Insper

Business and Economics
Working Papers

BEWP 201/2014

Temporal dependence in extremes with dynamic models

Fernando Ferraz do Nascimento
Dani Gamerman
Hedibert Freitas Lopes

Noname manuscript No.
(will be inserted by the editor)

Temporal dependence in extremes with dynamic models

Fernando Ferraz do Nascimento · Dani Gamerman · Hedibert Freitas Lopes

Received: date / Accepted: date

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, using a two component mixture model. Extremes beyond a threshold are assumed to follow a generalized Pareto distribution (GPD). Temporal dependence is induced by allowing to GPD parameter to vary with time. Temporal variation and dependence is introduced at a latent level via the novel use of dynamic linear models (DLM). Novelty lies in the time variation of the shape and scale parameter of the resulting distribution. These changes in limiting regimes as time changes reflect better the data behavior, with important gains in estimation and interpretation. 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: returns of Petrobrás stocks and SP500 index. Results show advantage of our proposal over currently entertained models such as stochastic volatility, with improved estimation of high quantiles and extremes.

Fernando Ferraz do Nascimento
Curso de Estatística, Universidade Federal do Piauí, Campus Ministro Petrônio Portela, CCN2, 64049-550, Teresina, Brazil, E-mail: fernandofn@ufpi.edu.br

Dani Gamerman
Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, 21945-970, Rio de Janeiro, Brazil E-mail: dani@im.ufrj.br

Hedibert Freitas Lopes
INSPER Institute of Education and Research, Rua Quata 300, Vila Olimpia, São Paulo-SP, Brazil, 04546-042 E-mail: hedibertFL@insper.edu.br

Keywords GPD · Bayesian · nonparametric · MCMC

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 (2011, 2012), 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 are the shape ξ and scale σ. 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 $\xi > 0$.

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 (2012) 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.

Finance and Environment Studies are key areas of application of extreme value theory. Both provide useful insight into the need for time-varying models.

The most famous environmental issue associated with time variation is global warming, where the behavior of extremes seem to suggest that trends are taking place. The El Nino is another example where seasonal effects are known to occur. Determination and quantification of these temporal variation is vital in understanding the climate pattern and establishing control policies. An early account of the relevance of time series modeling in the analysis of extremes is provided in a study of ground level ozone by Smith (1989). Similar ideas were also applied in the reliability context by Smith (1986). Finance studies are also affected by time variation. The returns of financial assets are known to behave according to non-linear time series. Failure to recognize and quantify these changes will inevitably lead to biased estimates of important decision quantities such as VaR and volatilities. It will be shown that time-varying extreme models are capable of capturing adequately the changes of financial markets, helping financial analysts and their investors with their portfolio allocations. See, for example, Zhao et al (2011), who uses a GARCH structure to model the time varying parameters of a GEV distribution for financial returns data.

The evidence in favor of time varying models in these settings is overwhelming. As an example, consider the time series of absolute returns of stocks of Figure 1. Details about this dataset are provided in Section 4. Assume the series is divided in 2 equal parts and separate static extreme value analyses are performed for each half with the approach of Nascimento et al (2012). Figure 2 shows the results of this exercise. Markedly different patterns are observed, with strong evidence of a light tail behavior ($\xi < 0$) in the first half and a heavy tail behavior ($\xi > 0$) in the second half. These results are expected and reflect the market volatility of the time frame considered. Although it is possible to adapt the model to consider one or multiple change points (see, for example, Dierckx and Teugels (2010)), it might be more convenient, either or both computationally or methodologically, to entertain the possibility of (quasi-)continuous variation in the model's parameters.

This paper extends 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 achieved by embedding the extreme parameters in a non-linear dynamic system (West and Harrison, 1997). A similar idea was proposed by Huerta and Sansó (2007), where dynamic models are used to describe temporal variation of the location of GEV distributions. The approach is used to model daily ozone levels in the US.

In our work, both GPD tail parameters are allowed to vary simultaneously in time. Reparametrizations are used to adequately accommodate time variations. Also, interest interpretations are introduced as the shape represents the tail weight. Data variations over time may lead to change in the tail behavior with light tails observed at some times, leading to finite upper data limits, while heavy tails may be observed in other ones. These features seem to capture well periods of turbulence when financial data is analyzed, as will be illustrated later in the applications section.

