Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 119 (2016) 108–127
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

Modeling volatility using state space models with heavy tailed distributions

Frank M. de Pinho[a], Glaura C. Franco[b,*], Ralph S. Silva[c]

[a] Departamento de Administração, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[b] Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[c] Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Received 30 October 2013; received in revised form 18 May 2015; accepted 1 August 2015
Available online 18 August 2015

Abstract

This article deals with a non-Gaussian state space model (NGSSM) which is attractive because the likelihood can be analytically computed. The paper focuses on stochastic volatility models in the NGSSM, where the observation equation is modeled with heavy tailed distributions such as Log-gamma, Log-normal and Weibull. Parameter point estimation can be accomplished either using Bayesian or classical procedures and a simulation study shows that both methods lead to satisfactory results. In a real data application, the proposed stochastic volatility models in the NGSSM are compared with the traditional autoregressive conditionally heteroscedastic, its exponential version, and stochastic volatility models using South and North American stock price indexes.

Keywords: Bayesian inference; Classical inference; Non-Gaussian state space model; Stochastic volatility; Stock price index

1. Introduction

The global financial crisis has generated a significant instability in the prices of financial assets and particularly in the stock market. For this reason, a major concern among economists, fund managers and investment researchers is how long this crisis will impact the variability of asset prices. For this reason, researches focusing on the study and modeling of volatility have been intensified in the last few years.

Relying on the fact that the unconditional distribution of daily returns has fatter tails than the normal distribution, the usual time series models that assume normality and homoscedasticity are not appropriate to model volatility. Thus, more adequate procedures, especially the ones presenting conditional variance evolving on time, have been proposed. The most known approaches are the ones concerning conditional heteroscedastic models, such as ARCH [1], GARCH [2], EGARCH [3], TGARCH [4] and multivariate GARCH [5].

Taylor [6] proposed the first stochastic volatility model, where the volatility is a stochastic function of the past volatility. Several studies on this approach have been developed, such as [7–11] and [12].

* Correspondence to: Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte, 31270-901, MG, Brazil. Tel.: +55 31 34095949; fax: +55 31 34095924.
E-mail address: glauraf@ufmg.br (G.C. Franco).

http://dx.doi.org/10.1016/j.matcom.2015.08.005

Recently, a non-Gaussian state space model was proposed by [13]. This procedure is a generalization of a result of [14], who proposed an exponential observation model with an exact evolution equation for the state. The work of [13] allows for analytical computation of the marginal likelihood, which increases the applicability of the model and enables its use with a wide class of distributions for observational time series. Additionally, this model allows the relaxation of the normality and heteroscedasticity assumptions.

According to [15], one of the main characteristics of volatility is that it evolves over time in a continuous way and it always varies within a fixed range. This means that volatility is often stationary. Due to the structure used in the model proposed by [13], the only stochastic component is the level of the series, and it is built in a way similar to the local level model of [16]. Thus, the model is highly recommended to be applied to stationary series. Any other component, such as seasonality or structural breaks should be inserted as covariates.

There are some recent contributions in the literature that employ the state space approach to handle non-linear and non-Gaussian time series. Some examples are the works of [17], extended by [18] for Bayesian estimation, that uses a local scale procedure for modeling volatility. Ferrante and Vidoni [19] and Vidoni [20] introduce non-linear and non-Gaussian state space models with analytic updating recursions for filtering and prediction.

Thus, the purpose of this work is to present new models in the non-Gaussian state space family that can be used to model volatility. Among them, there is the class of heavy tailed distributions, much employed in the volatility literature, as in the works of [21] and [22]. Then, the first contribution in this paper is to show that the Fréchet, Lévy, Log-gamma, Log-normal, and the Generalized Error Distribution (GED) can be written as a NGSSM to model volatility and they are used in many applications of the financial market. In addition, the Pareto and Weibull models, already considered in [13], are also presented.

The second contribution is on Monte Carlo results for Bayesian and classical methods of inference in the parameter estimation of the distributions cited above in the non-Gaussian state space model.

Additionally, as another contribution of the paper, the NGSSM addressed here is used to model the most known stock exchange indexes in North and South America and the fits are compared to the classical generalized autoregressive conditional heteroscedasticity (GARCH) [2] models, its exponential counterpart (EGARCH) [3] and the stochastic volatility model [10].

The paper is organized as follows. Section 2 defines the NGSSM and presents the inference procedures. Section 3 shows how to write the heavy tailed distributions cited above in the NGSSM form. Section 4 shows the results of the Monte Carlo simulation studies and Section 5 presents an application of heavy tailed models in the NGSSM to estimate the volatility of several stock exchange indexes. Section 6 concludes the work.

2. A non-Gaussian state space model

Gamerman et al. [13] define a new family of non-Gaussian state space models, which is a generalization of the works of [14] and [23]. Let $\{y_t\}_{t=1}^{n}$ be a time series with probability function given by

$$p\left(y_t \mid \mu_t, \boldsymbol{\varphi}\right) = \begin{cases} q\left(y_t, \boldsymbol{\varphi}\right) \mu_t^{r(y_t, \boldsymbol{\varphi})} \exp\left\{-\mu_t s\left(y_t, \boldsymbol{\varphi}\right)\right\}, & y_t \in H\left(\boldsymbol{\varphi}\right) \subset \mathbb{R} \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where $\boldsymbol{\varphi}$ is a p-dimensional parameter vector, $\boldsymbol{\varphi} = \left(\varphi_1, \ldots, \varphi_p\right)'$, and functions $q\left(y_t, \boldsymbol{\varphi}\right)$, $r\left(y_t, \boldsymbol{\varphi}\right)$, $s\left(y_t, \boldsymbol{\varphi}\right)$ and $H\left(\boldsymbol{\varphi}\right)$ are such that $p\left(y_t \mid \mu_t, \boldsymbol{\varphi}\right) \geqslant 0$ and the Lebesgue–Stieltjes integral $\int p\left(y_t \mid \mu_t, \boldsymbol{\varphi}\right) dy_t = 1$. If $r\left(y_t, \boldsymbol{\varphi}\right) = r\left(\boldsymbol{\varphi}\right)$, $s\left(y_t, \boldsymbol{\varphi}\right) = s\left(\boldsymbol{\varphi}\right)$ and $H\left(\boldsymbol{\varphi}\right)$ is a constant function (it does not depend on $\boldsymbol{\varphi}$), the distribution family becomes a special case of the exponential family.

The NGSSM considers $\{y_t\}_{t=1}^{n}$ following the distribution in Eq. (1) with the state given by

$$\mu_t = \lambda_t g\left(\boldsymbol{x}_t, \boldsymbol{\beta}\right), \quad \text{for } t = 1, \ldots, n,$$

where g is the link function, $\boldsymbol{x}_t$ is a vector of covariates and $\boldsymbol{\beta}$ (one of the components of $\boldsymbol{\varphi}$) is the regression coefficient vector.

The dynamic level λ_t is given by the evolution equation $\lambda_t = \omega^{-1}\lambda_{t-1}\varsigma_t$, with the *prior* specification $\lambda_0 \mid Y_0 \sim \text{Gamma}\left(a_0;\ b_0\right)$. In this case, $\varsigma_t \sim \text{Beta}\left(\omega a_{t-1},\ (1-\omega)\, a_{t-1}\right)$, that is

$$\omega \left.\frac{\lambda_t}{\lambda_{t-1}}\right| \lambda_{t-1}, \boldsymbol{Y_{t-1}} \sim \text{Beta}\left(\omega a_{t-1},\ (1-\omega)\, a_{t-1}\right), \quad \text{for } t = 1, \ldots, n,$$

where $\boldsymbol{Y_{t-1}} = \{y_{t-1}, \ldots, y_1\}$ for $t \geqslant 1$, $0 < \omega < 1$ and Y_0 is the initial information. Parameter ω has the function of increasing multiplicatively the variance over time.

Taking the logarithm of the evolution equation, λ_t, it can be seen that it is the random walk equation used for the local level model [16], that is

$$\ln(\lambda_t) = \ln(\lambda_{t-1}) + \xi_t,$$

where $\xi_t = \ln(\varsigma_t/\omega) \in \mathbb{R}$.

Theorem 1 in [13] presents the equations for the exact evolution of the dynamic level and the predictive density function for the NGSSM, which are as follows.

1. The *prior* distribution $\lambda_t \,|\, \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi} \sim \text{Gamma}\left(a_{t|t-1};\ b_{t|t-1}\right)$, where
$$a_{t|t-1} = \omega a_{t-1} \quad \text{and} \quad b_{t|t-1} = \omega b_{t-1}.$$
2. The *prior* distribution $\mu_t \,|\, \boldsymbol{Y_{t-1}}, \boldsymbol{\varphi} \sim \text{Gamma}\left(c_{t|t-1};\ d_{t|t-1}\right)$, where
$$c_{t|t-1} = \omega a_{t-1} \quad \text{and} \quad d_{t|t-1} = \omega b_{t-1}\left[g\left(\boldsymbol{x}_t,\ \boldsymbol{\beta}\right)\right]^{-1}.$$
They are easily obtained from Eq. (1) and the scale property of the Gamma distribution.
3. The *posterior* distribution $\lambda_t = \mu_t\left[g\left(\boldsymbol{x}_t,\ \boldsymbol{\beta}\right)\right]^{-1} \Big| \boldsymbol{Y_t}, \boldsymbol{\varphi} \sim \text{Gamma}\left(a_t;\ b_t\right)$ where
$$a_t = a_{t|t-1} + r\left(y_t,\ \boldsymbol{\varphi}\right) \quad \text{and} \quad b_t = b_{t|t-1} + s\left(y_t,\ \boldsymbol{\varphi}\right) g\left(\boldsymbol{x}_t,\ \boldsymbol{\beta}\right).$$
4. The *posterior* distribution $\mu_t \,|\, \boldsymbol{Y_t}, \boldsymbol{\varphi} \sim \text{Gamma}\left(c_t;\ d_t\right)$, where
$$c_t = c_{t|t-1} + r\left(y_t,\ \boldsymbol{\varphi}\right) \quad \text{and} \quad d_t = d_{t|t-1} + s\left(y_t,\ \boldsymbol{\varphi}\right).$$
5. The predictive density function is given by
$$p\left(y_t \,|\, \boldsymbol{Y_{t-1}},\ \boldsymbol{\varphi}\right) = \frac{\Gamma\left(r\left(y_t,\ \boldsymbol{\varphi}\right) + c_{t|t-1}\right) q\left(y_t,\ \boldsymbol{\varphi}\right) d_{t|t-1}^{c_{t|t-1}} I_{(y_t \in H(\boldsymbol{\varphi}))}}{\Gamma\left(c_{t|t-1}\right)\left[s\left(y_t,\ \boldsymbol{\varphi}\right) + d_{t|t-1}\right]^{r(y_t,\ \boldsymbol{\varphi}) + c_{t|t-1}}}. \tag{2}$$

2.1. Inference procedure

Parameter inference in the NGSSM can be performed either using Bayesian or classical procedures. Both are based on the likelihood function

$$L\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right) = \prod_{t=1}^{n} p\left(y_t \,|\, \boldsymbol{Y_{t-1}},\ \boldsymbol{\varphi}\right),$$

where $p\left(y_t \,|\, \boldsymbol{Y_{t-1}},\ \boldsymbol{\varphi}\right)$ is given in Eq. (2).

