Bayesian Analysis of Heavy-tailed Stochastic Volatility in Mean model using Scale Mixtures of Normal distributions

C. A. Abanto-Valle†1, Helio S. Migon† and V. H. Lachos‡

†Department of Statistics, Federal University of Rio de Janeiro, Brazil

‡Department of Statistics, Campinas State University, Brazil

Abstract

The stochastic volatility in mean (SVM) model using the class of symmetric scale mixtures of normal (SMN) distributions is introduced in this article. The SMN family distributions is an attractive class of symmetric distributions that includes the normal, Student-t, slash and contaminated normal distributions as special cases, providing a robust alternative to estimation in SVM models in the absence of normality. Using a Bayesian paradigm, an efficient method based on Markov chain Monte Carlo (MCMC) is developed for parameter estimation. Additionally, we develop a second-order approximation method to the usual Auxiliary Particle Filter (APF) in order to estimate efficiently the log-likelihood function to model comparison, as is the case of the Bayesian Predictive Information Criteria (BPIC). The methods developed are applied to analyze daily stock returns data on São Paulo Stock, Mercantile & Futures Exchange index (IBOVESPA). Bayesian model selection criteria as well as out-of- sample forecasting results reveal that the SVM model with slash distribution provides significant improvement in model fit as well as prediction to the IBOVESPA data over the usual normal model.

Keywords: Markov chain Monte Carlo, non-Gaussian and nonlinear state space models, scale mixture of normal distributions, stochastic volatility in mean.

1 Introduction

Over the last years the stochastic volatility (SV) models has been considered as an useful tool for modeling time-varying variances, mainly in financial applications where policies makers or stockholders are constantly facing decision problems usually dependent on measures of volatility and risk. SV models were introduced as an alternative approach to the family of GARCH (Bollerslev, 1986) models for describing time-varying volatilities. An attractive feature of the SV model is its close relationship to

[1]Correspondence to: Carlos A. Abanto-Valle, Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970, RJ, Brazil. E-mail: cabantovalle@im.ufrj.br

financial economic theories (Melino and Turnbull, 1990) and its ability to capture the main empirical properties often observed in daily series of financial returns in a more appropriate way (Carnero et al., 2004).

The estimation of the canonical SV model was at one time considered difficult since the likelihood function of these models is not easily calculable. This problem has fully resolved by the creative use of MCMC methods (see Jacquier et al., 1994; Shephard and Pitt, 1997; Kim et al., 1998; Mahieu and Schotman, 1998; Watanabe and Omori, 2004, among others). However, in all these aforementioned works the normal distribution (basic SV model) has been assumed as the basis for parameter inference. Although the basic SV model offers great flexibility in modeling data with time-varying variances, it may suffer from a lack of robustness in the presence of extreme outlying observations. Thus, it is of practical interest to explore frameworks with considerable flexibility in the distributional assumptions of the model in order to get more reliable inferences. There has been considerable work in SV models in this direction. See, for instance, Mandelbrot (1963); Fama (1965); Liesenfeld and Jung (2000); Chib et al. (2002); Jacquier et al. (2004); Chen et al. (2008). More recently Abanto-Valle et al. (2009b) have extended the basic SV model by assuming the flexible class of scale mixtures of normal (SMN) distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993; Fernández and Steel, 2000; Chow and Chan, 2008), where the popular SV model with Student-t errors is a particular member of this class. However, the volatility of daily stock index returns has been estimated with SV models but the results have relied on a extensive pre-modeling of these series in order to avoid the problem of simultaneous estimation of the mean and variance. The SV in mean (SVM) model introduced by Koopman and Uspensky (2002) deal with this serious problem by incorporates the unobserved volatility as explanatory variable in the mean equation of the returns and to propose a simulated maximum likelihood algorithm to parameters estimation. Despite this model having sound experimental results, it is also assumed that the innovations are normally distributed and hence parameters estimation could be unduly affected by observations that are atypical. This motivate us to develop a wider class of SVM models to accommodate long tails behavior.

Following Abanto-Valle et al. (2009b), in this article we propose to robustificate the specification of the innovation returns in SVM models by introducing SMN distributions. We refer to this generalization as SVM-SMN models. Interestingly, this rich class contains as proper elements the SVM with normal (SVM-N), Student-t (SVM-t), slash (SVM-S) and the contaminated normal (SVM-CN) distributions. Indeed, the flexibility of the SVM with SMN distributions could fit time varying features in the mean of the returns and heavy-tails simultaneously. The estimation of such intricate models is not straightforward since volatility now appears in both, the mean and the variance equation and hence intensive computational methods are call for. Inference in this new class of SVM–SMN 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. We simulate the log-volatilities and the shape parameters by using the block sampler algorithm (Shephard and Pitt, 1997; Watanabe and Omori, 2004; Abanto-Valle et al., 2009a) and the Metropolis-Hastings sampling, respectively.At the same time, is known that the stock returns may exhibit fairly frequent and extreme outliers, thus we develop an Auxiliary Particle Filter (APF) algorithm based on a second-order approximation **to** the density of returns, which appears to perform well in this circumstances. This algorithm is also useful to obtain the filtered distribution and the log-likelihood function used in model comparison.

The rest of the paper is structured as follows. Section 2 gives a brief description of the the SMN distributions and the risk and return dynamics. Section 3 outlines the general class of the SVM–SMN models as well as the Bayesian estimation procedure using MCMC methods. Additionally, we discuss some technical details about Bayesian model selection and out-of-sample forecasting of aggregated squared returns. Section 4 is devoted to application and model comparison among particular members of the SVM–SMN models using the IBOVESPA data set. Some concluding remarks as well as future developments are deferred to Section 5.

2 Preliminaries

In this section we present the SMN distributions and a brief discussion about the return and volatility dynamic.

2.1 SMN distributions

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

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

where μ is a location parameter, Z is a normal random variable with zero mean and variance σ^2, $\kappa(\lambda)$ is a positive weight function, λ is a mixing positive random variable with density $p(\lambda \mid \boldsymbol{\nu})$, $\boldsymbol{\nu}$

is a scalar or parameter vector indexing the distribution of λ. As in Lange and Sinsheimer (1993) and Chow and Chan (2008), we restrict our attention to the case in that $\kappa(\lambda) = 1/\lambda$. Thus, given λ, $Y \mid \lambda \sim \mathcal{N}(\mu, \lambda^{-1}\sigma^2)$ and the pdf of Y is given by

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

From a suitable choice of the mixing density $p(. \mid \boldsymbol{\nu})$, 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 $p(. \mid \boldsymbol{\nu})$. These are as follows:

- *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; Lange et al., 1989). For the Student-t distribution with location μ, scale σ and degrees of freedom ν, the pdf can be expressed in the following SMN form:

$$f(y \mid \mu, \sigma, \nu) = \int_0^\infty \mathcal{N}\left(y \mid \mu, \frac{\sigma^2}{\lambda}\right) \mathcal{G}(\lambda \mid \frac{\nu}{2}, \frac{\nu}{2}) d\lambda. \tag{3}$$

  where $\mathcal{G}(. \mid a, b)$ is the Gamma density function. That is, $Y \sim \mathcal{T}_p(\mu, \sigma^2, \nu)$ is equivalent to the following hierarchical form:

$$Y \mid \mu, \sigma^2, \nu, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda \mid \nu \sim \mathcal{G}(\nu/2, \nu/2). \tag{4}$$

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

  This distribution presents heavier tails than those of the normal distribution and it includes the normal case when $\nu \uparrow \infty$. Its pdf is given by

$$f(y \mid \mu, \sigma, \nu) = \nu \int_0^1 \lambda^{\nu-1} \mathcal{N}\left(y \mid \mu, \frac{\sigma^2}{\lambda}\right) d\lambda. \tag{5}$$

  Thus, the slash distribution is equivalent to the following hierarchical form:

$$Y \mid \mu, \sigma^2, \nu, \lambda \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right), \qquad \lambda \mid \nu \sim \mathcal{B}e(\nu, 1), \tag{6}$$

  where $\mathcal{B}e(.,.)$ denotes the beta distribution. The slash distribution has been mainly used in simulation studies because it represents some extreme situations depending on the value of ν, see for example Andrews et al. (1972), Gross (1973), Morgenthaler and Tukey (1991) and Wang and Genton (2006).

- *The contaminated normal distribution,* $Y \sim \mathcal{CN}(\mu, \sigma^2, \boldsymbol{\nu})$, $\boldsymbol{\nu}' = (\delta, \gamma)$.

