Stochastic Volatility Models with Heavy Tails: A Bayesian Approach

Carlos A. Abanto-Valle<sup>a,\*</sup>, Hernán B. Garrafa-Aragón<sup>b</sup>

<sup>a</sup>Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970, Rio de Janeiro, Brazil

✉ cabantovalle@im.ufrj.br \* Corresponding author

<sup>b</sup>Escuela de Ingeniería Estadística de la Universidad Nacional de Ingeniería, Lima, Perú

✉ hgarrafa@gmail.com

### Abstract

This paper extends the threshold stochastic volatility (THSV) model specification proposed in So et al. (2002) and Chen et al. (2008) by incorporating thick-tails in the mean equation innovation using the scale mixture of normal distributions (SMN). A Bayesian Markov Chain Monte Carlo algorithm is developed to estimate all the parameters and latent variables. Value-at-Risk (VaR) and Expected Shortfall (ES) forecasting via a computational Bayesian framework are considered. The MCMC-based method exploits a mixture representation of the SMN distributions. The proposed methodology is applied to daily returns of indexes from BM&F BOVESPA (BOVESPA), Buenos Aires Stock Exchange (MERVAL), Mexican Stock Exchange (MXX) and the Standar & Poors 500 (SP500). Bayesian model selection criteria reveals that there is a significant improvement in model fit for the returns of the data considered here, by using the THSV model with slash distribution over the usual normal and Student- $t$  models. Empirical results show that the skewness can improve VaR and ES forecasting in comparison with the normal and Student- $t$  models.

**Article History:** Received: 27 February 2019 / Revised: 8 May 2019 / Accepted: 10 May 2019

**Keywords:** Expected shortfall; Markov chain Monte Carlo; Non linear state space models; Scale mixtures of normal distributions; Stochastic volatility; Threshold; Value-at-Risk.

**JEL Classification:** C11, C15, C51, C52, C58

### Acknowledgements

The authors would like to thank the Editor Gabriel Rodríguez and an anonymous referee for their constructive comments which substantially improved the quality of this paper. The research of Carlos A. Abanto-Valle was partially supported by the CNPq and FAPERJ grants.

## 1. Introduction

A large literature in financial econometrics has documented many stylized facts for financial asset returns. As is well known, the return series exhibits significant non-Gaussian behavior such as heavy tails and asymmetry, in addition to volatility clustering. These properties are crucial not only for describing the return distributions but also for asset allocation, option pricing, forecasting and risk management.

Many empirical studies have shown strong evidence of heavy-tailed conditional mean errors in financial time series; see for example [Mandelbrot \(1963\)](#) and [Fama \(1965\)](#). In the stochastic volatility literature, [Liesenfeld and Jung \(2000\)](#), [Chib et al. \(2002\)](#), [Jacquier et al. \(2004\)](#) and [Abanto et al. \(2010\)](#), among others, have provided consistent evidence that leptokurtic distributions, such as the Student's  $t$ , the generalized error distribution or the scale mixture of normal (SMN) distributions, are more adequate to capture this empirical regularity by relaxing the restrictive normality assumption in the distribution of the returns.

Recently, asymmetric behavior in stock returns has received attention since the work of [Black \(1976\)](#), see also [Christie \(1982\)](#) for a discussion. Furthermore, this empirical evidence has provided the explanation that unexpected returns and innovations of the volatility process are negatively correlated, which is known as the leverage effect. [Black \(1976\)](#) asserted that a negative (positive) return shock leads to an increase (decrease) in the company's financial leverage ratio, which has an upward (downward) effect on the volatility of its stock returns.

[So et al. \(2002\)](#) introduced an alternative approach to incorporating asymmetry in stochastic volatility models. They developed a threshold stochastic volatility (THSV) model allowing the conditional mean and the log-volatility parameters' to change with the sign of the lagged returns. More recently, [Chen et al. \(2008\)](#) generalize the THSV models allowing the threshold variable to be different from zero and endowing the Student- $t$  distribution to the mean innovation. Note that the SV model ([Taylor, 1982, 1994](#)) is a special case of the THSV model when asymmetry does not exist in stock returns.

In this article, we extend the setup of [So et al. \(2002\)](#), [Chen et al. \(2008\)](#) and [Abanto et al. \(2010\)](#), in order to take account simultaneously for heavy-tails of the returns and volatility asymmetries, by considering the THSV model with SMN distributions. We refer to this generalization as THSV-SMN distributions. Interestingly, this rich class contains as proper elements the THSV with normal (THSV-N), Student- $t$  (THSV-T), slash (THSV-S) and the variance gamma (THSV-VG) distributions. The estimation of such intricate models is not straightforward, since volatility appears in both the mean and the variance equation and hence intensive computational methods are needed for estimating purposes. Inference in the THSV-SMN class of models is performed under a Bayesian paradigm via MCMC methods, which permits to obtain the posterior distribution of parameters by simulation starting from reasonable prior assumptions on the parameters. Based on the mixture sampler ([Kim et al., 1998](#); [Omori et al., 2007](#)), an efficient multi-move sampler is developed to sample the log-volatilities.

The remainder of this paper is organized as follows. Section 2 shows a brief review of the SMN distributions. Section 3 describes the THSV-SMN class of models as well as the Bayesian

estimation procedure using MCMC methods. We discuss some technical details about Bayesian model selection in Section 4. Section 5 is devoted to application and model comparison among particular members of the THSV-SMN models using the BOVESPA, MERVAL, MXX and SP500 returns. Finally, some concluding remarks as well as future developments are deferred to Section 6. Next, in Appendices A and B, we show some derivations for sampling from the full conditionals of parameters and states, respectively.

## 2. SMN Distributions

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

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

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

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

where  $\phi(\cdot | \mu, \sigma^2)$  denotes the density of the univariate  $\mathcal{N}(\mu, \sigma^2)$  distribution. From a suitable choice of the mixing density  $h(\cdot | \boldsymbol{\nu})$ , a rich class of continuous symmetric distributions can be described by the density given in (2) that can readily accommodate thicker-tails than the normal process. Note that when  $\lambda = 1$  (a degenerate random variable), we retrieve the normal distribution. Apart from the normal model, we explore three different types of heavy-tailed densities based on the choice of the mixing density  $h(\cdot | \boldsymbol{\nu})$ .

- *The Student-t distribution*,  $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$

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

$$Y | \mu, \sigma^2, \nu, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \quad \lambda | \nu \sim \mathcal{G}(\nu/2, \nu/2), \quad (3)$$

where  $\mathcal{G}(\cdot, \cdot)$  denotes the gamma distribution.

- *The Slash distribution*,  $Y \sim \mathcal{S}(\mu, \sigma^2, \nu)$ ,  $\nu > 0$ .

This distribution presents heavier tails than those of the normal distribution and it includes the normal case when  $\nu \rightarrow \infty$ . The slash distribution is equivalent to the following hierarchical form:

$$Y | \mu, \sigma^2, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \quad \lambda | \nu \sim \mathcal{Be}(\nu, 1), \quad (4)$$

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

- The variance gamma distribution,  $Y \sim \mathcal{VG}(\mu, \sigma^2, \nu)$ ,  $\nu > 0$ .

The symmetric variance gamma (VG) distribution was first proposed by Madan and Seneta (1990) to model share market returns. The VG distribution is controlled by the shape parameter  $\nu > 0$  and presents heavier tails than those of the normal distribution. The VG distribution is equivalent to the following hierarchical form:

$$Y|\mu, \sigma^2, \nu, \lambda \sim N\left(\mu, \frac{\sigma^2}{\lambda}\right), \quad \lambda|\nu \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \quad (5)$$

where  $\mathcal{IG}(\cdot, \cdot)$  denotes the inverse gamma distribution.

### 3. The Heavy-Tailed Threshold Stochastic Volatility Model

The THSV model with SMN distributions is defined by

$$y_t = \mu_{s_t} + \beta_{s_t} y_{t-1} + e^{\frac{h_t}{2}} \lambda_t^{-\frac{1}{2}} \epsilon_t, \quad (6a)$$

$$h_{t+1} = \alpha_{s_{t+1}} + \phi_{s_{t+1}} h_t + \sigma_{s_{t+1}} \eta_t, \quad (6b)$$

$$\lambda_t \sim p(\lambda_t | \boldsymbol{\nu}), \quad (6c)$$

where  $y_t$  and  $h_t$  are respectively the compounded return and the log-volatility at time  $t$ . The innovations  $\epsilon_t$  and  $\eta_t$  are assumed to be mutually independent and normally distributed with zero mean and unit variance. At time  $t - 1$ , when there is an unexpected drop in price due to the presence of bad news,  $y_{t-1} < 0$  and  $s_t = 0$ . In contrast, if there is good news at time  $t - 1$  then  $y_{t-1} \geq 0$  and  $s_t = 1$ . In the THSV-SMN class the parameters  $\mu, \beta, \alpha, \phi, \sigma^2$  switch between the two regimes corresponding to sign of  $y_{t-1}$ . In this setup,  $\lambda_t$  is a scale factor,  $p(\lambda_t | \boldsymbol{\nu})$  is the mixing density and  $\boldsymbol{\nu}$  the parameter that capture the heavy-tailness. The aim of the THSV-SMN class of models is to describe both mean and volatility asymmetry in the presence of outliers. The THSV-SMN class includes the THSV with normal (THSV-N), Student-t (THSV-T), slash (THSV-S) and variance gamma (THSV-VG) distributions as special cases. The first model is obtained with  $\lambda_t = 1$  for all  $t$ , and the other ones are obtained by choosing the mixing density as:  $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$ ,  $\lambda_t \sim \mathcal{Be}(\nu, 1)$  and  $\lambda_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$ , where  $\mathcal{G}(\cdot, \cdot)$ ,  $\mathcal{Be}(\cdot, \cdot)$  and  $\mathcal{IG}(\cdot, \cdot)$  denote the gamma, beta and inverse gamma distributions respectively.

