A BAYESIAN APPROACH FOR ESTIMATING EXTREME QUANTILES UNDER A SEMIPARAMETRIC MIXTURE MODEL

BY

STEFANO CABRAS AND MARÍA EUGENIA CASTELLANOS

ABSTRACT

In this paper we propose an additive mixture model, where one component is the Generalized Pareto distribution (GPD) that allows us to estimate extreme quantiles. GPD plays an important role in modeling extreme quantiles for the wide class of distributions belonging to the maximum domain of attraction of an extreme value model. One of the main difficulty with this modeling approach is the choice of the threshold u, such that all observations greater than u enter into the likelihood function of the GPD model. Difficulties are due to the fact that GPD parameter estimators are sensible to the choice of u. In this work we estimate u, and other parameters, using suitable priors in a Bayesian approach. In particular, we propose to model all data, extremes and non-extremes, using a semiparametric model for data below u, and the GPD for the exceedances over u. In contrast to the usual estimation techniques for u, in this setup we account for uncertainty on all GPD parameters, including u, via their posterior distributions. A Monte Carlo study shows that posterior credible intervals also have frequentist coverages. We further illustrate the advantages of our approach on two applications from insurance.

KEYWORDS

Extreme values; Generalized Pareto distribution; Jeffreys' prior; Lindsey method; Mixture distribution; Semiparametric density estimation.

1. INTRODUCTION

In the past two decades there has been an increasing interest in statistical modeling for estimating the probability of rare and extreme events. These models are of interest in numerous disciplines such as environmental sciences, engineering, finance and insurance, among others (see for instance Coles (2001) and Smith (2003)). In this paper, we mainly focus on insurance applications, see for instance Mikosh (2003), Chavez-Demoulin and V. Embrechts (2009), Donnelly and Embrechts (2010) and references therein. The Generalized Pareto

Astin Bulletin 41(1), 87-106. doi: 10.2143/AST.41.1.2084387

distribution is the most used statistical model to fit extreme exceedances, $X - u > 0$, over a high threshold, u. Pickands (1975) showed that if X is a continuous random variable with cumulative distribution function (cdf) $F(x)$, and $F(x)$ belongs to the maximum domain of attraction of some Extreme Value distribution, then $P(X \leq x + u \mid X > u)$ can be approximated, for $u \to \infty$, by the GPD, with cdf $G(x \mid \xi, \sigma, u)$, with shape ξ, scale $\sigma > 0$ and density (df)

$$g(x \mid \xi, \sigma, u) = \begin{cases} \sigma^{-1}\left(1 + \dfrac{\xi(x-u)}{\sigma}\right)^{-(1+\xi)/\xi}, & \left(1 + \dfrac{\xi(x-u)}{\sigma}\right) > 0 \\ \sigma^{-1}\exp(-(x-u)/\sigma), & \xi = 0 \end{cases}. \quad (1.1)$$

Note that the support is $X - u > 0$ for $\xi \geq 0$ and $0 < X - u < -\sigma/\xi$ for $\xi < 0$, meaning that $X - u$ is upper bounded for negative values of the shape parameter ξ, while $\xi > 0$ corresponds to a heavy tail distribution. A property of the GPD is that if $Y = X - u \sim G(\cdot \mid \xi, \sigma, u)$ and $u' > 0$ then $Y - u' \mid Y > u' \sim G(\cdot \mid \xi, \sigma', u')$, where $\sigma' = \sigma + \xi(u' - u)$. This produces a lack of identification addressed later in the paper.

In practical data analyses two steps are performed: first a threshold u is selected and then a GPD is fitted to the exceedances over u. Once u is selected there are different methods to estimate ξ and σ, for example maximum likelihood estimators (MLE) have been considered by many authors, including Davison (1984), Smith (1984) and Grimshaw (1993). Alternative methods have been considered due to likelihood irregularities (Smith, 1984), among these we have the probability-weighted moment and the elemental percentile method (Hosking and Wallis, 1987; Castillo and Hadi, 1997). In de Zea Bermudez and Turkman (2003) and Castellanos and Cabras (2007) Bayesian methods have been explored using default priors. In this paper, we are mainly interested in estimating extreme quantiles of X, $q_p : \Pr(X \leq q_p) = p$, $p \approx 1$. Such quantiles may be used to estimate the Value At Risk and to compute the premium based on the amount of the damage, q_p, that an insurance company may want to assure and want to face with probability $1 - p$.

Some papers show how estimates of ξ, σ and q_p depend significantly on the choice of u, see for example McNeil (1997) and Embrechts et al. (1997). The choice of u is a difficult task unless expert elicitation is available. In order to hold the conditions of the Pickand's Theorem, large values of u are usually selected, increasing the uncertainty on ξ and σ because few extreme observations are involved. One of the most popular tool for estimating u is the Mean Excess plot, see for instance Coles (2001), in which the empirical mean excess function is studied for different values of the threshold, looking at the minimum one that makes this function almost linear in u. The method in Reiss and Thomas (2007) considers the order statistics for a sample $x_{(1)} < x_{(2)} < \ldots, x_{(k)} < \ldots, x_{(n)}$ of size n and then set $u = u^* = x_{\hat{k}}$ where

$$\widehat{k} = \arg \min_{1 \leq k \leq n} \frac{1}{k} \sum_{i \leq k} i^{\eta} \left| \widehat{\xi}_i - median\left(\widehat{\xi}_1, \ldots, \widehat{\xi}_k\right) \right|, \quad 0 \leq \eta \leq 1/2 \qquad (1.2)$$

https://doi.org/10.2143/AST.41.1.2084387 Published online by Cambridge University Press

and $\hat{\xi}_i$ is the MLE of ξ based on the i upper order statistics. This method, that depends on the choice of η (Neves and Fraga Alves, 2004), provides a point estimation of u, but not an evaluation of the corresponding uncertainty. On the contrary, such uncertainty is considered with the Bayesian approach proposed in this paper. In fact, under a Bayesian perspective all parameters are regarded as random variables and their posterior distribution accounts for the uncertainty conditionally on the observed data.

In de Zea Bermudez et al. (2001) selection of u is based on the number of upper order statistics within a Bayesian predictive approach.

On the contrary, the majority of the methods used to estimate u, including the one here proposed, are based on a mixture model for all data where the GPD is a component. Differences among such methods lay on more or less restrictive assumptions about the sampling model for non-extreme data, referred in the sequel as the central model (Frigessi et al. (2002), Behrens et al. (2004), Cabras and Morales (2007), Tancredi et al. (2006)). In particular, Frigessi et al. (2002) introduce a dynamically weighted mixture model in which the central model is the Weibull distribution and MLE is used to estimate the unknown parameters. In Cabras and Morales (2007) the central model is the Normal distribution whose outliers are considered extremes and u is estimated as the smallest outlier. In Behrens et al. (2004) the central model is the truncated gamma model with a subjective prior on unknown parameters. In Tancredi et al. (2006) the central model is a mixture of uniform distributions and a Bayesian approach is implemented using vague priors.

Our proposal consists also in a mixture model, but as we want to make inference on extreme quantiles of X, say $q_{0.999}$ and $q_{0.9999}$, parameters of interest are those in (1.1), while parameters of the central model are regarded as nuisance parameters. In order to eliminate such nuisance parameters we use pseudo-likelihoods in an objective bayesian framework as recently proposed in Ventura et al. (2009). In particular we distinguish two situations:

i) the usual one where a parametric central model can be elicited. In this case we employ the integrated likelihood to eliminate the nuisance parameters;

ii) otherwise, we propose to use the profile likelihood estimating, semiparametrically, the central model conditionally on u.

