Generalized Pareto models with time-varying tail behavior*

Fernando F. Nascimento
Universidade Federal do Piauí

Dani Gamerman
Universidade Federal do Rio de Janeiro

Hedibert F. Lopes
University of Chicago

September 19, 2011

*Fernando F. Nascimento is Professor Adjunto, Departamento de Informática e Estatística, Universidade Federal do Piauí, Campus Ministro Petrônio Portela, SG2, 64049-550, Teresina, Brazil, fernandofn@ufpi.edu.br. Dani Gamerman is Professor Titular, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, 21945-970, Rio de Janeiro, Brazil, dani@im.ufrj.br. Hedibert F. Lopes is Associate Professor of Econometrics and Statistics, Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave, Chicago, IL 60637, hlopes@chicagobooth.edu.

Abstract

This paper is concerned with the analysis of time series data with temporal dependence through extreme events. This is achieved via a model formulation that considers separately the central part and the tail of the distributions. A two component mixture model is used for splitting the data into the extreme regime and the central part. Extremes beyond a threshold are assumed to follow a generalized Pareto distribution (GPD) and the parameters of the GPD are allowed to vary stochastically with time, thus inducing temporal dependence. Temporal variation and dependence is introduced at a latent level via the use of dynamic linear models (DLM). The central part follows a nonparametric, mixture approach. The uncertainty about the threshold is explicitly considered. Posterior inference is performed through Markov Chain Monte Carlo (MCMC) methods. A variety of scenarios can be entertained and include the possibility of alternation of presence and absence of a finite upper limit of the distribution for different time periods. Simulations are carried out in order to analyze the performance of our proposed model. We also apply the proposed model to financial time series: the returns of Petrobras stocks and Bovespa index, all of which exhibit several extreme events. Results show advantage of our proposal over currently entertained models such as stochastic volatility, with improved estimation of high quantiles and extremes.

1 Introduction

Extreme value theory was shown to provide a very useful tool in many areas of application where precise knowledge of the tail behavior of a distribution is of central interest. The areas where most impact was achieved are environmental science and finance. In financial applications, the main concern is an accurate assessment of the probability of huge losses or gains for any given stock. The same is valid for other financial indicators such as currencies, interest rates and futures. In environmental applications, Nascimento et al. (2011a,b), Parmesan et al. (2000) and Huerta and Sansó (2007) are just a sample of a large literature concerned with extremes in climatological features such as rain and temperature.

One approach of modelling extreme data is to consider the distribution of exceedances over a high threshold. Pickands (1975) showed that the limiting distribution of exceedances over suitably large thresholds behaves in a very stable fashion, converging to a the generalized Pareto distribution (GPD). Let x be the excess over a high threshold, say u. It is said that x follows a generalized Pareto distribution with tail parameters ξ and σ, denoted here by $GPD(x; \xi, \sigma)$, if its cumulative distribution function (cdf) can be written as

$$G(x; \xi, \sigma) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\sigma), & \text{if } \xi = 0 \end{cases}. \tag{1}$$

The tail parameters ξ and σ are shape and the scale, respectively. The support of a GPD is $x \geq 0$ when $\xi \geq 0$, $0 \leq x \leq \sigma/|\xi|$ when $\xi < 0$ and the data is said to exhibit heavy tail behavior when $\xi > 0$. The associated probability density function (pdf) is

$$g(x; \xi, \sigma) = \begin{cases} (1 + \xi x/\sigma)_+^{-(1+\xi)/\xi}/\sigma & \text{if} \quad \xi_t \neq 0 \\ \exp(-x/\sigma)/\sigma & \text{if} \quad \xi_t = 0 \end{cases}, \tag{2}$$

where $(a)_+$ denotes $max\{0, a\}$.

Traditional analysis of such a model is performed by fixing the threshold u, which is chosen either graphically by looking at the mean residual life plot (Coles, 2001; Embrechts et al., 1997), or by simply setting it at some high percentile of the data (DuMouchel, 1983). Nevertheless, the literature showed how the threshold selection influences the parameter estimation (Coles and Powell, 1996; Coles and Tawn, 1996a,b; Smith, 1984). Behrens et al. (2004) proposed a model to fit extreme data where the threshold itself is one of the model unknowns. More specifically, they proposed a parametric form to explain the data variability below the threshold and a GPD for the data above it. More recently, Nascimento et al. (2011b) generalized Behrens et al. (2004) allowing more flexibility below the threshold with a finite mixture of distributions.

However, all previous models consider static GPD parameters. In many situation, the tail behavior governing the extremes may change with the passage of time. Time trends and seasonality are only a few of the reasons for these changes and introduce temporal dependence in the evaluation of extremes.

We extend previous models by allowing the GPD parameters to be time dependent. These feature will be shown to be important when modeling real datasets, such as in financial time series data (see Section 4). Basically, this is done by embeding the extreme parameters in a non-linear dynamic system (West and Harrison, 1997). A similar idea was proposed by Huerta and Sansó (2007) who model daily ozone levels in the US via a generalized extreme value (GEV) distribution whose parameters evolve both in time and in space.

The remainder of the paper is organized as follows. Section 2 introduces our proposed model. We also discuss the prior specifications for model parameters and, in particular, introduce the DLM in the GPD context. We carry out simulation study and discuss the performance of our model in Section 3. In section 4 the model is applied to three financial time series from the Brazilian stock exchange data. The relevance of the changes introduced in this paper will be detailed here. Section 5 concludes the paper.

2 Time-varying tail behavior

Consider a time series y_t where, for a high threshold u_t, it is assume that $\{y_t | y_t < u_t\}$ is modeled by a finite mixture of distributions, while $\{y_t | y_t > u\}$ follows a GPD with time-varying parameters u_t, ξ_t and σ_t.

In this model, the parameters characterizing extremal behavior change with time. The model will be completed below with a mixture specification below the threshold. The nature of the mixture for the description below the threshold does not render any meaning to its components. The non-parametric nature of this specification enables the best possible marginal fit for this part of the model. Further dependence could be introduced in the central part eg via copulas. This is not pursued here since the main aim of the model is to estimate time-varying extremal behavior.

2.1 Modeling below and at the threshold

Given the lack of information below the threshold, non-parametric approximations seem a natural choice. Wiper et al. (2001) showed that a mixture of Gamma distributions, denoted here by MG_k, provides a good approximation for distributions with positive support. Nascimento et al. (2011b) used this approach with good results both in terms of density estimation and reliable threshold estimation. The probability density function (pdf) of this mixture is given by

$$h(x; \mu, \alpha, p) = \sum_{j=1}^{k} p_j f_G(x; \mu_j, \alpha_j), \tag{3}$$

where $\mu = (\mu_1, \ldots, \mu_k)$, $\alpha = (\alpha_1, \ldots, \alpha_k)$, $p = (p_1, \ldots, p_k)$ is the vector of mixture component weights and $f_G(x; \mu, \alpha)$ is the Gamma density with mean μ, variance μ/α and evaluated at x. The means μ_js and shapes α_js may take any positive value and weights p_js are positive and add up to one. The number of components k may be known and fixed or an additional model parameter and estimated. In this paper we compare models for different values of k

via the deviance information criterion (DIC) of Spiegelhalter et al. (2002). For more details see the applications in Sections 3 and 4.