Under a Bayesian paradigm, we use MCMC methods to conduct the posterior analysis in the next subsection. Conditionally to  $\lambda_t$ , some derivations are common to all members of the THSV-SMN family.

#### 3.1 Parameter Estimation via MCMC

Bayesian estimation in the THSV-SMN class of models defined by equations (6a)-(6c) is a non trivial task, because the stochastic volatility  $h_t$  is an unknown process. To overcome this difficulty, we propose an algorithm based on MCMC simulation to make the Bayesian analysis feasible.

First, let us to introduce some notation. Let  $\mathbf{h}_{1:T} = (h_1, \dots, h_T)'$  be the vector of log-volatilities,  $\boldsymbol{\lambda}_{1:T} = (\lambda_1, \dots, \lambda_T)'$  the mixing variables and  $\boldsymbol{\theta} = (\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \boldsymbol{\varphi}_0, \boldsymbol{\varphi}_1, \sigma_0^2, \sigma_1^2, \nu)'$  the entire parameter vector of the entire class of THSV-SMN models, where  $\boldsymbol{\psi}_{s_t} = (\mu_{s_t}, \beta_{s_t})'$ ,

$\varphi_{s_t} = (\alpha_{s_t}, \phi_{s_t})'$ , for  $s_t = 0, 1$  respectively. As the posterior density  $p(\boldsymbol{\theta}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T} \mid \mathbf{y}_{1:T})$  does not have closed form, we draw random samples from it using the Gibbs sampling. The sampling scheme is described by the following algorithm:

### Algorithm 1.

1. Set  $i = 0$  and get starting values for the parameters  $\boldsymbol{\theta}^{(i)}$  and the latent quantities  $\boldsymbol{\lambda}_{1:T}^{(i)}$ , and  $\mathbf{h}_{1:T}^{(i)}$ .
2. Generate  $\boldsymbol{\theta}^{(i+1)} \sim p(\boldsymbol{\theta} \mid \mathbf{h}_{1:T}^{(i)}, \boldsymbol{\lambda}_{1:T}^{(i)}, \mathbf{y}_{1:T})$ .
3. Draw  $\boldsymbol{\lambda}_{1:T}^{(i+1)} \sim p(\boldsymbol{\lambda}_{1:T} \mid \boldsymbol{\theta}^{(i+1)}, \mathbf{h}_{1:T}^{(i)}, \mathbf{y}_{1:T})$ .
4. Generate  $\mathbf{h}_{1:T} \sim p(\mathbf{h}_{1:T} \mid \boldsymbol{\theta}^{(i+1)}, \boldsymbol{\lambda}_{1:T}^{(i+1)}, \mathbf{y}_{1:T})$ .
5. Set  $i = i + 1$  and return to 2 until convergence is achieved.

As described by Algorithm 1, the Gibbs sampler requires to sample parameters and latent variables from their full conditionals. Sampling the log-volatilities  $\mathbf{h}_{1:T}$  in Step 4 is the more difficult task due to the nonlinear setup in the mean equation (6a). In order to avoid the higher correlations due to the Markovian structure of the  $h'_t$ s, we apply a multi-move sampler (Kim et al., 1998; Omori et al., 2007) in the next subsection to sample the  $\mathbf{h}_{1:T}$  at once. Multi-move algorithms are computationally efficient and convergence is achieved much faster than using a single move (Abanto et al., 2010, 2014). Details on the full conditionals of  $\boldsymbol{\theta}$  and the latent variable  $\boldsymbol{\lambda}_{1:T}$  are given in the Appendix A, some of them are easy to simulate from.

### Sampling the log-volatilities

Let,  $u_t = \log(y_t - \mathbf{r}_t \boldsymbol{\psi}_{s_t})^2$ , where  $\mathbf{r}_t = (1, y_{t-1})'$ . After a log-square transformation, (6a) can be re-written as.

$$u_t = -\log \lambda_t + h_t + \log \epsilon_t^2, \quad (7)$$

where  $\log \epsilon_t^2 \sim \log \chi_1^2$ . The model defined by equations (7) and (6b) is a non-Gaussian state space model. So, following the ideas of Kim et al. (1998) and Omori et al. (2007), we approximate  $\log \epsilon_t^2$  by a finite mixture of normals distributions,

$$f(\log \epsilon_t^2) \approx \sum_{i=1}^P q_i f(\log \epsilon_t^2 \mid k_t = i),$$

where,  $k_t$  is a discrete mixing variable,  $\log \epsilon_t^2 \mid k_t = i \sim \mathcal{N}(\vartheta_i, \varpi_i^2)$  and  $q_i = Pr(k_t = i)$ . So equation (7) can be written as

$$u_t = -\log \lambda_t + h_t + \xi_t, \quad (8)$$

where  $\xi_t \mid k_t = i \sim \mathcal{N}(\vartheta_i, \varpi_i^2)$ .

Omori et al. (2007) argued that  $P = 10$  gives a satisfactory approximation of the  $\log \chi_1^2$  density.

For the parameters of the mixture of normal distributions,  $\vartheta_i$  and  $\varpi_i^2$  for  $i = 1, \dots, 10$ , we refer to [Omori et al. \(2007\)](#). The discrete mixing variable  $k_t$  can be drawn by computing the conditional probability

$$Pr(k_t = i | u_t, \lambda_t, h_t) \propto \frac{1}{\varpi_i} \exp\left\{-\frac{1}{2\varpi_i^2}(u_t + \log \lambda_t - h_t - \vartheta_i)^2\right\} q_i, \quad (9)$$

for  $i = 1, \dots, 10$  and  $t = 1, \dots, T$ .

We use the simulation smoothing method proposed by [McCausland et al. \(2011\)](#) to simulate from the states in the system defined by equations (8) and (6b), which conditional to the  $k_t$ 's and  $s_t$ 's follows a linear Gaussian state space model. The method of [McCausland et al. \(2011\)](#)(MMP) is described in Result 1.

**Result 1.** *If  $\mathbf{h}_{1:T} | \mathbf{z}_{1:T} \sim \mathcal{N}(\Omega^{-1}\mathbf{c}, \Omega^{-1})$ , where  $\Omega$  is defined as in Appendix B, then*

$$h_t | h_{t+1}, \mathbf{y}_{1:T} \sim \mathcal{N}(m_t - \Sigma_t \Omega_{t,t+1} h_{t+1}, \Sigma_t) \quad \text{and} \quad E[\mathbf{h}_{1:T} | \mathbf{z}_{1:T}] = (\mu_1, \dots, \mu_T)'$$

where  $\Sigma_1 = \Omega_{11}^{-1}$ ,  $m_1 = \Sigma_1 c_1$ ,  $\Sigma_t = (\Omega_{tt} - \Omega_{t-1,t}^2 \Sigma_{t-1})^{-1}$ ,  $m_t = \Sigma_t (c_t - \Omega_{t-1,t} m_{t-1})$ ,  $\mu_T = m_T$  and  $\mu_t = m_t - \Sigma_{t,t+1} \mu_{t+1}$ .

Algorithm 2 describes the simulations scheme using the MMP procedure for the  $j$ -th MCMC iteration.

### Algorithm 2.

1. Simulate  $h_T^{(j)} \sim \mathcal{N}(m_T^{(j)}, \Sigma_T^{(j)})$
2. For  $t = T - 1, \dots, 1$ , simulate  $h_t^{(j)} \sim \mathcal{N}(m_t^{(j)} - \Sigma_t^{(j)} \Omega_{t,t+1}^{(j)} h_{t+1}^{(j)}, \Sigma_t^{(j)})$

## 3.2 Forecasting Returns, Volatility, Value-at-Risk and Expected Shortfall

The  $K$ -step ahead prediction densities can be calculated using the composition method via the following recursive procedure:

$$\begin{aligned} p(y_{T+K} | \mathbf{y}_{1:T}) &= \int \left[ p(y_{T+K} | h_{T+K}, \lambda_{T+K}) p(\lambda_{T+K} | \boldsymbol{\theta}) \right. \\ &\quad \left. \times p(h_{T+K} | \boldsymbol{\theta}, \mathbf{y}_{1:T}) p(\boldsymbol{\theta} | \mathbf{y}_{1:T}) \right] dh_{T+K} d\lambda_{T+K} d\boldsymbol{\theta}, \\ p(h_{T+K} | \boldsymbol{\theta}, \mathbf{y}_{1:T}) &= \int p(h_{T+K} | \boldsymbol{\theta}, h_{T+K-1}) p(h_{T+K-1} | \boldsymbol{\theta}, \mathbf{y}_{1:T}) dh_{T+K-1}. \end{aligned}$$

Numerical evaluation of the last integrals is straightforward. To initialize the recursion, we use  $h_T^{(i)}$  and  $\boldsymbol{\theta}^{(i)}$ , for  $i = 1, \dots, N$ , from the MCMC output. Given these  $N$  draws, we sample  $h_{T+k}^{(i)}$  from  $p(h_{T+k} | \boldsymbol{\theta}^{(i)}, h_{T+k-1}^{(i)})$  and  $\lambda_{T+k}^{(i)}$  from  $p(\lambda_{T+k} | \boldsymbol{\theta}^{(i)})$ , for  $i = 1, \dots, N$  and  $k = 1, \dots, K$ , by using (6b) and (6c), respectively. Finally, using (6a), we sample  $y_{T+k}^{(i)}$  from  $p(y_{T+k} | \boldsymbol{\theta}^{(i)}, \lambda_{T+k}^{(i)}, h_{T+k}^{(i)})$ , for  $i = 1, \dots, N$  and  $k = 1, \dots, K$ .

