A Bayesian Lasso for Tail Index Regression

Myung Won Lee [1] Miguel de Carvalho [1] Daniel Paulin [1]
Soraia Pereira[2] Ricardo Trigo [2] Carlos Da Camara [2]

[1]School of Mathematics, University of Edinburgh, United Kingdom [2]Faculdade de Ciências, Universidade de Lisboa, Portugal

Motivation

This study extends the Tail Index Regression [4] to a generalised additive framework with Bayesian regularisation that characterises the extreme behaviour of a response. Here, the response follows a conditional Pareto-type tail specification,

$$1 - F(y \mid \boldsymbol{x}) = \frac{\mathscr{L}(y \mid \boldsymbol{x})}{y^{\alpha(\boldsymbol{x})}}.$$

- $\alpha(\boldsymbol{x}) > 0$ is a covariate-adjusted tail index that controls the rate of decay of the tail
- $\mathscr{L}(y \mid \boldsymbol{x})$ is a family of covariate-adjusted slowly varying function, that is $\frac{\mathscr{L}(yt \mid \boldsymbol{x})}{\mathscr{L}(y \mid \boldsymbol{x})} \to 1$, as $y \to \infty$ for all $t > 0$.

With sufficiently large threshold, u, the conditional probability density of Y given $\boldsymbol{X}$ and $Y > u$ is,

$$f(y \mid \boldsymbol{x}, u) = \alpha(\boldsymbol{x})(y/u)^{\alpha(\boldsymbol{x})} y^{-1} + o(1).$$

Our approach,

1. Estimates the covariate-adjusted tail index and model with a flexible generalised additive framework.
2. Entails a regularisation on a fully semiparameteric framework by concentrating on learning about the regression and splines coefficients.

Regression for Extreme Value

Generalised Additive Framework for Tail Index Regression

$$\alpha(\boldsymbol{x}) = \exp\left\{\sum_{j=1}^{p} g_j(x_j)\right\}. \tag{1}$$

Each g_j is a smooth function for each covariate parameterised by a continuous thin plate splines basis function over a parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^q$.

$$g_j(x_j) = \beta_{j0} + x_j\beta_j + \sum_{\psi=1}^{\Psi} \gamma_{j\psi}|x_j - k_{j\psi}|^3. \tag{2}$$

- Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\intercal$ as the vector of linear component parameters and $\beta_0 = \sum_{j=1}^{p} \beta_{j0}$ as the overall intercept term,
- $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)$ with $\boldsymbol{\gamma}_j = (\gamma_{j1}, \ldots, \gamma_{j\psi})^\intercal$ as a matrix containing the nonlinear component parameters,
- $k_{j1} < \cdots < k_{j\Psi}$ represent the internal knots that are placed at every $1/\Psi$ quantile for each covariate, $\boldsymbol{x}_j$.

To regularise, we resort to shrinkage priors such as Bayesian Lasso and Bayesian group Lasso to impose simultaneous estimation and regularisation on the parameters of interest.

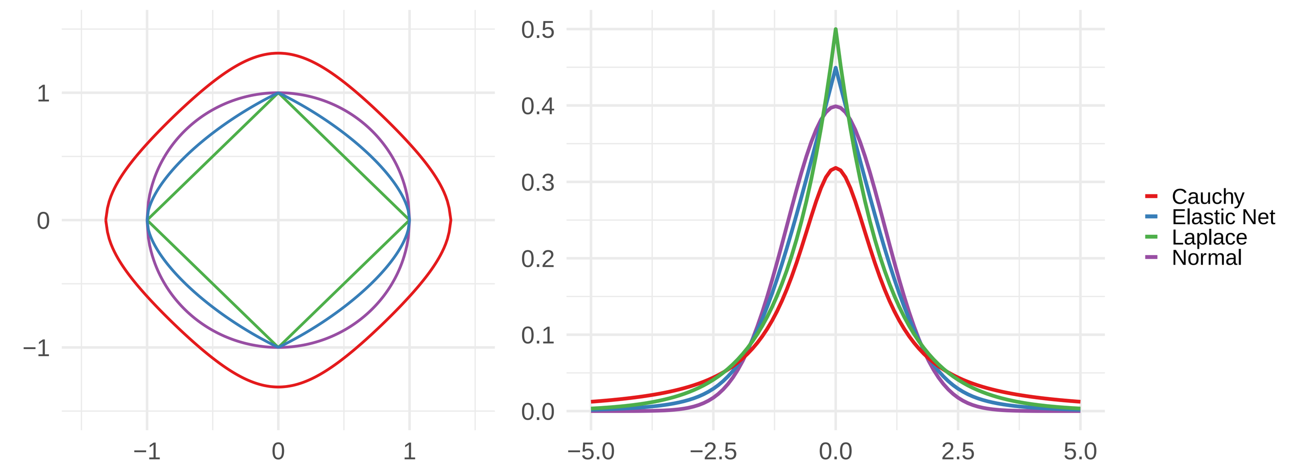

Fig 1. *Comparison of the geometry of a unit ball (left) induced by Cauchy (red), Elastic-Net (blue), Lasso (green) and Ridge (purple) penalties and their corresponding shrinkage priors (right).*