The case *i*) is mainly used for comparison purposes with respect to the second situation which is more challenging and where we propose a more general approach. In *i*) we will focus on Normal, Lognormal and Weibull distributions by using their corresponding default priors. For both situations we assume a uniform prior for u and a default prior for GPD parameters.

The rest of the paper is organized as follows: Section 2 illustrates the model approach for situations *i*) and *ii*); Section 3 discusses priors and Markov Chain Monte Carlo (MCMC) approximation of the posterior distribution; Section 4 includes two applications to insurance real data sets; Section 5 illustrates, trough simulations, coverage and length of the posterior credible intervals of q_p. Further remarks and conclusions are contained in Section 6.

2. Model

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common probability density function $f(\cdot|\boldsymbol{\theta})$ given by:

$$f(x|\boldsymbol{\theta}) = h(x|\gamma)\mathbf{1}_{\{x\leq u\}} + [1 - H(u|\gamma)]\, g(x|\xi,\sigma,u)\mathbf{1}_{\{x>u\}}, \quad (2.1)$$

where $\boldsymbol{\theta} = (\gamma, \xi, \sigma, u)$. Our model consists in an additive mixture model in which data below u are modeled with a truncated distribution with cdf $H(\cdot|\gamma)$ governed by parameters γ, while observations above u come from the GPD model in (1.1). In (2.1) $H(\cdot|\gamma)$ and $h(\cdot|\gamma)$ denote cdf and df, respectively. Further discussion of this model can be found in Behrens et al. (2004) where H is assumed to be the Gamma distribution.

If elicitation of a parametric class of models is available for H, then it can be estimated within this class. In this paper, and only for seek of comparison of our proposal with other methods, we consider the following three parametric models: Normal, Lognormal and Weibull. Any other parametric model can be used if it fits well the central part of the data. Our proposal is, however, to avoid such parametric model assumptions by considering a semiparametric estimation of h conditionally on u as described in Section 2.2.

In any case, parametric and semiparametric, the use of H, in (2.1), partially solves the identification problem commented in Section 1, because u acts as a separation point between the central and the extreme model.

For model (2.1) q_p is given by

$$q_p = \begin{cases} H^{-1}(p) & q_p \leq u \\ u + \dfrac{\left[(1-\tilde{p})^{-\xi} - 1\right]\sigma}{\xi} & q_p > u \end{cases}, \quad (2.2)$$

where $\tilde{p} = (p - H(u|\gamma)) / (1 - H(u|\gamma))$.

2.1. Parametric central model

For a random sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ of size n drawn from model (2.1), the full likelihood function for all parameters $\boldsymbol{\theta}$ with $\xi \neq 0$ is

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{\{i,\, x_i \leq u\}} h(x_i|\gamma) \prod_{\{i,\, x_i > u\}} (1 - H(u|\gamma)) \frac{1}{\sigma}\left(1 + \xi\frac{(x_i - u)}{\sigma}\right)^{-(1+\xi)/\xi} \mathbf{1}_A, \quad (2.3)$$

where $\mathbf{1}_A$ denotes the indicator function of $A \equiv \left\{\frac{\xi}{\sigma} > -\frac{1}{x_{(n)} - u}\right\}$.

For $\xi = 0$, the full likelihood takes the form

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{\{i,\, x_i \leq u\}} h(x_i|\gamma) \prod_{\{i,\, x_i > u\}} (1 - H(u|\gamma)) \frac{1}{\sigma} \exp(-(x_i - u)/\sigma)\mathbf{1}_{\{\sigma>0\}}.$$

When h is the Normal (Lognormal) density then $\gamma = (\mu, \tau^2)$ represents mean (location) and variance (squared scale), otherwise $\gamma = (\alpha, \beta)$ represents shape and scale parameters of the Weibull distribution.

2.2. Semiparametric central model

When we cannot assume a parametric model as the central model, we estimate its density h, conditionally on u, using the Lindsey method (Lindsey (1974a), Lindsey (1974b)) also discussed in Section 2 of Efron and Tibshirani (1996). For a given threshold u, the Lindsey's method provides a degree d polynomial approximation of the truncated df $hT(\cdot) = h(\cdot)\mathbf{1}_{\{x \leq u\}}/H(u)$, through the n_u^- non extreme data points $\{x_i : x_i \leq u\}$.

In this case we may rewrite function (2.3) as a profile likelihood function of u, ξ and σ only. For $\xi \neq 0$ it takes the form

$$L_p(u, \xi, \sigma \mid \boldsymbol{x}) = \prod_{\{i,\, x_i \leq u\}} \widehat{H}(u)\hat{h}_u(x_i) \prod_{\{i,\, x_i > u\}} (1 - \widehat{H}(u))\,\frac{1}{\sigma}\left(1 + \xi\frac{(x_i - u)}{\sigma}\right)^{-(1+\xi)/\xi} \mathbf{1}_A, \tag{2.4}$$

while for $\xi = 0$ it is

$$L_p(u, \xi, \sigma \mid \boldsymbol{x}) = \prod_{\{i,\, x_i \leq u\}} \widehat{H}(u)\hat{h}_u(x_i) \prod_{\{i,\, x_i > u\}} (1 - \widehat{H}(u))\,\frac{1}{\sigma}\exp(-(x_i - u)/\sigma)\,\mathbf{1}_{\{\sigma > 0\}},$$

where $\widehat{H}(u) = n_u^-/n$ is the proportion of observations below u. Note that it is necessary to multiply $\hat{h}_u$ by $\widehat{H}(u)$ as $\hat{h}_u$ is the density estimator of hT. In large samples we expect L_p to have the same properties of the full likelihood L in (2.3) and posterior distributions, based on L_p, can be interpreted as actual posterior distributions (Ventura et al., 2009).

2.2.1. *Danish fire loss data*

We consider Danish fire loss data to compare the assumptions of Weibull, Lognormal and the semiparametric density estimation. Danish data consist of 2157 insurance losses from 1980 to 1990 caused by industrial fires. Losses include damage to buildings, furniture and personal property as well as loss of profits. The unit is 1 million DKK and all data have been adjusted to the 1985 values.

McNeil (1997) analyses this data set using the GPD model, the author points out the difficulty of the selection of an appropriate threshold and several thresholds have been considered along with the corresponding MLE of GPD parameters.

These data are also analyzed in Frigessi et al. (2002) using a dynamic mixture model with the Weibull distribution as central model. As in Frigessi et al. (2002), we also translate data by -1 in order to have the minimum value at 0,

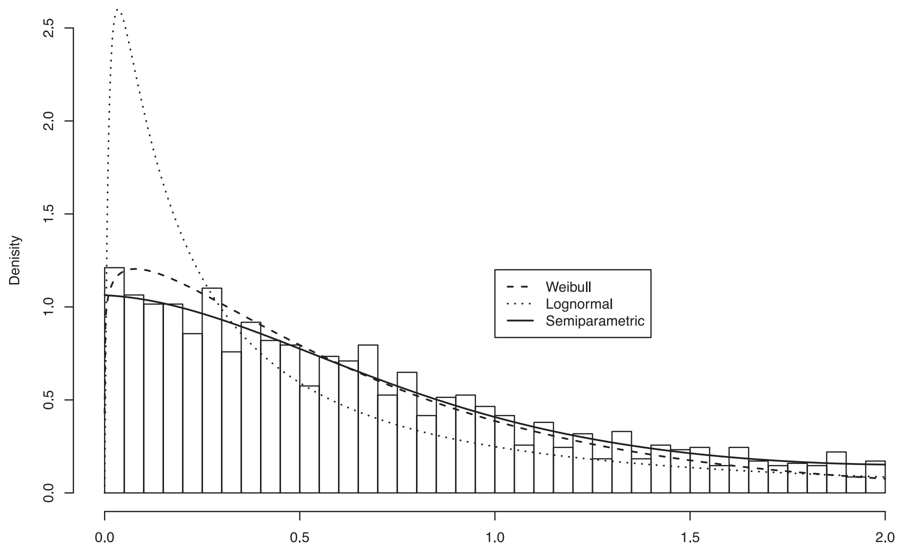

FIGURE 1: Density estimation for data below threshold $u = 2$ using Weibull model (dashed line), Lognormal (dotted line) and the semiparametric model (continuous line).