To check the performance of the volatility forecasts, we use an extra dataset as validation period to perform  $m$  one-step-ahead forecasts. By the moving window approach, we use the

first  $T$  observations in the initial period to estimate the model and to forecast the  $(T + 1)$ -th observation; the sample is then rolled forward by one observation so that the second to the  $(T + 1)$ -th observations are used to forecast the  $(T + 2)$ -th observation. This process is repeated until the end of the sample, the  $(T + m)$ -th observation. To evaluate the performance of the model on VaR prediction, the likelihood ratio test introduced in Kupiec (1995) is used to test that the null hypothesis that the expected proportion of the number of “beyond VaR” or “violation” during the test periods is equal to  $\alpha$ . The violation is formulated by  $I_t(\alpha) = I[y_{T+1} < \widehat{\text{VaR}}_t(\alpha)]$  for the left tail and  $I_t(\alpha) = I[y_t > \widehat{\text{VaR}}_t(\alpha)]$  for the right tail, where  $I[\cdot]$  is an indicator function and  $\widehat{\text{VaR}}_t(\alpha)$  is the estimated VaR at level  $\alpha$ , which can be obtained by simulation using the  $k$ -step ahead densities described below (See Chen et al., 2008; Fan and Wang, 2013, for a detailed review). Let  $x_\alpha$  be the number of violations, that is,  $x_\alpha = \sum_{t=T+1}^{T+m} I_t(\alpha)$  and  $\hat{\alpha} = x_\alpha/m$ . The unconditional test of Kupiec (1995) is a likelihood ratio test with the  $\chi_1^2$ -distributed test statistic defined as

$$LRuc = 2\{\log[\hat{\alpha}^{x_\alpha}(1 - \hat{\alpha})^{m-x_\alpha}] - \log[\alpha^{x_\alpha}(1 - \alpha)^{m-x_\alpha}]\}. \quad (10)$$

The Expected Shortfall (ES) is formally defined via  $ES_t(\alpha) = E[y_t | y_t < VaR_t(\alpha)]$  for the left tail and  $ES_t(\alpha) = E[y_t | y_t > VaR_t(\alpha)]$  for the right tail. Following Aas and Haff (2006) and Nakajima (2013), we compute the measure developed by Embrechts et al. (2004) for evaluating the performance of the predicted ES, denoted by  $\widehat{ES}_t(\alpha)$ . We define  $\delta_t(\alpha) = y_t - \widehat{ES}_t(\alpha)$  as an excess of return. Let  $\delta_\alpha$  be the  $\alpha$ th quantile  $\{\delta_t(\alpha)\}_{t=T+1}^{T+m}$ . Next, define  $S_t(\alpha) = I[\delta_t(\alpha) < \delta_\alpha]$  for the left tail and  $S_t(\alpha) = I[\delta_t(\alpha) > \delta_\alpha]$  for the right tail. Write  $s_\alpha = \sum_{t=T+1}^{T+m} S_t(\alpha)$ . The measure of Embrechts et al. (2004) is given by  $D(\alpha) = \frac{1}{2}(|D_1(\alpha)| + |D_2(\alpha)|)$ , where

$$D_1(\alpha) = \frac{1}{x_\alpha} \sum_{I_t(\alpha)=1} \delta_t(\alpha), \quad (11)$$

$$D_2(\alpha) = \frac{1}{s_\alpha} \sum_{S_t(\alpha)=1} \delta_t(\alpha). \quad (12)$$

As discussed in Aas and Haff (2006) and Nakajima (2013),  $D_1(\alpha)$  is the standard back-testing measure for expected shortfall estimates. Its weakness is that it depends strongly on the VaR estimates without adequately reflecting the correctness of these values.  $D_2(\alpha)$  is computed to correct it because  $D_2(\alpha)$  measures an average difference between the return and the estimated ES for the  $\alpha$ -level tail of that difference from all test periods. A smaller  $D(\alpha)$  implies more precise prediction of ES.

#### 4. Bayesian Model Comparison

To compare the goodness of the estimated models with different members of the SMN distributions, we calculate the Watanabe-Akaike information criterion (WAIC, Watanabe, 2010, 2013). The WAIC is defined as

$$\text{WAIC} = -2(\text{lppd} - p_{\text{WAIC}}), \quad (13)$$

where lppd is the log pointwise predictive density defined by

$$\text{lppd} = \sum_{t=1}^T \log \int p(y_t | \boldsymbol{\theta}, \mathbf{h}_{1:T}) p(\mathbf{h}_{1:T} | \mathbf{y}_{1:T}) d\boldsymbol{\theta} d\mathbf{h}_{1:T} \quad (14)$$

and

$$p_{\text{WAIC}} = 2 \sum_{t=1}^T \left\{ \log E_{\boldsymbol{\theta}, \mathbf{h}_{1:T} | \mathbf{y}_{1:T}} [p(y_t | \boldsymbol{\theta}, \mathbf{h}_{1:t})] - E_{\boldsymbol{\theta}, \mathbf{h}_{1:T} | \mathbf{y}_{1:T}} [\log p(y_t | \boldsymbol{\theta}, \mathbf{h}_{1:t})] \right\}. \quad (15)$$

We can compute both the lppd and the  $p_{\text{WAIC}}$  using the MCMC output. We label them as  $\boldsymbol{\theta}^{(i)}$  and  $\mathbf{h}_{1:T}^{(i)}$ ,  $i = 1, \dots, I$ . Approximated versions obtained for computation are given by

$$\begin{aligned} \hat{\text{lppd}} &= \sum_{t=1}^T \log \left( \frac{1}{I} \sum_{i=1}^I p(y_t | \boldsymbol{\theta}_t) \right) \\ p_{\text{WAIC}} &= 2 \sum_{t=1}^T \left( \log \left( \frac{1}{I} \sum_{i=1}^I p(y_t | \boldsymbol{\theta}_t) \right) - \frac{1}{I} \sum_{i=1}^I \log p(y_t | \Psi, \boldsymbol{\theta}_t) \right). \end{aligned}$$

The minimum value of the WAIC gives the best fit.

## 5. Empirical Application

We consider the daily closing prices for three Latin American indexes: BOVESPA (Brazil), Merval (Argentina) and MXX (Mexico)<sup>1</sup>. We also use the daily returns of S&P 500 stock market index in order to compare the results with Latin American stock markets. The datasets were obtained from the Yahoo finance web site available to download at <http://finance.yahoo.com>. The period of analysis is from January 5, 1998, until December 30, 2016. Stock returns are computed as  $y_t = 100 \times (\log P_t - \log P_{t-1})$ , where  $P_t$  is the (adjusted) closing price on day  $t$ . The sample size differs between countries due to holidays and stock market non-trading days. Table 1 reports a summary of descriptive statistics for the series of returns.

According with Table 1, Merval and SP500 returns are negatively skewed and the BOVESPA and MXX positively skewed. Regarding kurtosis, all the indexes considered here show kurtosis greater than three, confirming a well-known stylized fact for return series, namely the departure from normality. We analyze the datasets with the aim of providing robust inference by using the SMN class of distributions. In our analysis, we compare the THSV-N, THSV-T, THSV-S and THSV-VG models for each one of the series described above.

**Table 1**

Summary statistics for the BOVESPA, Merval, MXX and SP500 returns

Return	size	mean	SD	Minimum	Maximum	Skewness	Kurtosis	Return $\geq 0$
BOVESPA	4699	0.0369	2.0267	-17.2082	28.8325	0.5300	16.7908	2422
Merval	4652	0.0700	2.2123	-14.2986	16.1165	-0.2089	7.3432	2451
MXX	4761	0.0456	1.4281	-10.3410	12.1536	0.1447	8.7346	2522
SP500	4779	0.0174	1.2416	-9.4695	10.9572	-0.2079	10.2474	2537

<sup>1</sup>We do not consider the peruvian market due to difficulty to find stocks with large daily quotes series.



**Table 2**

Estimation results for the BOVESPA returns.

Parameter	THSV-N	THSV-T	THSV-S	THSV-VG
	0.0322	0.0340	0.0384	0.0326
$\mu_0$	(-0.0664,0.1334)	(-0.0688,0.1361)	(-0.0639,0.1393)	(-0.0651,0.1302)
	0.52	-1.02	-0.27	1.23
	0.0480	0.0403	0.0337	0.0402
$\mu_1$	(-0.0483,0.1405)	(-0.0522,0.1383)	(-0.0613,0.1272)	(-0.0567,0.1321)
	-0.04	0.35	-1.54	1.91
	-0.0174	-0.0200	-0.0213	-0.0201
$\beta_0$	(-0.0845,0.0487)	(-0.0844,0.0460)	(-0.0811,0.0461)	(-0.0848,0.0445)
	0.59	-1.91	-0.62	0.99
	0.0351	0.0371	0.0399	0.0375
$\beta_1$	(-0.0291,0.0977)	(-0.0245, 0.0986)	(-0.0209,0.1007)	(-0.0239,0.0998)
	1.14	-0.27	1.80	1.53
	0.0221	0.0171	0.0068	0.0196
$\alpha_0$	(0.0081,0.0378)	(0.0053,0.0308)	(0.0001,0.0014)	(0.0066,0.0341)
	0.37	-0.80	-0.17	-1.45
	0.0209	0.0160	0.0042	0.0189
$\alpha_1$	(0.0080,0.0362)	(0.0054,0.0281)	(-0.0020,0.0106)	(0.0065,0.0324)
	1.44	-0.68	1.70	-1.68
	0.9818	0.9851	0.9884	0.9841
$\phi_0$	(0.9689,0.9929)	(0.9737,0.9949)	(0.9793,0.9965)	(0.9726,0.9945)
	-1.52	0.84	0.18	1.65
	0.9761	0.9804	0.9852	0.9788
$\phi_1$	(0.9619,0.9874)	(0.9692,0.9904)	(0.9757, 0.9939)	(0.9670,0.9897)
	-1.70	0.49	-1.51	1.45
	0.0226	0.0175	0.0125	0.0191
$\sigma_0^2$	(0.0165, 0.0332)	(0.0117,0.0246)	(0.0089,0.0176)	(0.0133,0.0269)
	1.58	-0.44	0.20	-1.28
	0.0223	0.0173	0.0125	0.0189
$\sigma_1^2$	(0.0164,0.0328)	(0.0117, 0.0246)	(0.0087,0.0174)	(0.0132,0.0264)
	1.28	-0.51	0.18	-1.31
	–	24.2300	1.7587	27.9404
$\nu$	–	(16.9300,35.8700)	(1.6370,1.8480)	(18.9400,37.7400)
	–	-0.40	-0.30	-1.16