The priors for μ and α follow Wiper et al. (2001) and Nascimento et al. (2011b), and are given by

$$p(\mu_1,\ldots,\mu_k) \propto \prod_{i=1}^{k} f_{IG}(\mu_i; a_i/b_i, b_i) I(\mu_1 < \mu_2 < \ldots < \mu_k)$$

and

$$p(\alpha_1,\ldots,\alpha_k) = \prod_{i=1}^{k} f_{IG}(\alpha_i; c_i/d_i, d_i),$$

respectively. The prior distribution for the weights p is assumed to be a Dirichlet distribution $D_k(\gamma_1,\ldots,\gamma_k)$, whose density is proportional to $\prod_{i=1}^{k} p_i^{\gamma_i}$. The hyperparameters of the inverse gamma distributions, a_i, b_i, c_i and d_i, for $i = 1,\ldots,k$ are chosen to characterize little prior information regarding the μs and the αs. For the weights p, the hyperparameters in γ are chosen such that $\gamma_1 = \cdots = \gamma_k = \gamma$ *a priori*, where γ is large enough to induce large prior variability. More details can be found in the simulation exercise of Section 3.

2.2 Modeling above threshold

It is well known that the Gamma distribution belongs to the maximum domain of attraction of a Gumbel distribution. It can be easily shown that the result can be extended for mixture of Gamma distributions based on the results of Embrechts et al. (1997), page 156, which turns out to be a necessary condition for modeling the excess over the threshold via a GPD. Therefore, combining the two parts of the data, the cdf of y_t is

$$F_t(y_t;\Theta) = \begin{cases} H(y_t;\mu,\alpha,p) & \text{if } y_t < u \\ H(u;\mu,\alpha,p) + [1 - H(u;\mu,\alpha,p)]G(y_t - u;\xi,\sigma) & \text{if } y_t \geq u \end{cases}, \qquad (4)$$

where $\Theta_t = (\mu, \alpha, p, u, \xi, \sigma)$, $u = \{u_t\}$, $\xi = \{\xi_t\}$, $\sigma = \{\sigma_t\}$ and $H(\cdot; \mu, \alpha, p)$ is the cdf associated with the mixture of gammas pdf $h(\cdot; \theta, p)$ in (3). Consequently, for a sample of observations $y = (y_1, \cdots, y_T)$, the full likelihood function for the model can be written as

$$L(y; \Theta) = \prod_{\{t|y_t<u_t\}} h(y_t; \mu, \alpha, p) \prod_{\{t|y_t \geq u_t\}} (1 - H(u_t; \mu, \alpha, p))g(y_t - u_t; \xi_t, \sigma_t). \tag{5}$$

Notice that observations y_t below the threshold u_t only contribute to the likelihood via the first product on the right hand side of the above likelihood equation. In other words, learning μ, α and p is solely based on the set of y_ts below u_t. Similarly, learning ξ_t and σ_t is solely based on the set of y_ts above u_t, which is usually a much smaller set than the one below u_t. Learning about u_t is based on a combination of the two likelihood components. It is exactly the distinction between these two parts that enables the threshold estimation.

Another advantage of this class models is the case with which higher quantiles can be obtained. For values beyond the threshold, It is straightforward to obtain the p-quantile, by

$$q = u_t + \frac{((1 - p^*)^{-\xi_t} - 1)\sigma}{\xi_t}, \tag{6}$$

where $p^* = \{p - H(u_t \mid \mu, \eta, p)\}/\{1 - H(u_t \mid \mu, \eta, p)\}$, for quantiles beyond the threshold. Typically, one is interested in high quantiles well above the threshold but similar calculations can be performed for lower quantiles, even below the threshold.

2.3 Dynamic modeling

The model allows for time variation of the GPD parameters $\Psi_t = (u_t, \xi_t, \sigma_t)$ but this is too general for practical purposes. These parameters are expected to be related and can be described probabilistically in an evolution form

$$\Psi_t = g(\Psi_{t-1}, w_t), \tag{7}$$

where g is a possibly non-linear function and the w_t´s are random disturbance random vectors. The temporal relation (7) serves a number of purposes: a) it induces temporal correlation between observations; b) it allows for information to be borrowed from successive times thus strengthening the inference procedures; c) it establishes smoothness constraints in the GPD parameters thus avoiding unrealistic discontinuities in their temporal evolution.

A number of possibilities are available through (7). The most common ones are the inclusion of trend and seasonality in some model components. Obviously static parameters are contained in this class of models as the limiting case where g is the identity function and $w_t = 0$, for all t.

The simplest non-degenerate form of this model is provided by the first order dynamic linear model (DLM), as in West and Harrison (1997). It can be used to model directly the GPD parameters across time. A simple local evolution would assume that

$$\begin{aligned}
\xi_t &= \xi_{t-1} + w_{\xi,t} & \qquad w_{\xi,t} &\sim N(0, 1/W_\xi) \\
\sigma_t &= \sigma_{t-1} + w_{\sigma,t} & \qquad w_{\sigma,t} &\sim N(0, 1/W_\sigma) \\
u_t &= u_{t-1} + w_{u,t} & \qquad w_{u,t} &\sim N(0, 1/W_u)
\end{aligned}$$

where the precisions W_ξ, W_σ and W_u drive the local evolution and degree of smoothness of ξ, σ and u and can be user specified or estimated.

The above specification assumes that u, σ and ξ are real values. However, u and σ need to be positive, while Smith (1984) showed that maximum likelihood estimators are inexistent when $\xi < -1$. Therefore, the above model can be revised by applying a first order DLM to transformed parameters $lu_t = \log u_t$, $l\sigma_t = \log \sigma_t$ and $l\xi_t = \log(\xi_t + 1)$ as in

$$\begin{array}{rcll}
l\xi_t & = & \theta_{\xi,t} + v_{\xi,t} & v_{\xi,t} \sim N(0, 1/V_\xi) \\
\theta_{\xi,t} & = & \theta_{\xi,t-1} + w_{\xi,t} & w_{\xi,t} \sim N(0, 1/W_\xi) \\
l\sigma_t & = & \theta_{\sigma,t} + v_{\sigma,t} & v_{\sigma,t} \sim N(0, 1/V_\sigma) \\
\theta_{\sigma,t} & = & \theta_{\sigma,t-1} + w_{\sigma,t} & w_{\sigma,t} \sim N(0, 1/W_\sigma) \\
lu_t & = & \theta_{u,t} + v_{u,t} & v_{u,t} \sim N(0, 1/V_u) \\
\theta_{u,t} & = & \theta_{u,t-1} + w_{u,t} & w_{u,t} \sim N(0, 1/W_u),
\end{array} \tag{8}$$

for $t = 1, \dots, T$ and the initial information $\theta_{0,\xi} \sim N(m_{0,\xi}, C_{0\xi})$, $\theta_{0,\sigma} \sim N(m_{0,\sigma}, C_{0\sigma})$ and $\theta_{0,u} \sim N(m_{0,u}, C_{0u})$. The additional system disturbances $v_{\xi,t}$ $v_{\sigma,t}$ and $v_{u,t}$ have proved useful to provide further smoothing to the latent evolution of the GPD parameters and their use was supported by the data.