- **Classical inference**

Classical inference for the parameters of the NGSSM is performed through maximum likelihood estimation. The log-likelihood function is calculated as

$$\begin{aligned}\ell\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right) = &\sum_{t=1}^{n} \ln \Gamma\left(r\left(y_t,\ \boldsymbol{\varphi}\right) + c_{t|t-1}\right) + \sum_{t=1}^{n} \ln\left(q\left(y_t,\ \boldsymbol{\varphi}\right)\right) \\ &- \sum_{t=1}^{n} \ln \Gamma\left(c_{t|t-1}\right) + \sum_{t=1}^{n} c_{t|t-1} \ln\left(b_{t|t-1}\right) \\ &- \sum_{t=1}^{n}\left[r\left(y_t,\ \boldsymbol{\varphi}\right) + c_{t|t-1}\right] \ln\left[s\left(y_t,\ \boldsymbol{\varphi}\right) + d_{t|t-1}\right],\end{aligned}$$

where $a_0 > 0$ and $b_0 > 0$ (see [13]). Thus, the maximum likelihood estimator (MLE) for $\boldsymbol{\varphi}$ is given by

$$\hat{\boldsymbol{\varphi}}_{ML} = \arg\max_{\boldsymbol{\varphi}} \ell\left(\boldsymbol{\varphi}; \boldsymbol{Y_n}\right).$$

Due to the fact that $\ell\,(\boldsymbol{\varphi};\boldsymbol{Y_n})$ is a non-linear function of $\boldsymbol{\varphi}$, numerical procedures such as the BFGS algorithm [24] should be used.

The asymptotic confidence interval for $\boldsymbol{\varphi}$ is built based on a numerical approximation for the Fisher information matrix $I_n(\boldsymbol{\varphi})$, using $I_n(\boldsymbol{\varphi}) \cong -G(\boldsymbol{\varphi})$, where $-G(\boldsymbol{\varphi})$ is the matrix of second derivatives of the log-likelihood function with respect to the parameters. Let φ_i, $i = 1, \ldots, p$, be any component of $\boldsymbol{\varphi}$. Then, an asymptotic confidence interval of $100(1-\kappa)\%$ for φ_i is given by

$$\hat{\varphi}_i \pm z_{\kappa/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\varphi}_i)},$$

where $z_{\kappa/2}$ is the $\kappa/2$ percentile of the standard normal distribution and $\widehat{Var}(\hat{\varphi}_i)$ is obtained from the diagonal elements of the Fisher information matrix.

- **Bayesian inference**

The *posterior* distribution $\pi\,(\boldsymbol{\varphi}|\,\boldsymbol{Y_n})$ of the parameter vector $\boldsymbol{\varphi}$ is given by

$$\pi\,(\boldsymbol{\varphi}|\,\boldsymbol{Y_n}) = \frac{L\,(\boldsymbol{\varphi};\boldsymbol{Y_n})\,\pi\,(\boldsymbol{\varphi})}{\int L\,(\boldsymbol{\varphi};\boldsymbol{Y_n})\,\pi\,(\boldsymbol{\varphi})\,d\boldsymbol{\varphi}},$$

where $L\,(\boldsymbol{\varphi};\boldsymbol{Y_n})$ is the likelihood function and $\pi\,(\boldsymbol{\varphi})$ is the *prior* distribution for $\boldsymbol{\varphi}$. In this paper a proper and non informative Uniform distribution with respect to Bayes–Laplace is used. It is given by $\pi\,(\boldsymbol{\varphi}) = c$ for all possible values of $\boldsymbol{\varphi}$ in a determined range and 0 otherwise. The Bayesian estimates of the posterior mean (BE-Mean), the posterior median (BE-Median) and the credibility interval are obtained from a sample of the posterior distribution. The adaptive random walk Metropolis (ARWM) algorithm proposed by [25], (see also [26]) has been used to sample from the posterior distribution.

The ARWM works as follows. Suppose that given some initial $\boldsymbol{\varphi}_0$ from $\pi(\boldsymbol{\varphi}|\boldsymbol{Y}_n)$, the $j-1$ iterates $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{j-1}$ have been generated. The jth iterate $\boldsymbol{\varphi}_j$ is generated from the proposal density $\eta_j(\boldsymbol{\varphi}|\boldsymbol{\psi})$ which may also depend on some other value of $\boldsymbol{\varphi}$ which is called $\boldsymbol{\psi}$. Let $\boldsymbol{\varphi}_j^p$ be the proposed value of $\boldsymbol{\varphi}_j$ generated from $\eta_j(\boldsymbol{\varphi}|\boldsymbol{\varphi}_{j-1})$. Then $\boldsymbol{\varphi}_j = \boldsymbol{\varphi}_j^p$ is taken with probability

$$\alpha(\boldsymbol{\varphi}_j^p, \boldsymbol{\varphi}_{j-1}) = \min\left\{1, \frac{\pi(\boldsymbol{\varphi}_j^p|\boldsymbol{Y}_j)}{\pi(\boldsymbol{\varphi}_{j-1})}\frac{\eta_j(\boldsymbol{\varphi}_{j-1}|\boldsymbol{\varphi}_j^p)}{\eta_j(\boldsymbol{\varphi}_j^p|\boldsymbol{\varphi}_{j-1})}\right\}, \tag{3}$$

and $\boldsymbol{\varphi}_j = \boldsymbol{\varphi}_{j-1}$ otherwise. In adaptive sampling the parameters of $\eta_j(\boldsymbol{\varphi}|\boldsymbol{\psi})$ are estimated from the iterates $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{j-2}$. Under appropriate regularity conditions the sequence of iterates $\boldsymbol{\varphi}_j$, $j \geqslant 1$, converges to draws from the target distribution $\pi(\boldsymbol{\varphi}|\boldsymbol{Y}_n)$. The proposal distribution in the ARWM algorithm used in this paper is given by a mixture of two normal distributions with mean components given by $\boldsymbol{\varphi}_{j-1}$. The first component has a small weight and a fixed covariance matrix while the second component has more weight, say 0.95, and a covariance matrix that is updated as iteration goes. For more details see [25].

Credibility intervals for φ_i, $i = 1, \ldots, p$ are built as follows. Given a value $0 < \kappa < 1$, the interval $[c_1, c_2]$ satisfying

$$\int_{c_1}^{c_2} \pi(\varphi_i \mid \boldsymbol{Y}_n)\,d\varphi_i = 1 - \kappa$$

is a credibility interval for φ_i with level $100(1-\kappa)\%$.

- **Model selection**

The adequacy of the model should be checked after fitting a model to data set. There are many methods of diagnosis suggested in the literature, and some of them are described below.

Harvey and Fernandes [23] suggested a diagnosis method based on the standardized residuals, also known as Pearson residuals, which are defined as:

$$r_t^p = \frac{y_t - \mathrm{E}\,(y_t\,|\boldsymbol{Y_{t-1}}, \boldsymbol{\varphi})}{\sqrt{\mathrm{Var}\,(y_t\,|\boldsymbol{Y_{t-1}}, \boldsymbol{\varphi})}}. \tag{4}$$

The authors propose the following residual analysis:

1. Examine the plot of residuals vs. time and residuals vs. an estimate of the level component.
2. Verify if the sample variance of the standardized residuals is close 1. A value greater than 1 indicates overdispersion.

Another alternative is to use the deviance residuals [27], which are given by:

$$r_t^d = \left\{ 2\ln\left[\frac{p\left(y_t\,|y_t,\boldsymbol{\varphi}\right)}{p\left(y_t\left|\hat{\phi}_t,\boldsymbol{\varphi}\right.\right)}\right]\right\}^{\frac{1}{2}},$$

where $\hat{\phi}_t = \mathrm{E}\left(y_t\,|\boldsymbol{Y_{t-1}},\boldsymbol{\varphi}\right)$.

When two or more models present reasonable fits to the data, it is necessary to choose one of them. According to [16] the AIC and BIC criteria proposed, respectively, by [28] and [29], are suitable procedures. They are defined by:

$$AIC = -2l\left(\hat{\boldsymbol{\varphi}}\right) + 2k \quad \text{and} \quad BIC = -2l\left(\hat{\boldsymbol{\varphi}}\right) + k\ln(n),$$

where $l(\cdot)$ is the log-likelihood function, k the number of parameters and n the number of observations. For the Bayesian approach, $\hat{\boldsymbol{\varphi}}$ was considered as the posterior mean. Instead of AIC, it was used the AICc [30], given by

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}.$$

3. Heavy tailed distributions in the NGSSM

In this section, some of the most used heavy tailed distributions, such as the Fréchet, Lévy, Log-gamma, Log-normal, Pareto, Skew Generalized Normal (Skew GED), and Weibull, are discussed and they are proved to belong to the NGSSM.

The main characteristic of this kind of distribution is that it presents heavier tails than the normal distribution. The formal definition, found in [31], is as follows.

Definition 3.1. A distribution function, F_X, of a random variable X belongs to the class of heavy right tail if $\lim_{x\to\infty} e^{\lambda x}\left[1 - F_X(x)\right] = \infty$, for all $\lambda > 0$. This is equivalent to state that the moment generating function, $\widehat{F_X}(s)$, of F is infinite for all $s > 0$.

Teugels [32], Embrechts et al. [33] and Goldie and Klüppelberg [34], among others, presented a wide discussion about heavy tailed distribution properties and applications. Neyman and Scott [35] and Green [36] showed that there is a close relationship between the heavy tailed distribution family and the absolute or relative distribution outliers prone. That is, probability distributions that are contained in the heavy tailed distribution family are more likely to generate outliers.

For each model described below, the μ_t component, which corresponds to the state equation in the NGSSM, is either the precision or the scale parameter of the respective distribution. Note that the variance of y_t given μ_t depends directly on μ_t and not on all the past information $\boldsymbol{Y}_{t-1}$. Thus, given the prior $\mu_t|\boldsymbol{Y}_{t-1},\boldsymbol{\varphi}$, the variance of $y_t|\mu_t$ remains unchanged in terms of notation, i.e. $\mathrm{Var}(y_t|\mu_t) = \mathrm{Var}(y_t|\mu_t, Y_{t-1})$. This feature is highly appealing with respect of the use of these models to describe volatility.

3.1. Fréchet model

If a time series $\{y_t\}_{t=1}^n$ is generated from a Maximum Fréchet distribution with shape parameter $\alpha_t = \alpha$, location parameter $\gamma_t = \gamma$, unknown and invariant in time, and scale parameter μ_t^α, restricted to $\gamma < y_t$, $\alpha > 0$ and $\mu_t^\alpha > 0$, then

$$p\left(y_t|\,\mu_t,\,\boldsymbol{\varphi}\right) = \alpha\mu_t^{-1}\left(\frac{\mu_t}{y_t-\gamma}\right)^{\alpha+1}\exp\left\{-\left(\frac{\mu_t}{y_t-\gamma}\right)^{\alpha+1}\right\} I_{(\gamma<y_t<\infty)},$$

where $\mu_t^\alpha = \lambda_t g\left(\boldsymbol{x}_t,\,\boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi} = (\omega,\,\boldsymbol{\beta},\,\alpha,\,\gamma)'$.

The Maximum Fréchet model can be written in the NGSSM form as

$$q\left(y_t,\ \boldsymbol{\varphi}\right)=\alpha\left(y_t-\gamma\right)^{-\alpha-1},\qquad r\left(y_t,\ \boldsymbol{\varphi}\right)=1\quad\text{and}\quad s\left(y_t,\ \boldsymbol{\varphi}\right)=\left(y_t-\gamma\right)^{-\alpha}.$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)=\prod_{t=1}^{n}\left\{\frac{\Gamma\left(1+c_{t|t-1}\right)\alpha\left(y_t-\gamma\right)^{-\alpha-1}\left(d_{t|t-1}\right)^{c_{t|t-1}}I_{(\gamma<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(\left(y_t-\gamma\right)^{-\alpha}+d_{t|t-1}\right)^{1+c_{t|t-1}}}\right\}.$$

The Minimum Fréchet model can also be easily written in the NGSSM form, just changing $(y_t-\gamma)$ for $(\gamma-y_t)$ and using the restriction $\gamma>y_t$ instead of $\gamma<y_t$.