  Here, λ is a discrete random variable taking one of two states. The probability function of λ, given the parameter vector $\boldsymbol{\nu}' = (\delta, \gamma)$, is denoted by

$$p(\lambda \mid \boldsymbol{\nu}) = \delta \mathbb{I}_{(\lambda=\gamma)} + (1-\delta)\mathbb{I}_{(\lambda=1)}, \;\; 0 \leq \delta < 1, \, 0 < \gamma < 1, \tag{7}$$

  It follows then that

$$f(y) = \delta \mathcal{N}(y|\mu, \gamma^{-1}\sigma^2) + (1-\delta)\mathcal{N}(y|\mu, \sigma^2). \tag{8}$$

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

2.2 Returns and volatility dynamics

The relationships between expected returns and expected volatility have been extensively examined over the past years. Theory generally predicts a positive relation between expected stock returns and volatility if investors are risk averse. That is equity premium provides more compensation for risk when volatility is relatively high. In other words, investors require a larger expected return from a security that is riskier. Yet, empirical studies that attempt to test this important relation yield mixed results. French et al. (1987) found a positive and significant relationship and studies such as Baillie and DeGennaro (1990) and Theodossiou and Lee (1983) reported a positive but insignificant relationship between stock market volatility and stock returns. Consistent with the asymmetric volatility argument Nelson (1991), Glosten et al. (1993) and more recently Bekaert and Wu (2000), Wu (2001) and Brandt and Kang (2004) report negative and often significant relationship between the volatility and return. Overall, there appears to be stronger evidence of a negative relationship between unexpected returns and innovations to the volatility process which French et al. (1987) interpret as indirect evidence of a positive correlation between the expected risk premium and ex ante volatility. This theory, known as the feedback volatility states that stock price reactions to unfavorable events tend to be larger than reactions to favorable ones. More precisely, bad news decreases stock price and increases volatility, therefore determining a further decrease of the price. On the other hand, good news increases stock price and increases volatility, thus mitigating the initial effect on the price. An alternative explanation for asymmetric volatility where causality runs in the opposite direction is the leverage effect put forward by Black (1976) who asserts that a negative (positive) return shock lead to an increase (decrease) in the firm's financial leverage ratio which has an upward (downward) effect on the volatility of the stock returns. However, it has been argued by Black (1976), Christie (1982), French et al. (1987) and Schwert (1989) that leverage alone cannot account for the magnitude of the negative relationship. For example, Campbell and Hentschel (1992) find evidence of both volatility feedback and leverage effects, whereas Bekaert and Wu (2000) present results, which argued that the volatility feedback effect dominates the leverage effect empirically. From an empirical perspective the fundamental difference between the leverage and volatility feedback explanations lies in the causality: the leverage effect explains why a negative return leads to higher subsequent volatility, whereas the volatility feedback effect justifies how an increase in volatility may result in negative returns.

3 The heavy-tailed stochastic volatility in mean model

The SVM model (Koopman and Uspensky, 2002) formulated at the discrete-time is given by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \mathrm{e}^{h_t} + \mathrm{e}^{\frac{h_t}{2}} \varepsilon_t, \tag{9a}$$

$$h_t = \alpha + \phi h_{t-1} + \sigma_\eta \eta_t, \tag{9b}$$

where y_t and h_t are respectively the compounded return and the log-volatility at time t. The innovations ε_t and η_t are assumed to be mutually independent and normally distributed with mean zero and unit variance. As documented by Koopman and Uspensky (2002), the aim of the SVM model is to simultaneously estimate the ex ante relation between returns and volatility and the volatility feedback effect. We denote this basic model as the SVM-N.

In this article, we relax the normality assumption in equation (9a) by allowing the innovation ε_t to have a fat-tailed distribution. This extension is important to capture the leptokurtosis observed in many market returns. Here, we assume that ε_t follows a SMN distribution as follows:

$$\varepsilon_t \sim SMN(0,1,\nu), \;\; \eta_t \sim \mathcal{N}(0,1), \tag{10}$$

ε_t and η_t assumed to be independent. We refer to this generalization as SVM-SMN models. It follows from (1) that the set up defined in (9a)-(9b) and (10) can be written hierarchically as

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \mathrm{e}^{h_t} + e^{\frac{h_t}{2}} \lambda_t^{-\frac{1}{2}} \epsilon_t, \tag{11a}$$

$$h_t = \alpha + \phi h_{t-1} + \sigma_\eta \eta_t, \tag{11b}$$

$$\lambda_t \sim p(\lambda_t), \;\; \epsilon_t \sim \mathcal{N}(0,1), \;\; \eta_t \sim \mathcal{N}(0,1). \tag{11c}$$

As illustrated in Section 2.1, this class of models includes the SVM with student-t (SVM-t), with slash (SVM-S) and contaminated normal (SVM-CN) distributions as special cases. All these distributions have heavier tails than the normal density and thus provide an appealing robust alternative to the usual Gaussian process in SVM models. The SVM-t and SVM-S models are obtained chosen the mixing density as: $\lambda_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$ and $\lambda_t \sim \mathcal{B}e(\nu, 1)$. For the SVM-CN, λ_t follows a discrete distribution. Under a Bayesian paradigm, we use MCMC methods to conduct the posterior analysis in the next subsection. Conditionally to λ_t, some derivations are common to all members of the SVM-SMN family (see Appendix for details).

3.1 Parameter estimation via MCMC

A Bayesian approach to parameter estimation in the SVM-SMN class of models defined by equations (11a)-(11c) relies on MCMC techniques. We propose to construct an algorithm based on MCMC simulation methods to make the Bayesian analysis feasible.

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

$$p(\mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}, \boldsymbol{\theta} \mid \mathbf{y}_{0:T}) \quad \propto \quad p(\mathbf{y}_{1:T} \mid y_0, \boldsymbol{\theta}, \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T}) p(\mathbf{h}_{0:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\lambda}_{1:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{12}$$

where

$$p(\mathbf{y}_{1:T} \mid y_0, \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T}) = \prod_{t=1}^{T} p(y_t \mid y_{t-1}, \boldsymbol{\theta}, h_t, \lambda_t), \tag{13}$$

$$p(\mathbf{h}_{0:T} \mid \boldsymbol{\theta}) = p(h_0 \mid \boldsymbol{\theta}) \prod_{t=1}^{T} p(h_t \mid h_{t-1}, \boldsymbol{\theta}), \tag{14}$$

$$p(\boldsymbol{\lambda}_{1:T} \mid \boldsymbol{\theta}) = \prod_{t=1}^{T} p(\lambda_t \mid \boldsymbol{\theta}), \tag{15}$$

and $p(\boldsymbol{\theta})$ is the prior distribution. We set the prior distribution of the hyperparameters in the SVM–SMN class as: $\beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma^2_{\beta_0})$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(\bar{\beta}_1, \sigma^2_{\beta_1})$, $\beta_2 \sim \mathcal{N}(\bar{\beta}_2, \sigma^2_{\beta_2})$, $\alpha \sim \mathcal{N}(\bar{\alpha}, \sigma^2_{\alpha})$, $\phi \sim \mathcal{N}_{(-1,1)}(\bar{\phi}, \sigma^2_{\phi})$, and $\sigma^2_\eta \sim \mathcal{IG}(\frac{T_0}{2}, \frac{M_0}{2})$, where $\mathcal{N}_{(a,b)}(.,.)$ denotes the truncated normal distribution in the interval (a,b).

Since the posterior density $p(\mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}, \boldsymbol{\theta} \mid \mathbf{y}_{0:T})$ does not have closed form, we first sample the parameters $\boldsymbol{\theta}$, followed by the latent variables $\boldsymbol{\lambda}_{1:T}$ and $\mathbf{h}_{0:T}$ using Gibbs sampling. The sampling scheme is described by the following algorithm:

Algorithm 3.1

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

As described by Algorithm 3.1, the Gibbs sampler requires to sample parameters and latent variables from their full conditionals. Sampling the log-volatilities $\mathbf{h}_{0:T}$ in Step 4 due to the non linear setup in the mean equation (11a) is the most difficult task. In order to avoid the higher correlations due to the Markovian structure of the h_t's, we develop a multi-move block sampler (Shephard and Pitt 1997; Watanabe and Omori 2004; Abanto-Valle et al. 2009a) in the next subsection to sample the $\mathbf{h}_{0:T}$ by blocks. Details on the full conditionals of $\boldsymbol{\theta}$ and the latent variable $\boldsymbol{\lambda}_{1:T}$ are given in the appendix, some of them are easy to simulate from.

3.2 A block sampler algorithm

In order to simulate $\mathbf{h}_{0:T}$, we divide it into $K + 1$ blocks, $\mathbf{h}_{k_{l-1}+1:k_l-1} = (h_{k_{l-1}+1}, \ldots, h_{k_l-1})'$ for $l = 1, \ldots, K+1$, with $k_0 = 0$ and $k_{K+1} = T$, where $k_l - 1 - k_{l-1} \geq 2$ is the size of the $l-$th block. A suitable selection of K is important to obtain an efficient sampler that can reduce the correlation imposed by the model in the sampling process.

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

$$\begin{aligned}\log f(\boldsymbol{\eta}_{t+1:t+k}|h_t, h_{t+k+1}, \mathbf{y}_{t:t+k}, \boldsymbol{\lambda}_{t+1:t+k}, \boldsymbol{\theta}) &= \text{const} - \frac{1}{2\sigma_\eta^2}\sum_{r=t+1}^{t+k}\eta_r^2 + \sum_{r=t+1}^{t+k} l(h_r) \\ &- \frac{1}{2\sigma_\eta^2}(h_{t+k+1} - \alpha - \phi h_{t+k})^2. \qquad (16)\end{aligned}$$

Denotes the first and second derivatives of $l(h_r) = \log p(y_r \mid y_{r-1}, \beta_0, \beta_1, \beta_2, \lambda_r, h_r)$ with respect to h_r by l' and l''. As (16) does not have closed form, we use the Metropolis-Hastings acceptance-rejection algorithm (Tierney, 1994; Chib, 1995) to sampling from. We propose to use the following artificial Gaussian state space model as a proposal density to simulate the block $\boldsymbol{\eta}_{t+1:t+k}$

$$\hat{y}_r = h_r + \epsilon_r, \qquad \epsilon_r \sim \mathcal{N}(0, d_r), \quad r = t+1, \ldots, t+k \qquad (17)$$

$$h_r = \alpha + \phi h_{r-1} + \sigma_\eta \eta_r, \qquad \eta_r \sim \mathcal{N}(0,1). \qquad (18)$$