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 financial time series from stock exchange data. The relevance of the changes introduced in this paper will be illustrated 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 $\{y_t|y_t > u_t\}$ 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. Tancredi et al (2006) uses a mixture of uniform densities for the central part of the data and the number of observations beyond the threshold is a parameter to be estimated. Macdonald et al (2011) use a kernel density estimator to fit the data below the threshold. 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 (2012) 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

$$h(x;\mu,\alpha,p) = \sum_{j=1}^{k} p_j f_{Ga}(x;\mu_j,\alpha_j), \tag{2}$$

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 $f_{Ga}(x;\mu,\alpha)$ 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. There are a number of options for choosing the number of

components. These include the use of a summary criteria, estimation of this number and posterior predictive approaches. The number of components may be estimated by reversible jump MCMC schemes but this inevitably involves an increase in computational costs. Dey et al (1995) shows that a small number of components is enough to describe the main features of the density in many applications. Nascimento et al (2012) showed in simulations and real applications of analysis of extreme data with mixture models that inclusion of irrelevant components will lead to negligible weights for them, thus providing an indirect determination of the number of components. They also showed that standard summary criteria such as BIC (Schwarz, 1978) and DIC (Spiegelhalter et al, 2002) seem to work well in these settings. These ideas are used in our applications of Sections 3 and 4.

The priors for μ and α follow Wiper et al (2001), 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 order restriction on the μ's is imposed to ensure parameter identifiability. Nascimento et al (2012) discuss this issue for related models and provide theoretical and empirical evidence that the model is identified, based also on similar mixture results from Wiper et al (2001).

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 μ's and the α's. 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. This 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

$$F_t(y_t; \Theta) = \begin{cases} H(y_t; \mu, \alpha, p) & \text{if } y_t < u \\ H(u; \mu, \alpha, p) + [1 - H(u; \mu, \alpha, p)] G(y_t - u; \xi, \sigma) & \text{if } y_t \geq u \end{cases}, \tag{3}$$

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

$$L(y; \Theta) = \prod_{\{t|y_t<u_t\}} h(y_t; \mu, \alpha, p) \prod_{\{t|y_t \geq u_t\}} (1 - H(u_t; \mu, \alpha, p)) g(y_t - u_t; \xi_t, \sigma_t). \tag{4}$$

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. It is straightforward to obtain the p-quantile, by

$$q = u_t + \frac{((1 - p^*)^{-\xi_t} - 1)\sigma_t}{\xi_t}, \tag{5}$$

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 $\Psi_t = (u_t, \xi_t, \sigma_t)$. These parameters are expected to be related and can be described probabilistically in an evolution form

$$\Psi_t = g(\Psi_{t-1}, w_t), \tag{6}$$

where g is a possibly non-linear function and the w_t are random disturbance random vectors. The temporal relation (6) 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 (6). 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

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

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

$$\begin{aligned} l\xi_t &= \theta_{\xi,t} + v_{\xi,t} \;\; v_{\xi,t} \sim N(0, 1/V_\xi), \; \theta_{\xi,t} = \theta_{\xi,t-1} + w_{\xi,t} \;\; w_{\xi,t} \sim N(0, 1/W_\xi) \\ l\sigma_t &= \theta_{\sigma,t} + v_{\sigma,t} \; v_{\sigma,t} \sim N(0, 1/V_\sigma), \; \theta_{\sigma,t} = \theta_{\sigma,t-1} + w_{\sigma,t} \;\; w_{\sigma,t} \sim N(0, 1/W_\sigma) \\ lu_t &= \theta_{u,t} + v_{u,t} \; v_{u,t} \sim N(0, 1/V_u), \; \theta_{u,t} = \theta_{u,t-1} + w_{u,t} \; w_{u,t} \sim N(0, 1/W_u), \end{aligned} \quad (7)$$