3.2. Lévy model

If a time series $\{y_t\}_{t=1}^n$ is generated from a Lévy distribution with location parameter $\gamma_t=\gamma$, unknown and invariant in time, and precision parameter μ_t, restricted to $\mu_t>0$ and $y_t>\gamma$, then

$$p\left(y_t|\,\mu_t,\ \boldsymbol{\varphi}\right)=\frac{\mu_t^{\frac{1}{2}}}{\sqrt{2\pi\left(y_t-\gamma\right)^3}}\exp\left\{-\mu_t\left[2\left(y_t-\gamma\right)\right]^{-1}\right\}I_{(\gamma<y_t<\infty)},$$

where $\mu_t=\lambda_t g\left(\boldsymbol{x}_t,\ \boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi}=(\omega,\ \boldsymbol{\beta},\ \gamma)'$.

The Lévy model can be written in the NGSSM form as

$$q\left(y_t,\boldsymbol{\varphi}\right)=(2\pi)^{-1/2}(y_t-\gamma)^{-3/2},\qquad r\left(y_t,\boldsymbol{\varphi}\right)=\frac{1}{2}\quad\text{and}\quad s\left(y_t,\boldsymbol{\varphi}\right)=\left[2\left(y_t-\gamma\right)\right]^{-1}.$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)=\prod_{t=1}^{n}\left\{\frac{\Gamma\left(\frac{1}{2}+c_{t|t-1}\right)\left(y_t-\gamma\right)]^{-3/2}\left(d_{t|t-1}\right)^{c_{t|t-1}}I_{(\gamma<y_t<\infty)}}{(2\pi)^{1/2}\Gamma\left(c_{t|t-1}\right)\left(\left[2\left(y_t-\gamma\right)\right]^{-1}+d_{t|t-1}\right)^{\frac{1}{2}+c_{t|t-1}}}\right\}.$$

3.3. Log-gamma model

The Log-gamma distribution was presented by [37]. If a time series $\{y_t\}_{t=1}^n$ is generated from a Log-gamma distribution with shape parameter $\alpha_t=\alpha$, unknown and invariant in time, and scale parameter $\alpha\mu_t$, restricted to $\alpha>0$ and $\alpha\mu_t>0$, then

$$p\left(y_t|\,\mu_t,\ \boldsymbol{\varphi}\right)=\frac{\left(\alpha\mu_t\right)^{\alpha}\left[\ln\left(y_t\right)\right]^{\alpha-1}}{\Gamma\left(\alpha\right)y_t^{\alpha\mu_t+1}}I_{(1<y_t<\infty)},$$

where $\mu_t=\lambda_t g\left(\boldsymbol{x}_t,\ \boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi}=(\omega,\ \boldsymbol{\beta},\ \alpha)'$.

The Log-gamma model can be written in the NGSSM form as

$$q\left(y_t,\ \boldsymbol{\varphi}\right)=\alpha^{\alpha}\left[\ln\left(y_t\right)\right]^{\alpha-1}\left[\Gamma\left(\alpha\right)y_t\right]^{-1},\ r\left(y_t,\ \boldsymbol{\varphi}\right)=\alpha\quad\text{and}\quad s\left(y_t,\ \boldsymbol{\varphi}\right)=\alpha\ln\left(y_t\right).$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\boldsymbol{Y_n}\right)=\prod_{t=1}^{n}\left\{\frac{\Gamma\left(\alpha+c_{t|t-1}\right)\alpha^{\alpha}\left[\ln\left(y_t\right)\right]^{\alpha-1}\left[\Gamma\left(\alpha\right)y_t\right]^{-1}d_{t|t-1}^{c_{t|t-1}}I_{(1<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(\alpha\ln\left(y_t\right)+d_{t|t-1}\right)^{\alpha+c_{t|t-1}}}\right\}.$$

3.4. Log-normal model

If a time series $\{y_t\}_{t=1}^n$ is generated from a Log-normal distribution with location parameter $\delta_t=\delta$, shape parameter $\gamma_t=\gamma$, unknown and invariant in time, and precision parameter σ_t^{-2}, restricted to $\sigma_t^{-2}=\mu_t>0$

and $\gamma < y_t$, then

$$p\left(y_t|\,\mu_t,\,\boldsymbol{\varphi}\right) = \frac{\mu_t^{\frac{1}{2}}}{\left(y_t-\gamma\right)\sqrt{2\pi}} \exp\left\{-\mu_t \frac{\left[\ln\left(y_t-\gamma\right)-\delta\right]^2}{2}\right\} I_{(\gamma<y_t<\infty)},$$

where $\mu_t = \lambda_t g\left(\boldsymbol{x}_t,\,\boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi} = (\omega,\,\boldsymbol{\beta},\,\delta,\,\gamma)'$.

The Log-normal model can be written in the NGSSM form as

$$q\left(y_t,\,\boldsymbol{\varphi}\right) = \left(y_t-\gamma\right)^{-1}/\sqrt{2\pi},\; r\left(y_t,\,\boldsymbol{\varphi}\right) = \frac{1}{2} \quad \text{and} \quad s\left(y_t,\,\boldsymbol{\varphi}\right) = \frac{\left[\ln\left(y_t-\gamma\right)-\delta\right]^2}{2}.$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\,\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\,\boldsymbol{Y_n}\right) = \prod_{t=1}^{n}\left\{\frac{\Gamma\left(\frac{1}{2}+c_{t|t-1}\right)\left[\left(y_t-\gamma\right)\sqrt{2\pi}\right]^{-1} d_{t|t-1}^{c_{t|t-1}} I_{(\gamma<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(d_{t|t-1}+\left[\ln\left(y_t-\gamma\right)-\delta\right]^2/2\right)^{\frac{1}{2}+c_{t|t-1}}}\right\}.$$

3.5. Pareto model

If a time series $\{y_t\}_{t=1}^n$ is generated from a Pareto distribution with scale parameter μ_t, restricted to $y_t > 1$, then

$$p\left(y_t|\,\mu_t,\,\boldsymbol{\varphi}\right) = \mu_t y_t^{-\mu_t-1} I_{(1<y_t<\infty)},$$

where $\mu_t = \lambda_t g\left(\boldsymbol{x}_t,\,\boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi} = (\omega,\,\boldsymbol{\beta})'$.

The Pareto model can be written in the NGSSM form as

$$q\left(y_t,\,\boldsymbol{\varphi}\right) = y_t^{-1}, \qquad r\left(y_t,\,\boldsymbol{\varphi}\right) = 1 \quad \text{and} \quad s\left(y_t,\,\boldsymbol{\varphi}\right) = \ln\left(y_t\right).$$

Thus the likelihood function $L\left(\boldsymbol{\varphi};\,\boldsymbol{Y_n}\right)$ is given by

$$L\left(\boldsymbol{\varphi};\,\boldsymbol{Y_n}\right) = \prod_{t=1}^{n}\left\{\frac{\Gamma\left(1+c_{t|t-1}\right) y_t^{-1} d_{t|t-1}^{c_{t|t-1}} I_{(1<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(\ln\left(y_t\right)+d_{t|t-1}\right)^{1+c_{t|t-1}}}\right\}.$$

3.6. Skew GED model

The Skew Generalized Normal Distribution (Skew GED) is also known as the Skew Exponential Power Distribution. If a time series $\{y_t\}_{t=1}^n$ is generated from a Skew GED distribution with location parameter $\delta_t = \delta$, shape parameter $\alpha_t = \alpha$ and asymmetry parameter $\kappa_t = \kappa$, all of them unknown and invariant in time, and precision parameter μ_t, restricted to $\alpha > 0$, $\kappa > 0$ and $\mu_t > 0$, then

$$p\left(y_t|\,\mu_t,\,\boldsymbol{\varphi}\right) = \frac{\kappa \mu_t^{\frac{1}{\alpha}}}{\Gamma\left(1+\alpha^{-1}\right)\left(1+\kappa^2\right)} \exp\left\{-\mu_t\left\{\left[\kappa z_t^+\right]^\alpha + \left[z_t^-/\kappa\right]^\alpha\right\}\right\} I_{(y_t\in\Re)},$$

where $z_t = y_t - \delta$, $\mu_t = \lambda_t g\left(\boldsymbol{x}_t,\,\boldsymbol{\beta}\right)$ and $\boldsymbol{\varphi} = (\omega,\,\boldsymbol{\beta},\,\delta,\,\alpha,\,\kappa)'$,

$$u^+ = \begin{cases} u, & \text{if } u \geqslant 0 \\ 0, & \text{if } u < 0 \end{cases} \quad \text{and} \quad u^- = \begin{cases} -u, & \text{if } u \leqslant 0 \\ 0, & \text{if } u > 0. \end{cases}$$

The Skew GED includes the Skew Normal distribution ($\alpha = 2$, $\kappa \neq 1$), the Normal distribution ($\alpha = 2$, $\kappa = 1$), the Skew Laplace distribution ($\alpha = 1$, $\kappa \neq 1$), the Laplace distribution ($\alpha = 1$, $\kappa = 1$) and the Uniform distribution ($\alpha \to \infty$).

The Skew GED model can be written in the NGSSM form as

$$q\left(y_t,\,\boldsymbol{\varphi}\right) = \frac{\kappa}{\Gamma\left(1+\alpha^{-1}\right)\left(1+\kappa^2\right)}, \qquad r\left(y_t,\,\boldsymbol{\varphi}\right) = \frac{1}{\alpha} \quad \text{and} \quad s\left(y_t,\,\boldsymbol{\varphi}\right) = \left[\kappa z_t^+\right]^\alpha + \left[z_t^-/\kappa\right]^\alpha,$$

with $z_t = y_t - \delta$.

Thus the likelihood function $L(\boldsymbol{\varphi}; \boldsymbol{Y_n})$ is given by

$$L(\boldsymbol{\varphi}; \boldsymbol{Y_n}) = \prod_{t=1}^{n} \left\{ \frac{\Gamma\left(1/\alpha + c_{t|t-1}\right)\kappa\left[\Gamma\left(1+\alpha^{-1}\right)\left(1+\kappa^2\right)\right]^{-1} d_{t|t-1}^{c_{t|t-1}} I_{y_t \in \Re}}{\Gamma\left(c_{t|t-1}\right)\left(\left[\kappa z_t^{+}\right]^{\alpha} + \left[z_t^{-}/\kappa\right]^{\alpha} + d_{t|t-1}\right)^{1/\alpha + c_{t|t-1}}} \right\}.$$

For details about Skew GED random number generator see [38].

3.7. Weibull model

If a time series $\{y_t\}_{t=1}^{n}$ is generated from a Weibull distribution with location parameter $\upsilon_t = \upsilon$, unknown and invariant in time, and precision parameter μ_t, restricted to $\upsilon > 0$, $\mu_t > 0$ and $y_t > 0$, then

$$p(y_t | \mu_t, \boldsymbol{\varphi}) = \upsilon \mu_t y_t^{\upsilon-1} \exp\left\{-\mu_t y_t^{\upsilon}\right\} I_{(0<y_t<\infty)},$$

where $\mu_t = \lambda_t g(\boldsymbol{x}_t, \boldsymbol{\beta})$ and $\boldsymbol{\varphi} = (\omega, \boldsymbol{\beta}, \upsilon)'$.