The priors for V_ξ, V_σ and V_u are gammas $G(f_\xi, o_\xi)$, $G(f_\sigma, o_\sigma)$ and $G(f_u, o_u)$, while the priors for W_ξ, W_σ and W_u are $G(l_\xi, m_\xi)$, $G(l_\sigma, m_\sigma)$ and $G(l_u, m_u)$. Similar to the hyperparameters of the mixture components, the vector $(m_{0,\xi}, C_{0\xi}, m_{0,\sigma}, C_{0\sigma}, m_{0,u}, C_{0u}, f_\xi, o_\xi, f_\sigma, o_\sigma, f_u, o_u, l_\xi, m_\xi, l_\sigma, m_\sigma, l_u, m_u)$ are chosen to induce vague prior information. More details can be found in the simulation exercise of Section 3.

The logarithmic transformation of the threshold was applied in (8) but is typically unnecessary given this is a location parameter and given the negligible amount of probability below the lower limit of the data support, typically set at 0. So, lu_t can be replaced by u_t with appropriate care with the truncation below 0.

Many other possibilities are available with the broad spectrum of state space models rendered by (7). In particular, (8) can be further simplified by allowing some of the model parameters to be static. For example, if the threshold u is deemed to remain constant through time, it may be dropped from the evolutions in (8). In this case, a normal distribution with parameters (μ_u, σ_u^2), truncated below 0, may be assumed. The parameter μ_u is typically set at a high data percentile and σ_u^2 may be large if one wants to represent a fairly noninformative prior. A normal distribution to the logarithm of the threshold is another possibility to avoid the truncation, as in (8).

2.4 Posterior distribution

From the likelihood function and the prior distributions specified above, we use Bayes' theorem to obtain the posterior distribution, up to a normalizing constant, as follows

$$\begin{aligned}
\pi(\Theta|y) \;\propto\;& \prod_{\{t|y_t<u\}}\left[\sum_{j=1}^{k} p_j f_G(y_t;\mu_j,\eta_j)\right]\prod_{\{t|y_t\geq u\}}\left[\left(1-\sum_{j=1}^{k} p_j F_G(u;\mu_j,\eta_j)\right) g(y_t-u;\xi_t,\sigma_t)\right] \\
\times\;& \prod_{j=1}^{k}\left[\eta_j^{a_j-1}e^{-b_j\eta_j}\beta_j^{-(c_j+1)}e^{-d_j/\mu_j}\right] \\
\times\;& V_\xi^{T/2+f_\xi-1}\exp\left(-\frac{V_\xi}{2}\sum_{t=1}^{T}(l\xi_t-\theta_{\xi,t})^2-o_\xi V_\xi\right)\exp\left(-\frac{1}{2C_{\xi,0}}(\theta_{\xi,0}-m_{\xi,0})^2\right) \\
\times\;& W_\xi^{T/2+l_\xi-1}\exp\left(-\frac{W_\xi}{2}\sum_{t=1}^{T}(\theta_{\xi,t}-\theta_{\xi,t-1})^2-m_\xi W_\xi\right) \\
\times\;& V_\sigma^{T/2+f_\sigma-1}\exp\left(-\frac{V_\sigma}{2}\sum_{t=1}^{T}(l\sigma_t-\theta_{\sigma,t})^2-o_\sigma V_\sigma\right)\exp\left(-\frac{1}{2C_{\sigma,0}}(\theta_{\sigma,0}-m_{\sigma,0})^2\right) \\
\times\;& W_\sigma^{T/2+l_\sigma-1}\exp\left(-\frac{W_\sigma}{2}\sum_{t=1}^{T}(\theta_{\sigma,t}-\theta_{\sigma,t-1})^2-m_\sigma W_\sigma\right) \\
\times\;& V_\sigma^{T/2+f_u-1}\exp\left(-\frac{V_u}{2}\sum_{t=1}^{T}(lu_t-\theta_{u,t})^2-o_u V_u\right)\exp\left(-\frac{1}{2C_{u,0}}(\theta_{u,0}-m_{u,0})^2\right) \\
\times\;& W_u^{T/2+l_u-1}\exp\left(-\frac{W_u}{2}\sum_{t=1}^{T}(\theta_{u,t}-\theta_{u,t-1})^2-m_u W_u\right), \qquad (9)
\end{aligned}$$

with $\Theta=\left(\mu,\alpha,p,u,\{l\xi_t\}_{t=1}^{T},\{l\sigma_t\}_{t=1}^{T},\{\theta_{\xi,t}\}_{t=0}^{T},\{\theta_{\sigma,t}\}_{t=0}^{T},V_\xi,V_\sigma,W_\xi,W_\sigma\right)$. The likelihood in the first row, while the prior for the parameters below u is in the second row. Third to sixth rows are the priors for the DLM precisions and the latent state space variables. As expected, posterior inference is analytically infeasible, so Bayesian inference is performed through a customized Markov Chain Monte Carlo algorithm (Gamerman and Lopes, 2006), which is outlined in the appendix.

3 Simulation study

We test our model and posterior sampling algorithm on a simulated data set. The data set is generated as follows: i) starting from the given value of $\theta_{\xi,0}$ and $\theta_{\sigma,0}$, we generate an ordered sequence of $\{l\xi_t\}_{t=1}^T$ and $\{l\sigma_t\}_{t=1}^T$ using DLM equations as specified in Section 2, then ii) we draw an ordered sequence of T observations from $G(\mu, \alpha)$ and iii) define u as constant in time and set at a high quantile q_p from the generated sample. Let T_1 be the number of observations, out of T, that are below u. Retain these T_1 observations and keep their values and iv) replace $T - T_1$ observations above u with draws from $GPD(\xi_t, \sigma_t, u)$, with t corresponding to the location of the observation being replaced. The parameters used in the simulation are as follows: $T \in \{1000, 2500, 10000\}$, $u = q_{80}$, $\alpha = 1.0$, $\mu = 5.0$, $\theta_{\xi,0} = 0.2$ and $\theta_{\sigma,0} = 2.0$, where q_{80} is the 80th percentile of the sample, $V_\xi = 200$, $V_\sigma = 200$, $W_\xi = 1000$ and $W_\sigma = 1000$. These correspond to standard deviations of the order of 0.071 and 0.032. This specification is reasonable to mimic the behavior of daily data where abrupt changes are unlikely to occur between consecutive days. The sample size T_1 and p automatically define the value of the threshold parameter u. Obviously, the average number of observations above the threshold are in $\{200, 500, 2000\}$. We implement the MCMC algorithm detailed in the appendix.

The priors used are as described in Section 2. More specifically, $\mu_j \sim IG(2.1, 5.5)$ and $\eta_j \sim G(6, 0.5)$, for $j = 1, \ldots, k$. These distribution have mean around the actual parameter value but with large variance to represent lack of information: The prior variances of μ_js and η_js are 250 and 24, respectively. Additionally, $D_k(1, \ldots, 1)$ is the prior for the weights p and $N(u_0, 10)$ is the prior for the threshold parameters, u, for u_0 the true value. This distribution is relatively vague with a 95% probability interval from 3.7 and 16.3 for the threshold, covering more than half of the observations.