Note: First row: Posterior mean. Second row: Posterior 95% credible interval in parentheses. Third row: CD statistics

In all posterior calculations, we simulate the  $h_t^i$ 's using the mixture sampler described on Section 3. We set the priors distributions of the common parameters as  $\boldsymbol{\psi}_i \sim \mathcal{N}_{2[|\beta_i|<1]}(\bar{\boldsymbol{\psi}}_i^*, \mathbf{S}_{\boldsymbol{\psi}_i})$ ,  $\boldsymbol{\varphi}_i \sim \mathcal{N}_{2[|\phi_i|<1]}(\bar{\boldsymbol{\varphi}}_i^*, \mathbf{S}_{\boldsymbol{\varphi}_i})$ ,  $\sigma_i^2 \sim \mathcal{IG}(5, 0.5)$  for  $i = 0, 1$ , where  $\bar{\boldsymbol{\psi}}_i^* = (0, 0)'$ ,  $\mathbf{S}_{\boldsymbol{\varphi}_i} = \text{diag}(100, 100)$ ,  $\bar{\boldsymbol{\varphi}}_i^* = (0, 0.98)'$ ,  $\mathbf{S}_{\boldsymbol{\psi}_i} = \text{diag}(100, 100)$ .  $\mathcal{N}_{2[\cdot]}(\cdot, \cdot)$  and  $\mathcal{IG}(\cdot, \cdot)$  denote the truncated bivariate normal and inverse gamma distributions, respectively. For the shape parameters, we assume  $\nu \sim \mathcal{G}(2, 0.10)$  in the THSV-T (Juárez and Steel, 2004),  $\nu \sim \mathcal{G}(0.08, 0.04)$  for the THSV-S and THSV-VG respectively.

For all the models and datasets we considered, we generated 60000 MCMC iterations. In all the cases, the first 20000 draws were discarded as a “burn-in” period. In order to reduce the autocorrelation between successive values of the simulated chain, only every 20th values of the chain were stored. With the resulting 4000 values, we calculated the posterior means, the 95%

**Table 3**

Estimation results for the MERVAL returns.

Parameter	THSV-N	THSV-T	THSV-S	THSV-VG
$\mu_0$	0.0189	0.0059	0.0006	0.0096
	(-0.0833,0.1223)	(-0.0939,0.1049)	(-0.0995,0.1005)	(-0.0869,0.1078)
	-1.86	-0.16	-0.50	-0.73
$\mu_1$	0.1711	0.1716	0.1700	0.1698
	(0.0767,0.2656)	(0.0766,0.2642)	(0.0722,0.2646)	(0.0754,0.2624)
	-1.06	0.90	0.48	-0.02
$\beta_0$	-0.0033	-0.0245	-0.0323	-0.0253
	(-0.0675,0.0613)	(-0.0847,0.0351)	(-0.0923,0.0293)	(-0.0852,0.0350)
	-1.47	0.50	0.05	-0.10
$\beta_1$	0.0372	0.0291	0.0288	0.0273
	(-0.0234,0.0980)	(-0.0272, 0.0859)	(-0.0265,0.0861)	(-0.0289,0.0836)
	-0.31	0.55	-1.63	0.81
$\alpha_0$	0.0606	0.0343	0.0175	0.0387
	(0.0336,0.0934)	(0.0147,0.0575)	(0.0049,0.0323)	(0.0176,0.0623)
	-1.74	-1.75	0.67	1.85
$\alpha_1$	0.0572	0.0317	0.0133	0.0347
	(0.03201,0.0881)	(0.0135,0.0525)	(0.0023,0.0252)	(0.0191,0.0578)
	-1.62	1.72	1.17	-1.42
$\phi_0$	0.9539	0.9700	0.9739	0.9718
	(0.9292,0.9735)	(0.9521,0.9861)	(0.9563,0.9882)	(0.9553,0.9861)
	1.81	-1.38	-0.17	-1.42
$\phi_1$	0.9451	0.9644	0.9688	0.9665
	(0.9187,0.9651)	(0.9463,0.9807)	(0.9510, 0.9826)	(0.9509,0.9803)
	0.96	-1.91	-0.41	0.74
$\sigma_0^2$	0.0752	0.04146	0.0357	0.0384
	(0.0518, 0.1094)	(0.0247,0.0609)	(0.0242,0.0572)	(0.0268,0.0526)
	-1.51	1.85	0.28	1.56
$\sigma_1^2$	0.0721	0.0413	0.0355	0.0382
	(0.053,0.1076)	(0.0247, 0.0607)	(0.0241,0.0572)	(0.0266,0.0518)
	-1.62	1.82	0.29	1.47
$\nu$	–	10.1110	1.7690	6.5434
	–	(7.6580,13.5880)	(1.6650,1.8520)	(5.8740,8.9290)
	–	1.14	1.00	1.21

Note: First row: Posterior mean. Second row: Posterior 95% credible interval in parentheses. Third row: CD statistics

credibility intervals and the convergence diagnostic (CD) statistics. If the sequence of the recorded MCMC output is stationary, it converges in distribution to the standar normal. According to the CD, the null hypothesis that the sequence of 4000 draws is stationary is accepted at 5% level, i.e.  $CD \in (-1.96, 1.96)$ , for all the parameters in all the models and series of returns considered here (see Tables 2, 3, 4 and 5 for details.)

From Tables 2, 3, 4 and 5, we found for all the models and datasets, that posterior means and 95% credibility intervals of  $\phi_0$  and  $\phi_1$  were very close to the unity, which is consistent with the existing evidence of great persistence in the log-volatility process. Additionally, the posterior means of  $\phi_0$  and  $\phi_1$  under the THSV-N model were slightly smaller than those under the other three models. As expected, the posterior means of  $\sigma_0^2$  and  $\sigma_1^2$  under the THSV-N were higher than those under the THSV-T, THSV-S and the THSV-VG models, indicating that the log-

**Table 4**

Estimation results for the MXX returns.

Parameter	THSV-N	THSV-T	THSV-S	THSV-VG
$\mu_0$	0.0670	0.0676	0.0702	0.0692
	(0.0560,0.1255)	(0.0104,0.1270)	(0.0126,0.1266)	(0.0104,0.1289)
	1.21	0.71	1.27	-1.06
$\mu_1$	0.0623	0.0621	0.0624	0.0639
	(0.0059,0.1177)	(0.0089,0.1171)	(0.0092,0.1175)	(0.0106,0.1183)
	-0.36	0.29	0.23	-0.32
$\beta_0$	0.0671	0.0616	0.0601	0.0636
	(0.0044,0.1293)	(0.0004,0.1218)	(0.0017,0.1188)	(0.0036,0.1246)
	0.88	-0.82	0.36	-0.88
$\beta_1$	0.0881	0.0831	0.0803	0.0801
	(0.0304,0.1472)	(0.0246, 0.1409)	(0.0246, 0.1348)	(0.0274,0.1328)
	0.66	-1.68	0.36	1.21
$\alpha_0$	0.0079	0.0047	-0.0013	0.0062
	(-0.0012,0.0181)	(-0.0027,0.0124)	(-0.0088,0.0060)	(-0.0017,0.0145)
	0.64	-1.71	1.92	0.69
$\alpha_1$	0.0002	-0.0017	-0.0087	0.0007
	(-0.0081,0.0084)	(-0.0087,0.0053)	(-0.0167,0.0012)	(-0.0065,0.0076)
	-0.22	-1.46	1.74	-0.32
$\phi_0$	0.9857	0.9891	0.9903	0.9892
	(0.9739,0.9961)	(0.9803,0.9973)	(0.9817,0.9981)	(0.9804,0.9973)
	-1.26	0.61	1.75	1.26
$\phi_1$	0.9792	0.9840	0.9859	0.9840
	(0.9673,0.9897)	(0.9746,0.9926)	(0.9770,0.9938)	(0.9747,0.9921)
	-1.70	1.70	1.42	1.07
$\sigma_0^2$	0.0296	0.0214	0.0185	0.0213
	(0.0185, 0.0451)	(0.0154,0.0292)	(0.0134,0.0259)	(0.0150,0.0291)
	1.84	-0.96	-1.92	-1.17
$\sigma_1^2$	0.0292	0.0212	0.0183	0.0211
	(0.0184,0.0442)	(0.0151, 0.0284)	(0.0132,0.0257)	(0.0150,0.0288)
	1.91	-0.99	-1.86	-1.13
$\nu$	–	14.9193	1.7572	11.3888
	–	(11.1400,19.5100)	(1.6450,1.8400)	(8.5870,14.2820)
	–	0.16	0.16	0.41

Note: First row: Posterior mean. Second row: Posterior 95% credible interval in parentheses. Third row: CD statistics

volatility process of the last three models is less variable than that of the THSV-N model. We found that the persistence coefficients for the Merval index are lower than the coefficients of the other indexes. Under the THSV-T, THSV-S and THSV-VG models, the magnitude of the tail-heaviness is measured by the  $\nu$  parameter. We found that the posterior means of  $\nu$  were 24.2300, 1.7587 and 27.9404 (BOVESPA), 10.1110, 1.7690 and 6.5434 (Merval), 14.9193, 1.7572 and 11.3888 (MXX), 25.2610, 1.7532 and 9.3880 (SP500), respectively. This results seem to indicate the presence of heavy tails.