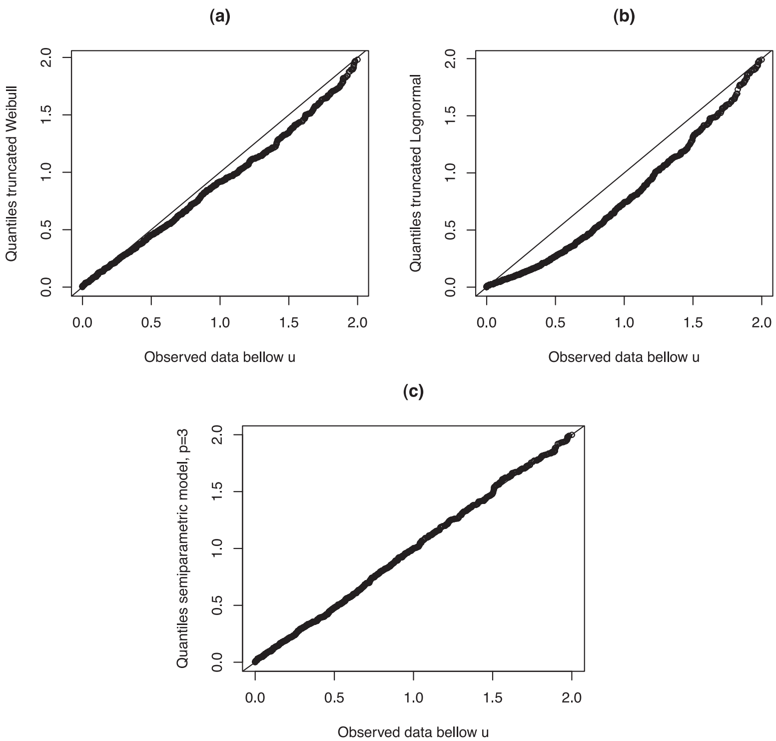

FIGURE 2: QQ plots comparing the observed data below threshold $u = 2$ and the theoretical quantiles for the Weibull model (a), Lognormal model (b) and the semiparametric model (c).

https://doi.org/10.2143/AST.41.1.2084387 Published online by Cambridge University Press

and we compare our results with those obtained in that article. Results reported in Table 12 and equation 4 from Frigessi et al. (2002) show that observations between 1.5 and 2 have probability between 0.95 and 0.98 of coming from the GPD, while the observation of 1 million DKK has probability 0.33. Although Frigessi et al. (2002) do not consider a point estimation for u, thresholds between 1.5 and 2 are appropriate according to their approach. We analyze the goodness-of-fit of the truncated Weibull and Lognormal for all observations below $u = 2$, using MLE of γ. We can see in Figures 1 and 2 a poor fit of both parametric models that affects also the fit of the complete mixture model (2.1). Instead, we consider the semiparametric model in (2.4) with $d = 3$. The fit of the semiparametric density estimator is considerably better than Weibull or Lognormal ones, as shown in Figures 1 and 2. Using $u = 1.5$ the comparison is even more favorable to the semiparametric density estimator.

3. Prior distribution and posterior inference

We use a Bayesian approach based on L or L_p to make inference on stochastic functionals of model (2.1) with particular attention to q_p. We assume that we have not information about the behavior of extreme and non-extreme data and thus we use minimum informative priors for all parameters. In particular, we assume for parametric models

$$\pi(\gamma, \xi, \sigma, u) = \pi(\gamma)\,\pi(\xi, \sigma \,|\, u)\,\pi(u),$$

and the reduced prior

$$\pi(\xi, \sigma, u) = \pi(\xi, \sigma \,|\, u)\,\pi(u),$$

for the semiparametric model with the profile-likelihood L_p.

3.1. Prior for GPD parameters

For parameters ξ and σ given u, we consider the prior distribution proposed in Castellanos and Cabras (2007),

$$\pi(\xi, \sigma | u) \propto \sigma^{-1}(1+\xi)^{-1}(1+2\xi)^{-1/2}, \; \xi > -1/2, \; \sigma > 0,$$

derived by the application of the Jeffreys' rule for the regular case of $\xi > -1/2$ and that leads to a proper posterior distribution (Castellanos and Cabras, 2007, Theorem 1).

As prior for u, we use $\pi(u) \propto \mathbf{1}_{[a,b]}$, where a and b are suitable limits in order to obtain a proper posterior distribution over the parameters when using a mixture model, see Roeder and Wasserman (1997). In particular, to satisfy such condition it must be $a \geq x_{(m+1)}$, where m is equal to d or the dimension of γ.

To obtain a proper posterior using the default prior on GPD parameters, it suffices $b \leq x_{(n-2)}$. We recommend that, if expert elicitation is available, other priors on the range $[x_{(m+1)}, x_{(n-2)}]$ should be used. In this paper we always set $a = x_{(m+1)}$ and $b = x_{(n-2)}$ because we do not have prior information. Therefore, we are assuming, a priori, that we have observed at least two extreme observations, say $x_{(n-1)}$ and $x_{(n)}$, and also m non-extreme values. It is worth to mention here that uniform prior is compatible with the usual default prior on location parameters and it has been also used in Tancredi et al. (2006).

3.2. Priors for parameters of the central model

In this paper we consider three parametric families of models: Normal, Lognormal and Weibull along with their corresponding usual default priors. In particular we have:

Normal (Lognormal) model: $\gamma = (\mu, \tau^2)$ representing mean (location) and variance (squared scale). The usual default prior for the Normal and Lognormal coincides (see Padgett and Johnson (1983) for the Lognormal case) in both models with $\pi(\gamma) \propto \tau^{-2}$;

Weibull model: parametrizing the Weibull model as follows $H(x \mid a, \beta) = 1 - \exp(-(x/\beta)^a)$, the reference prior for $\gamma = (\beta, a)$ is $\pi(\gamma) \propto \beta^{-1}$ (Sun and Berger 1994).

However, the main advantage of our semiparametric approach is that we do not need to specify such priors that can affect the final inference on q_p. Of course, we still have to use the priors for the GPD, $\pi(u)$ and $\pi(\xi, \sigma)$, as specified above.

3.3. Posterior distribution

The posterior distribution for all parameters in (2.1) is

$$\pi(\gamma, u, \xi, \sigma \mid \boldsymbol{x}) \propto L(\gamma, u, \xi, \sigma \mid \boldsymbol{x})\, \pi(\gamma)\, \pi(\xi, \sigma \mid u)\, \pi(u),$$

for the parametric case. When we cannot assume a parametric model for non extremes, then model h constitutes, itself, an infinite dimensional nuisance parameter and we eliminate it through an estimator $\hat{h}_u$ conditionally on u. Therefore, our posterior is the following pseudo-posterior

$$\pi(u, \xi, \sigma \mid \boldsymbol{x}) \propto L_p(u, \xi, \sigma \mid \boldsymbol{x})\, \pi(\xi, \sigma \mid u)\, \pi(u).$$

Both posteriors are approximated using MCMC, specifically a Metropolis-Hastings within Gibbs sampling. We update each parameter individually, using as proposals a Normal distribution for μ and truncated Normal distributions for all other parameters. The mean of each proposal distribution is the last state of the chain, while standard deviations are fixed in order to provide a good mix of the chain. Conditional distributions of each parameter given the rest, needed for the Gibbs sampling, are provided in the Appendix.