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

$$\begin{aligned} d_r &= -\frac{1}{l_F''(\hat{h}_r)}, \\ \hat{y}_r &= \hat{h}_r + d_r l'(\hat{h}_r), \qquad (19)\end{aligned}$$

For $r = t+k < T$

$$\begin{aligned} d_r &= \frac{\sigma_\eta^2}{\phi^2 - \sigma_\eta^2 l_F''(\hat{h}_{t+k})}, \\ \hat{y}_r &= d_r\left[l'(\hat{h}_r) - l_F''(\hat{h}_r)\hat{h}_r + \frac{\phi}{\sigma_\eta^2}(h_{r+1} - \alpha)\right]. \qquad (20)\end{aligned}$$

We obtain the measurement equation (17), by a second-order expansion of l_r around some preliminary estimate of η_r denoted by $\hat{\eta}_r$, where $\hat{h}_r$ is the estimate of h_r equivalent to $\hat{\eta}_r$. As $l''(h_r)$ being

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

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

$$l_F''(h_r) = E[l''(h_r)] = -\frac{1}{2} - \beta_2^2\lambda_r e^{h_r}, \qquad (21)$$

which is everywhere strictly negative. The expectation (21) is taken with respect to y_r conditional on h_r and λ_r, $\beta_0, \beta_1, \beta_2$ and y_{r-1}. Since (17)-(18) defines a gaussian state space model, we can apply the de Jong and Shephard's simulation smoother (de Jong and Shephard, 1995) to perform the sampling. We denote this density by g. Since f is not bounded by g, we use the Metropolis-Hastings acceptance-rejection algorithm to sample from f as recommended by Chib (1995). In the SVM-N case, we use the same procedure with $\lambda_t = 1$ for $t = 1, \ldots, T$.

We select the expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ as follows. Once an initial expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ is selected, we can calculate the auxiliary $\hat{\mathbf{y}}_{t+1:t+k}$ by using equations (19) and (20). In the MCMC implementation, the previous sample of $\mathbf{h}_{t+1:t+k}$ may be taken as an initial value of the $\hat{\mathbf{h}}_{t+1:t+k}$. Then, applying the Kalman filter and a disturbance smoother to the linear Gaussian state space model consisting of equations (17) and (18) with the artificial $\hat{\mathbf{y}}_{t+1:t+k}$ yields the mean of $\mathbf{h}_{t+1:t+k}$ conditional on $\hat{\mathbf{h}}_{t+1:t+k}$ in the linear Gaussian state space model, which is used as the next $\hat{\mathbf{h}}_{t+1:t+k}$. By repeating the procedure until the smoothed estimates converge, we obtain the posterior mode of $\mathbf{h}_{t+1:t+k}$. This is equivalent to the method of scoring to maximize the logarithm of the conditional posterior density. Although, we have just noted that iterating the procedure achieves the mode, this will slow our simulation algorithm if we have to iterate this procedure until full convergence. Instead we suggest to use only five iterations of this procedure to provide reasonably good sequence $\hat{\mathbf{h}}_{t+1:t+k}$ instead of an optimal one. The procedure is summarized in Algorithm 3.2.

Algorithm 3.2

1. Initialize $\hat{\mathbf{h}}_{t+1:t+k}$.

2. Evaluate recursively $l'(\hat{h}_r)$ and $l''_F(\hat{h}_r)$ for $r = t+1, \ldots, t+k$

3. Conditional on the current values of the vector of parameters $\boldsymbol{\theta}$, $\boldsymbol{\lambda}_{t+1:t+k}$, h_t and h_{t+k+1} define the auxiliary variables $\hat{y}_r$ and d_r using equations (19) or (20) for $r = t+1, \ldots, t+k$.

4. Consider the linear Gaussian state-space model in (17) and (18). Apply the Kalman Filter and a disturbance smoother Koopman (1993) and obtain the posterior mean of $\boldsymbol{\eta}_{t:t+k}$ ($\mathbf{h}_{t:t+k}$) and set $\hat{\boldsymbol{\eta}}_{t:t+k}$ ($\hat{\mathbf{h}}_{t:t+k}$) to this value.

5. Return to step 2 and repeat the procedure until achieve convergence.

Finally the knots conditions are updated from

$$
\begin{aligned}
h_0 \mid h_1, \theta &\sim \mathcal{N}(\alpha + \phi h_1, \sigma_\eta^2); &(22)\\
h_{k_l} \mid h_{k_l-1}, h_{k_l+1}, \boldsymbol{\theta} &\sim \mathcal{N}\left(\frac{\alpha(1-\phi) + \phi(h_{k_l-1} + h_{k_l+1})}{1+\phi^2}, \frac{\sigma_\eta^2}{1+\phi^2}\right), \text{ for } l = 1, \ldots, K &(23)
\end{aligned}
$$

3.3 Bayesian model selection

In this section, we describe two Bayesian model selection criteria: the deviance information criterion (Spiegelhalter et al. 2002; Berg et al. 2004; Celeux et al. 2006) and the Bayesian predictive information criterion (Ando, 2006, 2007).

3.3.1 Deviance information criterion

Spiegelhalter et al. (2002) introduced the deviance information criterion (DIC), defined as:

$$
\text{DIC} = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})] + p_D. \tag{24}
$$

The second term in (24) measures the complexity of the model by the effective number of parameters, p_D, defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$
p_D = 2[\log L(\mathbf{y}_{1:T} \mid \bar{\boldsymbol{\theta}}) - E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})]]. \tag{25}
$$

To calculate the DIC in the context of SVM-SMN models, the conditional likelihood $L(\mathbf{y}_{1:T} \mid \beta_0, \beta_1, \beta_2, \alpha, \phi, \sigma_\eta^2, \boldsymbol{\nu}, \boldsymbol{\lambda}_{1:T}, \mathbf{h}_{0:T})$, defined in (13), is used in equation (24), where $\boldsymbol{\theta}$ encompasses $(\beta_0, \beta_1, \beta_2, \alpha, \phi, \sigma_\eta^2, \boldsymbol{\nu}')'$, $\boldsymbol{\lambda}_{1:T}$ and $\mathbf{h}_{0:T}$.

As pointed by Stone (2002), Robert and Titterington (2002), Celeux et al. (2006) and Ando (2007), the DIC suffers from some theoretical aspects. First, in the derivation of DIC, Spiegelhalter et al. (2002, p.604) assumed that the specified parametric family of probability distributions that generate future observations encompasses the true model. This assumption may not always hold true. Secondly, the observed data are used both to construct the posterior distribution and to compute the posterior mean of the expected log likelihood. The bias estimate of DIC tends to underestimate the true bias considerably. To overcome these theoretical problems in DIC, recently Ando (2007) has proposed the Bayesian predictive information criterion (BPIC) as an improved alternative of the DIC.

3.3.2 Bayesian predictive information criterion

Let us consider $\mathbf{z}_{1:T} = (z_1, z_2, \ldots, z_T)'$ to be a new set of observations generated by the same mechanism as that of the observed data $\mathbf{y}_{1:T}$ drawn from the true model with unknown density $s(\mathbf{z}_{1:T})$. Concerning the concept of the future observations and the true model, we refer to Konishi and Kitagawa (1993). To evaluate the relative fit of the Bayesian model to the true model $s(\mathbf{z}_{1:T})$, Ando (2007) considered the maximization of the posterior mean of the expected log-likelihood

$$\eta = \int \left[\int \log L(\mathbf{z}_{1:T} \mid \boldsymbol{\theta}) p(\theta \mid \mathbf{y}_{1:T}) d\boldsymbol{\theta} \right] s(\mathbf{z}_{1:T}) d\mathbf{z}_{1:T},$$

where $\log L(. \mid \boldsymbol{\theta})$ denotes the log-likelihood function. It is obvious that η depends on the model fitted, and on the unknown true model $s(\mathbf{z}_{1:T})$. A natural estimator of η is the posterior mean of the log-likelihood,

$$\hat{\eta} = \int \log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) p(\theta \mid \mathbf{y}_{1:T}) d\boldsymbol{\theta},$$

where $L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) = \prod_{t=1}^{T} p(\mathbf{y}_t \mid \boldsymbol{\theta})$. As pointed by Ando (2006, 2007) the quantity $\hat{\eta}$ is generally a positively biased estimator of η, because the same data $\mathbf{y}_{1:T}$ are used both to construct the posterior distribution and to evaluate the posterior mean of the log-likelihood. Therefore, bias correction should be considered, where the bias b is defined as: $b = \int (\hat{\eta} - \eta) s(\mathbf{z}_{1:T}) d\mathbf{y}_{1:T}$. Ando (2007) evaluated the asymptotic bias as

$$T\hat{b} \approx E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})\}] - \log[L(\mathbf{y}_{1:T} \mid \hat{\boldsymbol{\theta}}) p(\hat{\boldsymbol{\theta}})] + \mathrm{tr}\{J_T^{-1}(\hat{\boldsymbol{\theta}}) I_T(\hat{\boldsymbol{\theta}})\} + 0.5q. \tag{26}$$

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

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

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

with $\eta_T(y_t, \boldsymbol{\theta}) = \log p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}) + \log p(\boldsymbol{\theta})/T$. Thus, correcting the asymptotic bias of the posterior mean of the log-likelihood, the Bayesian predictive information criterion (BPIC; Ando, 2006, 2007) can be written as

$$BPIC = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\}] + 2T\hat{b}. \tag{27}$$

The best model is chosen as the one that has the minimum BPIC. To calculate the BPIC, in the context of SV-SMN models, we use $\log\{L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\} = \sum_{t=1}^{T} \log p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ and $\boldsymbol{\theta} = (\alpha, \phi, \sigma_\eta^2, \nu)'$. Because $p(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ does not have closed form, it can be evaluated numerically by using the auxiliary particle filter method (see Kim et al., 1998; Pitt and Shephard, 1999; Chib et al., 2002), which is described next.