The Weibull model can be written in the NGSSM form as

$$q(y_t, \boldsymbol{\varphi}) = \upsilon y_t^{\upsilon-1}, \qquad r(y_t, \boldsymbol{\varphi}) = 1 \quad \text{and} \quad s(y_t, \boldsymbol{\varphi}) = y_t^{\upsilon}.$$

Thus the likelihood function $L(\boldsymbol{\varphi}; \boldsymbol{Y_n})$ is given by

$$L(\boldsymbol{\varphi}; \boldsymbol{Y_n}) = \prod_{t=1}^{n} \left\{ \frac{\Gamma\left(1 + c_{t|t-1}\right)\upsilon y_t^{\upsilon-1} d_{t|t-1}^{c_{t|t-1}} I_{(0<y_t<\infty)}}{\Gamma\left(c_{t|t-1}\right)\left(y_t^{\upsilon} + d_{t|t-1}\right)^{1+c_{t|t-1}}} \right\}.$$

4. Monte Carlo study

In this section the performance of the Fréchet, Lévy, Log-gamma, Log-normal, Pareto, Skew GED and Weibull models is evaluated through a Monte Carlo experiment, using the maximum likelihood estimator (MLE) and the Bayesian estimators, posterior mean (BE-Mean) and posterior median (BE-Median). Asymptotic confidence interval and credibility interval for the parameter vectors are also presented and they are compared with respect to the coverage rates, for a fixed level of 95%.

The number of Monte Carlo replications was set equal to 1000 for time series of size $n = \{100; 200; 500\}$, generated under the *prior* specification $\lambda_0 | Y_0 \sim \text{Gamma}(100.0; 1.0)$, with a covariate $x_t = \sin(2\pi t/12)$, $t = 1, \ldots, n$.

For all distributions $\beta = 1.0$ and $\omega = (0.90, 0.95)$ but only results for $\omega = 0.90$ are presented here, as they were very similar to the case $\omega = 0.95$.

Specific parameters were set as follows: Fréchet($\alpha = 5$); Log-gamma($\alpha = 5$); Log-normal($\delta = 5$); Skew GED($\delta = 5, \alpha = 1.5, \kappa = 1$); Weibull ($\upsilon = 5.0$). For the Fréchet, Lévy and Log-normal, models the parameter γ was fixed at 0.0. For the Skew GED model the parameter α was fixed at 1.5, thus, there is a distribution with a tail heavier than the Skew Normal ($\alpha = 2.0$) and lighter than the Skew Laplace (both are particular cases of the Skew GED).

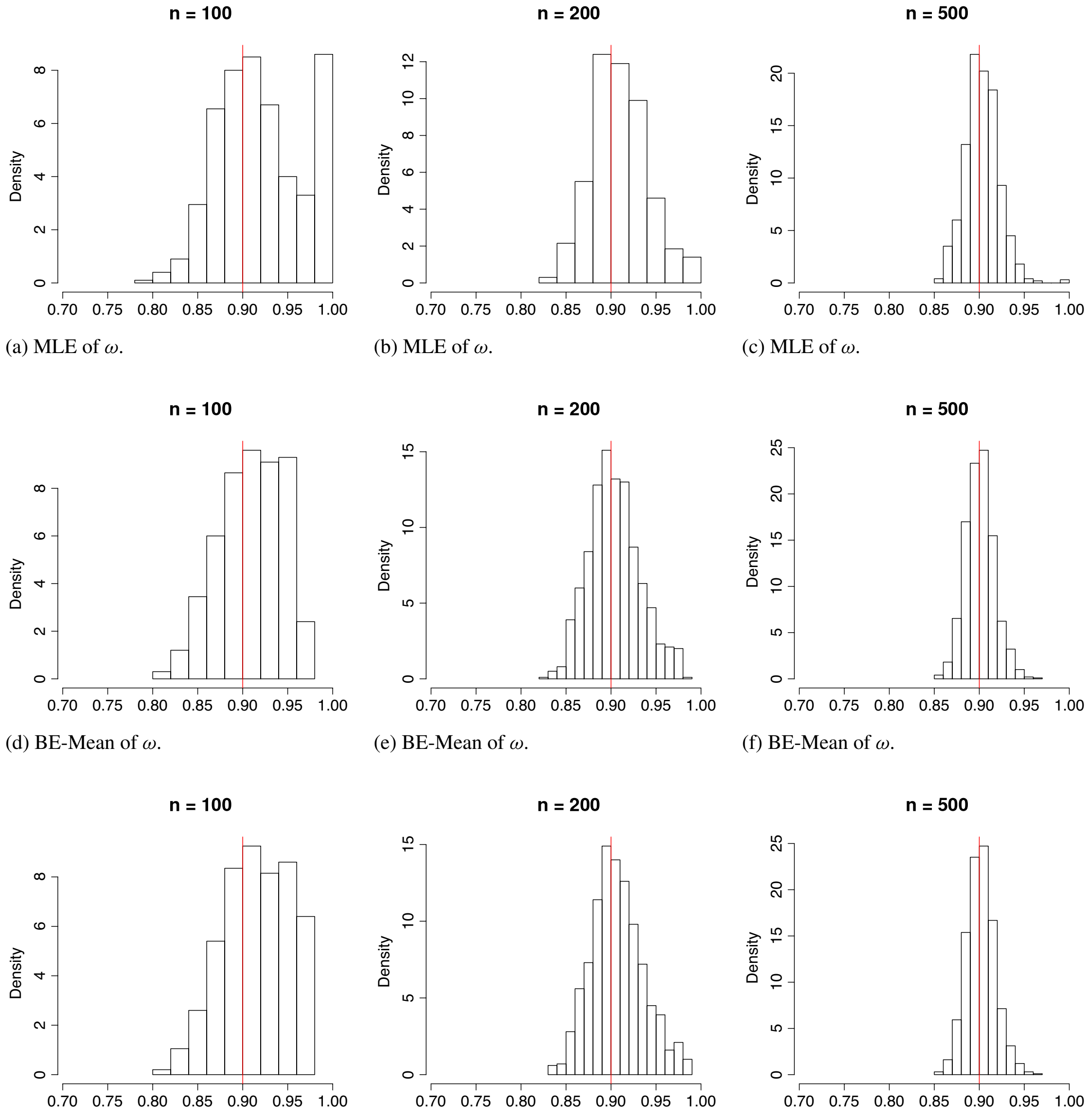
To calculate the maximum likelihood estimator, the BFGS algorithm assumed, as initial state condition $\lambda_0 | Y_0 \sim \text{Gamma}(0.01; 0.01)$, $\omega_0 = 0.50$ and $\beta_0 = \delta_0 = \alpha_0 = \upsilon_0 = \kappa_0 = 0.01$.

For the Bayesian estimation using the ARWM algorithm, chains of size 20,000 were generated with burn in of 5000. The Uniform (−5000; 5000) and Uniform (0; 10,000) are used as the *prior* distribution for the parameters that are defined in $\Re$ and $\Re^{+}$, respectively. More details about the initial conditions in the ARWM algorithm and the Bayesian approach are available from the authors upon request.

Most of the codes for NGSSM were developed by the authors in Ox Metrics®. A few others, specially for the real data applications, were done in R and MATLAB®.

4.1. Empirical distribution of the estimators

In this subsection, the empirical distributions of the MLE and Bayesian estimators for the parameters of the heavy tailed distribution in the NGSSM are investigated for time series of sizes $n = 100$, 200 and 500. As the empirical distribution of the estimators for ω, β and the third parameter (α for Fréchet and Log-gamma, δ for Log-normal and



(a) MLE of ω. (b) MLE of ω. (c) MLE of ω.

(d) BE-Mean of ω. (e) BE-Mean of ω. (f) BE-Mean of ω.

(g) BE-Median of ω. (h) BE-Median of ω. (i) BE-Median of ω.

Fig. 1. Histograms of the estimates (MLE, BE-Mean and BE-Median) of ω for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500.

Skew GED, and υ for Weibull) are very similar for all studied models, then only the results for the Log-normal model are presented here.

Fig. 1 shows the empirical distribution based on 1000 replications of the MLE, BE-Mean and BE-Median estimates for parameter ω. Series of small size shows an asymmetric behavior, always overestimating ω. It can be noted that the mode for the MLE is equal to 1.0. For larger series, the empirical distribution appears symmetric around the real value of the parameter. As expected, the variance decreases as the sample sizes increase.

Figs. 2 and 3 present the empirical distribution of the estimates of parameters β and δ, respectively, for the Log-normal model. The histograms are symmetric around the real value of the parameter for all sample sizes. For parameter δ, the MLE presents larger variability than the Bayesian estimators (this behavior only occurs in the Log-normal and Skew GED models). It can also be observed, as expected, that the variance of the estimates decreases with the increase of the sample size. (See Tables 1–3.)

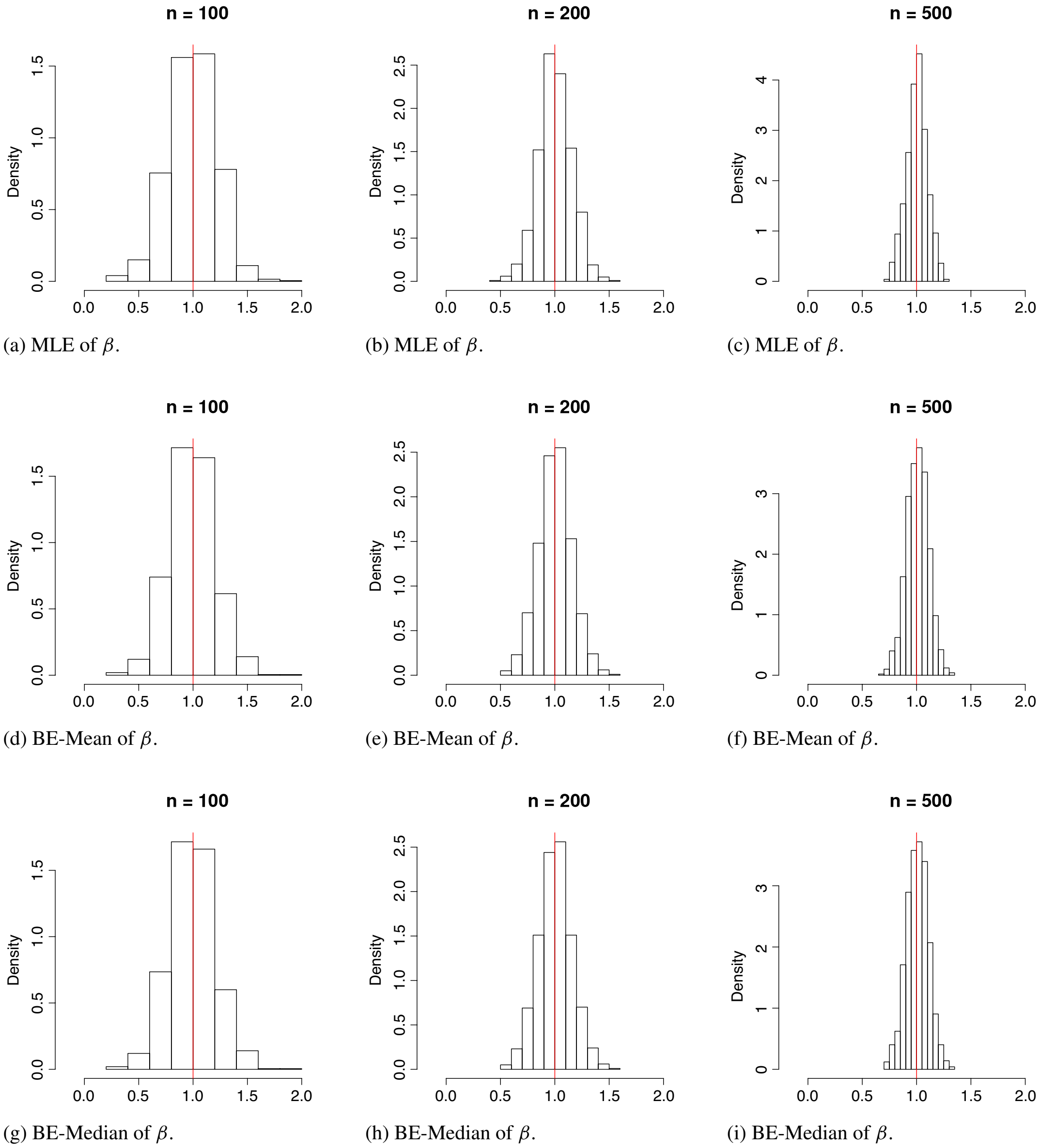
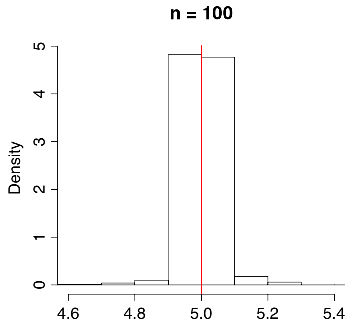