for $t = 1, \ldots, T$ 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 $Ga(f_\xi, o_\xi)$, $Ga(f_\sigma, o_\sigma)$ and $Ga(f_u, o_u)$, while the priors for W_ξ, W_σ and W_u are $Ga(l_\xi, m_\xi)$, $Ga(l_\sigma, m_\sigma)$ and $Ga(l_u, m_u)$. Care must be exercised as very large prior variances may cause no convergence of the MCMC algorithm. However, simulation studies showed that the prior information does not need to be very precise. Similar to the hyperparameters of the mixture components, the vector $(m_{0,\xi}, C_{0\xi}, m_{0,\sigma}, C_{0\sigma}, m_{0,u}, C_{0u}, f_\xi, o_\xi, f_\sigma, o_\sigma, f_u, o_u, l_\xi, m_\xi, l_\sigma, m_\sigma, l_u, m_u)$ is 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 (7) 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 (6). In particular, (7) 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 (7). 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. Nascimento et al (2012) performed a sensitivity analysis in simulations and showed that mild information is enough to ensure appropriate recovery of the true threshold. The mild information is in the form of

a prior variance value large enough to cover the data range with high probability. This issue is only relevant for small and moderate sample sizes. For large sample sizes, the threshold is well recovered with any prior variance. A normal distribution to the logarithm of the threshold is another possibility to avoid the truncation, as in (7), but the effect of the truncation is negligible in all our applications.

2.4 Posterior and Predictive 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

$$
\begin{aligned}
\pi(\Theta|y) \propto & \prod_{\{t|y_t<u\}} \left[\sum_{j=1}^{k} p_j f_{Ga}(y_t;\mu_j,\eta_j) \right] \\
& \times \prod_{\{t|y_t\geq u\}} \left[\left(1 - \sum_{j=1}^{k} p_j F_{Ga}(u;\mu_j,\eta_j) \right) g(y_t-u;\xi_t,\sigma_t) \right] \\
& \times \prod_{j=1}^{k} \left[p_j^{\gamma} \eta_j^{a_j-1} e^{-b_j\eta_j} \beta_j^{-(c_j+1)} e^{-d_j/\mu_j} \right] \\
& \times V_{\xi}^{T/2+f_{\xi}-1} \exp\left(-\frac{V_{\xi}}{2} \sum_{t=1}^{T} (l\xi_t - \theta_{\xi,t})^2 - o_{\xi} V_{\xi} \right) \exp\left(-\frac{1}{2C_{\xi,0}} (\theta_{\xi,0} - m_{\xi,0})^2 \right) \\
& \times W_{\xi}^{T/2+l_{\xi}-1} \exp\left(-\frac{W_{\xi}}{2} \sum_{t=1}^{T} (\theta_{\xi,t} - \theta_{\xi,t-1})^2 - m_{\xi} W_{\xi} \right) \\
& \times V_{\sigma}^{T/2+f_{\sigma}-1} \exp\left(-\frac{V_{\sigma}}{2} \sum_{t=1}^{T} (l\sigma_t - \theta_{\sigma,t})^2 - o_{\sigma} V_{\sigma} \right) \exp\left(-\frac{1}{2C_{\sigma,0}} (\theta_{\sigma,0} - m_{\sigma,0})^2 \right) \\
& \times W_{\sigma}^{T/2+l_{\sigma}-1} \exp\left(-\frac{W_{\sigma}}{2} \sum_{t=1}^{T} (\theta_{\sigma,t} - \theta_{\sigma,t-1})^2 - m_{\sigma} W_{\sigma} \right) \\
& \times V_{\sigma}^{T/2+f_{u}-1} \exp\left(-\frac{V_{u}}{2} \sum_{t=1}^{T} (lu_t - \theta_{u,t})^2 - o_{u} V_{u} \right) \exp\left(-\frac{1}{2C_{u,0}} (\theta_{u,0} - m_{u,0})^2 \right) \\
& \times W_{u}^{T/2+l_{u}-1} \exp\left(-\frac{W_{u}}{2} \sum_{t=1}^{T} (\theta_{u,t} - \theta_{u,t-1})^2 - m_{u} W_{u} \right), \qquad (8)
\end{aligned}
$$

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 and second row, while the prior for the parameters below u is in the third row. Fourth to seventh 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, $D_k(0.01, \ldots, 0.01)$ 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 $(f_\xi, o_\xi, f_\sigma, o_\sigma, l_\xi, m_\xi, l_\sigma, m_\sigma)$ 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 3 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 addition 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 4 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 4 shows that true maxima are well estimated by the model.