A Bayesian Lasso for Tail Index Regression

$$y_i \mid \boldsymbol{x}_i \overset{\text{i.i.d.}}{\sim} f(y \mid \boldsymbol{x}, u), \qquad \text{(Likelihood)}$$

$$\begin{aligned} \beta_0 &\sim N(0, \sigma^2); & \text{(Priors)} \\ \beta_j \mid \lambda_\beta &\overset{\text{ind}}{\sim} \text{Laplace}(\lambda_\beta), \\ \lambda_\beta &\sim \text{Gamma}(a_\beta, b_\beta); \\ \boldsymbol{\gamma}_j \mid \sigma^2, \tau_j^2 &\overset{\text{ind}}{\sim} N_\Psi(\boldsymbol{0}, \mathbf{I}_\Psi/\tau_j^2), \\ \tau_j^2 \mid \lambda_\gamma &\overset{\text{ind}}{\sim} \text{Gamma}\left\{(\Psi+1)/2, \lambda_\gamma^2/2\right\}, \\ \lambda_\gamma &\sim \text{Gamma}(a_\gamma, b_\gamma); \end{aligned}$$

Monte Carlo Simulation Study

Consider a conditional Pareto distribution scenario:
$\mathscr{L}(y \mid \boldsymbol{x}) \propto 1$, with
$\boldsymbol{\beta} = (-0.5, 0, -0.5, -0.5, 0, 0)^\intercal$ and $\boldsymbol{\gamma} = (\boldsymbol{0}, t\boldsymbol{1}, t\boldsymbol{1}, \boldsymbol{0}, \boldsymbol{0})$, where $t = -25$.

- Retrieved the 95% quantile from $15\,000$ random samples and repeated the study 250 times with $\Psi = 20$ using `stan`.
- An uninformative diffuse gamma prior, $\Gamma(1, 10^{-3})$ was used, and in terms of MCMC, we took $2\,000$ burn-in iterations and collected $3\,000$ posterior samples.

