Generalized Pareto models with time-varying tail behavior

Hedibert F. Lopes, Fernando F. Nascimento and Dani Gamerman

Abstract

In this paper we analyze the extremal events using generalized Pareto distributions (GPD), allowing the parameters of GPD to vary with time. We use a mixture model that combines a nonparametric approach for the center and a GPD for the tail of distributions, in which the uncertainty about the threshold is explicitly considered. We introduce the use of dynamic linear model (DLM), a very general class of time series models, to model the shape and scale parameters changes across time. Posterior inference is performed through Markov Chain Monte Carlo (MCMC) methods. Simulations are carried out in order to analyze the performance of our proposed model. We also apply the proposed model to three real financial time series: the Brazilian Vale do Rio Doce, Petrobrás and BOVESPa index, all of which exhibit several extreme events.

1 Introduction

One approach of modelling extreme data is to consider the distribution of exceedances over a high threshold. Pickands (1975) shows that under certain conditions, this distribution can be approximated by a generalized pareto distribution (GPD). Let x be the excess over a high threshold, u. It is said that x follows a generalized Pareto distribution, $GPD(x;\xi,\sigma)$, if its cumulative distribution function is

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

where ξ and σ are shape and the scale parameters, respectively. It can be readily seen that the support of a GPD is $x\geq 0$ when $\xi\geq 0$, $0\leq x\leq \sigma/|\xi|$ when $\xi<0$ and that the data 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). A lot of the literature has shown how the threshold selection influences the parameter estimation (Coles and Powell, 1996; Coles and Tawn, 1996a; Coles and Tawn, 1996b; Smith, 1987). Recently, Behrens *et al* (2004) proposed a model to fit extreme data where the threshold u is also one of the model parameters. 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* (2009) used this idea to generalize Behrens *et al* (2004) and to use a mixture of Gammas below the threshold and a GPD above it.

In the last two cited works, however, all observations above the threshold are assumed to follow the GPD with same parameters, even when dealing with time series data. In this article, we extend their model, with immediate and important consequences for time series data, by allowing the GPD parameters to be time dependent. This is done by means of a dynamic linear model (DLM) evolution equation. Huerta and Sansó (2007) used a similar idea when modeling daily ozone levels by means of generalized extreme value (GEV) distributions with both time and space evolutions.

The paper is organized as follows. In Section 2, we present the model that considers all the observations in the estimation process. In Section 3 we discuss the prior specifications for model parameters and in particular, introduce the DLM in the GPD context. The details of implementing Markov Chain Monte Carlo (MCMC) algorithm based on the proposed model are presented in Section 4. We carry simulation study and discuss the performance of our model in Section 5. In section 6, we apply the proposed model to three real datasets from financial markets. We conclude with Seciton 7, discussing the implications of our model and pointing out possible future work.

2 Tail dynamics

Consider a time series y_t where, for a high threshold u, it is assume that $\{y_t|y_t < u\}$ has a nonparametric form, and $\{y_t|y_t > u\}$ follows a generalized Pareto distribution with parameters ξ_t and σ_t, $GPD(\xi_t, \sigma_t)$. In this model, the shape and scale parameters of GPD is changing with time, while the threshold and the parameters of the nontail are assumed to be unknown but fixed.

Given the lack of information below the threshold, non-parametric approximations seem a natural choice. Wiper *et al* (2001) showed that mixture of Gamma distributions, denoted MG_k, provide good approximations for distributions with positive support. Their density is

$$h(x \mid \theta, \mathbf{p}) = \sum_{j=1}^{k} p_j f_G(x \mid \mu_j, \alpha_j), \tag{1}$$

where $\theta = (\mu, \alpha)$, $\mu = (\mu_1, ..., \mu_k)$ and $\alpha = (\alpha_1, ..., \alpha_k)$, $\mathbf{p} = (p_1, ..., p_k)$ is the vector of weights and f_G is the Gamma density

$$f_G(x|\mu, \alpha) = \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-(\alpha/\mu)x), \text{ for } x > 0, \tag{2}$$

where μ is the mean of f_G is this parametrization. The means μ_j´s and shapes α_j´s may take any positive value and weights p_j´s are positive and sum to 1. The number k of components may be known, may be fixed according to a choice based on some optimality criteria or may be assumed to be an additional model parameter and estimated. Using the properties showed in Embrechts *et al* (1997), can be possible show that the finite mixture Gammas model belongs to domain of attraction to Generalized Extreme Value distribution (GEV), that is a necessary condition to application of GPD distribution to exceedance up a threshold.