3.4 The Auxiliary Particle Filter

In this subsection, we revised the auxiliary particle filtering (APF) method of Pitt and Shephard (1999), which allows us to draw samples from the filtering distribution $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$ by numerical approximation.

Let us consider the nonlinear dynamic representation of the SVM-SMN models. Suppose that the parameter vector $\boldsymbol{\theta}$ is known. Therefore, the evolution, updating and the predictive equations at each t are given, respectively, by

$$p(h_t \mid \mathbf{y_{0:t-1}}, \boldsymbol{\theta}) = \int p(h_t \mid h_{t-1}, \boldsymbol{\theta}) p(h_{t-1} \mid \mathbf{y}_{0:t-1}, \boldsymbol{\theta}) dh_{t-1} \tag{28}$$

$$p(h_t \mid \mathbf{y}_{0:t}, \boldsymbol{\theta}) = \frac{p(y_t \mid h_t) p(h_t \mid h_{t-1}, \boldsymbol{\theta})}{p(y_t \mid \mathbf{y}_{0:t-1}, \boldsymbol{\theta})}, \tag{29}$$

$$p(y_t \mid \mathbf{y}_{0:t-1}, \boldsymbol{\theta}) = \int p(y_t \mid h_t) p(h_t \mid \mathbf{y}_{0:t-1}, \boldsymbol{\theta}) dh_t \tag{30}$$

Let $\{(h_{t-1}^{(1)}, w_{t-1}^{(1)}), \ldots, (h_{t-1}^{(N)}, w_{t-1}^{(N)})\} \sim p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ be an approximate sample of $p(h_{t-1}|\boldsymbol{\theta}, \mathbf{y}_{1:t-1})$, i.e., the pdf $p(h_{t-1}|\boldsymbol{\theta}, \mathbf{y}_{1:t-1})$, of the continuous random variable, h_{t-1}, is approximated by a discrete variable with random support. It then follows that the one-step ahead predictive distribution $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ can be approximated as:

$$\hat{p}(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{0:t-1}) = \sum_{i=1}^{N} p(h_t \mid \boldsymbol{\theta}, h_{t-1}^{(i)}) w_{t-1}^{(i)}, \tag{31}$$

where $h_{t-1}^{(i)}$ is a sample from $p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ with weight $w_{t-1}^{(i)}$ and $\hat{p}(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ is the "empirical prediction density". The equation (31) can be combined with the measurement density to produce, up to proportionality

$$\hat{p}(h_t \mid \mathbf{y}_{0:t}, \boldsymbol{\theta}) \propto p(y_t \mid h_t, \boldsymbol{\theta}) \sum_{i=1}^{N} p(h_t \mid \boldsymbol{\theta}, h_{t-1}^{(i)}) w_{t-1}^{(i)} \tag{32}$$

Pitt and Shephard (1999) pointed out that using $p(h_t \mid \boldsymbol{\theta}, h_{t-1})$ as a density approximating $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{0:t})$ is not generally efficient because it constitutes a *blind* proposal that does not take into account the information contained in y_t. To improve the efficiency, we include this observation in the approximating density. The nonlinear/non-Gaussian component of the measurement equation then starts to play an important role, and certain algebraic manipulations need to be carried out to use standard approximations. This can be accomplished by sampling and index k is sampled on the mixture (32), which defines the particles at $t-1$ that are propagated to t. This correspond to sampling from

$$p(h_t, k \mid \boldsymbol{\theta}, \mathbf{y}_{0:t}) \quad \propto \quad p(y_t \mid \boldsymbol{\theta}, h_t)p(h_t \mid \boldsymbol{\theta}, h_{t-1}w_{t-1}^{(k)}). \tag{33}$$

We can sample first from $p(k \mid \mathbf{y}_{0:t})$ and then from $p(h_t \mid \mathbf{y}_{0:t})$, obtaining a sample $\{(h_t^{(i)}, k^{(i)}), i = 1, \ldots, N\}$. The marginal density $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{0:t})$ is obtained by dropping the index k. If the information contained in $p(h_t, k \mid \boldsymbol{\theta}, \mathbf{y}_{0:t})$ and the information in y_t is carried forward by w_{t-1}. One of the simplest approach, described by Pitt and Shephard (1999) is to define

$$p(h_t, k \mid \boldsymbol{\theta}, \mathbf{y}_{0:t}) \simeq g(h_t, k) \propto g(y_t \mid h_t, \vartheta_t^k)p(h_t \mid \boldsymbol{\theta}, h_{t-1}^k)w_{t-1}^k \tag{34}$$

where $\vartheta_t^{(k)} = \alpha + \phi h_{t-1}^{(k)}$.

This recursive procedure needs to draw h_t sequentially from the filtered distribution, $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$, which is updated as described in Algorithm 3.3.

Algorithm 3.3

1. Posterior at $t-1$:
   $\{(h_{t-1}^{(1)}, w_{t-1}^{(1)}), \ldots, (h_{t-1}^{(i)}, w_{t-1}^{(i)}), \ldots, (h_{t-1}^{(N)}, w_{t-1}^{(N)})\} \sim p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$
2. For $i = 1, \ldots, N$, calculate $\vartheta_t^{(i)} = \alpha + \phi h_{t-1}^{(i)}$
3. Sampling (k, h_t):
   For $i = 1, \ldots, N$
   Indicator: k^i such that $P(k^i = k) \propto g(y_t \mid \vartheta_t^{(k^i)})w_{t-1}^{(k^i)}$
   Evolution:
   $$h_t^{(i)} \sim g(h_t \mid y_t, \vartheta_t^{(i)})$$
   Weights: compute $w_t^{(i)}$ as follows
   $$w_t^{(i)} \propto \frac{p(y_t \mid \boldsymbol{\theta}, h_t^{(i)})p(h_t^{(i)} \mid \theta, h_{t-1}^{(k^i)})}{g(y_t \mid \vartheta_t^{(k^i)})g(h_t^{(i)} \mid \boldsymbol{\theta}, y_t, \vartheta_t^{(k^i)})} = \frac{p(y_t \mid \boldsymbol{\theta}, h_t^i)}{g(y_t \mid h_t^{(i)}, \vartheta_t^{(k^i)})}$$

4. Posterior at t:
$\{(h_t^{(1)}, w_t^{(1)}), \ldots, (h_t^{(i)}, w_t^{(i)}), \ldots, (h_t^{(N)}, w_t^{(N)})\} \sim p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$

Next, we give some technical details related to the out-of-sample forecasting of aggregated squared returns in SV-SMN models. We refer to the reader to see Tauchen and Pitts (1983) for more details..

3.5 Out-of-sample forecasting of aggregated returns

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

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

Evaluation of the last integrals is straightforward, by using Monte Carlo approximation. To initialize a recursion, we use N draws $\{h_T^{(i)}, \boldsymbol{\theta}^{(i)}\}_{i=1}^N$ from the MCMC sample. Then given these N draws, sample $h_{T+k}^{(i)}$ from $p(h_{T+k} \mid \boldsymbol{\theta}^{(i)}, h_{T+k-1}^{(i)})$ and $\lambda_{T+k}^{(i)}$ from $p(\lambda_{T+k} \mid \boldsymbol{\theta}^{(i)})$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$, by using equations (11b) and (11c), respectively. Finally, sample $y_{T+k}^{(i)}$ from $p(y_{T+k} \mid \boldsymbol{\theta}^{(i)}, \lambda_{T+k}^{(i)}, h_{T+k}^{(i)})$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$. With draws from $h_{T+k}^{(i)}$ and $y_{T+k}^{(i)}$, the aggregated daily squared return (a common model-free indicator of volatility) can be calculated as $V_K^{(i)} = \sum_{k=1}^K y_{T+k}^{2(i)}$ and the aggregated volatility as, $S_K^{(i)} = \sum_{k=1}^K \mathrm{e}^{h_{T+k}{}^{(i)}}$, for $i = 1, \ldots, N$, respectively.

4 Empirical Application

This section analyzes the daily closing prices for 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 by a theoretical portfolio with the stocks that accounted for 80% of the volume traded in the last 12 months and that were traded at least on 80% of the trading days. It is revised quarterly, in order to keep its representativeness of the volume traded and in average the components of IBOVESPA represent 70% of all the stock value traded. The data set was obtained from the Yahoo finance web site available to download at http://finance.yahoo.com. The period of analysis is January 5, 1998 - October 3, 2005 which yields 1917 observations. Throughout, we will work with the compounded return expressed as a percentage, that is

$$y_t = 100(\log P_t - \log P_{t-1}),$$

where P_t is the closing price on day t.

Table 1 summarize descriptive statistics for the IBOVESPA corrected compounded returns with the time series plot in Figure 1. For the returns series, the basic statistics viz. the mean, standard deviation, skewness and kurtosis are calculated to be 0.0579, 2.3345, 0.8346 and 19.1847, respectively. Some extreme observations, explained by turbulences in financial markets occurred by August-Setembro, 1998 and January, 1999 (the Russian and Brazilian exchange rate crisis, respectively), contribute to extent to the large kurtosis of the IBOVESPA returns. As a result, the IBOVESPA returns likely shows a departure from the underlying normality assumption. Thus, we reanalyze this data with the aim of providing robust inference by using the SMN class of distributions. In our analysis, we compare between the SVM-N, SVM-t, SVM-S and SVM-CN models.