Fig. 2. Histograms of the estimates (MLE, BE-Mean and BE-Median) of β for time series generated from the Log-normal model with ($\omega = 0.90$; $\beta = 1.0$; $\delta = 5.0$) with sizes 100, 200 and 500.

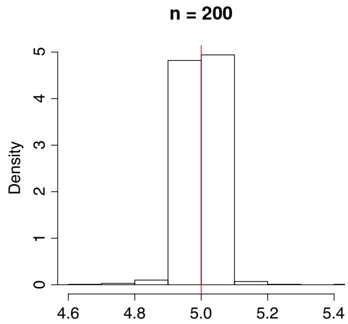
4.2. Point and interval estimation

In this section, point and interval estimation for parameters of the models described in Section 3 are presented. Tables 4–7 show, respectively, the results for the Fréchet, Lévy, Log-gamma, Log-normal, Pareto, Skew GED and Weibull models. The average of 1,000 Monte Carlo replications of the MLE, BE-Mean and BE-Median, along with the mean square error (MSE), are presented. The tables also show the lower and upper limits and coverage rates (Cov Rate) of the asymptotic confidence intervals (Conf Int) and of the confidence credibility intervals (Cred Int). Parameter γ for Féchet, Lévy and Log-normal, and parameter α for the Skew GED distributions were kept fixed in the estimation stage.

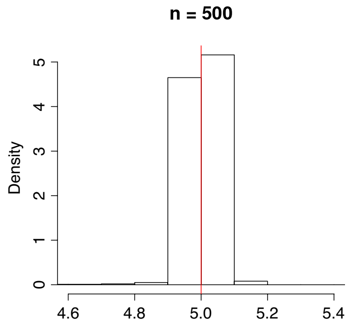
The patterns are very similar for the parameter estimation in all models and therefore the conclusions will be jointly summarized for all cases. It can be observed that the estimation procedures seem consistent, as the MSE decreases as the sample sizes increase.



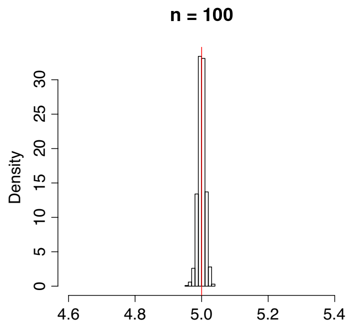
(a) MLE of δ.



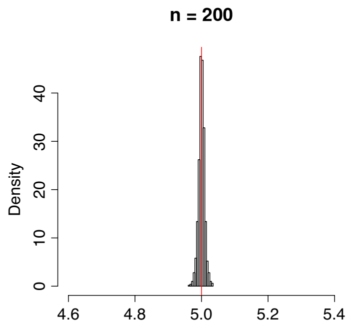
(b) MLE of δ.



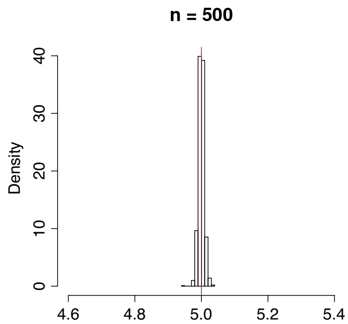
(c) MLE of δ.



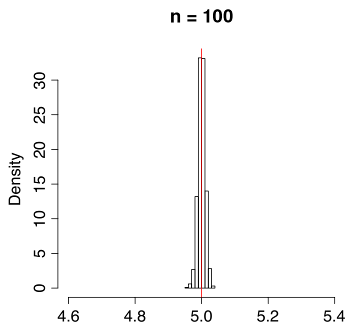
(d) BE-Mean of δ.



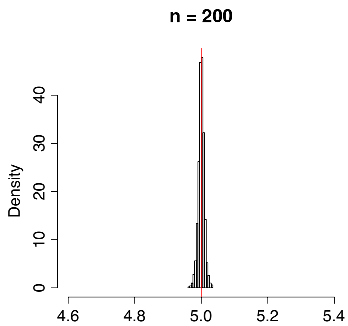
(e) BE-Mean of δ.



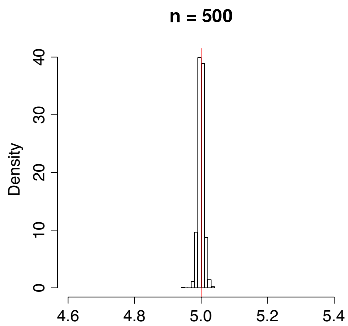
(f) BE-Mean of δ.



(g) BE-Median of δ.



(h) BE-Median of δ.



(i) BE-Median of δ.

Fig. 3. Histograms of the estimates (MLE, BE-Mean and BE-Median) of δ for time series generated from the Log-normal model with ($\omega = 0.90;\ \beta = 1.0;\ \delta = 5.0$) with sizes 100, 200 and 500.

Concerning parameter ω (the first line in all tables and all sample sizes), the MLE seems to consistently overestimate the true value, presenting larger bias and MSE than the Bayesian estimators, for small sample sizes. With respect to the Bayesian estimators, there is not much difference between BE-Mean and BE-Median and they are quite close to the true value of ω even for small samples. Concerning the intervals, it is interesting to note that, for all series of size $n = 100$, the coverage rate of the asymptotic confidence intervals is below the nominal rate and the coverage rate of the credibility intervals is above the nominal rate. For larger sample sizes, the coverage rates of both intervals are close to the 95% level, except the confidence interval for the Log-gamma model with $n = 200$.

Estimates of parameter β (the second parameter in all tables and all sample sizes) do not differ for the MLE and Bayesian estimators and are very close to the real value $\beta = 1.0$ for all models. The Lévy and Log-normal models present the largest MSE values for all sample sizes, while the Log-gamma possesses the smallest ones. Therefore, the limits of the asymptotic confidence and credibility intervals are larger for the Lévy and Log-normal models. The Fréchet, Pareto, Skew GED and Weibull models show the same pattern for the MSE, which are smaller than the values

Table 1
Monte Carlo study for the Fréchet model with ($\omega = 0.9;\ \beta = 1;\ \alpha = 5$).

n	φ	MLE (MSE)	BE-Mean (MSE)	BE-Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
	ω	0.9204	0.9021	0.9096	[0.7391; 0.9681]	[0.7880; 0.9740]
		(0.0029)	(0.0016)	(0.0018)	0.920	0.983
100	β	1.0093	1.0157	1.0145	[0.6752; 1.3433]	[0.6834; 1.3544]
		(0.0312)	(0.0288)	(0.0287)	0.938	0.957
	α	5.0368	5.1230	5.1143	[4.2355; 5.8381]	[4.3475; 5.9506]
		(0.1741)	(0.1719)	(0.1698)	0.940	0.944
	ω	0.9102	0.8988	0.9024	[0.8199; 0.9519]	[0.8263; 0.9509]
		(0.0012)	(0.0010)	(0.0010)	0.954	0.955
200	β	1.0046	1.0141	1.0134	[0.9518; 1.2407]	[0.7776; 1.2543]
		(0.0137)	(0.0161)	(0.0161)	0.956	0.935
	α	5.0106	5.0677	5.0631	[4.4404; 5.5808]	[4.5087; 5.6565]
		(0.0865)	(0.0892)	(0.0889)	0.956	0.946
	ω	0.9028	0.9002	0.9017	[0.8589; 0.9331]	[0.8592; 0.9328]
		(0.0004)	(0.0004)	(0.0004)	0.945	0.941
500	β	1.0004	1.0046	1.0044	[0.8514; 1.1494]	[0.8559; 1.1543]
		(0.0057)	(0.0059)	(0.0059)	0.949	0.949
	α	5.0062	5.0212	5.0190	[4.6437; 5.3688]	[4.6653; 5.3879]
		(0.0336)	(0.0352)	(0.0354)	0.957	0.947

Table 2
Monte Carlo study for the Lévy model with ($\omega = 0.9;\ \beta = 1$).

n	φ	MLE (MSE)	BE-Mean (MSE)	BE-Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
100	ω	0.9188	0.9115	0.9174	[0.7438; 0.9638]	[0.8155; 0.9740]
		(0.0026)	(0.0014)	(0.0017)	0.925	0.987
	β	0.9917	0.9897	0.9900	[0.5671; 1.4164]	[0.5607; 1.4176]
		(0.0496)	(0.0480)	(0.0480)	0.949	0.954
200	ω	0.9090	0.9040	0.9068	[0.8299; 0.9482]	[0.8481; 0.9364]
		(0.0010)	(0.0007)	(0.0008)	0.959	0.953
	β	0.9961	0.9454	0.9455	[0.6966; 1.2956]	[0.9508; 1.2283]
		(0.0238)	(0.0218)	(0.0218)	0.938	0.963
500	ω	0.9035	0.9015	0.9027	[0.8658; 0.9308]	[0.8658; 0.9306]
		(0.0003)	(0.0003)	(0.0003)	0.950	0.948
	β	0.9989	0.9938	0.9938	[0.8102; 1.1875]	[0.8049; 1.1827]
		(0.0100)	(0.0089)	(0.0089)	0.944	0.962

in the Log-normal but larger than the ones in the Log-gamma models. Nevertheless, the coverage rates are all very close to the 95% fixed level, for all models and all sample sizes.

The third parameter, which depends on the distribution employed, was set equal to 5.0 for all cases, except in the Lévy and Pareto models, where there is no extra parameter. For the Log-normal model, the behavior is the same for all methods and the estimates are very close to 5.0, with very small MSE. The intervals show coverage rates very close to 95% and small width. For the Log-gamma model, the MLE presents a better performance compared to the Bayesian estimators, with smaller MSE. The coverage rates of the intervals are below the 95% nominal level and the widths are the largest ones. The Fréchet and Weibull models present a very similar behavior, with the same magnitude for the estimates. In this case, the MLE is again the procedure with the best performance (smaller bias and MSE).