All computations have been implemented under the open source software R (R Development Core Team (2009)) using some functions from the library `POT` (see Appendix).

4. Applications

In this section we apply the proposed semiparametric mixture to two datasets from insurance. We compare our approach to that based on parametric assumption on the central model and with the usual one based on MLE for u fixed.

4.1. Danish fire loss data (cont.)

This data has been used in Subsection 2.2.1 to show that Weibull and Lognormal densities do not fit non-extreme data, for this reason we relax parametric assumptions by using the semiparametric density estimator introduced in Subsection 2.2. Summaries of the posterior distribution for parameters u, ξ and σ in model (2.1) appear in Table 1. The 95% credible interval for the shape parameter reflects a heavy tail behavior. The median of u is 5.29 millions of DKK while its mean is 7.48 and it is interesting to notice that the marginal posterior distribution for u is multimodal, as we can see in Figure 3. The main modes are around 5 and 9 millions of DKK, reflecting that several subsets of exceedances are compatible with the GPD. This result is consistent with findings in McNeil (1997), where it appears that several thresholds make compatible the GPD model with the observed data. For this data McNeil (1997) and Frigessi et al. (2002) showed that the predicted quantiles are very sensible to the choice of u. In our approach uncertainty on u is automatically accounted through its posterior distribution. Table 2 reports the posterior median of some extreme quantiles using the mixture with semiparametric model. For comparison purposes, Table 2 shows the estimated quantiles according to: the Lognormal-GPD mixture, the Dynamic Mixture Model (DMM) in Frigessi et al. (2002) (very similar to those obtained with the Weibull-GPD mixture), the POT-MLE model for the value u^* in Reiss and Thomas (2007) and the two pointed thresholds in McNeil (1997), namely $u = 4$ and $u = 9$. As we can see, the estimated quantiles $q_{0.95}$ and $q_{0.99}$ are similar, while much higher quantiles tend to diverge and are generally larger than those based on MLE with $u = 9$. This is due to the fact that the choice of u affects the estimation of ξ, namely $\hat{\xi}$ in Table 2, which represents, with an abuse of notation, the posterior median for the semiparametric approach and Lognormal-GPD while these are the MLEs for the rest. Very large values of u, say $u = 9$, tend to produce less thick tails ($\hat{\xi} = 0.5$) and consequently smaller extreme quantiles and vice versa. In fact,

TABLE 1

POSTERIOR ESTIMATES OF THE PARAMETERS BASED ON 200000 SIMULATIONS WITH A BURN-IN OF 50000 DRAWS AND SAVING ONE MCMC STEP EVERY 10 STEPS.

Parameter	Median	Mean	$q_{0.025}$	$q_{0.975}$
u	5.296	7.476	0.991	23.345
σ	5.921	5.63	1.5	11.007
ξ	0.583	0.601	0.298	1.138

TABLE 2

QUANTILE PREDICTION FOR THE DANISH LOSS DATA, USING OUR BAYESIAN SEMIPARAMETRIC-GPD MIXTURE, THE LOGNORMAL-GPD MIXTURE, THE MIXTURE MODEL IN FRIGESSI ET AL. (2002), THE POT-ML MODEL IN MCNEIL (1997) FOR $u = u^*$, ACCORDING TO REISS AND THOMAS (2007), AND u FIXED AT 4 AND 9. $\hat{\xi}$ REPRESENTS THE POSTERIOR MEDIAN FOR THE SEMIPARAMETRIC AND LOGNORMAL GPD MIXTURES AND THE MLE FOR THE OTHER APPROACHES.

Quantile	Semiparametric	Lognormal	Frigessi	POT-MLE		
p	GPD mixture	GPD mixture	DMM	$u^* = 2$	$u = 4$	$u = 9$
0.95	9	9	8	8	8	9
0.99	26	30	25	26	26	26
0.999	106	156	112	125	120	93
0.9999	412	798	474	574	521	304
0.99999	1572	4091	1987	2619	2238	965
$\hat{\xi}$	0.58	0.71	0.62	0.68	0.63	0.50

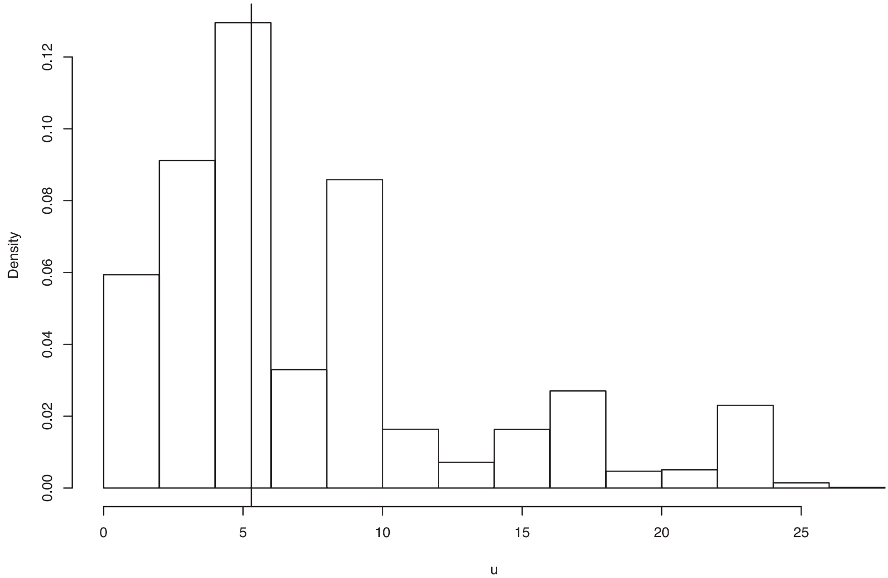

FIGURE 3: Histogram of simulations from the marginal posterior distribution for u.

https://doi.org/10.2143/AST.41.1.2084387 Published online by Cambridge University Press

according to u^* we obtain the largest estimations of q_p for fixed u. The approach in Frigessi et al. (2002) tends to estimate lower values of u and it is much similar to the MLE approach conditionally on $u = 4$. However, the estimation of ξ, reported in Frigessi et al. (2002), is slightly greater reflecting a heavier tail than the one estimated with our model. Although differences seems to be not dramatic, 0.62 against 0.58, they matter in extreme quantiles estimation. On the contrary, our estimation of q_p is based on averaging over all parameters according to the posterior distribution and our results are in between $u = 2$ and $u = 9$.

However, using the Bayesian approach with the Lognormal-GPD mixture, results are sensible to this choice of the central model. In fact, the poor fit of this mixture induces a severe underestimation of u, the posterior median is about 0.73 million DKK meaning that more than 50% of the data are extremes. This underestimation of u produces similar results to that commented for u^*, namely large values of ξ and q_p.

Figure 4 shows the quantile-quantile plot comparing the empirical ones with the posterior predicted values of our semiparametric-GPD fitted model, the plot reflects a good fit of the semiparametric mixture model. In fact, the Anderson Darling (AD) test statistic to check the goodness-of-fit of the GPD to a sample (Choulakian and Stephens (2001)) is 0.55, for observations above $u = 5.29$, with $\xi = \hat{\xi}$ and $\sigma = \hat{\sigma}$ (first column in Table 1). The observed AD is smaller than 0.83, which is the 0.05 upper-tail asymptotic point for $\xi = 0.5$ (Table 2 of Choulakian and Stephens (2001)).