Table 1: Summary statistics for the IBOVESPA return series

	mean	s.d.	max	min	skewness	kurtosis
Returns	0.0579	2.3345	28.8324	-17.2082	0.8346	19.1847

In all cases, we simulated the h_t's in a multi-move fashion with stochastic knots based on the method described in Section 3.1. We set the prior distributions of the common parameters as: $\beta_0 \sim \mathcal{N}(0.0, 100)$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(0.1, 100.0)$, $\beta_2 \sim \mathcal{N}(-0.1, 100.0)$, $\alpha \sim \mathcal{N}(0.0, 100.0)$, $\phi \sim \mathcal{N}_{(-1,1)}(0.95, 100.0)$, $\sigma_\eta^2 \sim \mathcal{IG}(2.5, 0.025)$. The prior distributions on the shape parameters were chosen as: $\nu \sim \mathcal{G}(12.0, 0.8)$ and $\nu \sim \mathcal{G}(2.0, 0.25)$ for the SVM-t model and the SVM-S model, respectively. For the SVM-CN, we set $\delta \sim \mathcal{B}e(2, 2)$ and $\gamma \sim \mathcal{B}e(2, 4)$. The initial values of the parameters are randomly generated from the prior distributions. We set all the log-volatilities, h_t, to be zero. Finally the initial $\boldsymbol{\lambda}_{1:T}$ are generated from the prior $p(\lambda_t \mid \nu)$. All the calculations were performed running stand alone code developed by the authors using an open source C++ library for statistical computation, the Scythe statistical library (Pemstein et al., 2007), which is available for free download at http://scythe.wustl.edu.

For the block sampler algorithm, we set the number of blocks K to be 60 in such a way that each block contained 32 h_t's on average. For the SVM-N, SVM-t and the SVM-S models, we conducted

Table 2: Estimation results for the IBOVESPA returns. The first row: Posterior mean. The second row: Posterior 95% credible interval in parentheses. The third row: CD statistics

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

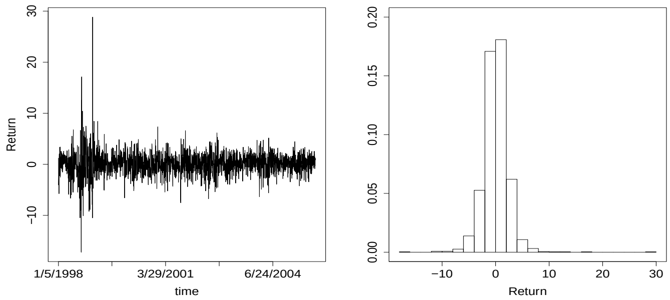

Figure 1: IBOVESPA compounded returns with sample period from January 5, 1998 to September 03, 2005. The left panel shows the plot of the raw series and the right panel plots the histogram of returns.

the MCMC simulation for 50000 iterations. However, for the SVM-CN model, we use 210000 iterations. In both cases, the first 10000 draws were discarded as a burn-in period. In order to reduce the autocorrelation between successive values of the simulated chain, only every $10th$ (SVM-N, SVM-t and SVM-S models) and $100th$ (SVM-CN model) values of the chain are stored. With the resulting 4000 (2000) values, we calculated the posterior means, the 95% credible intervals and the convergence diagnostic (CD) statistics (Geweke, 1992). Table 2 summarizes the results. According to the CD values, the null hypothesis that the sequence of 4000 (2000) draws is stationary is accepted at the 5% level for all the parameters in all the models considered here.

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

The posterior means together with the posterior 95% intervals of the three parameters, which govern the mean process for each one of the four models are reported in Table 2. We observed that in all the cases the posterior mean of β_0 is always positive and statistically significant, because the 95% does not contain zero. We found that the posterior mean of β_1 is positive and similar to the first order autocorrelation not reported here. As the 95% posterior interval contains zero, so this coefficient could be not significant. The β_2 parameter, which measures both *ex ante* relationship between returns and volatility and the volatility feedback effect, has a negative posterior mean for all the models. We found β_2 is statistically significant because in all the cases the 95% posterior credibility interval does not contain zero. This result confirms previous results in the literature and it indicates that when investors expect higher persistent levels of volatility in the future they require compensation for this in the form higher expected returns.

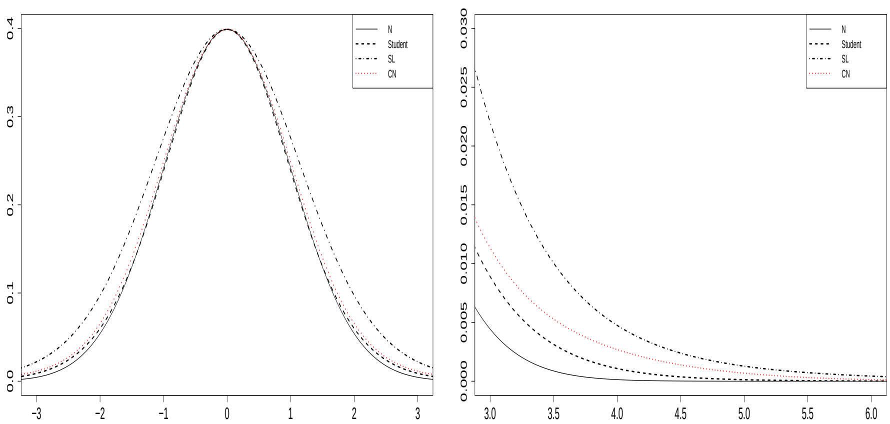

Figure 2: Density curves of the univariate normal, Student-t, slash, variance gamma and contaminated normal distributions using the estimated tail-fatness parameter from the respective SVM model.

The magnitude of the tail fatness is measured by the shape parameter ν in the SVM-t and SVM-S models. In the SVM-CN case it is measured by δ. The posterior mean of ν in the SVM-t model is 16.2892. In the SVM-S model, the posterior mean of ν is 2.4657. In the SVM-CN the posterior mean of δ is 0.1188, which can be interpreted as the proportion of outlier present in the data set and γ as an scale factor has posterior mean of 0.2952. These results seem to indicate that the measurement error of the stock returns are better explained by heavy-tailed distributions.

The reason why the volatility of the SVM-SMN models is estimated to be more persistent and

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

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

less variable can be understood by comparing the densities of this distributions consider here. To illustrate the tail behavior, we plot the normal ($\mathcal{N}(0,1)$) density, Student's-t ($\mathcal{T}(0,1,\nu)$) density with ν degrees of freedom, the slash ($\mathcal{S}(0,1,\nu)$) density with shape parameter ν and the contaminated normal ($\mathcal{CN}(0,1,\delta,\gamma)$). We set ν, δ and γ as the posterior mean of the respective SVM model (see Table 2 for details). Figure 2 depicts the four density curves (the student-t, slash and contaminated normal have been rescaled to be comparable). All the distributions have fatter tail than the normal distribution. Note that the slash distribution has fatter tail than the other distributions that we have considered (see Figure 2 right panel). Therefore, the SVM of models considered here attributes a relatively larger proportion of extreme return values to ε_t instead of η_t than of the SVM-N model, making the volatility of the SVM-t, the SVM-S and the SVM-CN models less variables. It also increases the persistence in volatility of these models. This interpretation is confirmed by comparing the volatility estimates. In Figure 3, we plot the smoothed mean of $\mathrm{e}^{\frac{h_t}{2}}$. The posterior smoothed mean of $\mathrm{e}^{\frac{h_t}{2}}$ of the SVM-t, SVM-S and SVM-CN exhibit smoother movements than this from the SVM-N model (solid line). Extreme returns such a Brazilian currency crisis in January, 1999 make the differences clear. The models with heavy tails accommodate possible outliers in a somewhat different way by inflating the variance $\mathrm{e}^{\frac{h_t}{2}}$ by $\lambda_t^{-1}\mathrm{e}^{\frac{h_t}{2}}$. This can have a substantial impact, for instance, in the valuation of derivative instruments and several strategic or tactical asset allocation topics.

Next,we use the deviance information criterion (DIC) and the Bayesian predictive information criterion (BPIC) to compare between all the competing models. In both cases, the best model has the smallest DIC (BPIC). From Table 3, the BPIC criterion indicates that the SVM-SMN models with heavy tails present better fit than the basic SVM-N model, with the SVM-S model relatively better among all the considered models, suggesting that the IBOVESPA returns data demonstrate sufficient departure from underlying normality assumptions. As expected, the DIC also selects the SVM-S model as the best model.