4 Application to stock returns

The dynamic extreme value model is now applied to the study of two financial time series: i) Petrobrás[1] returns and ii) SP500 Returns, is a stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. 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 analysis reported here are based on a constant threshold with normal prior distribution and time varying scale and form parameters of the GPD. The threshold parameter, either constant or time-varying, is the hardest to form posterior inference about and it particularly susceptible to identification problems. Nascimento et al (2012) proposes a middy informative prior for the (constant) threshold parameters. See also Cooley et al (2012).

The same convergence criteria of the previous section were applied in this Section to ensure appropriateness of the results. In order to avoid excess return

[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.

clustering, maxima of sets of k days are considered. When analyzing SP500 and Petrobrás data, we considered groups of $k = 5$ and $k = 3$ days, respectively.

Petrobrás. We analyzed daily absolute returns for the Petrobrás company from October 8st 2000 to February 18th 2014 or 1131 3-day maxima. 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 the left panel in Figure 6.

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 $MGPDLM_k$. The proposed model was also implemented when ξ or σ are constant. The subscript k denotes to the number of components in the Gamma mixture for all models entertained. 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.

Table 2 shows that GDP with time-varying parameters fits well the data beyond the threshold, better than all alternative models. The best model according to table 2 is indeed the proposed one, where ξ is time-varying and σ constant.

The left panel of Figure 5 presents the time-varying behavior of ξ for Petrobrás. The figure shows 2 time periods when these become larger. The most prominent one corresponds to the 2nd semester 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.

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. The left panel of Figure 6 shows that both the 95th and the 99th percentiles of Petrobrás' absolute returns follow the data pattern, specially so during the crisis in the 2nd semester of 2008.

Figure 7 shows the estimation of the maximum and high quantiles at any time point, with posterior probabilities of $\xi < 0$, which implies a sharp upper tail. It also shows that higher quantiles get larger and the upper limit vanishes during periods of crisis. During calm market periods, these tend to allow 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.

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

In order to investigate how excess returns are affected by changes in ξ, the left panel of Figure 8 shows the expected return plot over t. The return curve has higher levels around 2008. For example, a return level of about 10% was expected once every 50 time periods in 2003, but once every 18 time periods in 2008.

SP500. We analyzed daily absolute returns for the SP500 Index. Here we analyzed data from Jaguary 2st 1950 to February 18th 2014 or 3345 5-day maxima. Similar to the analysis of the Petrobrás 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. Once again, our proposed model turned out to be the best model, with time varying σ and constant ξ. The right panel of Figure 5 show that for a long period of 64 years the time series, the scale parameters exhibits large variability, with visible peaks in 1973, 1982, 1999 and 2008, all of which corresponding to well established market crisis (Carvalho and Lopes, 2007).

The right panel of figure 6 presents the time-varying behavior of high quantiles, which are fairly synchronized with periods of higher returns. The right panel of Figure 8 shows the expected return plot over t. The two outer courts correspond to 1988 and 2008, which are periods of low and high volatility, respectively.

5 Conclusions

In this paper we propose an extension to the mixture model used by Nascimento et al (2012), 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 that 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.

Acknowledgements The research of the 1st and 2nd authors was supported by grants from FAPERJ and CNPq-Brazil. The research of the 3rd author was supported by the University of Chicago Booth School of Business.

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

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

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 $l\sigma_t^*$ 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

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

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

References

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

Carvalho C, Lopes HF (2007) Simulation-based sequential analysis of markov switching stochastic volatility models. Computational Statistics and Data Analysis 51(9):4526–4542

Coles SG (2001) Extreme Value Theory an Applications. Kluver Academic Publishers

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

Coles SG, Tawn JA (1996a) A bayesian analysis of extreme rainfall data. Applied Statistics 45:463–78

Coles SG, Tawn JA (1996b) Modelling extremes of the areal rainfall process. Journal of the Royal Statistical Society B 58:329–347