The hyperparameters of the prior distributions for the parameters driving the dynamic model are $m_{0,\xi} = m_{0,\sigma} = 0.2$, $C_{0,\xi} = C_{0,\sigma} = 1000$. The other hyperparameters ($f_\xi, o_\xi, f_\sigma, o_\sigma, l_\xi, m_\xi$, l_σ, m_σ) are chosen such that the priors are centered around the true values with standard deviations around 100. The chains are initialized from the respective prior distributions. After a burn-in of 150,000 iterations, the remaining 50,000 iterations are used for inference. The posterior mean and variance are computed after thinning at every 100 steps, leading to posterior approximations based on 500 draws. Convergence was established for all model parameters based on standard convergence diagnostic tests. Two parallel chains are run from distinct initial values.

Table 1 gives the posterior mean and the 95% posterior credibility interval for every model parameter except $\{l\xi_t\}_{t=1}^T$, $\{l\sigma_t\}_{t=1}^T$, $\{\theta_{\xi,t}\}_{t=1}^T$ and $\{\theta_{\sigma,t}\}_{t=1}^T$. The posterior mean of each parameter is close to its respective true value, and the true value lies within the corresponding 95% posterior credibility interval. As expected, estimates are more accurate for larger samples sizes for both parameters below and above the threshold. Figure 1 shows the difficulty in estimating tail parameters when the (tail) sample size is relatively small $T = 1,000$ ($T_1 = 200$). The results are more stable and reliable when $T = 2,500$ and $T = 10,000$.

In additional to posterior inference of the parameters that define the model, it is commonly desirable to learn about the extreme quantiles. This paper adds to the literature by introducing the ability to learn the time-varying behavior of extreme events. Figure 2 shows estimates for the 95th and 99th percentiles of the data over time, for $T = 1,000$ and $T = 2,500$, respectively. In both examples, the estimated percentiles virtually match the actual, simulated ones.

Recall that when ξ is negative, the support of the GPD has a finite maximum, given by $u - \sigma/\xi$. When dealing with time varying ξ, it is possible that a finite maximum exists at a given time but not at subsequent times depending on whether ξ_t remains greater than zero or not. Figure 2 shows that true maxima are well estimated by the model.

4 Application to Brazilian stock returns

The dynamic extreme value model is now applied to the study of two well-known financial Brazilian time series: i) Petrobrás[1] returns and ii) Ibovespa returns, where Ibovespa is the index of the Brazilian stock exchange (Bovespa). The data sets are thus chosen given their importance to financial market and the presence of many extreme events. For each data set, the original data, daily closing prices p_t, are first converted to daily returns via $x_t = p_t/p_{t-1} - 1$, and then converted into the data we use by $y_t = 100 \times |x_t - \bar{x}|$, where $\bar{x}$ is the sample average of x_t. The subtraction of $\bar{x}$ from each x_t is used to avoid zeros in the converted data and the multiplication of 100 is introduced for convenience of presentation. Absolute values are used since financial data sets usually exhibit clusters of high volatility, caused by either positively or negatively large returns. Both positive and negative large returns are important in most practical volatility evaluations by risk analysts. The analyses reported here are based on a constant threshold with normal prior distribution and time varying scale and form parameters of the GPD.

Petrobras. We analyzed daily absolute returns for the Petrobrás company from July 1st 2002 to June 30th 2011 or 2267 observations. Petrobrás is the state-run Brazilian oil company and is ranked among the top 50 biggest companies in the world. This period includes the credit crunch that affected major financial markets around the globe (Lopes and Polson, 2010). This is highlighted by the larger volatility for the second half of the observed time series (the credit crunch of 2007-2008) as indicated by Figure 3.

The following models were entertained: i) stochastic volatility AR(1) model[2], ii) mixture of Gamma distributions, denote here by MG_k, iii) mixture of gammas below threshold and GPD beyond threshold with static parameters, denoted $MGPD_k$, and iv) mixture of gammas below threshold and GPD beyond threshold and time-varying tail parameters, denoted

[1] Similar results, not presented here but available upon request, were obtained when analyzing Vale do Rio Doce, which is, along with Petrobrás, the biggest (non-banking) company in Brazil.

[2] See, for instance, Lopes and Polson (2010) and their references for further details on stochastic volatility models.

$MGPDLM_k$. Table 2 compare these models for each time series via the deviance information criterion (DIC) of Spiegelhalter et al. (2002). Our proposed model clearly outperforms the existing alternatives. The bottom panels of Figure 3 presents the time-varying behavior of both tail parameters, i.e. ξ_t and σ_t. The figure shows 2 time periods when these become larger. The most prominent one corresponds to the 2nd semestre of 2008, a period of crisis with larger returns in absolute value. The scale parameter ξ_t reflects the increase in the weight of the tail, typical of more volatile market conditions. This will also impact the estimation of higher quantiles, that will typically increase as the scale of the GPD gets larger.

Table 2 shows that GDP with time-varying parameters fits well the data beyond the threshold, better than all other models considered. As discussed earlier in the simulated example, one of the key features of extreme value models is the possibility of computing high quantiles of the data, which is exacerbated when dealing with financial time series data with time-varying variance components. Figure 3 shows that both the 95th and the 99th percentiles of Petrobras' absolute returns follow the data pattern, specially so during the crisis in the 2nd semester of 2008.

Ibovespa. We analyzed daily absolute returns for the Bovespa Index. The Bovespa Index, or simply Ibovespa, is an index of about 50 stocks that are traded on the São Paulo Stock, Mercantile and Futures Exchange. Petrobras is included in the Ibovespa. Here we analyzed data from July 1st 2002 to June 30th 2011 or 2229 observations. Similar to the analysis of the Petrobras dataset, Table 2 shows that a MGDP with time-varying parameters fits well the data beyond the threshold, while a single gamma distribution is not enough for the data below the threshold. Similarly to the simulated exercise, Figure 4 shows the estimation of the maximum and high quantiles at any time point. It also shows that higher quantiles get larger and the upper limit vanishes during periods of crisis. During calm market periods, these tend to get smaller to the point of allowing for a finite upper limit to the values of the series with probability close to 1. In these cases, the very high 99.9999% quantile and the upper limit become close.

5 Conclusions

In this paper we propose an extension to the mixture model used by Nascimento et al. (2011b), by allowing parameters to vary across time. Dynamic linear models are introduced to model the time-varying behavior of tail parameters of the GPD. Posterior inference is accomplished approximately with MCMC methods are extensively applied, with particular emphasis on the Metropolis-Hastings and Gibbs types. A simulation study was performed and the results have shown that we obtained good estimates of the parameters and successfully caught the major time-varying patterns of the shape parameter. The real applications have also been encouraging in which they all point out towards the existence of time varying tail behavior in common financial time series data, even to the point of favoring this approach when compared against standard procedures such as stochastic volatility models

An immediate extension of our proposal, is to consider more complex dynamic structures for ξ, σ and u. Huerta and Sansó (2007), for example, used second order dynamic to model growth of the location parameter of a GEV distribution. Another extension involves dynamic regression structures for ξ and/or σ, where external, possibly exogenous, information might be combined with standard dynamic structures.

Appendix - MCMC scheme

In this section we detail the MCMC algorithm we designed to perform posterior inference regarding Θ. Following from the updated $l\xi_t\{y_t > u\}, l\sigma_t\{y_t > u\}, u, \alpha, \mu$ and p are drawn using standard Metropolis-Hastings steps. Simultaneously, also following from $l\xi_t\{y_t < u\}, l\sigma_t\{y_t < u\}, V_\xi, V_\sigma, W_\xi, W_\sigma, \theta_\xi$ and θ_σ are drawn via Gibbs steps. Suppose that at iteration s, the chain is at $\Theta^{(s)}$. Then, at iteration $s+1$, the algorithm cycles through the following steps.