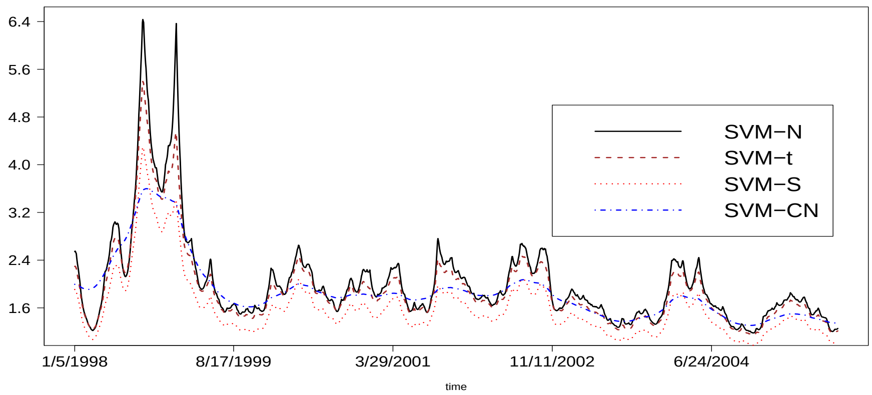

Figure 3: IBOVESPA data set. Posterior smoothed mean of $e^{\frac{h_t}{2}}$

We evaluate the SVM-SMN models by using the out-of-sample forecasting of the squared returns aggregated over certain period of time. Based on the 1917 observations of returns used previously, we calculate the forecast over the following $1, 2, \ldots, 10$ days as described in Section 3.5. Figure 4 plots the posterior means and 95% posterior credibility interval of the aggregated squared returns together with the realized values. The 95% posterior interval of the aggregated volatility, e^{h_t}, are also plotted. For all models (a)-(d), the 95% intervals of the aggregated squared returns are much wider than those for the aggregated volatility. The 95% posterior credibility interval of the aggregated squared returns for the SVM-S model does not include the realized values for days from 3, 4 and 10. The SVM-t model shows different forecasts and, days 3, 4, 5, 7 and 10 are outside the 95% credibility intervals. The SVM-CN include all the realized values of the aggregated squared returns for days from 1 to 10. The SVM-N shows the worst behavior, it includes only the realized values for day 1.

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

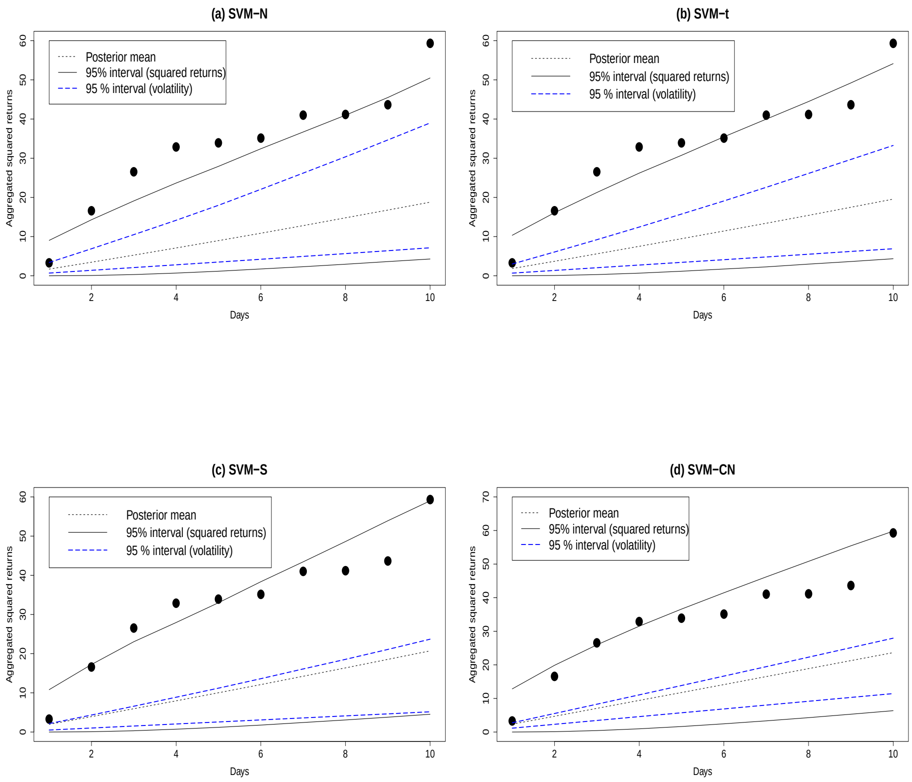

Figure 4: IBOVESPA data set. Out-of-sample forecast of the aggregated squared returns for (a) SVM-N, (b) SVM-t, (c) SVM-S and (d) SVM-CN models.

posterior estimates of mean and 95% credible interval of model parameters. The observations in $t = 1566, 1582$ and 1599, which corresponds to March 5, 2005, April 20, 2005 and May 16, 2005, respectively, are contaminated by ky_t, where k varies from -6 and 6 with increments of 0.5 units. In Figures 5 and 6, we plot the posterior mean and 95% credible interval of ϕ and σ^2_η, respectively, for the SVM-N, the SVM-t and the SVM-S models. Clearly, the SVM-S and the SVM-t models are less affected by variations of k than the SVM-N model, meaning substantial robustness of the estimates over the usual normal process in presence of outlying observations.

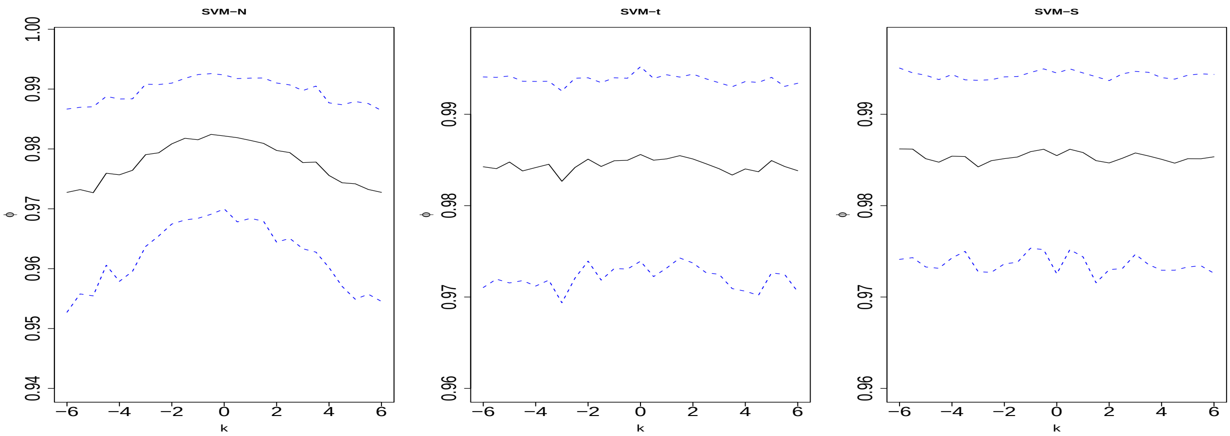

Figure 5: Posterior mean (dashed line) and 95% credible interval (solid line) for ϕ of fitting the SVM-N, SVM-t and SVM-S models for the IBOVESPA data set.

5 Conclusions

This article discusses a Bayesian implementation of a robust alternative to estimation in the stochastic volatility in mean model, proposed by Koopman and Uspensky (2002), via MCMC methods. The SVM enables us to investigate the dynamic relationship between returns and its time-varying volatility. The Gaussian assumption of the mean innovation was replaced by univariate thick-tailed processes, known as scale mixtures of normal distributions. We study three specific sub-classes, viz. the Student-t, slash and the contaminated normal distributions and compare parameter estimates and model fit with the default normal model. Under a Bayesian perspective, we constructed an algorithm based

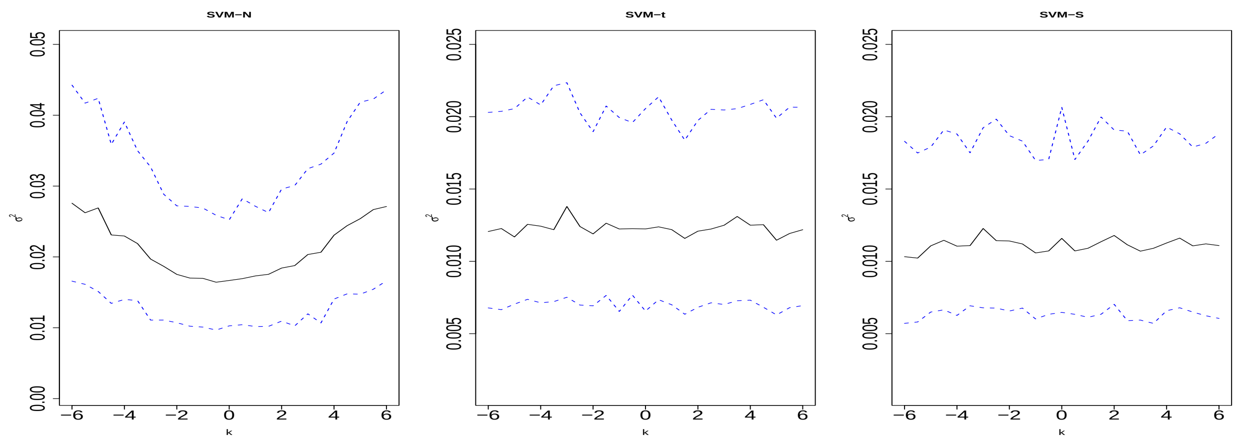

Figure 6: Posterior mean (dashed line) and 95% credible interval (solid line) for σ^2 of fitting the SVM-N, SV-t and SVM-S models for the IBOVESPA data set.

on Markov Chain Monte Carlo (MCMC) simulation methods to estimate all the parameters and latent quantities in our proposed SVM-SMN model. We illustrate our methods through an empirical application of the IBOVESPA return series, which shows that the SVM-S model provide better model fitting than the SVM-N model in terms of parameter estimates, interpretation, robustness aspects and out-of-sample forecast of the aggregated squared returns. The β_2 estimate which measures both the ex ante relationship between returns and volatility and the volatility feedback effect is found to be negative. The results fall in line with those of French et al. (1987), who found similar relationship between unexpected volatility dynamics and returns and confirm the hypothesis that investors require higher expected returns when unanticipated increase in future volatility are highly persistent. This is consistent with our findings of higher values of ϕ combined with larger negative values for the in-mean parameter.