To assess the goodness-of-fit of the estimated models, we calculate the WAIC criterion for the SV-N, SV-T, SV-S, SV-VG<sup>2</sup>, THSV-N, THSV-T, THSV-S and THSV-VG models, respectively.

<sup>2</sup>We fit the basic class of SV models with SMN distributions, but we do not report the results for parameter estimation.

**Table 5**

Estimation results for the SP500 returns

Parameter	THSV-N	THSV-T	THSV-S	THSV-VG
$\mu_0$	0.0770	0.0796	0.0825	0.0817
	(0.0264,0.1264)	(0.0299,0.1276)	(0.0336,0.1311)	(0.0339,0.1317)
	-0.96	-1.11	-0.23	0.08
$\mu_1$	0.0224	0.0206	0.0183	0.0179
	(-0.0237,0.0669)	(-0.0234,0.0657)	(-0.0261,0.0616)	(-0.0260,0.0606)
	-0.69	0.34	0.11	-0.67
$\beta_0$	-0.0619	-0.0638	-0.0680	-0.0672
	(-0.1208,-0.0006)	(-0.1216,-0.0067)	(-0.1256,-0.0101)	(-0.1250,-0.0067)
	0.37	-1.84	-0.59	0.43
$\beta_1$	0.0069	0.0089	0.0130	0.0125
	(-0.0540,0.0676)	(-0.0482, 0.0658)	(-0.0404,0.0668)	(-0.0421,0.0665)
	-1.16	-0.34	-1.30	-0.11
$\alpha_0$	0.0037	0.0020	-0.0062	0.0032
	(-0.0060,0.0135)	(-0.0066,0.0108)	(-0.0166,0.0668)	(-0.0051,0.0120)
	-0.31	-1.57	-1.19	-1.86
$\alpha_1$	-0.0085	0.0085	-0.0154	-0.0058
	(-0.0181,0.0007)	(-0.0174,0.0000)	(-0.0267,-0.0049)	(-0.0137,0.0020)
	1.24	-1.77	0.27	-0.20
$\phi_0$	0.9814	0.9840	0.9865	0.9861
	(0.9706,0.9917)	(0.9731,0.9938)	(0.9766,0.9959)	(0.9764,0.9950)
	1.73	1.90	-1.01	-0.84
$\phi_1$	0.9799	0.9827	0.9854	0.9848
	(0.9693,0.9892)	(0.9635,0.9885)	(0.9761, 0.9941)	(0.9756,0.9936)
	0.96	1.36	-1.18	-0.84
$\sigma_0^2$	0.0352	0.0299	0.0250	0.0256
	(0.0270, 0.0354)	(0.0205,0.0408)	(0.0177,0.0334)	(0.0188,0.0331)
	-0.86	-1.31	1.58	0.58
$\sigma_1^2$	0.0354	0.0300	0.0251	0.0257
	(0.0273,0.0447)	(0.0205, 0.0411)	(0.0177,0.0336)	(0.0187,0.0330)
	-0.81	-0.42	1.53	0.38
$\nu$	-	25.2610	1.7532	9.3880
	-	(17.2200,36.0200)	(1.6410,1.8400)	(7.0260,12.1850)
	-	-1.64	1.00	1.21

Note: First row: Posterior mean. Second row: Posterior 95% credible interval in parentheses. Third row: CD statistics

From Table 6 the WAIC indicates the THSV-S model gives the best in-sample fit among all the models considered here, for BOVESPA, Merval, MXX and SP500 returns, suggesting a sufficient departure from the underlying assumption of normality and the presence of asymmetric behavior in the returns.

In Figure 1, we plot the absolute returns (gray line), the posterior smoothed mean of  $e^{\frac{h_t}{2}}$  obtained from the MCMC output for the SV-N model (dotted black line) and the THSV-S model (dotted red line), which are the worst and the best model fit according to the WAIC criterion, for BOVESPA, Merval, MXX and SP500 returns. It is important to note that returns from Latin American indexes show a volatile period at the beginning of 1998, which is caused by Brazilian crisis. All the indexes showed higher volatility between 2007 and 2008 as a consequence of the subprime crisis. MXX and SP500 showed a similar behaviour. From a practical point

**Table 6**

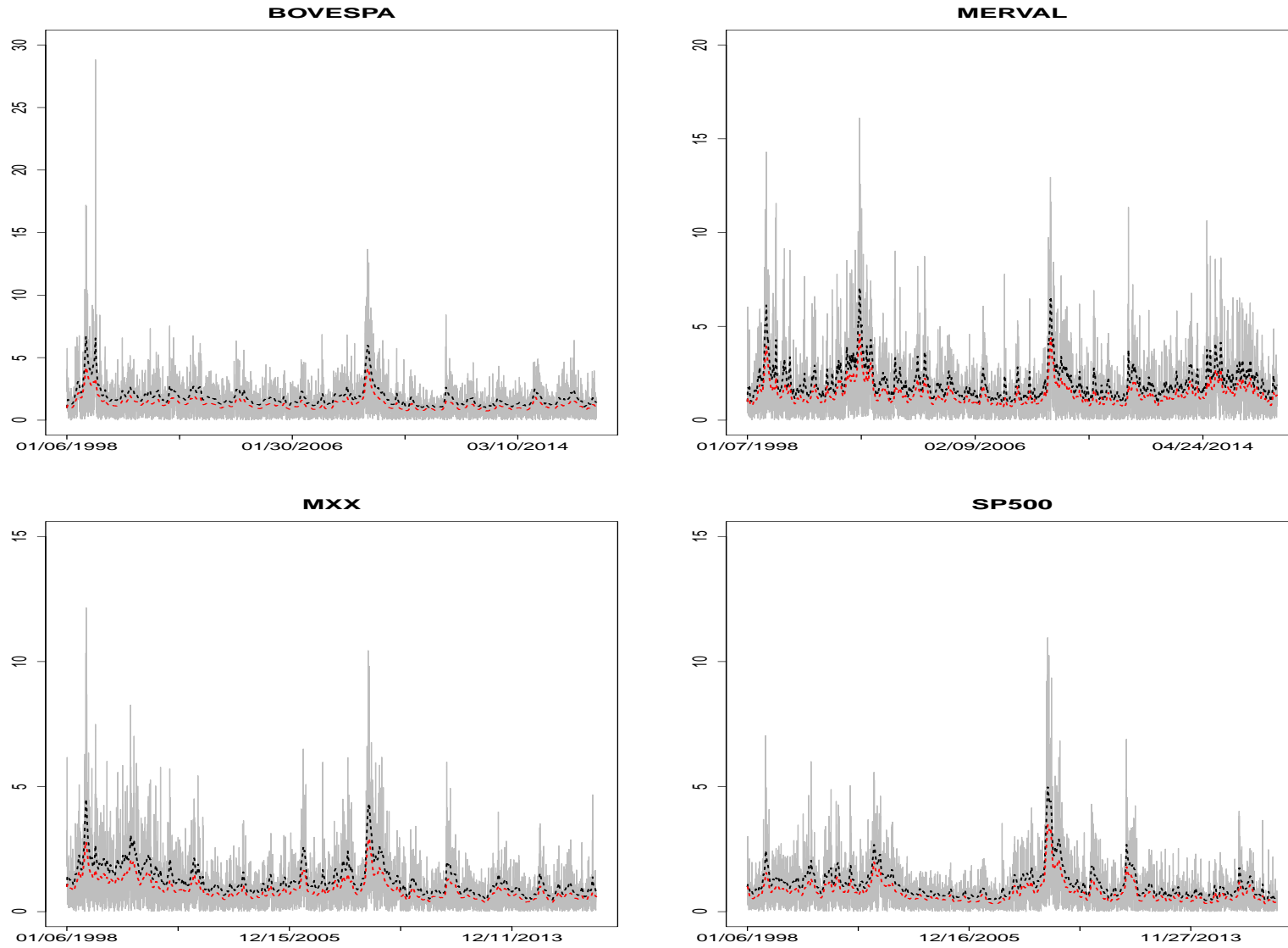
WAIC criterion for BOVESPA, Merval, MXX and SP500 returns

BOVESPA			Merval		
Model	Value	Ranking	Model	Value	Ranking
SV-N	18367.3	7	SV-N	19006.8	8
SV-T	18297.7	6	SV-T	18786.2	6
SV-S	18236.6	2	SV-S	18775.7	4
SV-VG	18261.3	3	SV-VG	18779.7	5
THSV-N	18438.7	8	THSV-N	18939.5	7
THSV-T	18282.3	4	THSV-T	18732.3	3
<b>THSV-S</b>	<b>18220.3</b>	<b>1</b>	<b>THSV-S</b>	<b>18715.5</b>	<b>1</b>
THSV-VG	18286.0	5	THSV-VG	18725.8	2
MXX			SP500		
Model	Value	Ranking	Model	Value	Ranking
SV-N	14946.0	8	SV-N	15950.3	8
SV-T	14769.6	6	SV-T	15880.4	6
SV-S	14716.3	3	SV-S	15808.2	2
SV-VG	14723.7	5	SV-VG	15818.7	3
THSV-N	14872.8	7	THSV-N	15926.3	7
THSV-T	14721.3	4	THSV-T	15843.6	5
<b>THSV-S</b>	<b>14652.3</b>	<b>1</b>	<b>THSV-S</b>	<b>15780.8</b>	<b>1</b>
THSV-VG	14695.8	2	THSV-VG	15841.3	4