The distribution function F_t of an observation y_t at time t is

$$F_t(y_t; \theta, \mathbf{p}, u, \xi_t, \sigma_t) = \begin{cases} H(y_t; \theta, \mathbf{p}) & \text{for } y_t < u \\ H(u; \theta, \mathbf{p}) + [1 - H(u; \theta, \mathbf{p})]G(y_t - u; \xi_t, \sigma_t) & \text{for } y_t \geq u \end{cases},$$

where $H(\cdot; \alpha, \beta)$ is the cumulative distribution function of mixture of Gammas distribution. Consequently, for a sample of observations $\mathbf{y} = (y_1, \cdots, y_T)$ and the parameter vector $\Theta = \big(\mathbf{p}, \mu, \alpha, u, \{\xi_t\}_{t=1}^T, \{\sigma_t\}_{t=1}^T, \{\theta_{\xi,t}\}_{t=0}^T, \{\theta_{\sigma,t}\}_{t=0}^T, V_\xi, V_\sigma, W_\xi, W_\sigma\big)$, the likelihood function can be written as

$$L(\mathbf{y}; \Theta) = \prod_{\{t:\ y_t < u\}} h(y_t; \theta, \mathbf{p}) \prod_{\{t:\ y_t \geq u\}} (1 - H(u; \theta, \mathbf{p}))g(y_t - u; \xi_t, \sigma_t) \tag{3}$$

where $h(\cdot; \alpha, \beta)$ is the probability density function of mixture of Gammas, $g(y_t - u; \xi_t, \sigma_t) = (1 + \xi_t(y_t - u)/\sigma_t)_+^{-(1+\xi_t)/\xi_t}/\sigma_t$ for $\xi_t \neq 0$ and $f(y_t - u; \xi_t, \sigma_t) = \exp(-(y_t - u)/\sigma_t)/\sigma_t$ for $\xi_t = 0$, and $(x)_+ = max(0, x)$.

2.1 Shape and scale dynamics

A first order DLM, as in West and Harrison (1997), is used here to model either or both shape and scale parameters changes across time. Consider a dymanic first model to the parameters:

$$\begin{array}{lcll} \xi_t & = & \theta_{\xi,t} + v_{\xi,t} & v_{\xi,t} \sim N(0, 1/V_\xi), \\ \theta_{\xi,t} & = & \theta_{\xi,t-1} + w_{\xi,t} & w_{\xi,t} \sim N(0, 1/W_\xi), \\ \sigma_t & = & \theta_{\sigma,t} + v_{\sigma,t} & v_{\sigma,t} \sim N(0, 1/V_\sigma), \\ \theta_{\sigma,t} & = & \theta_{\sigma,t-1} + w_{\sigma,t} & w_{\sigma,t} \sim N(0, 1/W_\sigma), \end{array} \tag{4}$$

where $\theta_{\xi,t}, \theta_{\sigma,t}, t = 0, \ldots, n$, V_ξ, W_ξ, V_σ and W_σ are the hyperparameters of model.

However, σ is restrict to positive values, and Smith (1984) show that maximum likelihood estimators to ξ doesn's exist to $\xi < -1$. The dynamic form proposed in (4) not have carefull with the restrictions in σ and ξ. Then, was proposed a reparametrization in the tail, with the dynamic in $\sigma_t = \exp(l\sigma_t)$ and $\xi_t = \exp(l\xi_t) - 1$. The dynamic model to the tail parameters are given by

$$\begin{array}{rcll} l\xi_t &=& \theta_{\xi,t} + v_{\xi,t} & v_{\xi,t} \sim N(0, 1/V_\xi), \\ \theta_{\xi,t} &=& \theta_{\xi,t-1} + w_{\xi,t} & w_{\xi,t} \sim N(0, 1/W_\xi), \\ l\sigma_t &=& \theta_{\sigma,t} + v_{\sigma,t} & v_{\sigma,t} \sim N(0, 1/V_\sigma), \\ \theta_{\sigma,t} &=& \theta_{\sigma,t-1} + w_{\sigma,t} & w_{\sigma,t} \sim N(0, 1/W_\sigma), \end{array} \tag{5}$$