Our SVM-SMN models have shown considerable flexibility to accommodate outliers, however its robustness aspects could be seriously affected by the presence of skewness. Lachos et al. (2009) have recently proposed a remedy to incorporate skewness and heavy-tailedness simultaneously using scale mixtures of skew-normal (SMSN) distributions. We conjecture that the methodology presented in this paper can be undertaken under univariate and multivariate setting of SMSN distributions and should yield satisfactory results in situations where data exhibit non-normal behavior, although at the expense of additional complexity in its implementation. Nevertheless, a deeper investigation of those modifications is beyond the scope of the present paper, but provides stimulating topics for further research.

Appendix A: The Full conditionals

In this appendix, we describe the full conditional distributions for the parameters and the mixing latent variables $\boldsymbol{\lambda}_{1:T}$ of the SV-SMN class of models.

Full conditional distribution of β_0, β_1 and β_2

For parameters β_0, β_1 and β_2, we set the priors distributions as: $\beta_0 \sim \mathcal{N}(\bar{\beta}_0, \sigma^2_{\beta_0})$, $\beta_1 \sim \mathcal{N}_{(-1,1)}(\bar{\beta}_1, \sigma^2_{\beta_1})$, $\beta_2 \sim \mathcal{N}(\bar{\beta}_2, \sigma^2_{\beta_2})$. Using equation (9a), we have the full conditionals are given by

$$\beta_0 \mid \mathbf{y}_{0:T}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T}, \beta_1, \beta_2 \quad \sim \quad \mathcal{N}(\frac{b_{\beta_0}}{a_{\beta_0}}, \frac{1}{a_{\beta_0}}) \tag{35}$$

$$\beta_1 \mid \mathbf{y}_{0:T}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T}, \beta_0, \beta_1 \quad \sim \quad \mathcal{N}(\frac{b_{\beta_1}}{a_{\beta_1}}, \frac{1}{a_{\beta_1}})\mathbb{I}_{|\beta_2|<1} \tag{36}$$

$$\beta_2 \mid \mathbf{y}_{0:T}, \mathbf{h}_{1:T}, \boldsymbol{\lambda}_{1:T}, \beta_0, \beta_1 \quad \sim \quad \mathcal{N}(\frac{b_{\beta_2}}{a_{\beta_2}}, \frac{1}{a_{\beta_2}}) \tag{37}$$

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

Full conditional distribution of α, ϕ and σ^2_η

The prior distributions of the common parameters are set as: $\alpha \sim N(\bar{\alpha}, \sigma^2_\alpha)$, $\phi \sim \mathcal{N}_{(-1,1)}(\bar{\phi}, \sigma^2_\phi)$, $\sigma^2_\eta \sim \mathcal{IG}(\frac{T_0}{2}, \frac{M_0}{2})$. Together with (14), we have the following full conditional for α:

$$p(\alpha \mid \mathbf{h}_{0:T}, \phi, \sigma^2_\eta) \quad \propto \exp\{-\tfrac{a_\alpha}{2}(\alpha - \tfrac{b_\alpha}{a_\alpha})^2\}, \tag{38}$$

which is the normal distribution with mean $\frac{b_\alpha}{a_\alpha}$ and variance $\frac{1}{a_\alpha}$, where $a_\alpha = \frac{1}{\sigma^2_\alpha} + \frac{T}{\sigma^2_\eta} + \frac{1+\phi}{\sigma^2_\eta(1-\phi)}$ and $b_\alpha = \frac{\bar{\alpha}}{\sigma^2_\alpha} + \frac{(1+\phi)}{\sigma^2_\eta} h_0 + \frac{\sum_{t=1}^{T}(h_t - \phi h_{t-1})}{\sigma^2_\eta}$. Similarly, by using (14), we have that the conditional posterior of ϕ is given by

$$p(\phi \mid \mathbf{h}_{0:T}, \alpha, \sigma^2_\eta) \quad \propto Q(\phi) \exp\{-\tfrac{a_\phi}{2\sigma^2_\eta}(\phi - \tfrac{b_\phi}{a_\phi})^2\}\mathbb{I}_{|\phi|<1} \tag{39}$$

where $Q_\phi = \sqrt{1-\phi^2}\exp\{-\frac{1}{2\sigma_\eta^2}[(1-\phi^2)(h_0-\frac{\alpha}{1-\phi})^2\}$, $a_\phi = \sum_{t=1}^T h_{t-1}^2 + \frac{\sigma_\eta^2}{\sigma_\phi^2}$, $b_\phi = \sum_{t=1}^T h_{t-1}(h_t-\alpha) + \bar{\phi}\frac{\sigma_\eta^2}{\sigma_\phi^2}$ and $\mathbb{I}_{|\phi|<1}$ is an indicator variable. As $p(\phi \mid \mathbf{h}_{0:T}, \alpha, \sigma_\eta^2)$ in (39) does not have closed form, we sample from using the Metropolis-Hastings algorithm with truncated $\mathcal{N}_{(-1,1)}(\frac{b_\phi}{a_\phi}, \frac{\sigma_\eta^2}{a_\phi})$ as the proposal density.

From (14), the conditional posterior of σ_η^2 is $\mathcal{IG}(\frac{T_1}{2}, \frac{M_1}{2})$, where $T_1 = T_0 + T + 1$ and $M_1 = M_0 + [(1-\phi^2)(h_0 - \frac{\alpha}{1-\phi})^2] + \sum_{t=1}^T (h_t - \alpha - \phi h_{t-1})^2$.

Full conditional of λ_t and ν

- **SV-t case**

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

$$p(\lambda_t \mid y_t, y_{t-1}, h_t, \nu) \quad \propto \lambda_t^{\frac{\nu+1}{2}-1} \mathrm{e}^{-\frac{\lambda_t}{2}[(y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 \mathrm{e}^{h_t})^2 \mathrm{e}^{-h_t}+\nu]}, \tag{40}$$

which is the gamma distribution, $\mathcal{G}(\frac{\nu+1}{2}, \frac{[y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 \mathrm{e}^{h_t}]^2 \mathrm{e}^{-h_t}+\nu}{2})$.

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

$$p(\nu \mid \boldsymbol{\lambda}_{1:T}) \quad \propto \quad \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu-1} \mathrm{e}^{-\frac{\nu}{2}[\sum_{t=1}^T(\lambda_t - \log\lambda_t)+2b_\nu]}}{[\Gamma(\frac{\nu}{2})]^T} \mathbb{I}_{2<\nu\leq 40}. \tag{41}$$

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

$$\log\Gamma(\nu) \quad = \quad \frac{\log(2\pi)}{2} + \frac{2\nu-1}{2}\log(\nu) - \nu + \frac{\theta}{12\nu}, \qquad 0<\theta<1. \tag{42}$$

Taking the second derivative of $\ell(\nu)$ from (41) and using (42), we have that

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

because in practical applications $T \geq 2$.

- **SV-S case**

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

$$p(\lambda_t \mid y_t, y_{t-1}, h_t, \nu) \quad \propto \lambda_t^{\nu+\frac{1}{2}-1} \mathrm{e}^{-\frac{\lambda_t}{2}[(y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 \mathrm{e}^{h_t})^2 \mathrm{e}^{-h_t}]} \mathbb{I}_{0<\lambda_t<1}, \tag{43}$$

that is $\lambda_t \sim \mathcal{G}_{(0<\lambda_t<1)}(\nu+\frac{1}{2}, \frac{1}{2}[y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 \mathrm{e}^{h_t}]^2 \mathrm{e}^{-h_t})$, the right truncated gamma distribution. Assuming that a prior distribution of $\nu \sim \mathcal{G}(a_\nu, b_\nu)$, the full conditional distribution of ν is given by

$$p(\nu \mid \mathbf{h}_{0:T}, \boldsymbol{\lambda}_{1:T}) \quad \propto \nu^{T+a_\nu-1} \mathrm{e}^{-\nu(b_\nu-\sum_{t=1}^T \log\lambda_t)} \mathbb{I}_{\nu>1}. \tag{44}$$

Then, the full conditional of ν is $\mathcal{G}_{\nu>1}(T+a_\nu, b_\nu-\sum_{t=1}^T \log\lambda_t)$, i.e. the left truncated gamma distribution. We simulate from the right and left truncated gamma distributions using the algorithm proposed by Philippe (1997).

- **SVM-CN case**

Here λ_t is a discrete random variable and $\boldsymbol{\nu} = (\delta, \gamma)'$. To sample from λ_t, we introduce an auxiliary variable, S_t, such that $P(S_t = 1) = \delta$ and $\lambda_t = \gamma S_t + 1 - S_t$. Using (8) with $\sigma^2 = 1$, we have that the full conditional of S_t is given by

$$p(S_t \mid \delta, \gamma, \beta_0, \beta_1, \beta_2, h_t, y_t, y_{t-1}) \quad \propto \quad \delta^{S_t}(1-\delta)^{1-S_t}\gamma^{\frac{S_t}{2}} \mathrm{e}^{-\frac{1}{2}[\mathrm{e}^{-h_t}(\gamma S_t+1-S_t)(y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 \mathrm{e}^{h_t})^2]}. \tag{45}$$

That is, $S_t \mid \delta, \gamma, \beta_0, \beta_1, \beta_2, h_t$ has a Bernoulli distribution. We assume that $\delta \sim \mathcal{B}e(\delta_0, \delta_1)$ and $\gamma \sim \mathcal{B}e(\gamma_0, \gamma_1)$. Then, the full conditional of δ is given by

$$p(\delta \mid \gamma, \mathbf{S}_{1:T}) \quad \propto \quad \delta^{\delta_0-1}(1-\delta)^{\delta_1-1} \prod_{t=1}^T \delta^{S_t}(1-\delta)^{1-S_t} \tag{46}$$

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

$$p(\gamma \mid \beta_0, \beta_1, \beta_2, \mathbf{S}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{0:T}) \quad \propto \quad (1-\gamma)^{\gamma_1-1}\gamma^{\gamma_0-1+\sum_{t=1}^T \frac{S_t}{2}} \mathrm{e}^{-\frac{\gamma}{2}\sum_{t=1}^T \mathrm{e}^{-h_t} S_t (y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 \mathrm{e}^{h_t})^2}. \tag{47}$$

As (47) does not have closed form, we can sample from using the Metropolis-Hastings algorithm. The the right truncated gamma distribution $\mathcal{G}_{0<\gamma<1}(\gamma_0+\sum_{t=1}^T \frac{S_t}{2}, -\frac{1}{2}\sum_{t=1}^T \mathrm{e}^{-h_t} S_t(y_t-\beta_0-\beta_1 y_{t-1}-\beta_2 \mathrm{e}^{h_t})^2)$ can be used as a proposal density.