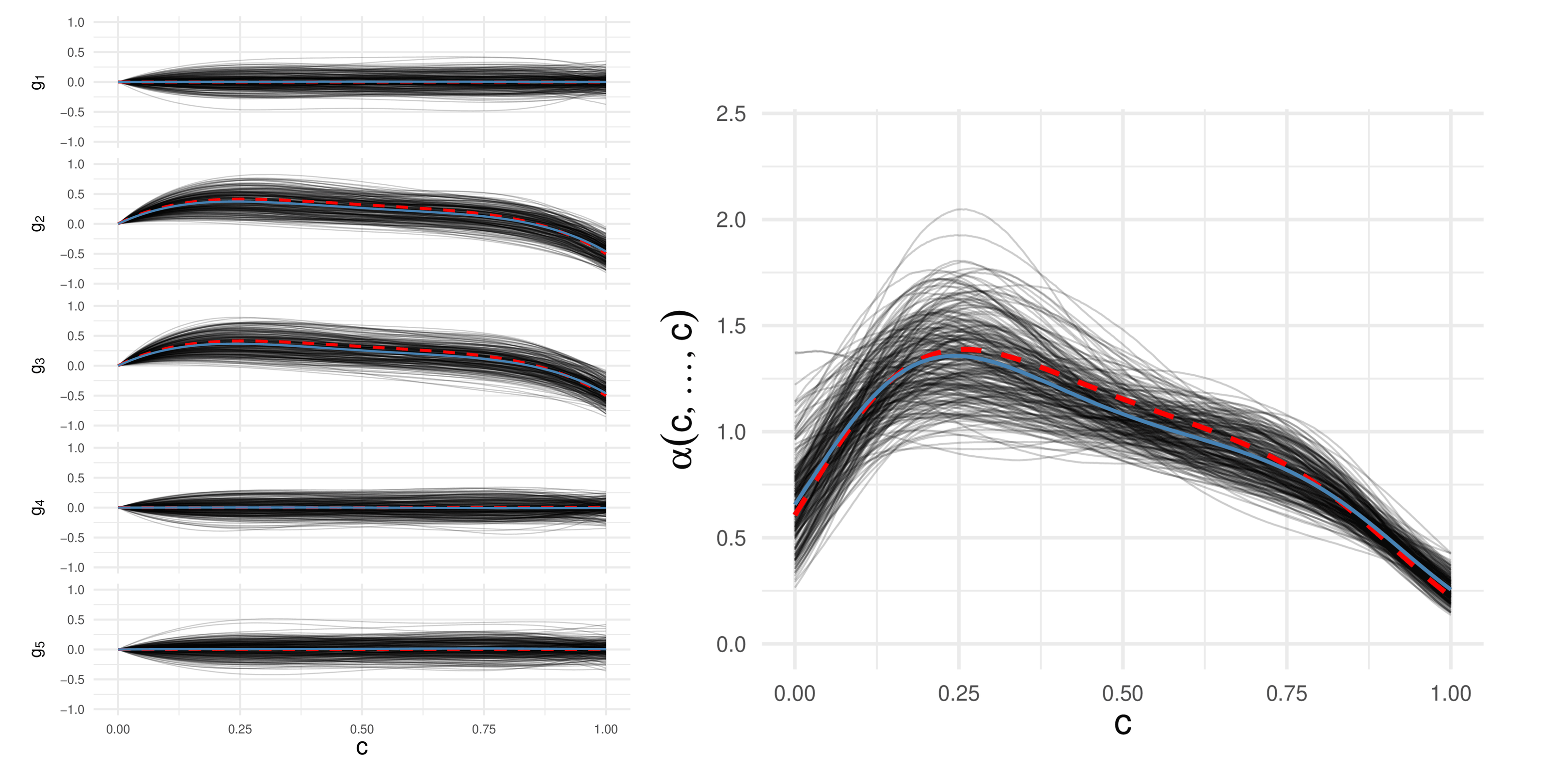

Fig 2. *Monte Carlo analysis: Posterior median estimates and the Monte Carlo mean (blue solid) of smooth functions (left) and covariate-adjusted tail index (right) against the true (red dash) where $c \in [0, 1]$.*

Extreme Wildfires in Portugal

We illustrate the proposed method on the data from *Instituto Dom Luiz* consisting daily burnt area and the Fire Weather Index in Portugal between 1980 to 2019 [3].

- Filtered 366 (top 97.5%) observations out of 14 609 and standardised the covariates.
- Daily Severity Rating (DSR) and Fine Fuel Moisture Code (FFMC) are the significant drivers of the extreme wildfires by putting a stronger mass at the tail.
- From Table 1, model with $\Psi = 30$ shows lowest expected log pointwise predictive density (ELPD) score and indicates the nonlinear relationship in the exceedances.

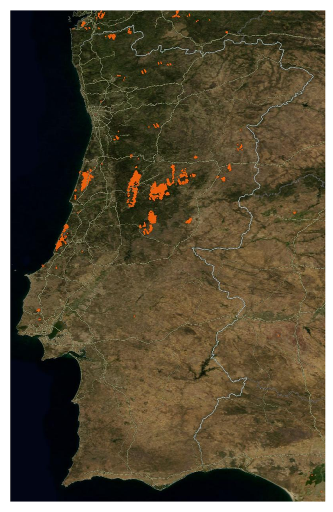

Fig 3. *Burnt area (orange) caused by extreme wildfires in 15th Oct 2017 in Portugal.*