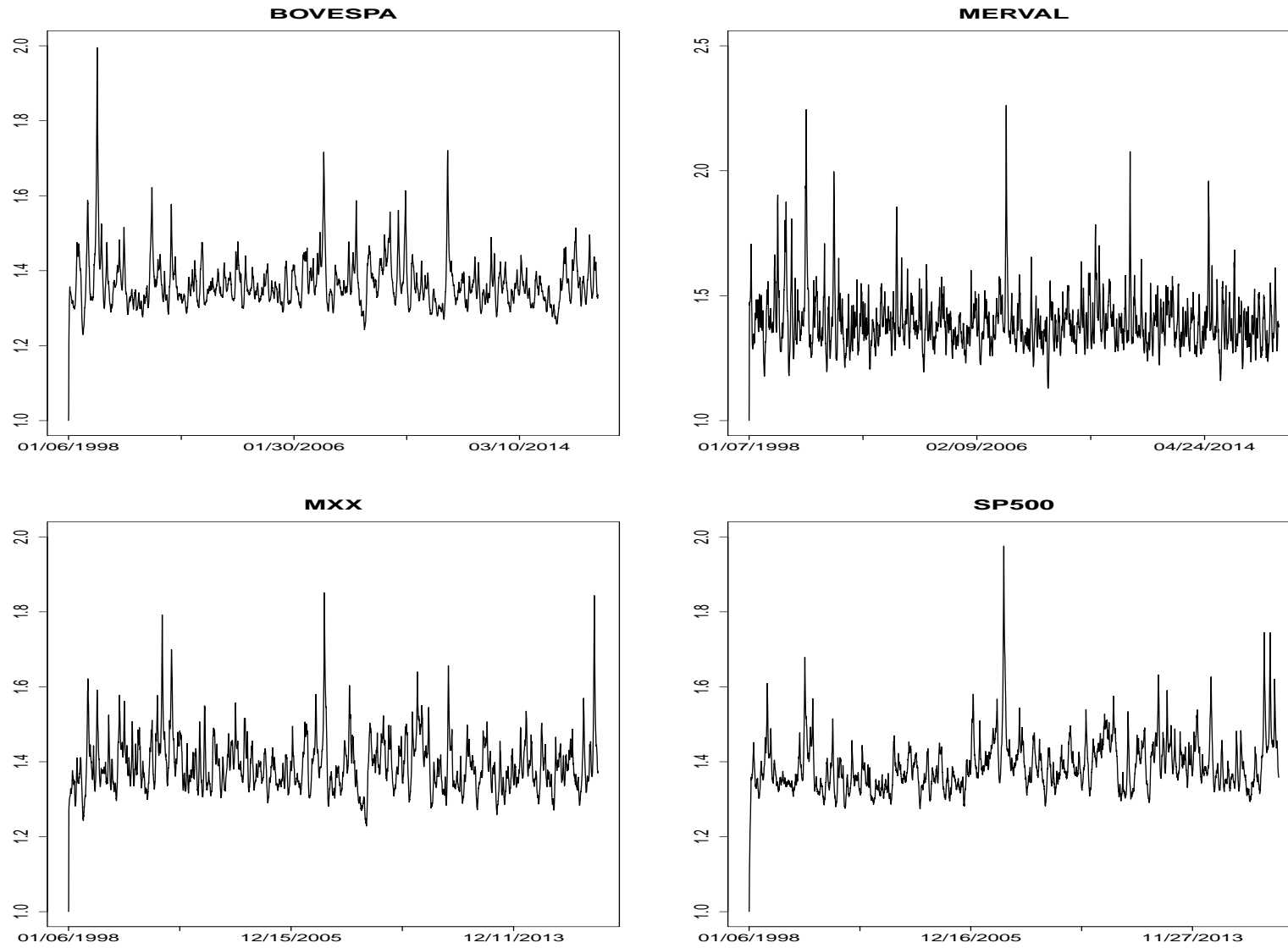
of view, we are mainly interested in whether we find a significant difference between the two series. Therefore, in the bottom panel of [Figure 2](#), we plot the smoothed mean of the ratio of  $e^{\frac{h_t}{2}}$  obtained from the SV-N and THSV-S models for BOVESPA, Merval, MXX and SP500 returns. Some extreme returns make the differences clear. This can have a substantial impact, for instance, in the evaluation of derivative instruments and several strategic or tactical asset allocation topics.

The magnitudes of the mixing parameter  $\lambda_t$  are associated with extremeness of the corresponding observations. In the Bayesian paradigm, the posterior mean of the mixing parameter can be used to identify a possible outlier (see [Rosa et al., 2003](#), for instance). The SV and THSV models with Student-t, slash and variance gamma distributions can accommodate an outlier by inflating the variance component for that observation in the conditional distribution with smaller  $\lambda_t$  value. This fact is showed in [Figure 3](#) where we plot the posterior mean of the mixing variable  $\lambda_t$  for the THSV-S model (the best in-sample fit).

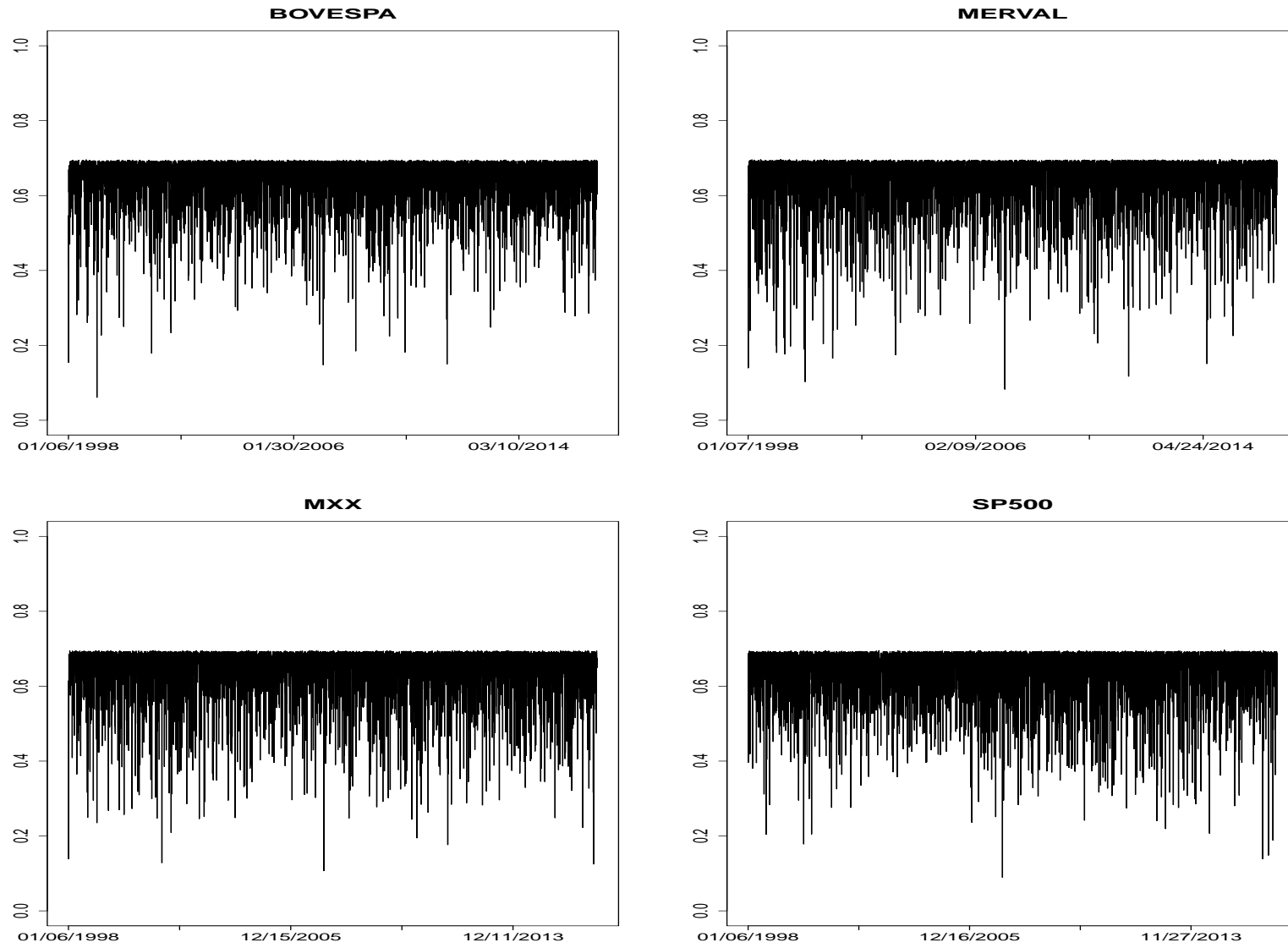
In order to examine the performance of VaR and ES forecast for the competing models, we use the data from January 3, 2017 to February 1, 2019 as validation period, giving  $m$  trading days ( $m = 516, 508, 525, 524$  trading days for BOVESPA, Merval MXX and SP500 returns, respectively). In the moving window approach, we use the first  $T$  observations in the period January 5, 1998 - December 30, 2016 to estimate the model and to forecast the  $(T + 1)$ -th observation; the sample is then rolled forward by one observation, so that the second to the  $(T + 1)$ -th observations are used to forecast the  $(T + 2)$ -th observation. This process is repeated until the end of the sample, i.e., the  $(T + m)$ -th observation.



**Figure 1.** BOVESPA, Merval, MXX, SP500 returns datasets: absolute returns (gray line) and the posterior smoothed mean of  $e^{\frac{h_t}{2}}$  obtained from the MCMC output, for the SV-N model (dotted black line) and THSV-S model (dotted red line), respectively.