Appendix B: Additionals details related to the APF algorithm

As documented by Smith and Santos (2006), for series such as stock returns, which exhibit fairly frequent and extreme outliers, filters based on this first order approximation can be easliy break down. However, the APF based on the much rarely used second order approximation appears to perform better in this circumstances.

The main modification considered is the definition of a second-order approximation around ϑ_t for $\log p(y_t \mid \beta_0, \beta_1, \beta_2, y_{t-1}, h_t) = l(h_t)$, which we designate as

$$\log g(y_t \mid h_t, \vartheta_t^k) \quad = \quad l(\vartheta_t^k) + (h_t - \vartheta_t^k) l'(\vartheta_t^k) + \frac{1}{2} l''_F(\vartheta_t^k)(h_t - \vartheta_t^k)^2 \tag{48}$$

We define $l''_F(\vartheta_t^k) = -(l'(\vartheta_t^k))^2$, we have $l''_F(\vartheta_t^k) < 0$ for all the values of h_t. Rearranging equation (34), we can express this as

$$g(h_t, k \mid \mathbf{y}_{0:t}) \propto g(y_t \mid \vartheta_t^k) g(h_t \mid y_t, \vartheta_t^k) w_{t-1}^k,$$

where the factors are

$$g(y_t \mid \vartheta_t^k) = \exp\{l(\vartheta_t^k) - l'(\vartheta_t^k)\vartheta_t^k + \frac{1}{2} l''_F(\vartheta_t^k)\vartheta_t^{2k} - \frac{1}{2\sigma^2}\vartheta_t^{2k} + \frac{b_t^{2k}}{2a_t^k} + \frac{1}{2}\log a_t^k\}$$

and

$$g(h_t \mid y_t, \vartheta_t, \mu_t^k) = \mathcal{N}(h_t \mid \frac{b_t^k}{a_t^k}, \frac{1}{a_t^k})$$

where $a_t^k = \frac{1}{\sigma_\eta^2} - l''_F(\vartheta_t^k)$ and $b_t^k = l'(\vartheta_t^k) - l''_F(\vartheta_t^k)\vartheta_t^k + \frac{\vartheta_t^k}{\sigma_\eta^2}$ and $\mathcal{N}(h_t \mid \frac{b_t^k}{a_t^k}, \frac{1}{a_t^k})$ denotes that the random variable h_t follows a normal distribution with mean $\frac{b_t^k}{a_t^k}$ and variance $\frac{1}{a_t^k}$. Then, we could simulate the index with probability proportional to $g(y_t \mid \vartheta_t^k)$ and then draw from $g(h_t \mid h_{t-1}^k, y_t, \vartheta_t^k)$. The resulting reweighted sample's second-stage weights are proportional to the, hopefully fairly even, weights

$$w_t^{(i)} \propto \frac{p(y_t \mid \boldsymbol{\theta}, h_t^{(i)}) p(h_t^{(i)} \mid \theta, h_{t-1}^{(k^i)})}{g(y_t \mid \vartheta_t^{(k^i)}) g(h_t^{(i)} \mid \boldsymbol{\theta}, y_t, \vartheta_t^{(k^i)})} = \frac{p(y_t \mid \boldsymbol{\theta}, h_t^i)}{g(y_t \mid h_t^{(i)}, \vartheta_t^{(k^i)})}.$$

References

Abanto-Valle CA, Bandyopadhyay D, Lachos VH, Enriquez I. 2009a. Robust bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics & Data Analysis, doi:10.1016/j.csda.2009.06.011* .

Abanto-Valle CA, Migon HS, Lopes HF. 2009b. Bayesian modeling of financial returns: A relationship between volatility and trading volume. *Applied Stochastic Modeling in Business and Industry, doi:10.1002/asmb.789* .

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

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

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

Andrews DF, Bickel PJ, Hampel FR, Huber PJ, Rogers WH, Tukey J. 1972. *Robust Estimates of Location: Survey and Advances.* Princeton, NJ: Princeton University Press.

Andrews DF, Mallows SL. 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36**: 99–102.

Baillie RT, DeGennaro RP. 1990. Stock returns and volatility. *Journal of Financial and Quantitative Analysis* **25**: 203–214.

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

Berg A, Meyer R, Yu J. 2004. Deviance Information Criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics* **22**: 107–120.

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

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticy. *Journal of Econometrics* **31**: 307–327.

Brandt MW, Kang Q. 2004. Generalized autoregressive conditional heteroskedasticy. *Journal of Financial Economics* **72**: 217–257.

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

Carnero MA, Peña D, Ruiz E. 2004. Persistence and kurtosis in GARCH and Stochastic volatility models. *Journal of Financial Econometrics* **2**: 319–342.

Celeux G, Forbes F, Robert CP, Titterington DM. 2006. Deviance information criteria for missing data models. *Bayesian Analysis* **1**: 651–674.

Chen CWS, Liu FC, So MKP. 2008. Heavy-tailed-distributed threshold stochastic volatility models in financial time series. *Australian & New Zeland Journal of Statistics* **50**: 29–51.

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

Chib S, Nardari F, Shepard N. 2002. Markov Chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* **108**: 281–316.

Chow STB, Chan JSK. 2008. Scale mixtures distributions in statistical modelling. *Australian & New Zeland Journal of Statistics* **50**: 135–146.

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

Fama E. 1965. Portfolio analysis in a stable paretian market. *Managament Science* **11**: 404–419.

Fernández C, Steel MFJ. 2000. Bayesian regression analysis with scale mixtures of normals. *Econometric Theory* **16**: 80–101.

French KR, Schert WG, Stambugh RF. 1987. Expected stock return and volatility. *Journal of Financial Economics* **19**: 3–29.

Geweke J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds.) *Bayesian Statistics. Vol. 4.* 169–193.

Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess returns on stocks. *Journal of Finance* **48**: 1791–1801.

Gross AM. 1973. A Monte Carlo swindle for estimators of location. *Journal of the Royal Statistical Society, Series C. Applied Statistics* **22**: 347–353.

Jacquier E, Polson N, Rossi P. 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**: 371–418.

Jacquier E, Polson N, Rossi P. 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* **122**: 185–212.

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

Konishi S, Kitagawa G. 1993. Generalised information criteria in model selection. *Biometrika* **83**: 875–890.

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

Koopman SJ, Uspensky EH. 2002. The stochastic volatility in mean model: empirical evidence from international tock markets. *Journal of Applied Econometrics* **17**: 667–689.

Lachos VH, Ghosh P, Arellano-Valle RB. 2009. Likelihood based inference for skew–normal/independent linear mixed models. *Statistica Sinica* **(to appear)**.

Lange KL, Little R, Taylor J. 1989. Robust statistical modeling using t distribution. *Journal of the American Statistical Association* **84**: 881–896.

Lange KL, Sinsheimer JS. 1993. Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat* **2**: 175–198.

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

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

Mahieu R, Schotman PC. 1998. Am empirical application of stochastic volatility models. *Journal of Applied Econometrics* **13**: 333–360.

Mandelbrot B. 1963. The variation of certain speculative prices. *Journal of Business* **36**: 314–419.

Melino A, Turnbull SM. 1990. Pricing foreign options with stochastic volatility. *Journal of Econometrics* **45**: 239–265.

Morgenthaler S, Tukey J. 1991. *Configural Polysampling: A Route to Practical Robustness.* New York: Wiley.

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

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

Philippe A. 1997. Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing* **7**: 173–181.

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

Robert CP, Titterington DM. 2002. Discussion on "Bayesian measures of model complexity and fit". *Biometrical Journal* **64**: 573–590.

Schwert GW. 1989. Why does stock market volatility change over time? *Journal of Finance* **44**: 1115–1153.

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

Smith JQ, Santos AAF. 2006. Second-order filter distribution approximations for financial time series with extreme outliers. *Journal of Business and Economic Statistics* **24**: 329–337.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**: 621–622.

Stone M. 2002. Discussion on "Bayesian measures of model complexity and fit". *Journal of the Royal Statistical Society, Series B* **64**: 621.

Tauchen GE, Pitts M. 1983. The price variability-volume relationshis in speculative markets. *Econometrica* **51**: 485–506.

Theodossiou P, Lee U. 1983. Relationship between volatility and expected return across international stock markets. *Journal of Business Finance and Accounting* **22**: 289–300.

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

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

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

Wu G. 2001. The determinants of asymmetric volatility. *The Review of Financial Studies* **14**: 837–859.