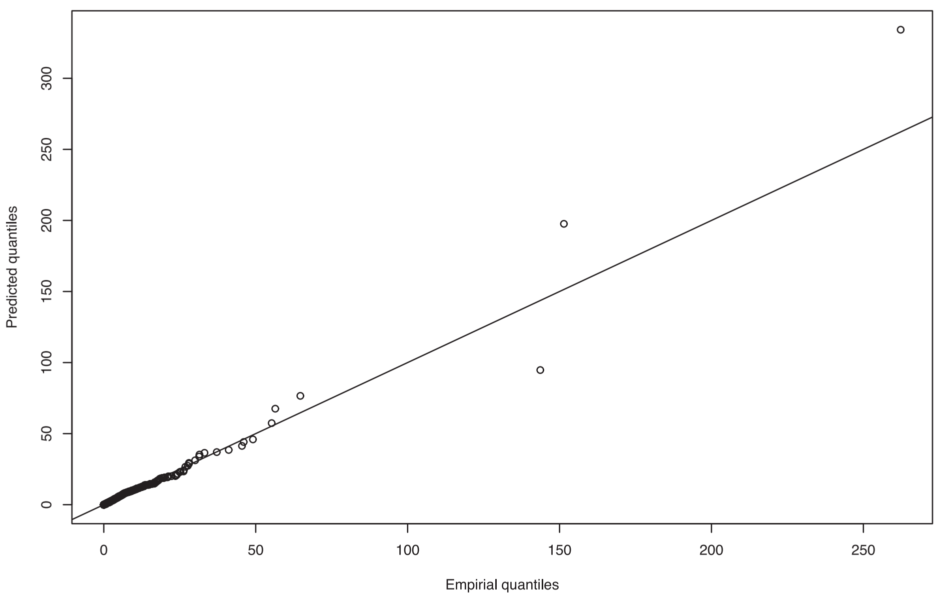

FIGURE 4: Quantile-quantile comparing Danish loss data with posterior predictive quantiles for model in equation (2.1) when the semiparametric estimation is used.

4.2. Norwegian fire claims

The Norwegian fire claims data (Figure 5(a)), presented in the Appendix I of Beirlant et al. (1996), represent the total damage done by 142 fires in Norway for the year 1975. As in the previous example, we subtract half million Krones to shift all data towards 0.

According to Brazakauskas and Serfling (2003) the Pareto distribution is compatible with all observations. Our approach indicates that three values of u have high posterior probability as shown in Figure 5(b) and also by the Mean Excess plot (Figure 5d). The most probable thresholds suggest that different sets of exceedances are compatible with the GPD. This fact is compatible with the fit of this distribution to all observations, as used in Brazakauskas and Serfling (2003), because of the reproductivity property of the GPD mentioned in Section 1. Again, we do not compromise with one u, or one set of parameters of the GPD, and averaging over the posterior distribution for all parameters leads to the posterior predictive distribution for future observations. This predictive distribution fits quite well the observed data as shown in Figure 5(c). In this case, the AD statistic, for observations above the modal $u = 7.3$ with posterior medians $\hat{\xi} = 0.52$ and $\hat{\sigma} = 6.23$, is 0.71 smaller than the 0.05 upper-tail quantile, 0.83. Posterior median of some q_ps are reported in Table 3 and are compared with the estimated quantiles using the POT-MLE model for $u = 0$ and $u = u^* = 1.32$ according to Reiss and Thomas (2007). The

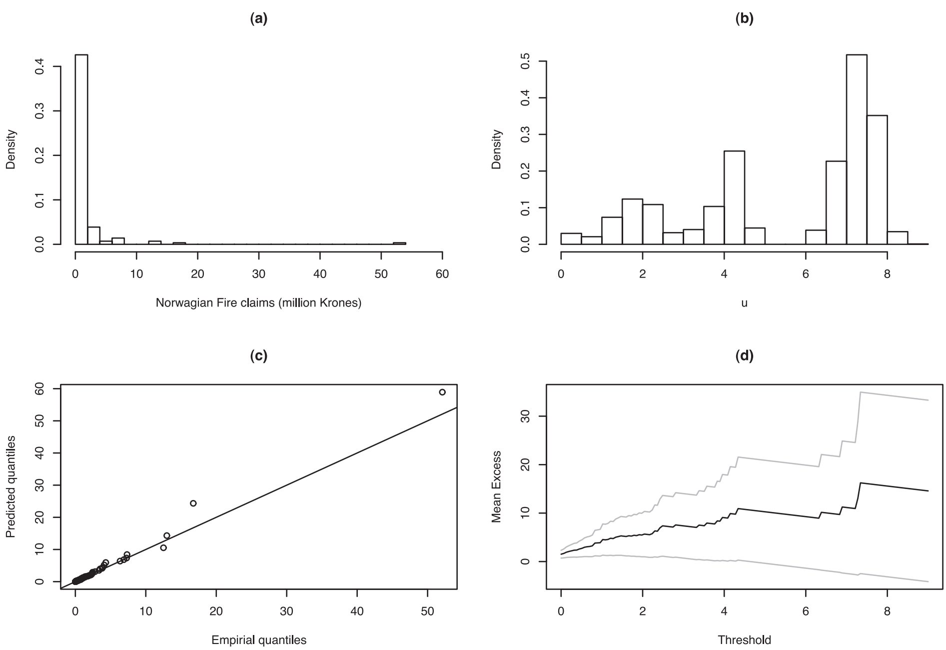

FIGURE 5: (a) Norwegian fire loss data. (b) Posterior distribution of u. (c) Quantile-quantile plot comparing Norwegian loss data with posterior predictive quantiles using the Bayesian semiparametric-GPD mixture. (d) Mean Excess Plot.

TABLE 3

POSTERIOR MEDIAN OF EXTREME QUANTILES FOR THE NORWEGIAN LOSS DATA, USING OUR BAYESIAN SEMIPARAMETRIC-GPD MIXTURE, AND ESTIMATED QUANTILES USING POT-MLE FOR $u = 0$ AND $u = u^* = 1.32$ ACCORDING TO REISS AND THOMAS (2007).

p	0.95	0.99	0.999	0.9999	0.99999
$q_p \mid \boldsymbol{x}$	6	26	96	303	987
q_p POT-MLE, $u = 0$	5	22	153	1041	7075
q_p POT-MLE, $u^* = 1.32$	6	21	115	625	3369

estimated quantiles using the POT-MLE for $u = 0$ and $u = u^*$, for large values of p, tends to be very large compared to those obtained by averaging over all possible values of u, as occurred in the previous example. Also in this case the assumption of the GPD mixture with the Weibull, Normal or Lognormal, as central models, is not compatible with the data.

5. SIMULATION STUDY

We used a simulation study in order to investigate the frequentist properties of the credible intervals for extreme quantiles when these are estimated under the true parametric model and with the semiparametric density estimation $\hat{h}_u$, with $d = 2$ and $d = 3$. We simulated data separately from two types of mixture models with 10% of extreme values, $\xi \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, $\sigma = 1$,

i) Normal-GPD mixture, $0.9 \cdot N(\mu, \tau^2)\mathbf{1}_{\{x \leq u\}} + 0.1 \cdot GPD(\xi, \sigma)\mathbf{1}_{\{x > u\}}$, where $\mu = 0$, $\tau^2 = 1$;

ii) Weibull-GPD mixture, $0.9 \cdot Wei(a, \beta)\mathbf{1}_{\{x \leq u\}} + 0.1 \cdot GPD(\xi, \sigma)\mathbf{1}_{\{x > u\}}$, where $\alpha = 2$, $\beta = 1/\Gamma(1.5)$ in order to have unit mean and variance.