Sampling (μ, α). The components of (μ, α) are sampled separately for each mixture component. The α_j's and μ_j's must be positive. Therefore, α_j^* is proposed from $\alpha_j^*|\alpha_j^{(s)} \sim G(\alpha_j^{(s)}, \alpha_j^{(s)^2}/V_{\alpha_j})$. Note that, $E(\alpha_j^* \mid \alpha_j^{(s)}) = \alpha_j^{(s)}$, and $Var(\alpha_j^* \mid \alpha_j^{(s)}) = V_{\alpha_j}$, for $j = 1, \ldots, k$. Same procedure is adopted for the proposal for μ_j, given by $\mu_j^*|\mu_j^{(s)} \sim G(\mu_j^{(s)}, \mu_j^{(s)^2}/V_{\mu_j})I_A$, where $I_A = I(\mu_1^{(s+1)} < \ldots < \mu_{j-1}^{(s+1)} < \mu_j^{(s)} < \mu_{j+1}^{(s)} < \ldots < \mu_k^{(s)})$ and the difference that they must also obey the order constraint. The values $\alpha_j^{(s+1)} = \alpha_j^*$ and $\mu_j^{(s+1)} = \mu_j^*$ are accepted with probability

$$\min\left\{1, \frac{\pi(\Theta^*|y) f_G(\mu_j^{(s)}|\mu_j^*, \mu_j^{*2}/V_{\mu_j}) f_G(\alpha_j^{(s)}|\alpha_j^*, \alpha_j^{*2}/V_{\alpha_j}) I(\mu_1^{(s+1)} < \ldots < \mu_j^* < \ldots < \mu_k^{(s)})}{\pi(\tilde{\Theta}|y) f_G(\mu_j^*|\mu_j^{(s)}, \mu_j^{(s)2}/V_{\mu_j}) f_G(\alpha_j^*|\alpha_j^{(s)}, \alpha_j^{(s)2}/V_{\alpha_j}) I(\mu_1^{(s+1)} < \ldots < \mu_j^{(s)} < \ldots < \mu_k^{(s)})}\right\},$$

where $\Theta^* = (\alpha_{<j}^{(s+1)}, \alpha_j^*, \alpha_{>j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_j^*, \mu_{>j}^{(s)}, p^{(s)}, u^{(s)}, \{l\xi_t^{(s)}\}_{t=1}^T, \{l\sigma_t^{(s)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$ and $\tilde{\Theta} = (\alpha_{<j}^{(s+1)}, \alpha_{\geq j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_{\geq j}^{(s)}, p^{(s)}, u^{(s)}, \{l\xi_t^{(s)}\}_{t=1}^T, \{l\sigma_t^{(s)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$, with $y_{<l} = (y_1, \ldots, y_{l-1})$ and $y_{\geq l} = (y_l, \ldots, y_k)$, for any vector $y = (y_1, \ldots, y_k)$.

Sampling p. p^* is sampled from a Dirichlet with parameters $(V_p p_1^{(s)}, \ldots, V_p p_k^{(s)})$, where V_p is a tuning constant that determines the variance of the proposal distribution. Then, set $p^{(s+1)} = p^*$ with probability $\min\{1, \pi(\Theta^*|y) f_D(p^{(s)}|p^*)/[\pi(\tilde{\Theta}|y) f_D(p^*|p^{(s)})]\}$, where $\Theta^* = (\alpha^{(s+1)}, \mu^{(s+1)}, p^*, u^{(s)}, \{l\xi_t^{(s)}\}_{t=1}^T, \{l\sigma_t^{(s)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$ and $\tilde{\Theta} = (\alpha^{(s+1)}, \mu^{(s+1)}, p^{(s)}, u^{(s)}, \{l\xi_t^{(s)}\}_{t=1}^T, \{l\sigma_t^{(s)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$,

Sampling $\{l\xi_t\}_{t=1}^T$. For a value t between 1 and T, if $x_t < u^{(s)}$, then the parameter can be samplet by: $l\xi_t^{(s+1)} \sim N(\theta_{\xi,t}^{(s)}, 1/V_\xi^{(s)})$. Now, if $x_t \geq u^{(s)}$ its necessary sampling $l\xi_t$ by Metropolis algorithm. The proposal distribution to $l\xi_t^*$ is a truncated Normal $N(l\xi_t^{(s)}, K_{\xi,t})I(\xi_U, \infty)$, where $\xi_U = \log(-\sigma_t^{(s)}/(x_t - u^{(s)}) + 1)$, $\sigma_t^{(s)} = \exp(l\sigma_t^{(s)})$. Then, $l\xi_t^{(s+1)} = l\xi_t^*$ with probability

$$\min\left\{1, \frac{\pi(\Theta^*|y)\Phi((l\xi_t^{(s)} - \xi_U)/\sqrt{K_{\xi,t}}))}{\pi(\tilde{\Theta}|y)\Phi((l\xi_t^* - \xi_U)/\sqrt{K_{\xi,t}}))}\right\},$$

where $\Theta^* = (\alpha^{(s+1)}, \mu^{(s+1)}, p^{(s+1)}, u^{(s)}, l\xi_{<t}^{(s+1)}, l\xi_t^*, l\xi_{>t}^{(s)}, \{l\sigma_t^{(s)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$ and $\tilde{\Theta} = (\alpha^{(s+1)}, \mu^{(s+1)}, p^{(s+1)}, u^{(s)}, l\xi_{<t}^{(s+1)}, l\xi_{\geq t}^{(s)}, \{l\sigma_t^{(s)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$.

Sampling $\{l\sigma_t\}_{t=1}^T$. For a value t between 1 and T, if $x_t < u^{(s)}$, then the parameter can be samplet by: $l\sigma_t^{(s+1)} \sim N(\theta_{\sigma,t}^{(s)}, 1/V_\sigma^{(s)})$. Now, if $x_t \geq u^{(s)}$ its necessary sampling $l\sigma_t$ by Metropolis algorithm. If $\xi_t^{(s+1)} = \exp(l\xi_t^{(s+1)}) - 1 > 0$, sampling $l\sigma_t^*$ by $N(l\sigma_t^{(s)}, K_{\sigma,t})$. Then, $l\sigma_t^{(s+1)} = l\sigma_t^*$ with probability $\min\{1, \pi(\Theta^*|\mathbf{x})/\pi(\tilde{\Theta}|\mathbf{x})\}$. If $\xi_t^{(s+1)} < 0$, the proposal distribution to $l\sigma_t^*$é is a truncated Normal $N(l\sigma_t^{(s)}, K_{\sigma,t})I(\sigma_U, \infty)$, where $\sigma_U = \log(-\xi_t^{(s)}) + \log(x_t - u^{(s)})$. Then, $l\sigma_t^{(s+1)} = l\sigma_t^*$ with probability

$$\min\left\{1, \frac{\pi(\Theta^*|y)\Phi((l\sigma_t^{(s)} - \sigma_U)/\sqrt{K_{\sigma,t}}))}{\pi(\tilde{\Theta}|y)\Phi((l\sigma_t^* - \sigma_U)/\sqrt{K_{\sigma,t}}))}\right\},$$

where $\Theta^* = (\alpha^{(s+1)}, \mu^{(s+1)}, p^{(s+1)}, u^{(s)}, \{l\xi_t^{(s+1)}\}_{t=1}^T, l\sigma_{<t}^{(s+1)}, l\sigma_t^*, l\sigma_{>t}^{(s)}, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$ and $\tilde{\Theta} = (\alpha^{(s+1)}, \mu^{(s+1)}, p^{(s+1)}, u^{(s)}, \{l\xi_t^{(s+1)}\}_{t=1}^T, l\sigma_{<t}^{(s+1)}, l\sigma_{\geq t}^{(s)}, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$.