**Figure 2.** BOVESPA, Merval, MXX, SP500 returns datasets: posterior smoothed mean of the ratio of  $e^{\frac{h_t}{2}}$  obtained from the MCMC output for the SV-N and THSV-S models.



**Figure 3.** Posterior smoothed mean of the mixture variable  $\lambda_t$  obtained from the MCMC output for the THSV-S model for BOVESPA, Merval, MXX and SP500 returns.



**Table 7**  
BOVESPA, Merval, MXX and SP500 return datasets.

	BOVESPA				Merval			
	Violation	$LR_{uc}$	D(0.05)	D(0.95)	Violation	$LR_{uc}$	D(0.05)	D(0.95)
	Rate (%)				Rate (%)			
RiskMetrics	0.0821	0.0021	0.3572	0.5512	0.0942	0.0001	0.5511	0.2543
SV-N	0.0310	0.0337	0.1974	0.1336	0.2717	0.0000	1.2976	1.3591
SV-T	0.0329	0.0549	0.1674	0.1320	0.0490	0.9174	0.3984	0.4007
SV-S	0.0271	0.0098	0.1418	0.1311	0.0334	0.0684	0.3142	1.0340
SV-VG	0.0349	0.0968	0.1465	0.1322	0.0490	0.9174	0.1580	0.2775
THSV-N	0.0329	0.0549	0.1739	0.1689	0.0511	0.9097	0.3249	0.2954
THSV-T	0.0349	0.0968	0.1461	0.1317	0.0531	0.7508	0.1947	0.4393
THSV-S	0.0349	0.0968	<b>0.1289</b>	<b>0.1311</b>	0.0511	0.9097	<b>0.0493</b>	<b>0.1150</b>
THSV-VG	0.0349	0.0968	0.1510	0.1320	0.0551	0.6036	0.1958	0.4387
	MXX				SP500			
	Violation	$LR_{uc}$	D(0.05)	D(0.95)	Violation	$LR_{uc}$	D(0.05)	D(0.95)
	Rate (%)				Rate (%)			
RiskMetrics	0.0630	0.1882	0.5511	0.2543	0.073	0.1022	0.5511	0.2543
SV-N	0.0495	0.9578	0.2122	0.4045	0.0611	0.2279	0.4398	0.0956
SV-T	0.0438	0.5058	0.1899	0.1619	0.0592	0.3097	0.3984	0.0669
SV-S	0.0343	0.0809	0.1203	0.4495	0.0553	0.5296	<b>0.1418</b>	0.0641
SV-VG	0.0457	0.6467	0.1834	0.1450	0.0591	0.3097	0.3946	0.0641
THSV-N	0.0514	0.8837	0.2411	0.1653	0.0649	0.1141	0.5033	0.0728
THSV-T	0.0533	0.7313	0.1827	0.1960	0.0649	0.1141	0.4629	0.0599
THSV-S	0.0476	0.7992	<b>0.0445</b>	<b>0.1221</b>	0.0553	0.5296	0.3289	<b>0.0456</b>
THSV-VG	0.0533	0.7313	0.1759	0.2209	0.0668	0.1141	0.4382	0.0881

Note: Violation rate in  $m$  one-step-ahead forecast, P-values of the unconditional coverage test,  $D(0.05)$  and  $D(0.95)$  measures, ( $m = 516, 508, 525, 524$  for the BOVESPA, Merval, MXX and SP500 returns, respectively).

We thus obtain  $m$  volatility forecasts, VaR at level 5%, and ES estimates with confidence levels of 5% and 95%. The competing models were: Riskmetrics, SV-N, SV-T, SV-S, SV-VG, THSV-N, THSV-T, THSV-S and THSV-VG.

The results of  $m$  one-step-ahead forecasts are presented in Table 7. For the BOVESPA returns, according to the unconditional coverage test, we reject the null hypothesis that the achieved violation rate is equal to 5% for the Riskmetrics, SV-N and SV-S models at 5% level. For the Merval returns, we reject the null hypothesis at 5% level for Riskmetrics and SV-N model. For MXX and SP500 returns, we accept the null hypothesis for all models. According to the violation rate the SV-S and THSV-S give better performance than the other competing models, it is the  $P$ -values of SV-S and THSV-S models are greater than the other competing models. When, we compare  $D(0.05)$  and  $D(0.95)$ , the THSV-S models show the smaller measures than the competing models for the BOVESPA, Merval and MXX returns. For the SP500 returns the SV-S show the smallest for the  $D(0.05)$  measure (see the values in bold in Table 7 for details).

In order to evaluate the prediction accuracy of the SV-N, SV-T, SV-S, SV-VG, THSV-N, THSV-T, THSV-S, THSV-VG models, we use the same moving window approach as in the

**Table 8**

MSPE for BOVESPA, Merval, MXX and SP500 returns

Series	SV-N	SV-T	SV-S	SV-VG	THSV-N	THSV-T	THSV-S	THSV-VG
BOVESPA	0.6378	0.6225	0.6035	0.6037	0.6227	0.6019	<b>0.6022</b>	0.6047
Merval	0.4625	0.4592	0.4478	0.4489	0.4485	0.4481	<b>0.4475</b>	0.4481
MXX	0.5181	0.4935	0.4931	0.4940	0.5024	0.4925	<b>0.4914</b>	0.4917
SP500	0.4378	0.4381	0.4374	0.4375	0.4372	0.4374	<b>0.4240</b>	0.4375

VaR estimation. We define the Mean Square Predictive Error (MSPE) as

$$MSPE^g = \sum_{t=T+1}^{T+m} \frac{1}{m} \sum_{i=1}^N \frac{1}{N} (y_t^{(i,g)} - y_t)^2$$

where  $y_t^{(i,g)}$  is obtained by simulation using the MCMC procedure described in Section 3.2 and  $g \in \{SV-N, SV-T, SV-S, SV-VG, THSV-N, THSV-T, THSV-S, THSV-VG\}$  denotes the model.

From Table 8, we find that the THSV-S model outperforms the other models using the MSPE. So, according to the MSPE, the THSV-S model gives the best out-of-sample fit for all series of returns analyzed here. It is also important to emphasize that, in general, we do not advocate the use of the THSV-S model in all situations but recommend using the model discussed here to assess the robustness of the conclusions, replacing the normal assumption with a more flexible model if this provides a more appropriate analysis.

## 6. Discussion

In this article, we have proposed the threshold stochastic volatility model with scale mixture of normal distributions (THSV-SMN) errors as an alternative to the normal (symmetric) assumption in the conditional distribution of the returns. The THSV-SMN class of models allows a parsimonious yet flexible treatment of both the skewness and the heavyness of the tails of the error distribution. Within the Bayesian framework, we have developed a fast and efficient MCMC sampling procedure to estimate all the parameters and latent quantities in our proposed THSV-SMN class of models. As a by-product of the MCMC algorithm, we were able to produce an estimate of the latent information process which can be used in financial modeling. The use of mixing variables,  $\lambda_{1:T}$  not only simplifies the full conditional distributions required for the Gibbs sampling algorithm, but also provides a mean for the outlier diagnostics. We applied our methods to the analysis of the BOVESPA, Merval, MXX and SP500 returns, which showed that the THSV-S model provides a better fit than the SV-N, SV-T, SV-S, SV-VG, THSV-N, THSV-T and THSV-VG terms of parameter estimates, interpretation and robustness aspects. We found that the THSV-S model outperforms the other models using the MSPE given the best out-of-sample fit and it can be used to VaR and ES forecast.

A potential interesting future research topic is the further investigation of the large observations by introducing jump components or correlation between perturbations of returns and volatilities.

## Appendix A

### Full Conditionals of $\psi_0, \psi_1, \varphi_0, \varphi_1, \sigma_{\eta,0}^2$ and $\sigma_{\eta,1}^2$

Let  $r'_t = (1, y_{t-1})$  and  $H'_{t+1} = (1, h_t)$ . We assume that  $\psi_i \sim \mathcal{N}_{2[|\beta_i|<1]}(\bar{\psi}_i^*, \Sigma_{\psi_i})$ ,  $\varphi_i \sim \mathcal{N}_{2[|\phi_i|<1]}(\bar{\varphi}_i^*, \Sigma_{\varphi_i})$ ,  $\sigma_i^2 \sim \mathcal{IG}(5, 0.5)$ , for  $i = 0, 1$ . Then the full conditionals are given by

$$\psi_i \sim \mathcal{N}_{2[|\beta_i|<1]}(\bar{\psi}_i^*, \Sigma_{\psi_i}), \quad (\text{A.1})$$

$$\varphi_i \sim \mathcal{N}_{2[|\phi_i|<1]}(\bar{\varphi}_i^*, \Sigma_{\varphi_i}), \quad (\text{A.2})$$

$$\sigma_i^2 \sim \mathcal{IG}\left(\frac{n_i^*}{2}, \frac{S_i^*}{2}\right), \quad (\text{A.3})$$

where:

$$\begin{aligned} \bar{\psi}_i^* &= \Sigma_{\psi_i} \left[ S_{\psi_i}^{-1} \bar{\psi}_i + \sum_{\{t:s_t=i\}} \lambda_t e^{-h_t} r_t y_t \right], \quad \Sigma_{\psi_i} = \left[ S_{\psi_i}^{-1} + \sum_{\{t:s_t=i\}} \lambda_t e^{-h_t} r_t r_t' \right]^{-1}, \\ \bar{\varphi}_i^* &= \Sigma_{\varphi_i} \left[ S_{\varphi_i}^{-1} \bar{\varphi}_i + \sum_{\{t:s_{t+1}=i\}} \frac{h_{t+1} H_{t+1}}{\sigma_{\eta,i}^2} \right], \quad \Sigma_{\varphi_i} = \left[ S_{\varphi_i}^{-1} + \sum_{\{t:s_{t+1}=i\}} \frac{H_{t+1} H_{t+1}'}{\sigma_{\eta,i}^2} \right]^{-1}, \\ n_i^* &= n_i + \sum_{\{t:s_{t+1}=i\}} \mathbb{I}(s_{t+1} = i), \quad S_i^* = S_i + \sum_{\{t:s_{t+1}=i\}} (h_{t+1} - H'_{t+1} \varphi_i)^2 \text{ for } i = 0, 1, \end{aligned}$$

and  $\mathcal{N}_{2[\cdot]}(\cdot, \cdot)$ ,  $\mathcal{IG}(\cdot, \cdot)$  denote the bivariate truncated normal and inverse gamma distributions, respectively.