Table 3
Monte Carlo study for the Log-gamma model with ($\omega = 0.9;\ \beta = 1;\ \alpha = 5$).

n	φ	MLE (MSE)	BE-Mean (MSE)	BE-Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
	ω	0.9245	0.8844	0.8935	[0.7673; 0.9687]	[0.7506; 0.9669]
		(0.0044)	(0.0026)	(0.0026)	0.794	0.960
100	β	0.9977	0.9983	0.9984	[0.8705; 1.1249]	[0.8695; 1.1273]
		(0.0043)	(0.0041)	(0.0041)	0.949	0.954
	α	5.1396	5.3720	5.3265	[3.6782; 6.6009]	[3.9632; 7.0443]
		(0.6493)	(0.7823)	(0.7375)	0.936	0.941
	ω	0.9128	0.8921	0.8964	[0.8286; 0.9536]	[0.8110; 0.9487]
		(0.0020)	(0.0012)	(0.0012)	0.869	0.952
200	β	0.9987	0.9975	0.9975	[0.9084; 1.0890]	[0.9066; 1.0883]
		(0.0021)	(0.0023)	(0.0023)	0.943	0.947
	α	5.0630	5.1783	5.1577	[4.0494; 6.0765]	[4.1986; 6.2794]
		(0.3097)	(0.3310)	(0.3213)	0.937	0.939
	ω	0.9026	0.8970	0.8987	[0.8559; 0.9343]	[0.8523; 0.9320]
		(0.0004)	(0.0004)	(0.0004)	0.952	0.952
500	β	0.9995	1.0000	1.0000	[0.9425; 1.0565]	[0.9430; 1.0570]
		(0.0008)	(0.0008)	(0.0008)	0.948	0.953
	α	5.0292	5.0667	5.0591	[4.3923; 5.6661]	[4.4519; 5.7283]
		(0.1085)	(0.1151)	(0.1139)	0.949	0.938

Concerning the fourth parameter in the Skew GED model, the MSE is larger for the MLE compared to the Bayesian estimators for all sample sizes, although its bias is smaller for sample sizes 100 and 500. The coverage rates are close to the 95% fixed level for all sample sizes.

5. Application to South and North American stock exchange indexes

Heavy tailed models in the NGSSM were fitted to the volatility of the following stock exchange indexes: S&P 500 and NASDAQ (USA), INMEX (Mexico), IBOVESPA (Brazil), MERVAL (Argentina) and IPSA (Chile) comprising the period 02/01/2007 to 05/16/2011. Considering only working days, each series possesses 1101, 1101, 1098, 1078, 1074 and 1092 observations, respectively. Let r_t be the log-return for each of the stock exchange indexes. Then, the square of the log returns $y_t = r_t^2$ were fitted for each NGSSM with the own series r_t with an one-day delay as a covariate ($x_t = r_{t-1}$) and the exponential link function, $g(r_{t-1}, \beta) = \exp\{\beta r_{t-1}\}$.

With the purpose of comparing the models in the NGSSM with some known procedures in the literature, GARCH [2], EGARCH [3] and stochastic volatility models [10] were also fitted to the series.

The GARCH is defined as follows,

$$y_t = \sigma_t \epsilon_t, \quad t = 1, \ldots, n, \tag{5}$$

$$\sigma_t^2 = \eta + \sum_{j=1}^{p} \phi_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \theta_i \epsilon_{t-1}^2, \tag{6}$$

where $\eta > 0$, $\theta_i \geq 0$, $\phi_j \geq 0$ and $\sum_{k=1}^{r} (\theta_k + \phi_k) < 1$ with $i = 1, \ldots, p$, $j = 1, \ldots, q$ and $r = max\,(p, q)$.

The EGARCH has the same Eq. (5), but with the Eq. (6) replaced by

$$\log(\sigma_t^2) = \eta + \sum_{j=1}^{p} \phi_j \log(\sigma_{t-j}^2) + \sum_{i=1}^{q} \theta_i [|\epsilon_t| - \mathrm{E}(|\epsilon_t|)] + \sum_{i=1}^{q} \xi_i \epsilon_t .$$

The following distributions were assumed for ϵ_t: Gaussian, Skew Gaussian, Student-t and Skew Student-t, but only results for the skew versions are presented for brevity.

Table 4
Monte Carlo study for the Log-normal model with ($\omega = 0.9;\ \beta = 1;\ \delta = 5$).

n	φ	MLE (MSE)	BE-Mean (MSE)	BE-Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
	ω	0.9206	0.9090	0.9149	[0.7407; 0.9644]	[0.8121; 0.9728]
		(0.0028)	(0.0013)	(0.0016)	0.916	0.983
100	β	0.9955	0.9915	0.9922	[0.5619; 1.4291]	[0.5575; 1.4223]
		(0.0507)	(0.0436)	(0.0436)	0.948	0.962
	δ	5.0006	5.0001	5.0001	[4.9441; 5.0570]	[4.9792; 5.0209]
		(0.0024)	(0.0001)	(0.0001)	0.932	0.951
	ω	0.9098	0.9039	0.9067	[0.8325; 0.9484]	[0.8429; 0.9490]
		(0.0011)	(0.0008)	(0.0009)	0.958	0.944
200	β	1.0032	1.0029	1.0030	[0.7031; 1.3033]	[0.7011; 1.3045]
		(0.0239)	(0.0246)	(0.0247)	0.944	0.940
	δ	4.9980	5.0002	5.0002	[4.9489; 5.0471]	[4.9832; 5.0171]
		(0.0020)	(0.0001)	(0.0001)	0.946	0.951
	ω	0.9038	0.9006	0.9018	[0.8659; 0.9311]	[0.8651; 0.9296]
		(0.0003)	(0.0003)	(0.0003)	0.949	0.953
500	β	1.0021	1.0076	1.0074	[0.8136; 1.1906]	[0.8183; 1.1968]
		(0.0090)	(0.0102)	(0.0102)	0.951	0.937
	δ	4.9996	4.9999	4.9999	[4.9586; 5.0406]	[4.9847; 5.0151]
		(0.0025)	(0.0001)	(0.0001)	0.944	0.948

Table 5
Monte Carlo study for the Pareto model with ($\omega = 0.9;\ \beta = 1$).

n	φ	MLE (MSE)	BE-Mean (MSE)	BE-Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
100	ω	0.9183	0.9048	0.9115	[0.7351; 0.9655]	[0.8004; 0.9721]
		(0.0026)	(0.0014)	(0.0017)	0.937	0.991
	β	0.9990	0.9941	0.9943	[0.7065; 1.2915]	[0.6967; 1.2899]
		(0.0227)	(0.0221)	(0.0221)	0.952	0.959
200	ω	0.9079	0.9016	0.9049	[0.8239; 0.9486]	[0.8346; 0.9500]
		(0.0011)	(0.0008)	(0.0009)	0.964	0.961
	β	0.9961	0.9995	0.9996	[0.7893; 1.2028]	[0.7914; 1.2073]
		(0.0110)	(0.0108)	(0.0108)	0.950	0.958
500	ω	0.9043	0.8996	0.9009	[0.8640; 0.9329]	[0.8609; 0.9307]
		(0.0003)	(0.0003)	(0.0003)	0.952	0.959
	β	1.0014	1.0013	1.0013	[0.8713; 1.1315]	[0.8709; 1.1318]
		(0.0043)	(0.0046)	(0.0046)	0.955	0.942

The stochastic volatility (SV) model for y_t is defined as follows,

$$
\begin{aligned}
(y_t|h_t) &\sim \text{Log-normal}(\xi, h_t\lambda_t^{-1}), \quad t = 1, \ldots, n \\
h_t &= \alpha + \phi(h_{t-1} - \alpha) + \varepsilon_t, \quad \text{where } \varepsilon_t \sim \text{Normal}(0, \sigma^2) \\
h_1 &\sim \text{Normal}(\alpha, \sigma^2/(1-\phi^2)) \\
\xi &\sim \text{Normal}(0, 10^4) \\
\alpha &\sim \text{Normal}(0, 10^4) \\
\phi &\sim \text{Beta}(20, 1.5) \\
\sigma^{-2} &\sim \text{Gamma}(0.001, 0.001) \\
\lambda_t &\sim \text{Gamma}(\nu/2, \nu/2) \\
\nu &\sim \text{Uniform}(2, 100),
\end{aligned}
$$

Table 6
Monte Carlo study for the Skew GED model with ($\omega = 0.9;\ \beta = 1;\ \delta = 5;\ \kappa = 1$).

n	$\boldsymbol{\varphi}$	MLE (MSE)	BE-Mean (MSE)	BE-Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
	ω	0.9330	0.9051	0.9075	[0.7359; 0.9728]	[0.8321; 0.9631]
		(0.0031)	(0.0012)	(0.0015)	0.913	0.975
	β	1.0113	1.0043	1.0051	[0.6468; 1.3758]	[0.8554; 1.1494]
100		(0.0344)	(0.0057)	(0.0062)	0.945	0.969
	δ	5.0000	4.9998	4.9998	[4.9897; 5.0103]	[4.9981; 5.0016]
		(0.00003)	(0.00000)	(0.00000)	0.931	0.946
	κ	1.0058	1.0206	1.0226	[0.8152; 1.1963]	[0.9618; 1.0474]
		(0.0100)	(0.0035)	(0.0044)	0.945	0.944
	ω	0.9131	0.9045	0.9057	[0.8284; 0.9516]	[0.8527; 0.9539]
		(0.0011)	(0.0006)	(0.0009)	0.962	0.982
	β	1.0063	1.0037	0.0039	[0.7491; 1.2636]	[0.9151; 1.0933]
200		(0.0190)	(0.0038)	(0.0043)	0.934	0.949
	δ	4.9998	4.9999	4.9999	[4.9918; 5.0079]	[4.9988; 5.0013]
		(0.00002)	(0.00000)	(0.00000)	0.945	0.947
	κ	0.9986	1.0119	0.0124	[0.8755; 1.1217]	[0.9860; 1.0377]
		(0.0041)	(0.0012)	(0.0014)	0.943	0.938
	ω	0.9039	0.9011	0.9014	[0.8650; 0.9319]	[0.8773; 0.9235]
		(0.0003)	(0.0003)	(0.0004)	0.9440	0.958
	β	0.9989	1.0028	1.0027	[0.8374; 1.1605]	[0.9755; 1.0406]
500		(0.0067)	(0.0010)	(0.0011)	0.9560	0.968
	δ	5.0000	5.0001	5.0001	[4.9938; 5.0061]	[4.9990; 5.0012]
		(0.00001)	(0.00000)	(0.00000)	0.9320	0.941
	κ	1.0015	1.0108	1.0112	[0.9327; 1.0703]	[0.9941; 1.0255]
		(0.0014)	(0.0004)	(0.0004)	0.9440	0.939

Table 7
Monte Carlo study for the Weibull model with ($\omega = 0.9;\ \beta = 1;\ \upsilon = 5$).

n	$\boldsymbol{\varphi}$	MLE (MSE)	BE-Mean (MSE)	BE-Median (MSE)	Conf Int Cov Rate	Cred Int Cov Rate
	ω	0.9233	0.8969	0.9041	[0.7409; 0.9684]	[0.7823; 0.9711]
		(0.0034)	(0.0017)	(0.0019)	0.892	0.972
100	β	1.0018	1.0294	1.0282	[0.6689; 1.3347]	[0.6943; 1.3711]
		(0.0284)	(0.0318)	(0.0317)	0.953	0.942
	υ	5.0204	5.1499	5.1412	[4.2224; 5.8183]	[4.3678; 5.9844]
		(0.1706)	(0.1939)	(0.1913)	0.949	0.944
	ω	0.9083	0.9008	0.9045	[0.8163; 0.9504]	[0.8285; 0.9521]
		(0.0012)	(0.0010)	(0.0010)	0.961	0.951
200	β	0.9979	1.0054	1.0049	[0.7620; 1.2338]	[0.7697; 1.2444]
		(0.0142)	(0.0149)	(0.0149)	0.952	0.949
	υ	5.0100	5.0490	5.0444	[4.4404; 5.5795]	[4.4940; 5.6320]
		(0.0872)	(0.0839)	(0.0835)	0.944	0.952
	ω	0.9035	0.8991	0.9005	[0.8599; 0.9337]	[0.8581; 0.9317]
		(0.0004)	(0.0003)	(0.0003)	0.939	0.960
500	β	1.0020	1.0058	1.0054	[0.8531; 1.1509]	[0.8574; 1.1557]
		(0.0056)	(0.0061)	(0.0061)	0.949	0.946
	υ	5.0133	5.0244	5.0222	[4.6503; 5.3764]	[4.6696; 5.3921]
		(0.0352)	(0.0389)	(0.0389)	0.951	0.935

which defines an SV model with Log-Student-t distribution through a scale mixture of Log-normal and Gamma distributions, and the degrees of freedom also being estimated.