Sampling u. The proposal threshold u^* is sampled from a $N(u^{(s)}, V_u)I(u_L^{(s)}, \infty)$, where

$$u_L^{(s)} = \max\left\{\min(x_1, \ldots, x_T), \max_{\{t:\xi_t^{(s+1)}<0, x_t>u^{(s)}\}} (x_t + \sigma^{(s+1)}/(\xi_t^{(s+1)}(1 + \xi_t^{(s+1)})))\right\},$$

V_u is the proposal variance distribution to the threshold. The value $u^{(s+1)} = u^*$ is accept with probability

$$\min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})\Phi((u^{(s)} - u_L^{(s)})/\sqrt{V_u})}{\pi(\tilde{\Theta}|\mathbf{x})\Phi((u^* - u_L^{(s)})/\sqrt{V_u})}\right\},$$

where $\Theta^* = (\alpha^{(s+1)}, \mu^{(s+1)}, p^{(s+1)}, u^{(s)}, \{l\xi_t^{(s+1)}\}_{t=1}^T, \{l\sigma_t^{(s+1)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$ and $\tilde{\Theta} = (\alpha^{(s+1)}, \mu^{(s+1)}, p^{(s+1)}, u^*, \{l\xi_t^{(s+1)}\}_{t=1}^T, \{l\sigma_t^{(s+1)}\}_{t=1}^T, \{\theta_{\xi,t}^{(s)}\}_{t=0}^T, \{\theta_{\sigma,t}^{(s)}\}_{t=0}^T, V_\xi^{(s)}, V_\sigma^{(s)}, W_\xi^{(s)}, W_\sigma^{(s)})$.

Gibbs steps for the parameters of the dynamic linear models. It is relatively simple to show that V_ξ, W_ξ, $\theta_{\xi,t}$, V_σ, W_σ and $\theta_{\sigma,t}$ can be updated via Gibbs steps

$$
\begin{aligned}
V_\xi^{(s+1)} &\sim G\left(\frac{f_\xi + \frac{T}{2}}{o_\xi + \sum_{t=1}^T (l\xi_t^{(s+1)} - \theta_{\xi,t}^{(s)})^2}, f_\xi + \frac{T}{2}\right), \\
W_\xi^{(s+1)} &\sim G\left(\frac{l_\xi + \frac{T}{2}}{m_\xi + \sum_{t=1}^T (\theta_{\xi,t}^{(s)} - \theta_{\xi,t-1}^{(s)})^2}, l_\xi + \frac{T}{2}\right), \\
\theta_{\xi,0}^{(s+1)} &\sim N\left(\frac{W_\xi^{(s+1)}\theta_{\xi,1}^{(s)} + m_{\xi,0}/C_{\xi,0}}{W_\xi^{(s+1)} + 1/C_{\xi,0}}, \frac{1}{W_\xi^{(s+1)} + 1/C_{\xi,0}}\right), \\
\theta_{\xi,t}^{(s+1)} &\sim N\left(\frac{V_\xi^{(s+1)} l\xi_t^{(s+1)} + W_\xi^{(s+1)}(\theta_{\xi,t+1}^{(s)} - \theta_{\xi,t-1}^{(s+1)})}{V_\xi^{(s+1)} + 2W_\xi^{(s+1)}}, \frac{1}{V_\xi^{(s+1)} + 2W_\xi^{(s+1)}}\right), \\
t = 1, & \quad \ldots \quad, T-1, \\
\theta_{\xi,T}^{(s+1)} &\sim N\left(\frac{V_\xi^{(s+1)} l\xi_T^{(s+1)} + W_\xi^{(s+1)}\theta_{\xi,T-1}^{(s+1)}}{V_\xi^{(s+1)} + W_\xi^{(s+1)}}, \frac{1}{V_\xi^{(s+1)} + W_\xi^{(s+1)}}\right), \\
V_\sigma^{(s+1)} &\sim G\left(\frac{f_\sigma + \frac{T}{2}}{o_\sigma + \sum_{t=1}^T (l\sigma_t^{(s+1)} - \theta_{\sigma,t}^{(s)})^2}, f_\sigma + \frac{T}{2}\right), \\
W_\sigma^{(s+1)} &\sim G\left(\frac{l_\sigma + \frac{T}{2}}{m_\sigma + \sum_{t=1}^T (\theta_{\sigma,t}^{(s)} - \theta_{\sigma,t-1}^{(s)})^2}, l_\sigma + \frac{T}{2}\right), \\
\theta_{\sigma,0}^{(s+1)} &\sim N\left(\frac{W_\sigma^{(s+1)}\theta_{\sigma,1}^{(s)} + m_{\sigma,0}/C_{\sigma,0}}{W_\sigma^{(s+1)} + 1/C_{\sigma,0}}, \frac{1}{W_\sigma^{(s+1)} + 1/C_{\sigma,0}}\right), \\
\theta_{\sigma,t}^{(s+1)} &\sim N\left(\frac{V_\sigma^{(s+1)} l\sigma_t^{(s+1)} + W_\sigma^{(s+1)}(\theta_{\sigma,t+1}^{(s)} - \theta_{\sigma,t-1}^{(s+1)})}{V_\sigma^{(s+1)} + 2W_\sigma^{(s+1)}}, \frac{1}{V_\sigma^{(s+1)} + 2W_\sigma^{(s+1)}}\right), \\
t = 1, & \quad \ldots \quad, T-1, \\
\theta_{\sigma,T}^{(s+1)} &\sim N\left(\frac{V_\sigma^{(s+1)} l\sigma_T^{(s+1)} + W_\sigma^{(s+1)}\theta_{\sigma,T-1}^{(s+1)}}{V_\sigma^{(s+1)} + W_\sigma^{(s+1)}}, \frac{1}{V_\sigma^{(s+1)} + W_\sigma^{(s+1)}}\right).
\end{aligned}
$$

References

Behrens, C., D. Gamerman, and H. F. Lopes (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling 4*, 227–244.

Coles, S. G. (2001). *Extreme Value Theory an Applications*. Kluver Academic Publishers.

Coles, S. G. and E. Powell (1996). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review 64*, 119–136.

Coles, S. G. and J. A. Tawn (1996a). A Bayesian analysis of extreme rainfall data. *Applied Statistics 45*, 463–78.

Coles, S. G. and J. A. Tawn (1996b). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society B 58*, 329–347.

DuMouchel, M. H. (1983). Estimating the stable index α in order to measure tail thickness: a critique. *The Annals of Statistics 11*, 1019–1031.

Embrechts, P., C. Küppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. New York: Springer.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd edition)*. Baton Rouge: Chapman and Hall/CRC.

Huerta, G. and B. Sansó (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics 14(3)*, 285–299.