The value of u is such that $H(u \mid \gamma) = 0.9$, where H can be the Normal or Weibull cdf. We consider two reasonable sample sizes for usual applications, say $n = 500$ and $n = 1000$. Note that Normal as well as Lognormal distributions belong to the POT-domain of attraction of the exponential df, therefore, as we are mainly interested in tail behavior, we only provide results of the simulation study for the mixture Normal-GPD model.

The actual coverages of the (nominal) 95% credible intervals calculated over 500 Monte Carlo (MC) replications along with their respective median length appear in Table 4. Results under the Normal-GPD model are shown in the first four columns, while the corresponding ones with $\hat{h}_u$ in the remaining columns.

We can see that the actual coverages are between 0.93 and 0.97 and these values are compatible with the nominal 95% if considering the MC standard error. The median length of the obtained intervals increases with ξ as the

TABLE 4

COVERAGES (ODD COLUMNS) AND MEDIAN LENGTH INTERVAL (EVEN COLUMNS) FOR 500 95% CREDIBLE INTERVALS FOR POSTERIOR QUANTILES WITH $n = 500$ AND $n = 1000$ SAMPLES WITH $p = 0.9$. THE TRUE h IS THE STANDARD NORMAL DENSITY. TRUE QUANTILES ARE BETWEEN 3 ($\xi = -0.4$) AND 15 ($\xi = 0.4$) FOR q_{999} AND BETWEEN 4 AND 38 FOR q_{9999}. COVERAGE STANDARD ERRORS ARE AROUND 0.01.

ξ	Normal model				$d = 2$				$d = 3$			
	$q_{999}\|\boldsymbol{x}$		$q_{9999}\|\boldsymbol{x}$		$q_{999}\|\boldsymbol{x}$		$q_{9999}\|\boldsymbol{x}$		$q_{999}\|\boldsymbol{x}$		$q_{9999}\|\boldsymbol{x}$	
	$n = 500$											
0.4	0.94	43	0.94	347	0.93	58	0.94	805	0.93	59	0.95	853
0.2	0.93	17	0.94	85	0.94	21	0.93	173	0.93	21	0.94	206
0	0.93	7	0.93	22	0.94	9	0.94	54	0.94	10	0.95	60
–0.2	0.94	2	0.96	6	0.95	3	0.97	7	0.93	4	0.94	16
–0.4	0.97	1	0.97	3	0.97	1	0.96	2	0.96	2	0.97	6
	$n = 1000$											
0.4	0.93	25	0.94	161	0.93	27	0.93	215	0.94	29	0.95	246
0.2	0.93	10	0.93	43	0.95	11	0.95	57	0.94	11	0.95	61
0	0.94	4	0.95	11	0.93	4	0.94	14	0.93	4	0.93	16
–0.2	0.93	2	0.93	3	0.93	2	0.95	4	0.93	2	0.93	4
–0.4	0.95	1	0.95	1	0.95	1	0.96	1	0.96	1	0.97	1

TABLE 5

COVERAGES (ODD COLUMNS) AND MEDIAN LENGTH INTERVAL (EVEN COLUMNS) FOR 500 95% CREDIBLE INTERVALS FOR POSTERIOR QUANTILES WITH $n = 500$ AND $n = 1000$ SAMPLES WITH $p = 0.9$. THE TRUE h IS THE WEIBULL DENSITY WITH $\alpha = 2$ AND MEAN 1. TRUE QUANTILES ARE BETWEEN 4 ($\xi = -0.4$) AND 15 ($\xi = 0.4$) FOR q_{999} AND BETWEEN 4 AND 39 FOR q_{9999}. COVERAGE STANDARD ERRORS ARE AROUND 0.01.

ξ	Weibull model				$d = 2$				$d = 3$			
	$q_{999}\|\boldsymbol{x}$		$q_{9999}\|\boldsymbol{x}$		$q_{999}\|\boldsymbol{x}$		$q_{9999}\|\boldsymbol{x}$		$q_{999}\|\boldsymbol{x}$		$q_{9999}\|\boldsymbol{x}$	
	$n = 500$											
0.4	0.93	55	0.93	519	0.92	64	0.93	795	0.94	66	0.95	868
0.2	0.94	19	0.93	108	0.92	23	0.93	156	0.94	26	0.95	241
0	0.93	7	0.93	24	0.92	8	0.92	35	0.93	10	0.96	59
–0.2	0.94	3	0.95	6	0.93	3	0.94	7	0.94	4	0.96	19
–0.4	0.96	1	0.97	2	0.96	1	0.97	2	0.96	2	0.97	6
	$n = 1000$											
0.4	0.95	26	0.96	177	0.96	27	0.95	197	0.95	29	0.95	235
0.2	0.94	10	0.94	43	0.93	10	0.93	49	0.94	10	0.94	49
0	0.94	4	0.94	11	0.94	4	0.94	11	0.93	4	0.94	13
–0.2	0.93	1	0.93	3	0.93	1	0.93	3	0.93	1	0.93	3
–0.4	0.94	1	0.96	1	0.96	1	0.96	1	0.94	1	0.96	1

process is more extreme. Their length also increases with polynomial degrees while coverages seems to be robust with respect to ξ. Comparing the interval length obtained with the parametric Normal model against the corresponding ones obtained with the semiparametric mixture, we can see that the latter are larger than the former. This is due to the greater model flexibility when considering the semiparametric estimation instead of a fixed parametric model. Using large samples, say $n = 1000$, the usual case in finance and actuarial problems, the length of posterior intervals significantly decreases with respect to the case of $n = 500$. Moreover, as expected, the semiparametric and the Normal model tend to provide credible intervals that share the same features.

Similar results appear in Table 5, where we consider the Weibull model to generate the data, instead of the Normal one. In this case coverages are slightly smaller than in the Normal model, also when we use the parametric mixture model. With larger values of n differences tend to vanish.

6. Concluding remarks

Estimation of extreme quantiles is of great interest in many applied fields. It can be generally dealt with model (2.1). In this setup, the Pickand's theorem allows to fix the parametric family of extremal component of the mixture, but the central model, for non-extreme data, is still unspecified. Elicitation of this model affects the final inference as shown in the application examples.

When subjective information is available on the central model, this can be used in the proposed setup and extreme quantiles can be estimated with more precision. On the contrary, when a parametric model cannot be assumed, as in the cases of fire claim data here analyzed, one can consider the central model a nuisance parameter and use a semiparametric estimation along with the profile likelihood, L_p, as proposed here. With this latter approach is still possible to draw Bayesian inference on extreme quantiles without recurring to a parametric model. We would like to remark that we have used the semiparametric estimation method for the central model instead of non-parametric density estimators, such as a kernel density estimator for several reasons:

1. we can control the amount of fitting of non-extreme observations by fixing d, the degree of the polynomial Poisson regression;
2. it is faster to be calculated;
3. it may solve the identifiability issue pointed out in the introduction section.

Density estimation, based on kernel methods, fits the tail of the data better than the GPD and essentially leads to an inferential process that does not account for the Pickand's Theorem and so it is useless for estimating extreme quantiles.