### Full Conditional of $\lambda_t$ and $\nu$

#### • THSV-T Case

As  $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$ , then  $\lambda_t \mid y_t, y_{t-1}, h_t, \theta \sim \mathcal{G}(\frac{\nu+1}{2}, \frac{[y_t - r'_t \psi_{s_t}]^2 e^{-h_t + \nu}}{2})$ . As in Juárez and Steel (2004), we take a  $\mathcal{G}(2, 1/10)$  prior for  $\nu$  with mass covering a large range of relevant values (prior mean 20 and variance 200) for the THSV-T model. Then the full conditional of  $\nu$  is given by

$$p(\nu \mid \lambda_{1:T}) \propto \frac{\left[ \frac{\nu}{100} e^{-\frac{\nu}{10}} \right] \left[ \frac{\nu}{2} \right]^{\frac{T\nu}{2}} e^{-\frac{\nu}{2} [\sum_{t=1}^T (\lambda_t - \log \lambda_t)]}}{[\Gamma(\frac{\nu}{2})]^T} \mathbb{I}_{[2 < \nu \leq 40]}, \quad (\text{A.4})$$

where  $\mathbb{I}_{[\cdot]}$  denotes an indicator function. As (A.4) does not have closed form, we sample  $\nu$  using and adaptive Metropolis-Hastings algorithm based on Shaby and Wells (2011).

#### • THSV-S Case

Using the fact that  $\lambda_t \sim \mathcal{Be}(\nu, 1)$ , then  $\lambda_t \mid y_t, y_{t-1}, h_t, \theta \sim \mathcal{G}_{(0 < \lambda_t < 1)}(\nu + \frac{1}{2}, \frac{1}{2} [y_t - r'_t \psi_{s_t}]^2 e^{-h_t})$ , the right truncated gamma distribution. Assuming that a prior distribution of  $\nu \sim \mathcal{G}(a_\nu, b_\nu)$ , the full conditional distribution of  $\nu$  is  $\mathcal{G}_{(\nu > 1)}(T + a_\nu, b_\nu - \sum_{t=1}^T \log \lambda_t)$ , i.e. the left truncated gamma distribution.

### • THSV-VG Case

As  $\lambda_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$ , then  $\lambda_t \mid y_t, y_{t-1}, h_t, \boldsymbol{\theta} \sim \mathcal{GIG}(\frac{-\nu+1}{2}, \frac{[y_t - r'_t \boldsymbol{\psi}_{s_t}]^2 e^{-h_t} + 1}{2}, \nu)$ , where  $\mathcal{GIG}(\cdot, \cdot, \cdot)$  denotes the generalized inverse Gaussian distribution. In the THSV-VG, we assume the prior  $\nu \sim \mathcal{G}(a_0, b_0)$ . Then the full conditional is given

$$p(\nu \mid \boldsymbol{\lambda}_{1:T}) \propto \frac{\nu^{a_0-1} e^{-b_0 \nu} e^{-\frac{\nu}{2} [\sum_{t=1}^T (\log \lambda_t) + \frac{1}{\lambda_t}]} }{[\Gamma(\frac{\nu}{2})]^T} \mathbb{I}_{2 < \nu \leq 40}. \quad (\text{A.5})$$

As (A.5) does not have closed form, we sample  $\nu$  using an adaptive Metropolis-Hastings algorithm based on Shaby and Wells (2011).

## Appendix B

First we define the matrix

$$A_t = \begin{bmatrix} A_{11,t} & 0 \\ 0 & A_{22,t} \end{bmatrix},$$

where  $A_{11,t} = 1/\varpi_{k_t}^2$  and  $A_{22,t} = 1/[(1-s_t)\sigma_0^2 + s_t\sigma_1^2]$ . We define  $\Omega$  as follows

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & 0 & \dots & 0 \\ \Omega_{12} & \Omega_{22} & \Omega_{23} & \ddots & \vdots \\ 0 & \Omega_{23} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \Omega_{T-1,T-1} & \Omega_{T-1,T} \\ 0 & \dots & 0 & \Omega_{T-1,T} & \Omega_{T,T} \end{bmatrix}, \quad (\text{B.1})$$

where

$$\begin{aligned} \Omega_{11} &= A_{11,1} + [((1-s_1)\phi_0 + s_1\phi_1) \times A_{22,1} + (1/P_1)] \\ \Omega_{tt} &= A_{11,t} + [((1-s_t)\phi_0 + s_t\phi_1) \times A_{22,t} + A_{22,t-1}] \\ \Omega_{TT} &= (1/\varpi_{k_T}^2) + A_{22,T-1} \\ \Omega_{t,t+1} &= -[(1-s_t)\phi_0 + s_t\phi_1] A_{22,t} \end{aligned} \quad (\text{B.2})$$

and

$$\begin{aligned} c_1 &= A_{11,1} \times (z_1 + \log(\lambda_1) - \vartheta_{k_1}) - [((1-s_1)\phi_0 + s_1\phi_1) \times A_{22,1} \times [((1-s_1)\alpha_0 + s_1\alpha_1)] + (a_1/P_1)] \\ c_t &= A_{11,t} \times (z_t + \log(\lambda_t) - \vartheta_{k_t}) - [((1-s_t)\phi_0 + s_t\phi_1) \times A_{22,t} \times [((1-s_t)\alpha_0 + s_t\alpha_1)] \\ &\quad + A_{22,t-1} \times [((1-s_{t-1})\alpha_0 + s_{t-1}\alpha_1)] \\ c_T &= (1/\varpi_{k_T}^2) [z_T + \log(\lambda_T) - \vartheta_{k_T}] + A_{22,T-1} \times [((1-s_{T-1})\alpha_0 + s_{T-1}\alpha_1)] \end{aligned} \quad (\text{B.3})$$

## References

- Aas, K., and Haff, I. (2006). The Generalized Hyperbolic Skew Student's t-Distributions. *Journal of Financial Econometrics* 4, 275-309.
- Abanto-Valle, C., Bandyopadhyay, D., Lachos, V., and Enriquez, I. (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis* 54, 2883-2898.
- Abanto-Valle, C., Wang, C., Wang, X., Wang, F., and Chen, M. (2014). Bayesian Inference for Stochastic Volatility Models Using the Generalized Skew-t Distribution with Applications to the Shenzhen Stock Exchange Returns. *Statistics and Its Interface* 7, 487-502.
- Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the Business and Economical Statistics Section*. American Statistical Association, 177-181.
- Chen, C., Liu, F., and So, M. (2008). Heavy-tailed-distributed threshold stochastic volatility models in financial time series. *Australian & New Zealand Journal of Statistics* 50, 29-51.
- Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108(2), 281-316.
- Choy, S., and Chan, J. (2008). Scale mixtures distributions in statistical modelling. *Australian & New Zealand Journal of Statistics* 50, 135-146.
- Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage, and interest rates effects. *Journal of Financial Economics* 108, 281-316.
- Embrechts, P., Kaufmann, R., and Patie, P. (2004). Strategic Long-term Financial Risks: Single Risk Factors. *Computational Optimization and Applications* 22, 61-90.
- Fama, E. (1965). Portfolio analysis in a stable paretian market. *Management Science* 11, 404-419.
- Fan, T., and Wang, Y. (2013). An empirical Bayesian forecast in the threshold stochastic volatility models. *Journal of Statistical Computation and Simulation* 83, 486-500.
- Jacquier, E., Polson, N., and Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122(1), 185-212.
- Juárez, M., and Steel, M. (2004). Strategic Long-term Financial Risks: Single Risk Factors. *Computational Optimization and Applications* 22, 61-90.
- Kim, S., Shepard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361-393.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3, 73-84.
- Lange, K., and Sinsheimer, J. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* 2, 175-198.
- Liesenfeld, R., and Jung, C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* 15(2), 137-160.
- Little, R. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics* 37, 23-38.
- Madan, D., and Seneta, E. (1990). The variance gamma (V.G) model for share market return. *Journal Business* 63, 511-524.

- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* 36, 314-419.
- McCausland, W., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state space models: A computational efficiency analysis. *Computational Statistics and Data Analysis* 55, 199-212.
- Nakajima, J. (2013). Stochastic volatility model with regime-switching skewness in heavy-tailed errors for exchange rate returns. *Studies Nonlinear Dynamics Economics* 175(5), 499- 520.
- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* 36, 425-449.
- Rosa, G., Padovani, C., and Gianola, D. (2003). Robust linear mixed models with Normal/Independent distributions and Bayesian MCMC implementation. *Biometrical Journal* 45, 573-590.
- Shaby, B., and Wells, M. (2011). Exploring an adaptive Metropolis algorithm. Technical Report, Duke University Department of Stastical Science. TR 2011-14.
- So, M., Li, W., and Lam, K. (2002). A threshold stochastic volatility model. *Journal of Forecasting* 21, 473-500.
- Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. In O. Anderson (Ed.). *Time Series Analysis: Theory and Practice, Vol 1,*. Amsterdam: North-Holland, 203-226.
- Taylor, S. (1994). Modelling stochastic volatility: a review and comparative study. *Mathematical Finance* 4, 183-204.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 14, 867-897.
- Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research* 14, 867-897.