Lopes, H. F. and N. Polson (2010). Extracting SP500 and NASDAQ volatility: The credit crisis of 2007-2008. *The Oxford Handbook of Applied Bayesian Analysis Oxford: Oxford University Press*, 319–342.

Nascimento, F. F., D. Gamerman, and H. F. Lopes (2011a). Regression models for exceedance data via the full likelihood. *Environmental and Ecological Statistics 18*, 495–512.

Nascimento, F. F., D. Gamerman, and H. F. Lopes (2011b). Semiparametric Bayesian approach to extreme estimation. *Statistics and Computing Accepted for publication.*

Parmesan, C., T. L. Root, and M. R. . Willing (2000). Impacts of extreme weather and climate on terrestrial biota. *Bulletin of the American Meteorological Society 81*, 443–450.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 3*, 119–131.

Smith, R. L. (1984). Threshold models for sample extremes. *Statistical extremes and applications*, 621–638.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B 64*, 583–639.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer - Second Edition.

Wiper, M., D. Rios Insua, and F. Ruggeri (2001). Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics 10*, 440–454.

	T=1,000			T=2,500			T=10,000		
	True	Mean	95% C. I.	True	Mean	95% C. I.	True	Mean	95% C. I.
μ	5.0	4.8	(4.4,5.1)	5.0	4.8	(4.7,5.0)	5.0	4.94	(4.82,5.05)
α	1.0	1.0	(0.96,1.14)	1.0	1.02	(0.97,1.08)	1.0	1.01	(0.98,1.03)
u	7.5	7.6	(7.5,8.1)	7.8	7.8	(7.7,7.9)	7.9	7.9	(7.8,8.0)
$\theta_{\xi,0}$	0.2	0.26	(0.1,0.4)	0.2	0.20	(0.0,0.4)	0.2	0.16	(0.0,0.4)
$\theta_{\sigma,0}$	2.0	2.0	(1.9,2.2)	2.0	2.0	(1.8,2.2)	2.0	2.0	(1.8,2.1)
V_ξ	200	184	(34, 434)	200	214	(79,442)	200	228	(74,473)
V_σ	200	196	(46,402)	200	193	(74,404)	200	155	(53,390)
W_ξ	1000	998	(814,1189)	1000	1025	(847,1220)	1000	1006	(841,1178)
W_σ	1000	997	(810,1199)	1000	1020	(839,1219)	1000	1017	(840,1216)

Table 1: *Simulated data:* True values, posterior means and 95% posterior credibility intervals of model parameters.

	Model				
Time series	$MGPDLM_1$	$MGPDLM_2$	MG_2*	$MGPD_1*$	SV
Petrobras	3061	**2996**	3106	3008	3165
Ibovespa	6563	**6495**	6684	6683	6808

Table 2: *Real data:* Deviance information criterion for Petrobras and Ibovespa. MG_k* and $MGPD_k*$ are the models within each class with the smallest deviance. SV is the standard stochastic volatility model with AR(1) dynamics for log-volatilities.

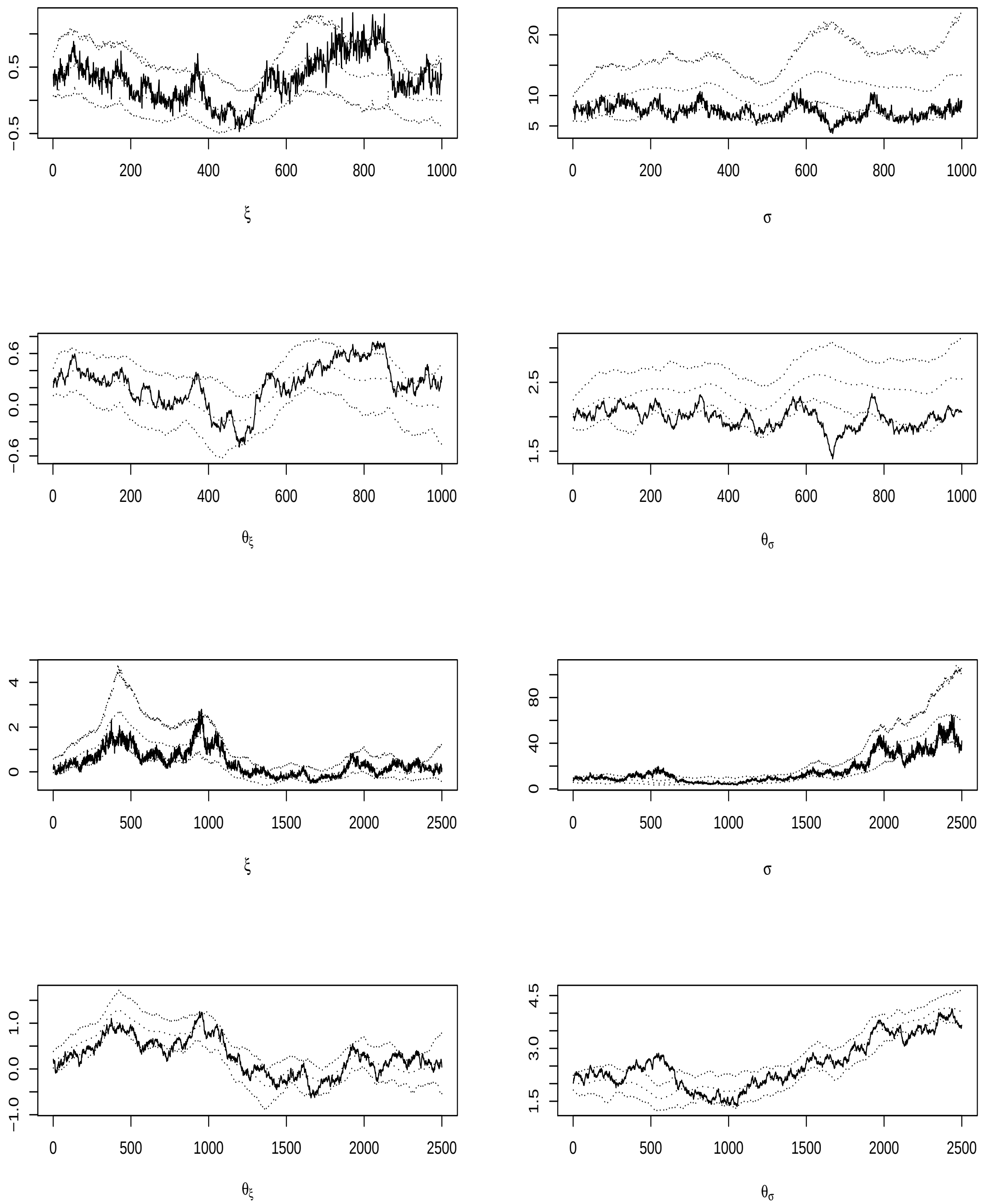

Figure 1: *Simulated data:* Posterior means and 95% credibility intervals for ξ_t, σ_t, $\theta_{\xi,t}$ and $\theta_{\sigma,t}$. Sample size $T = 1,000$ (top two rows) and $T = 2,500$ (bottom two rows). True values are the solid lines.