As we mentioned in Section 1 the parameters σ and u are related and this could cause an identifiability problem. This can be avoided using subjective priors over u as done in Behrens et al. (2004), or a central model that is not excessively flexible, as in Frigessi et al. (2002) and also in this paper. In order

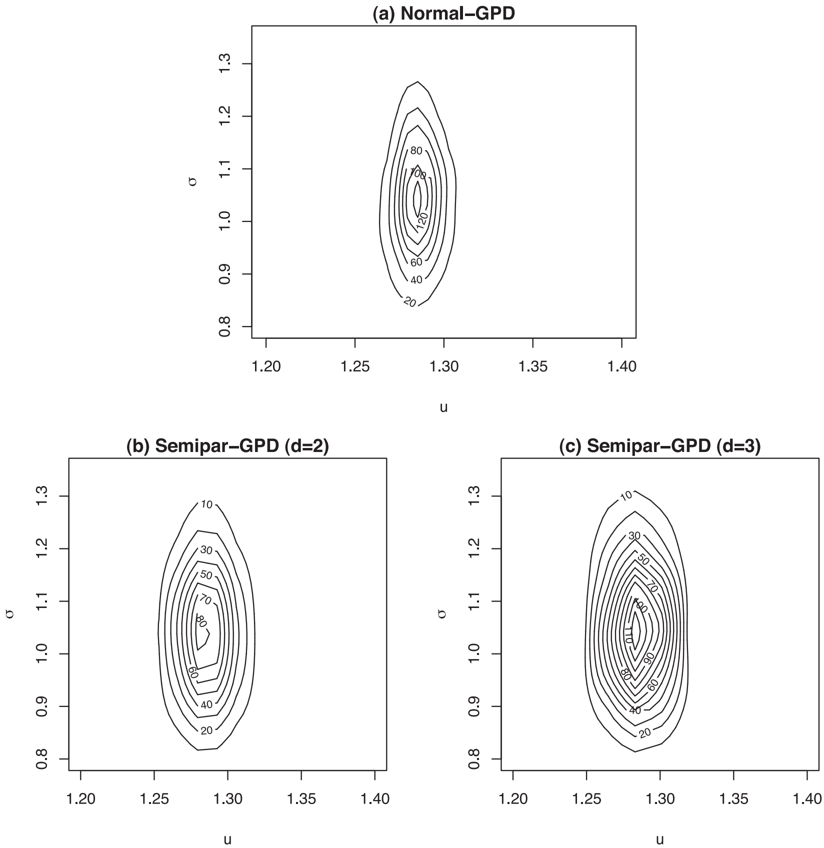

FIGURE 6: Bivariate density estimation of the approximate posterior distribution of (u, σ) for a simulated data set: (a) corresponds to posterior simulations obtained with the Normal-GPD model, while (b) and (c) correspond to those obtained with the semiparametric-GPD model with $d = 2$ and $d = 3$ respectively.

to support this latter statement, we present results of the approximated posterior distribution for (u, σ) under three scenarios where the central model is increasingly flexible. In particular for a simulated data set from the Normal-GPD, with parameters $\xi = 0.2$, $\sigma = 1$, $\mu = 0$, $\tau = 1$ and $u = \Phi^{-1}(0.9) = 1.28$, we fit the Normal-GPD mixture and the semiparametric-GPD with $d = 2$ and $d = 3$. Figure 6 shows the approximated density of the joint posterior distribution of (u, σ). Although the three posterior distributions share the same mode, their dispersion increases with model flexibility, but model (2.1) is clearly identifiable.

Other possible approaches for estimating the central model are mixture models such as that considered in Tancredi et al. (2006). The approach of mixtures, however, need always an elicitation of a prior over the number of mixture components. As we are interested in the fitting of the tail in order to estimate or predict extreme quantiles, we think that the semiparametric-GPD mixture model proposed here simplifies the whole modeling approach.

Acknowledgements

The authors thank the Editor, the Associate Editor and two anonymous Referees for their useful comments that made possible to improve the original version of the paper. This research has been supported by M.I.U.R. (Ministero dell'Istruzione, dell'Univesità e della Ricerca) of Italy, by the Spanish Ministry of Science and Technology, under grants MTM2010-19528, TIN2008-0679-C04-01 and by grant S2009/esp-1594 from the CAM.

Appendix

We show in this section the conditional distributions needed for the Gibbs sampling used to simulate from the posterior distribution of $\boldsymbol{\theta}$. We consider different sets of conditional distributions depending on the central parametric model: Normal, Lognormal, Weibull or semiparametric density estimator. In each case $\boldsymbol{\theta}$ denotes the vector of parameters of interest, in the case of using the semiparametric density estimator, these are only u, ξ and σ. Note that the conditional distributions for ξ and σ depend on u only and not on γ. In the following formulas n_u^+ and n_u^- denote the number of observations above and below u, respectively.

Last subsection is devoted to some comments on the used R code.

Conditional distributions of GPD parameters

$$\pi(\xi \mid \boldsymbol{\theta} \backslash \xi, \boldsymbol{x}) \propto (1+\xi)^{-1}(1+2\xi)^{-1/2} \prod_{\{i,\, x_i > u\}} \left(1 + \xi \frac{(x_i - u)}{\sigma}\right)^{-(1+\xi)/\xi} \times$$
$$\times \mathbf{1}_{\max\{-0.5,\, -\sigma/(x_{(n)} - u)\}}$$
$$\rho(\sigma \mid \boldsymbol{\theta} \backslash \xi, \boldsymbol{x}) \propto \sigma^{-(1+n_u^+)} \prod_{\{i,\, x_i > u\}} \left(1 + \xi \frac{(x_i - u)}{\sigma}\right)^{-(1+\xi)/\xi} \mathbf{1}_{\max\{0,\, -\xi(x_{(n)} - u)\}}$$

where $\boldsymbol{\theta} \setminus \xi$ and $\boldsymbol{\theta} \setminus \sigma$ denote all parameters in $\boldsymbol{\theta}$ without ξ and σ, respectively.

Conditional distributions of Normal parameters

$$\pi(\mu \mid \boldsymbol{\theta} \backslash \mu, \boldsymbol{x}) \propto (1 - \Phi(u \mid \mu, \tau^2))^{n_u^+} \prod_{\{i,\, x_i \le u\}} \phi(x_i \mid \mu, \tau^2)$$
$$\pi(\tau^2 \mid \boldsymbol{\theta} \backslash \tau^2, \boldsymbol{x}) \propto \frac{1}{\tau^2}(1 - \Phi(u \mid \mu, \tau^2))^{n_u^+} \prod_{\{i,\, x_i \le u\}} \phi(x_i \mid \mu, \tau^2),$$

where $\Phi(\cdot \mid \mu, \tau^2)$ and $\phi(\cdot \mid \mu, \tau^2)$ denote, respectively, the cdf and df of a Normal model with mean μ and variance τ^2. Again, $\boldsymbol{\theta} \setminus \mu$ and $\boldsymbol{\theta} \setminus \tau^2$ denote all parameters in $\boldsymbol{\theta}$ without μ and τ^2, respectively.

https://doi.org/10.2143/AST.41.1.2084387 Published online by Cambridge University Press

Conditional distributions of Lognormal parameters

$$\pi(\mu \mid \boldsymbol{\theta} \setminus \mu, \boldsymbol{x}) \propto (1 - H_{LN}(u \mid \mu, \tau^2))^{n_u^+} \prod_{\{i,\, x_i \leq u\}} h_{LN}(x_i \mid \mu, \tau^2)$$

$$\pi(\tau^2 \mid \boldsymbol{\theta} \setminus \tau^2, \boldsymbol{x}) \propto \frac{1}{\tau^2}(1 - H_{LN}(u \mid \mu, \tau^2))^{n_u^+} \prod_{\{i,\, x_i \leq u\}} h_{LN}(x_i \mid \mu, \tau^2),$$

where $H_{LN}(\cdot \mid \mu, \tau^2)$ and $h_{LN}(\cdot \mid \mu, \tau^2)$ denote, respectively, the cdf and df of a Lognormal model with location μ and scale τ^2. Again, $\boldsymbol{\theta} \setminus \mu$ and $\boldsymbol{\theta} \setminus \tau^2$ denote all parameters in $\boldsymbol{\theta}$ without μ and τ^2, respectively.