Cooley D, Cisewski J, Erhardt RJ, Jeon S, Mannshardt E, Omolo BO, Sun Y (2012) A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. REVSTAT-Statistical Journal 10:135–165

Dey D, Kuo L, Sahu S (1995) A bayesian predictive approach to determining the number of components in a mixture distribution. Statistics and Computing 5:297–305

Dierckx G, Teugels JL (2010) Change point analysis of extreme values. Environmetrics 21:661–686

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

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

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

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

Lopes HF, Polson N (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

Macdonald A, Scarrott CJ, Lee D, Darlow B, Reale M, Russell G (2011) A flexible extreme value mixture model. Computational Statistics and Data Analysis 55:2137–2157

Nascimento FF, Gamerman D, Lopes HF (2011) Regression models for exceedance data via the full likelihood. Environmental and Ecological Statistics 18:495–512

Nascimento FF, Gamerman D, Lopes HF (2012) Semiparametric bayesian approach to extreme estimation. Statistics and Computing 22:661–675

Parmesan C, Root TL, Willing MR (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

Schwarz G (1978) Estimating the dimension of a model. Annals of Statistics 6:461–4

Smith RL (1984) Threshold models for sample extremes. Statistical extremes and applications pp 621–638

Smith RL (1986) Extreme value statistics and reliability applications. Reliability Engineering 15:161–170

Smith RL (1989) Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. Statistical Science 14:367–377

Spiegelhalter DJ, Best NG, Carlin BP, Linde A (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B 64:583–639

Tancredi A, Anderson C, O'Hagan A (2006) Accounting for threshold uncertainty in extreme value estimation. Extremes 9:87–106

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

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

Zhao X, Scarrott CJ, Oxley L, Reale M (2011) Garch dependence in extreme value models with bayesian inference. Mathematics and Computers in Simulation 81:1430–1440

Tables and Figures

	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	MG_k	$MGPD_k$	SV	$Prop$	$Prop_\sigma$	$Prop_\xi$
Petrobrás	4529	4524	4885	4467	**4444**	4525
SP500	9412	9438	9338	9297	9320	**9295**

Table 2 *Real data:* Deviance information criterion for Petrobrás 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. $Prop$ is the proposed model and $Prop_\sigma$ and $Prop_\xi$ are respectively the proposed model with σ and ξ constant.

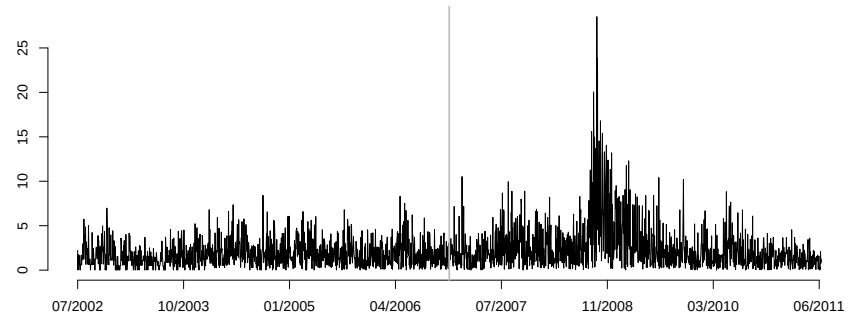

Fig. 1 Series of absolute returns of Vale stocks. The grey vertical mine indicates the point where the series was split into 2 parts.

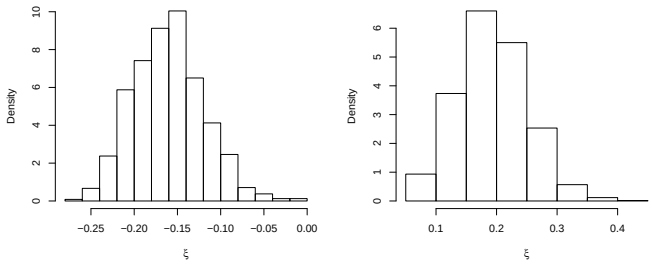

Fig. 2 Posterior histograms of parameter ξ of the GPD distribution for Vale stock series. Left: 1st half of data; right: 2nd half of data.