Two chains of 500,000 iterations each were generated with convergence properly checked. The first 100,000 draws of each chain were discarded and the remaining 400,000 were kept with thinning of 400. The summary of the posterior distribution are then based on 2000 draws.

According to the results of the simulation study in Section 4, for large sample sizes the MLE and Bayesian estimators are very similar in the NGSSM. Thus, for the comparison with GARCH, EGARCH and SV models (Table 8), only the results of the MLE are presented.

The programs developed by the authors in Ox Metrics® were used to estimate the NGSSM, while the stochastic volatility model was implemented in OpenBUGS. For GARCH and EGARCH models, it was employed the rugarch package in R. For more details see [39].

For the Fréchet, Lévy and Log-normal models the parameter γ was fixed at 0.0 and, consequently, not estimated. For the Log-gamma and Pareto models there is a constraint that the series should have values greater than 1.0. Thus, for these models a constant value 1.0 was added to the observations of all series.

Fig. 4 presents the indexes and the log-returns of the six series. It can be observed, in all cases, an increase in the volatility around observations 400 and 500, which corresponds to the second semester of 2008, period of the Global Financial Crisis in 2008.

For in-sample analysis, model comparison was performed using the AICc, BIC and log-likelihood (LN LIKE) criteria, shown in Table 8. According to the three criteria, the Log-gamma and Weibull NGSSM are the best models, except for the MERVAL index, where the SV model is slightly better. The GARCH and EGARCH models present worse results than the NGSSM with Log-gamma and Weibull distributions, but perform better than the other NGSSM models.

The advantage of the NGSSM over the SV model, which is based on Gaussian state evolution, is that it does not require approximations in the estimation process, as it allows for exact computation of the marginal likelihood function. Thus, it renders significant reduction in the computational time taking just a few seconds to compute the maximum likelihood estimator. The Bayesian estimation of the NGSSM is usually ten times faster than the SV-Log-t model. Moreover, for the NGSSM, the convergence of the Markov chain Monte Carlo scheme, employed in this work, is much faster since the states are integrated out, whilst for the SV-Log-t model it is slower, due to the sampling from the highly correlated states and parameters at the same time.

For the out-of-sample analysis, model comparison was accomplished only for the best fitted models, through the square root of the prediction mean square error (SRPMSE). Thus, Table 9 shows the SRPMSE for the three best fitted models (NGSSM-Log-gamma, NGSSM-Weibull and SV-Log-t). The SRPMSE was computed using one-step ahead forecast $\hat{y}_{t+1}$, through the Bayesian approach, where the parameters were estimated along with $\hat{y}_{t+1}$ leaving the last five observations out, then the last four observations out, and so on until the last observation out. Finally, the SRPMSE was computed as $((1/5)\sum_{j=1}^{5}(y_{t+j} - \hat{y}_{t+j})^2)^{1/2}$, where the index j varies over the last five observations. From the results, it can be observed that the NGSSM with Log-gamma or Weibull distribution is better than the SV-Log-t model, with the NGSSM Log-gamma providing the best one-step-ahead predictions in most cases.

Table 10 presents the MLE and BE-Mean for parameters, along with their respective standard errors, of the Log-gamma model fitted to the volatility series of all indexes. Bayesian and MLE produce almost the same results due to the large sample size, but the standard errors for β in the Bayesian estimation are usually larger than the MLE. Most of the β's are statistically significant. The fit of the Log-gamma model was assessed by the Pearson residual for all series and it was not observed any evidence of inadequacy.

It is interesting to note that the parameter estimates are relatively close for all models, except for IPSA. Values of ω are between 0.89 and 0.94 for all indexes and this indicates a smaller impact of the crisis in the variance of the level of this series, as can be visualized in Fig. 4.

6. Conclusion

Due to the recent instability in the global economic scenario, a great variety of procedures to model volatility are being proposed in the econometric literature. In order to accommodate the main characteristics of this kind of series, the models need to, necessarily, incorporate heteroscedasticity and non-normality assumptions.

Thus, the main objective of this work was to present some particular models in a non-Gaussian state space family (NGSSM), proposed by [13], whose distribution function is contained in the family of heavy tailed distributions, such

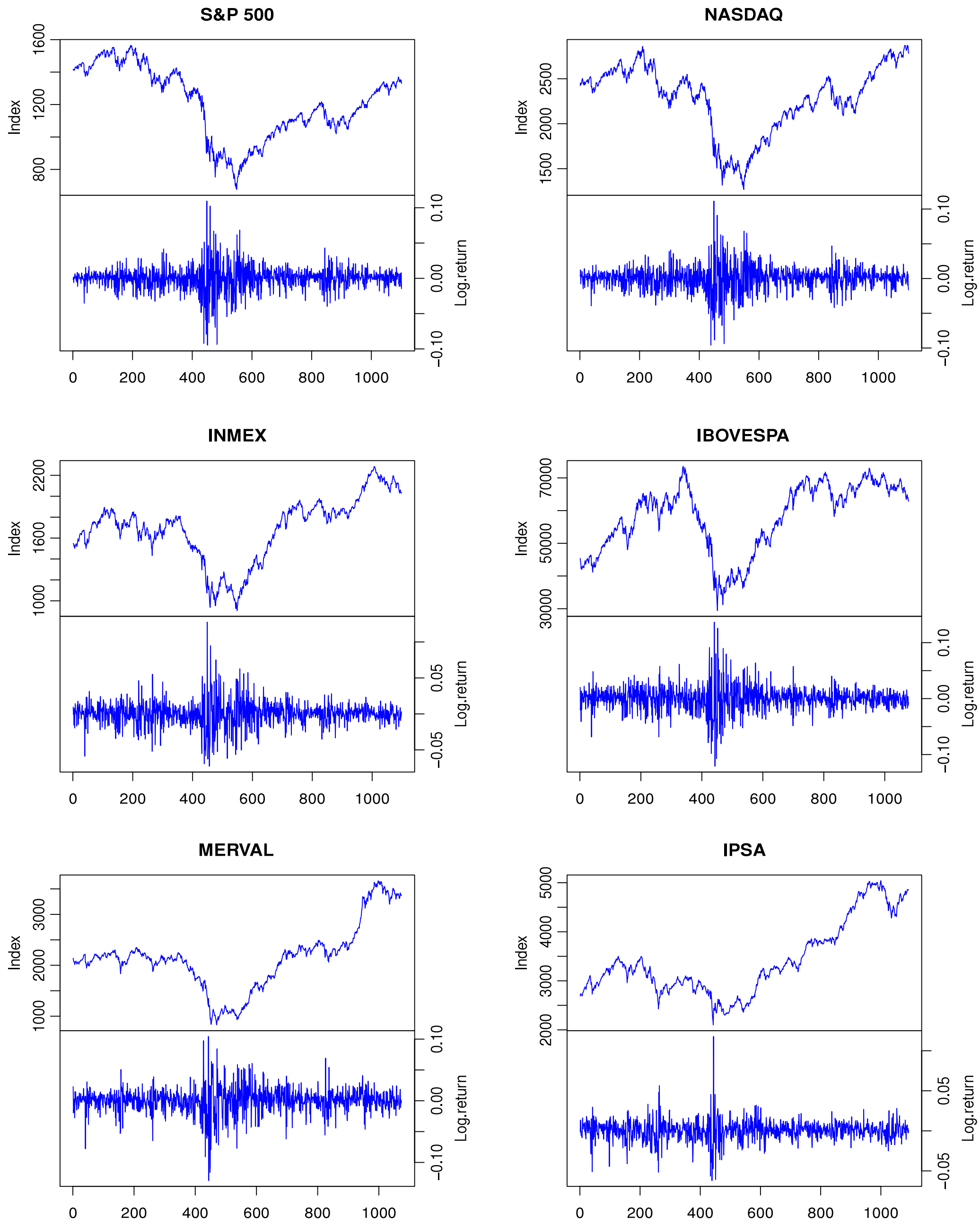


Fig. 4. The index and the log-return of S&P 500, NASDAQ, INMEX, IBOVESPA, MERVAL and IPSA, in the period from 02/01/2007 to 05/16/2011.

as the Fréchet, Lévy, Log-gamma, Log-normal, Pareto, Skew GED and Weibull. The NGSSM, when combined with heavy tailed distributions, can produce better results than the classical methodologies often employed in econometric studies, such as the GARCH and EGARCH like families.

The superiority of the method addressed here was confirmed through the fit of the methodology to the main return indexes of North and South America, when compared to different GARCH and EGARCH models and the Log-*t* SV model. The paper also presents the results of a Monte Carlo study comparing classical and Bayesian estimation for some heavy tailed distributions in the NGSSM. In general, the estimation procedures show very satisfactory results.

Table 8
Fitted models for the North and South American stock indexes.