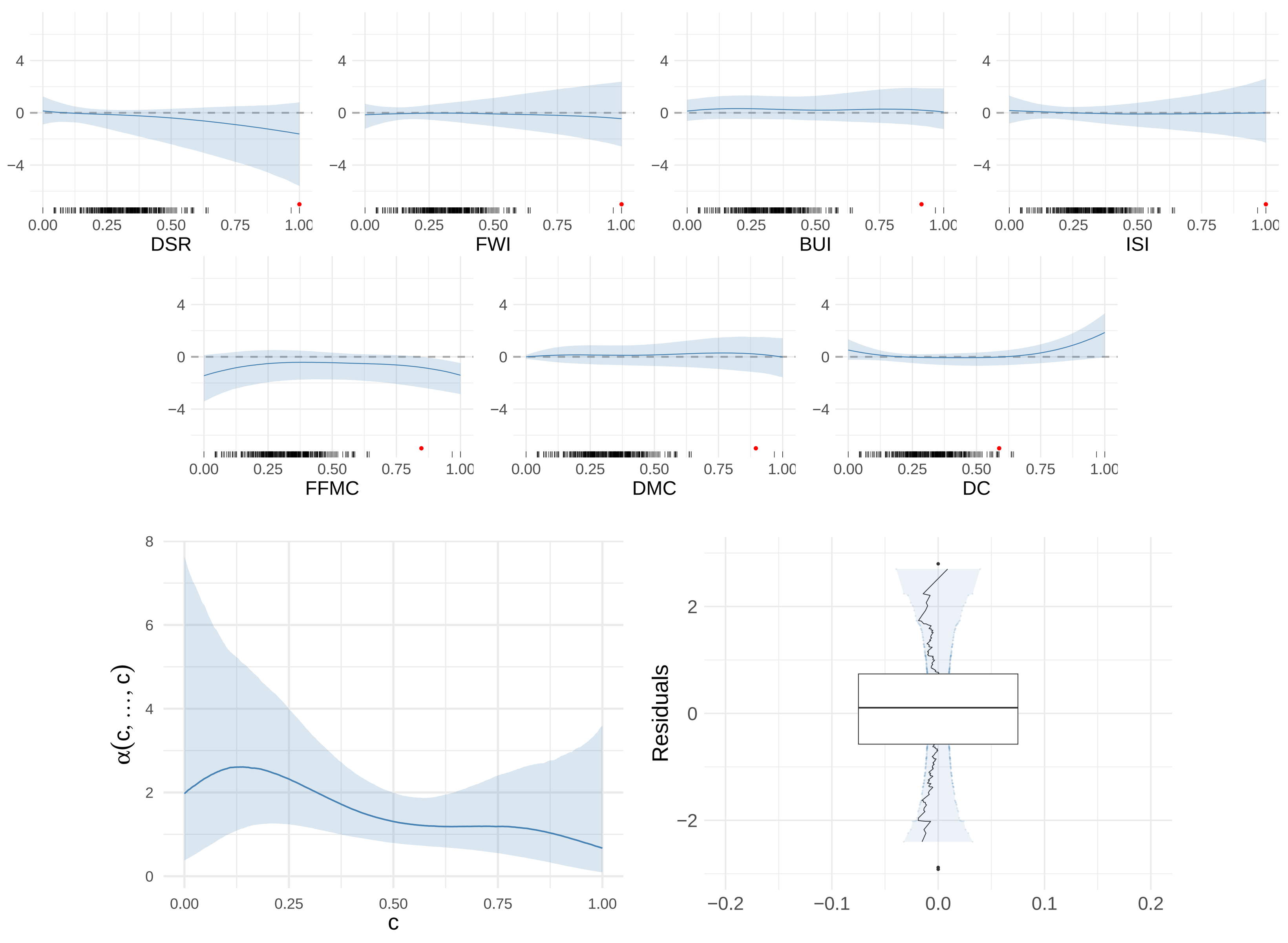

Fig 4. *Credible intervals for each smooth function (top) and covariate-adjusted tail index (bottom left) at 97.5% quantile and posterior median estimates (solid blue). QQ-boxplot showing quantile residuals against theoretical standard normal quantiles with box-and-whisker plot (bottom right).*

Quantile	g_j	Ψ	Information Criteria ELPD	ΔELPD	p
0.975	Full	10	$-3379.4(31.1)$	$-0.3(0.3)$	$8.7(1.3)$
	Full	20	$-3379.4(31.1)$	$-0.3(0.3)$	$8.7(1.3)$
	Full	30	$\mathbf{-3379.1}(31.1)$	$0.0(0.0)$	$8.7(1.3)$
	Linear	-	$-3385.7(30.8)$	$-6.5(2.9)$	$3.8(0.4)$

Table 1. *Model comparison between different versions of proposed model.*

References

[1] M. de Carvalho, S. Pereira, P. Pereira, and P. de Zea Bermudez. An extreme value Bayesian lasso for the conditional left and right tails. *Journal of Agricultural, Biological and Environmental Statistics*, pages 1–18, 2021.

[2] T. Park and G. Casella. The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

[3] M. Turco, S. Jerez, S. Augusto, P. Tarín-Carrasco, N. Ratola, P. Jiménez-Guerrero, and R. M. Trigo. Climate drivers of the 2017 devastating fires in portugal. *Scientific Reports*, 9(1):13886, 2019.

[4] H. Wang and C.-L. Tsai. Tail index regression. *Journal of the American Statistical Association*, 104(487):1233–1240, 2009.

tiny.cc/mwl johnny.myungwon.lee@ed.ac.uk