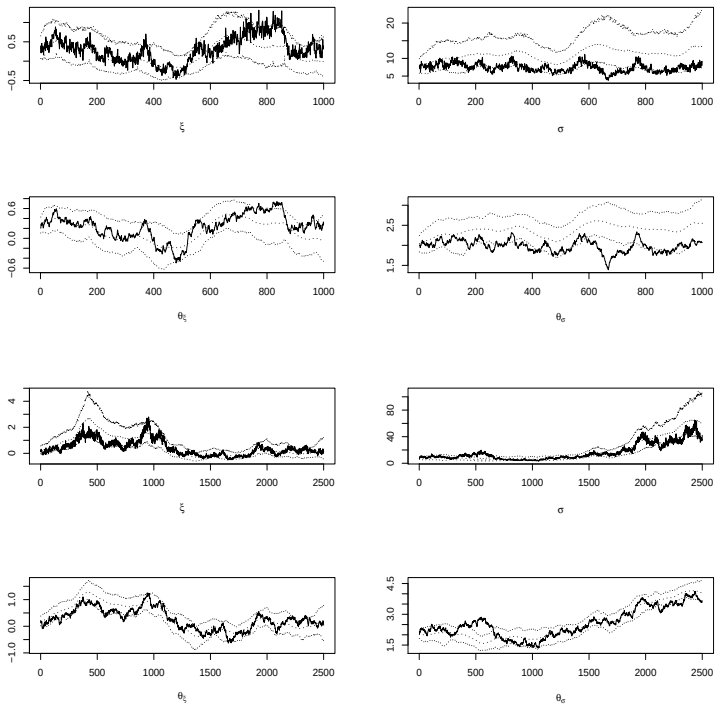

Fig. 3 *Simulated data:* Posterior means and 95% credibility intervals for ξ_t, σ_t, $\theta_{\xi,t}$ and $\theta_{\sigma,t}$. Sample size $T = 1,000$ (top two rows) and $T = 2,500$ (bottom two rows). True values are the solid lines.

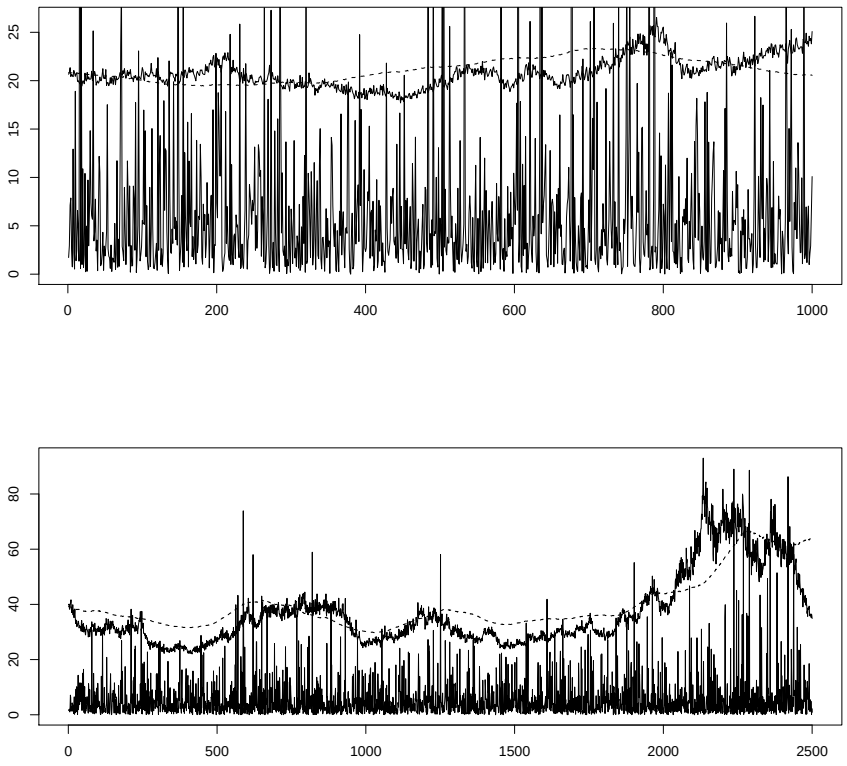

Fig. 4 *Simulated data:* Simulated data along with simulated (solid line) and posterior mean (dotted line) for extreme quantiles. *Top frame:* $T = 1,000$ observations and 95th quantiles. *Bottom frame:* $T = 2,500$ observations and 99th quantiles.