SERIES	NGSSM	AICc	BIC	LN LIKE	E/GARCH and SV	AICc	BIC	LN LIKE
	FRÉCHET	−15.54	−15.52	8556.7	GARCH-SKEW NORMAL	−16.05	−16.02	8834.5
	LÉVY	−15.02	−15.01	8269.3	GARCH-SKEW *t*	−16.05	−16.01	8836.0
	LOG-GAMMA	−16.22	−16.21	8934.4	EGARCH-SKEW NORMAL	−16.11	−16.07	8870.3
S&P 500	LOG-NORMAL	−15.88	−15.87	8744.7	EGARCH-SKEW *t*	−16.12	−16.07	8873.6
	PARETO	−15.64	−15.63	8612.3	SV-LOG-*t*	−16.08	−16.05	8855.0
	SKEW GED	−15.60	−15.58	8592.1				
	WEIBULL	−16.23	−16.22	8938.9				
	FRÉCHET	−15.12	−15.11	8327.5	GARCH-SKEW NORMAL	−15.66	−15.63	8621.0
	LÉVY	−14.64	−14.63	8060.3	GARCH-SKEW *t*	−15.67	−15.63	8627.4
	LOG-GAMMA	−15.83	−15.81	8715.0	EGARCH-SKEW NORMAL	−15.73	−15.70	8661.6
NASDAQ	LOG-NORMAL	−15.47	−15.46	8521.3	EGARCH-SKEW *t*	−15.75	−15.71	8672.9
	PARETO	−15.28	−15.27	8411.3	SV-LOG-*t*	−15.70	−15.67	8645.0
	SKEW GED	−15.25	−15.23	8398.4				
	WEIBULL	−15.82	−15.80	8708.9				
	FRÉCHET	−14.94	−14.92	8202.3	GARCH-SKEW NORMAL	−15.49	−15.46	8504.3
	LÉVY	−14.42	−14.41	7918.0	GARCH-SKEW *t*	−15.53	−15.49	8527.2
	LOG-GAMMA	−15.72	−15.70	8631.8	EGARCH-SKEW NORMAL	−15.56	−15.52	8540.8
INMEX	LOG-NORMAL	−15.34	−15.32	8421.9	EGARCH-SKEW *t*	−15.61	−15.57	8572.3
	PARETO	−15.26	−15.26	8381.9	SV-LOG-*t*	−15.72	−15.70	8635.0
	SKEW GED	−15.24	−15.22	8372.0				
	WEIBULL	−15.71	−15.70	8627.5				
	FRÉCHET	−14.01	−13.99	7551.8	GARCH-SKEW NORMAL	−14.60	−14.57	7872.0
	LÉVY	−13.57	−13.56	7315.1	GARCH-SKEW *t*	−14.63	−14.59	7887.6
	LOG-GAMMA	−14.75	−14.74	7955.5	EGARCH-SKEW NORMAL	−14.67	−14.63	7908.1
IBOVESPA	LOG-NORMAL	−14.44	−14.42	7784.3	EGARCH-SKEW *t*	−14.71	−14.66	7928.4
	PARETO	−14.31	−14.31	7717.2	SV-LOG-*t*	−14.70	−14.68	7930.0
	SKEW GED	−14.30	−14.28	7711.3				
	WEIBULL	−14.75	−14.73	7951.5				
	FRÉCHET	−14.29	−14.27	7674.4	GARCH-SKEW NORMAL	−14.92	−14.88	8010.2
	LÉVY	−13.69	−13.68	7351.7	GARCH-SKEW *t*	−14.93	−14.90	8019.7
	LOG-GAMMA	−15.04	−15.03	8081.8	EGARCH-SKEW NORMAL	−14.95	−14.91	8028.5
MERVAL	LOG-NORMAL	−14.73	−14.72	7915.0	EGARCH-SKEW *t*	−14.98	−14.93	8043.8
	PARETO	−14.47	−14.46	7771.5	SV-LOG-*t*	−15.06	−15.03	8090.0
	SKEW GED	−14.45	−14.44	7765.9				
	WEIBULL	−15.04	−15.02	8078.1				
	FRÉCHET	−16.05	−16.04	8768.2	GARCH-SKEW NORMAL	−16.62	−16.59	9075.1
	LÉVY	−15.63	−15.62	8534.3	GARCH-SKEW *t*	−16.64	−16.60	9083.1
	LOG-GAMMA	−16.74	−16.72	9141.9	EGARCH-SKEW NORMAL	−16.69	−16.65	9112.7
IPSA	LOG-NORMAL	−16.45	−16.44	8986.5	EGARCH-SKEW *t*	−16.72	−16.67	9129.0
	PARETO	−16.35	−16.34	8928.1	SV-LOG-*t*	−16.69	−16.67	9120.0
	SKEW GED	−16.32	−16.30	8916.1				
	WEIBULL	−16.73	−16.72	9138.7				

Future research encompasses the improvement of the maximum likelihood method to properly estimate ω for small samples and hypothesis test for the parameters, and also other features of stochastic volatility models such as leverage effects (which would result in a likelihood in known closed form).

Table 9
Comparison for the out-of-sample forecasts of the three best fitted models for the North and South American stock indexes. SRPMSE stands for the square root of the prediction mean square error and was computed using one-step ahead forecast for the last five observations of each series.

SERIES	MODEL	SRPMSE	SERIES	MODEL	SRPMSE
	NGSSM-LOG-GAMMA	0.000167		NGSSM-LOG-GAMMA	0.000217
S&P 500	NGSSM-WEIBULL	0.000172	**NASDAQ**	NGSSM-WEIBULL	0.000198
	SV-LOG-t	0.011710		SV-LOG-t	0.023418
	NGSSM-LOG-GAMMA	0.000138		NGSSM-LOG-GAMMA	0.000259
INMEX	NGSSM-WEIBULL	0.000146	**IBOVESPA**	NGSSM-WEIBULL	0.000289
	SV-LOG-t	0.340632		SV-LOG-t	0.037606
	NGSSM-LOG-GAMMA	0.000341		NGSSM-LOG-GAMMA	0.000098
MERVAL	NGSSM-WEIBULL	0.000368	**IPSA**	NGSSM-WEIBULL	0.000104
	SV-LOG-t	0.715625		SV-LOG-t	0.014071

Table 10
Parameter estimates of the Log-gamma models for the volatility of the indexes. The standard errors are between brackets.

NGSSM	Parameter	MLE	BE-Mean	NGSSM	Parameter	MLE	BE-Mean
	ω	0.9210	0.9180		ω	0.9315	0.9276
		(0.0110)	(0.0115)			(0.0107)	(0.0109)
S&P 500	β	6.7938	4.0215	**NASDAQ**	β	8.1367	4.9696
		(3.5117)	(7.0122)			(2.6491)	(7.2484)
	α	0.4388	0.4389		α	0.4497	0.4492
		(0.0138)	(0.0156)			(0.0162)	(0.0160)
	ω	0.9189	0.9147		ω	0.9256	0.9210
		(0.0116)	(0.0123)			(0.0117)	(0.0120)
INMEX	β	6.7530	6.1239	**IBOVESPA**	β	8.4150	6.8293
		(2.7749)	(5.0525)			(2.4464)	(4.7735)
	α	0.4744	0.4750		α	0.4811	0.4795
		(0.0156)	(0.0174)			(0.0168)	(0.0177)
	ω	0.9098	0.9066		ω	0.8963	0.8869
		(0.0118)	(0.0134)			(0.0111)	(0.0129)
MERVAL	β	6.5492	5.2754	**IPSA**	β	15.2041	2.8329
		(2.3339)	(4.3531)			(2.9900)	(5.4234)
	α	0.4307	0.4304		α	0.4947	0.4924
		(0.0159)	(0.0161)			(0.0178)	(0.0183)

Acknowledgments

The authors wish to acknowledge Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) for financial support.

References

[1] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations, Econometrica 50 (1982) 987–1007.
[2] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Appl. Econometrics 31 (1986) 307–327.
[3] D.B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59 (1991) 347–370.
[4] J.M. Zakoian, Threshold heteroscedastic models, J. Econom. Dynam. Control 18 (1994) 931–955.
[5] L. Bauwens, S. Laurent, J.V.K. Rombouts, Multivariate GARCH models: A survey, J. Appl. Econom. 21 (2006) 79–109.
[6] S.J. Taylor, Modelling Financial Time Series, John Wiley & Sons, 1986.

[7] A. Melino, S.M. Turnbull, Pricing foreign currency options with stochastic volatility, J. Appl. Econom. 45 (1990) 239–265.
[8] S.J. Taylor, Modelling stochastic volatility: A review and comparative study, Math. Finance 4 (1994) 183–204.
[9] A.C. Harvey, E. Ruiz, N. Shephard, Multivariate stochastic variance models, Rev. Econ. Stud. 61 (1994) 247–264.
[10] E. Jacquier, N.G. Polson, P. Rossi, Bayesian analysis of stochastic volatility models (with discussion), J. Bus. Econom. Statist. 12 (1994) 371–417.
[11] B. Eraker, M. Johanners, N.G. Polson, The impact of jumps in returns and volatility, J. Financ. 53 (2003) 1269–1330.
[12] D. Raggi, S. Bordignon, Comparing stochastic volatility models through Monte Carlo simulations, Comput. Statist. Data Anal. 50 (2006) 1678–1699.
[13] D. Gamerman, T.R. Santos, G.C. Franco, A non-Gaussian family of state-space models with exact marginal likelihood, J. Time Ser. Anal. 34 (2013) 625–645.
[14] R.L. Smith, J.E. Miller, A non-Gaussian state space model and application to prediction of records, J. Roy. Stat. Soc. B 48 (1986) 79–88.
[15] R.S. Tsay, Analysis of Financial Time Series, John Wiley & Sons, New Jersey, 2005.
[16] A.C. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1989.
[17] N. Shephard, Local scale model: state space alternative to integrated GARCH processes, J. Appl. Econom. 60 (1994) 181–202.
[18] P.K. Deschamps, Bayesian estimation of an extended local scale stochastic volatility model, J. Appl. Econom. 162 (2011) 369–382.
[19] M. Ferrante, P. Vidoni, Finite dimensional filters for nonlinear stochastic difference equations with multiplicative noises, Stochastic Process. Appl. 77 (1998) 69–81.
[20] P. Vidoni, Exponential family state space models based on conjugate latent process, J. Roy. Statist Soc. Ser. B 61 (1999) 213–221.
[21] J. Anderson, On the normal inverse Gaussian stochastic volatility model, J. Bus. Econom. Statist. 19 (2001) 44–54.
[22] S. Chib, F. Nardari, N. Shephard, Markov chain Monte Carlo methods for stochastic volatility models, J. Appl. Econom. 108 (2002) 281–316.
[23] A.C. Harvey, C. Fernandes, Time series models for count or qualitative observations, J. Bus. Econom. Statist. 7 (1989) 407–417.
[24] M. Avriel, Nonlinear Programming: Analysis and Methods, Dover Publications, 2003.
[25] G.O. Roberts, J.S. Rosenthal, Examples of adaptive MCMC, J. Comput. Graph. Statist 18 (2009) 349–367.
[26] H. Haario, F. Saksman, J. Tamminen, An adaptive Metropolis algorithm, Bernoulli 7 (2001) 223–242.
[27] P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman and Hall, London, 1989.
[28] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control 19 (1974) 716–723.
[29] G.E. Schwarz, Estimating the dimension of a model, Ann. Statist 6 (1978) 461–464.
[30] C.M. Hurvich, C.L. Tsai, A corrected Akaike information criterion for vector autoregressive model selection, J. Time Ser. Anal. 14 (1993) 271–279.
[31] S. Asmussen, Applied Probability and Queues, Springer, Berlin, 2003.
[32] J.L. Teugels, The class of subexponential distributions, Ann. Probab. 3 (1975) 1000–1011.
[33] P. Embrechts, C. Klüppelberg, T. Milosch, Modelling Extremal Events, Springer, New York, 1997.
[34] C.M. Goldie, C. Klüppelberg, Subexponential Distributions. A Practical Guide to Heavy Tails: Statistical Techniques and Applications, Birkhauser Boston, Cambridge, 1998, pp. 435–459.
[35] J. Neyman, E.T. Scott, Outliers Proneness of Phenomena and Related Distributions, in: Optimizing Methods in Statistics, Academic Press, New York, 1971, pp. 413–430.
[36] R.F Green, Outlier-prone and outlier-resistant distributions, J. Amer. Statist. Assoc. 71 (1976) 502–505.
[37] P.C. Consul, G.C. Jain, On the log-gamma distribution and its properties, Statist. Papers 12 (1971) 100–106.
[38] A. Ayebo, T.J. Kozubowski, An asymmetric generalization of Gaussian and Laplace laws, JPSS J. Probab. Stat. Sci. 1 (2003) 187–210.
[39] T. Bollerslev, J.M. Wooldridge, Quasi-Maximum likelihood estimation and inference in dynamic models with time-varying covariance, Econometric. Rev. 11 (1992) 143–172.