where t runs from 1 to T and the initial information is $\theta_{0,\xi} \sim N(m_{0,\xi}, C_{0\xi})$ and $\theta_{0,\sigma} \sim N(m_{0,\sigma}, C_{0\sigma})$. The prior for V_ξ and V_σ is an gamma distribution $G(f_\xi, o_\xi)$ and $G(f_\sigma, o_\sigma)$. The prior for W_ξ and W_σ is $G(l_\xi, m_\xi)$ and $G(l_\sigma, m_\sigma)$. We can assume that u follows a truncated normal distribution with parameters (μ_u, σ_u^2), with μ_u set at some high data percentile and σ_u^2 large enough to represent a fairly noninformative prior. The different model components θ, **p** and β are assumed to be independent a priori. The prior for θ follows from Wiper *et al* (2001) and is given by

$$\begin{aligned} p(\mu_1, \ldots, \mu_k) &= K \prod_{i=1}^{k} f_{IG}(\mu_i \mid a_i/b_i, b_i) I(\mu_1 < \mu_2 < \ldots < \mu_k), \\ p(\alpha_1, \ldots, \alpha_k) &= \prod_{i=1}^{k} f_{IG}(\alpha_i \mid c_i/d_i, d_i), \end{aligned}$$

where $K^{-1} = \int \prod_{i=1}^{k} p(\mu_i) d(\mu_1, \ldots, \mu_k)$ and f_{IG} is the inverse Gamma density . The prior distribution for the weights **p** is assumed to be a Dirichlet distribution $D_k(\gamma_1, \ldots, \gamma_k)$, with density proportional to $\prod_{i=1}^{k} p_i^{w_i}$.

3 Simulation-based posterior inference

After introducing DLM for the time-varying shape parameter, our parameter vector is now $\Theta = \big(\mathbf{p}, \mu, \alpha, u, \{l\xi_t\}_{t=1}^T, \{l\sigma_t\}_{t=1}^T, \{\theta_{\xi,t}\}_{t=0}^T, \{\theta_{\sigma,t}\}_{t=0}^T, V_\xi, V_\sigma, W_\xi, W_\sigma\big)$. 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

$$
\begin{aligned}
\pi(\Theta|\mathbf{x}) \quad \propto \quad & \prod_{t:x_t<u} \left[\sum_{j=1}^{k} p_j f_G(x_t|\mu_j,\eta_j) \right] \prod_{x_t \geq u} \left[\left(1 - \sum_{j=1}^{k} p_j F_G(u|\mu_j,\eta_j) \right) g(x_t|\xi_t,\sigma_t,u) \right] \\
\times \quad & \prod_{j=1}^{k} \left[\eta_j^{a_j-1} e^{-b_j\eta_j} \beta_j^{-(c_j+1)} e^{-d_j/\mu_j} \right] \exp\left(-\frac{(u-\mu_u)^2}{2\sigma_u^2} \right) \\
\times \quad & V_\xi^{T/2+f_\xi-1} \exp\left(-\frac{V_\xi}{2} \sum_{t=1}^{T} (l\xi_t - \theta_{\xi,t})^2 - o_\xi V_\xi \right) \exp\left(-\frac{1}{2C_{\xi,0}} (\theta_{\xi,0} - m_{\xi,0})^2 \right) \\
\times \quad & W_\xi^{T/2+l_\xi-1} \exp\left(-\frac{W_\xi}{2} \sum_{t=1}^{T} (\theta_{\xi,t} - \theta_{\xi,t-1})^2 - m_\xi W_\xi \right) \\
\times \quad & V_\sigma^{T/2+f_\sigma-1} \exp\left(-\frac{V_\sigma}{2} \sum_{t=1}^{T} (l\sigma_t - \theta_{\sigma,t})^2 - o_\sigma V_\sigma \right) \exp\left(-\frac{1}{2C_{\sigma,0}} (\theta_{\sigma,0} - m_{\sigma,0})^2 \right) \\
\times \quad & W_\sigma^{T/2+l_\sigma-1} \exp\left(-\frac{W_\sigma}{2} \sum_{t=1}^{T} (\theta_{\sigma,t} - \theta_{\sigma,t-1})^2 - m_\sigma W_\sigma \right). \qquad (6)
\end{aligned}
$$

as expected, posterior inference is analytically infeasible, so modern Bayesian inference is performed through a customized Markov Chain Monte Carlo (MCMC) algorithm (see Gamerman and Lopes, 2006 for more details).