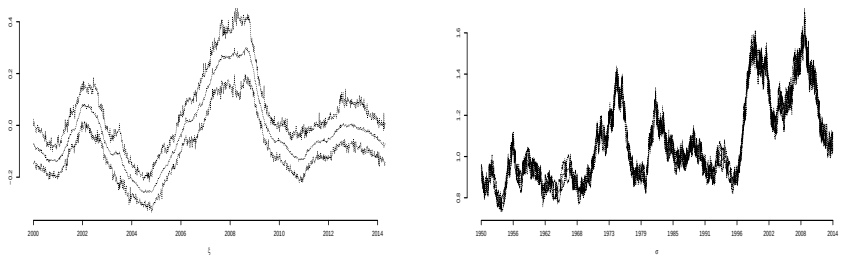

Fig. 5 Posterior means and 95% credibility intervals for ξ_t for Petrobrás (left) and σ_t for SP500 (right).

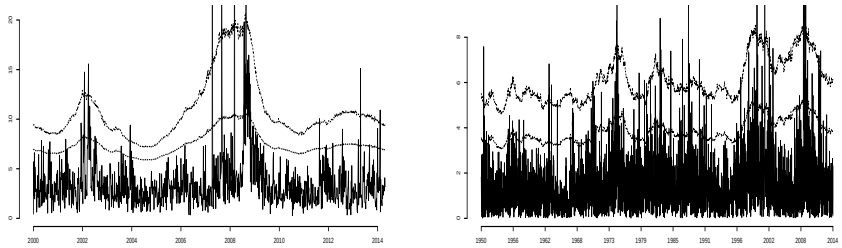

Fig. 6 Time series of absolute returns along with 95th and 99th percentiles for Petrobrás (left panel) and SP500 (right panel)

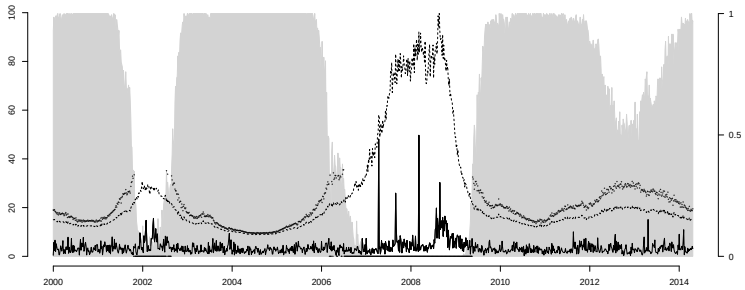

Fig. 7 Time series of absolute returns of Petrobrás along 99.9999th quantiles and maximum when the posterior median of $\xi < 0$. The grey area represents the posterior probability of existence of a finite maximum, $P(\xi_t < 0|y)$, for all t.

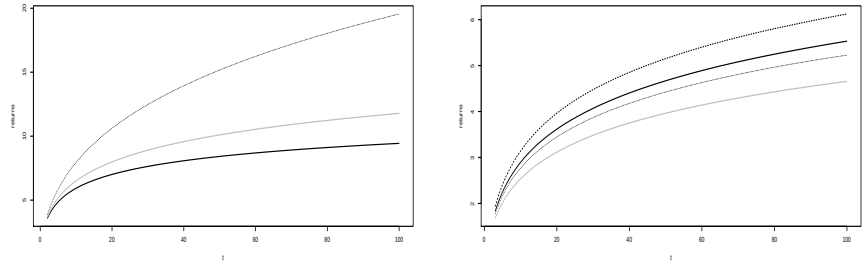

Fig. 8 RETURN LEVEL PLOT FOR APPLICATIONS, IN DIFFERENT POINTS IN THE TIME. LEFT: Petrobrás IN October 2000 (Full), February 2003 (Grey) and July 2008 (Dashed). RIGHT: SP500 IN JANUARY 1950 (Full), April 1988 (Grey), March 2008 (Dashed) and February 2014 (Dotted).