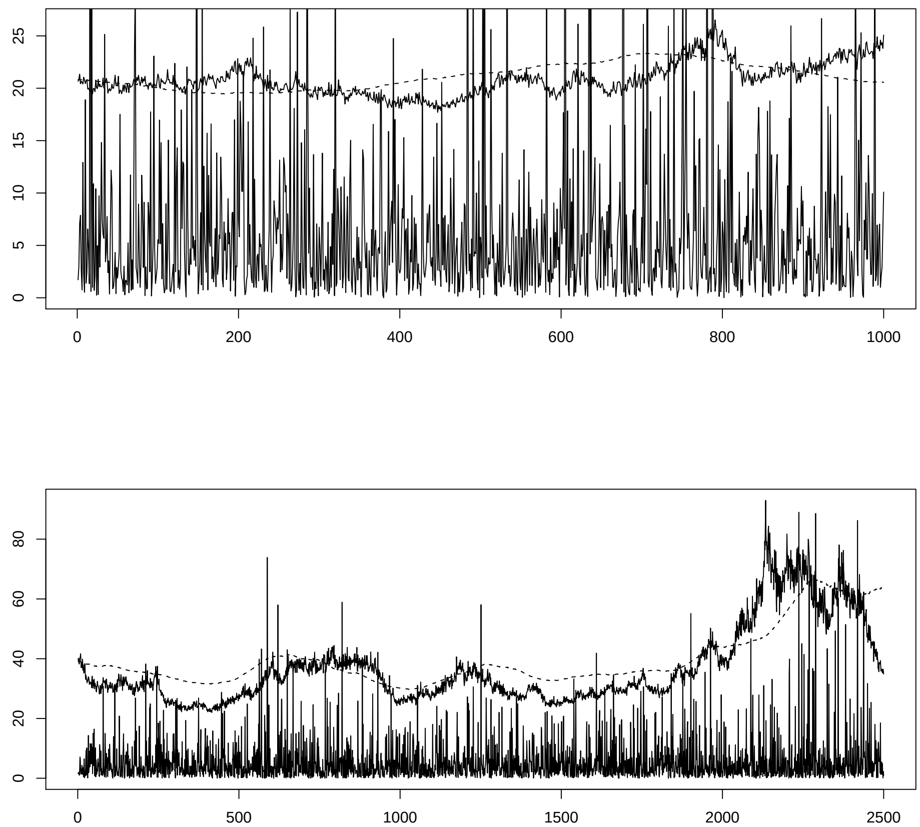

Figure 2: *Simulated data:* Simulated data along with simulated (solid line) and posterior mean (dotted line) for extreme quantiles. *Top frame:* $T = 1,000$ observations and 95th quantiles. *Bottom frame:* $T = 2,500$ observations and 99th quantiles.

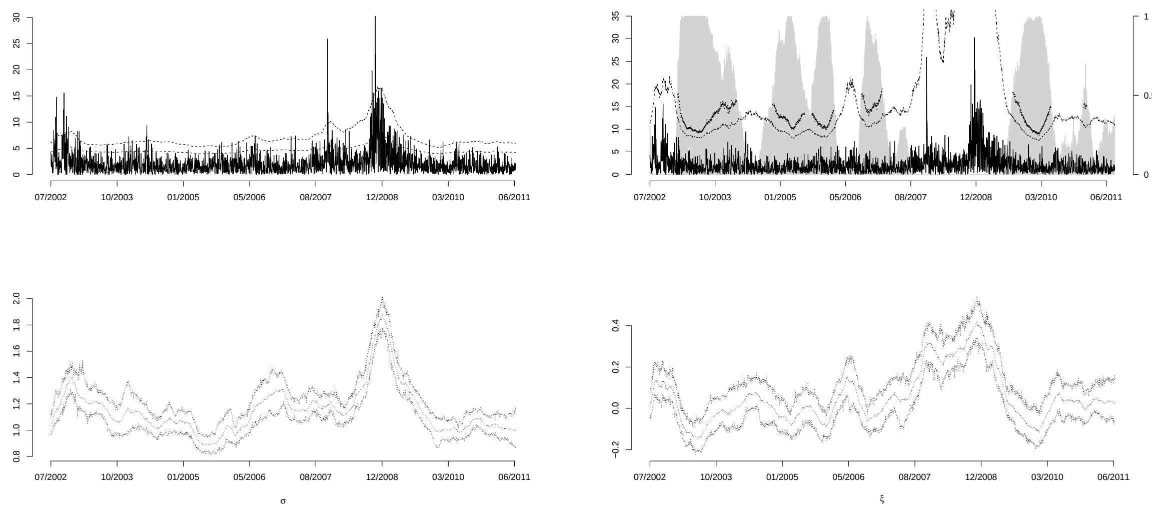

Figure 3: *Petrobras time series:* Top frame: Time series of absolute returns along with 95th and 99th percentiles (left panel) and along with 99.9999th quantiles and maximum when the posterior median of $\xi < 0$ (right panel). The grey area represents the posterior probability of existence of a finite maximum, $P(\xi_t < 0|y)$, for all t. Bottom frame: Posterior means and 95% credibility intervals for σ_t (left panel) and ξ_t (right panel).

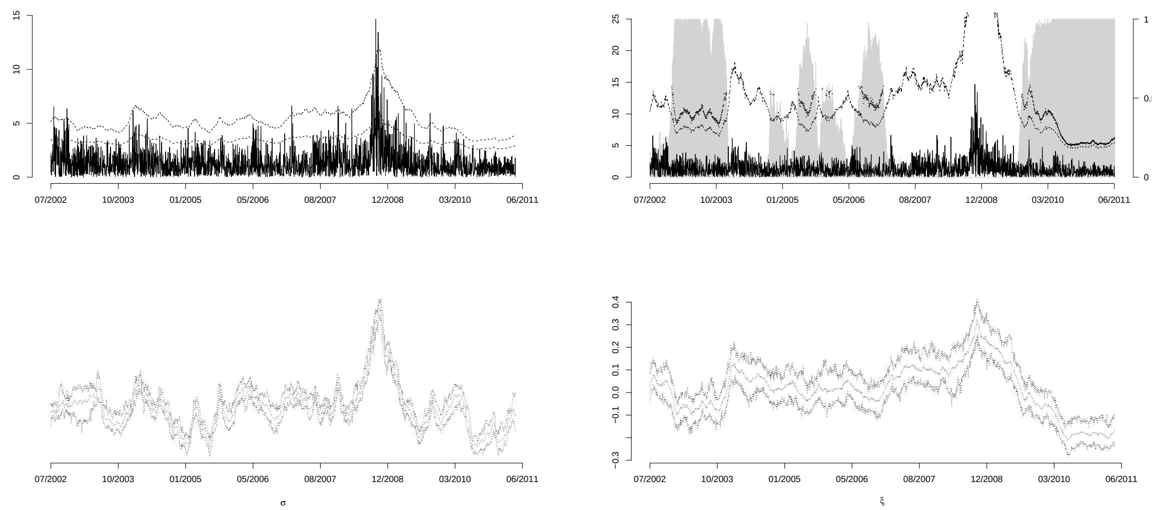

Figure 4: *Ibovespa time series:* Top frame: Time series of absolute returns along with 95th and 99th percentiles (left panel) and along with 99.9999th quantiles and maximum when the posterior median of $\xi < 0$ (right panel). The grey area represents the posterior probability of existence of a finite maximum, $P(\xi_t < 0|y)$, for all t. Bottom frame: Posterior means and 95% credibility intervals for σ_t (left panel) and ξ_t (right panel).