Conditional distributions of Weibull parameters

$$\pi(\alpha \mid \boldsymbol{\theta} \setminus \alpha, \boldsymbol{x}) \propto (1 - H_{Wei}(u \mid \alpha, \beta))^{n_u^+} \prod_{\{i,\, x_i \leq u\}} h_{Wei}(x_i \mid \alpha, \beta)$$

$$\pi(\beta \mid \boldsymbol{\theta} \setminus \beta, \boldsymbol{x}) \propto \frac{1}{\beta}(1 - H_{Wei}(u \mid \alpha, \beta)^{n_u^+} \prod_{\{i,\, x_i \leq u\}} h_{Wei}(x_i \mid \alpha, \beta),$$

where $H_{Wei}(\cdot \mid \alpha, \beta)$ and $h_{Wei}(\cdot \mid \alpha, \beta)$ denote the cdf and df of a Weibull model with shape parameter a and scale parameter β.

Conditional distribution of *u*

When a parametric model is considered for non extreme data, Normal, Lognormal or Weibull, the conditional distribution for u is,

$$\pi(u \mid \boldsymbol{\theta} \setminus u, \boldsymbol{x}) \propto (1 - H(u \mid \gamma))^{n_u^+} \sigma^{-n_u^+} \prod_{\{i,\, x_i \leq u\}} h(x_i \mid \gamma) \times$$

$$\times \prod_{\{i,\, x_i > u\}} \left(1 + \xi \frac{(x_i - u)}{\sigma}\right)^{-(1+\xi)/\xi} \mathbf{1}_B,$$

where $B = \{[(\xi < 0) \cap (u > x_{(n)} + \sigma/\xi) \cup (\xi \geq 0)\,] \cap (a \leq u \leq b)\}$ and $h(\cdot \mid \gamma)$ and $H(\cdot \mid \gamma)$ denote the df and cdf of the parametric model.

Instead, when we use the semiparametric density estimator described in Section 2.2, the conditional distribution for u is:

$$\pi(u \mid \boldsymbol{\theta} \setminus u, \boldsymbol{x}) \propto (1 - \widehat{H}(u))^{n_u^+} \sigma^{-n_u^+} \widehat{H}(u)^{n_u^-} \prod_{\{i,\, x_i \leq u\}} \hat{h}_u(x_i) \times$$

$$\times \prod_{\{i,\, x_i > u\}} \left(1 + \xi\, \frac{(x_i - u)}{\sigma}\right)^{-(1+\xi)/\xi} \mathbf{1}_B,$$

We replace the estimator of the cdf $\widehat{H}(\cdot)$ by the empirical cdf, and the estimator $\hat{h}_u(\cdot)$ by the semiparametric density estimator described in Section 2.2.

R implementation

Density $\hat{h}_u(x)$ has been implemented using the `glm()` function with the orthogonal polynomials obtained with the `poly()` function. The following code provides an estimation with df degrees of freedom.

```
# Assume z contains sorted observations below u
zh = hist(z, plot = F)
zb=zh$mids #Middle class points
y = zh$counts #Histogram counts
mp=poly(zb, degree = df) #Orthogonal polynomial
mc=glm(y~mpoly, poisson)$coefficients # Poisson Regression
kh = function(x,mp,mc) exp(cbind(1,predict.poly(mp,x))%*%mc)
#Kernel of h
nc = integrate(kh,lower = z[1],upper = u,mp = mp,mc = mc)$value
#Normalizing constant of kh()
h=function(x,mp,mc,nc) kh(x,mp,mc)/nc #Semiparametric estimation
```

Density of GPD, $g(x|\xi, \sigma, u)$, is implemented in `dgpd()` (library `POT`). The implemented MCMC algorithm uses the above conditional distributions that use the functions `h()`, `dgpd()` and the above prior densities.

References

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

Beirlant, J., Teugels, J. and Vynckier, P. (1996) *Practical Analysis of Extreme Values*. Leuven University Press, Leuven, Belgium.

Brazauskas, V. and Serfling, R. (2003) Favorable estimators for fitting pareto models. A study using goodness-of-fit measures with actual data. *ASTIN Bulletin* **33(2)**, 365-381.

Cabras, S. and Morales, J. (2007) Extreme value analysis within a parametric outlier detection framework. *Applied Stochastic Models in Business and Industry* **23(2)**, 157-164.

Castellanos, M. and Cabras, S. (2007) A default bayesian procedure for the generalized pareto distribution. *Journal of Statistical Planning and Inference* **137(2)**, 473-483.

Castillo, E. and Hadi, A. (1997) Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association* **92**, 1609-1620.

Chavez-Demoulin and Embrechts, P.V. (2009) *An EVT primer for credit risk*. Oxford University Press.

Choulakian, V. and Stephens, M. (2001) Goodness-of-fit tests for the generalized pareto distribution. *Technometrics* **43(4)**, 478-484.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer-Verlag.

Davison, A. (1984) *Statistical extremes and Applications*, Chapter Modelling excesses over high thresholds, with an application, pp. 461-482. Reidel, Dordrecht.

de Zea Bermudez, P. and Turkman, M.A. (2003) Bayesian approach to parameter estimation of the generalized pareto distribution. *Test* **12**, 259-277.

de Zea Bermudez, P., Turkman, M.A. and Turkman, K. (2001) A predictive approach to tail probability estimation. *Extremes* **4**, 295-314.

Donnelly, C. and Embrechts, P. (2010) The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *ASTIN Bulletin* **40(1)**, 1-33.

Efron, B. and Tibshirani, R. (1996) Using specially designed exponential families for density estimation. *The Annals of Statistics* **24(6)**, 2431-2461.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997) *Modeling extremal events for insurance and finance*. New York: Springer.

FRIGESSI, A., HAUG, O. and RUE, H. (2002) A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* **5**, 219-235.

GRIMSHAW, S.D. (1993) Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics* **35(2)**, 185-191.

HOSKING, J. and WALLIS, J. (1987) Parameter and quantile estimation for the generalized pareto distribution. *Technometrics* **29(3)**, 339-349.

LINDSEY, J. (1974a) Comparison of probability distributions. *Journal of the Royal Statistical Society, Series B* **36**, 38-47.

LINDSEY, J. (1974b) Construction and comparison of statistical models. *Journal of the Royal Statistical Society, Series B* **36**, 418-425.

MCNEIL, A.J. (1997) Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 117-137.

MIKOSCH, T. (2003) Modeling dependence and tails of financial time series. In B. Finkenstaedt and H. Rootzen (Eds.), *Extreme Values in Finance, Telecommunications, and the Environment*, pp. 185-286. Chapman & Hall/CRC.

NEVES, C. and FRAGA ALVES, M.I. (2004) Reiss and thomas' automatic selection of the number of extremes. *Computational Statistics and Data Analysis* **47**, 689-704.

PADGETT, W. and JOHNSON, M. (1983) Some bayesian lower bounds on reliability in the lognormal distribution. *The Canadian Journal of Statistics* **11**, 137-147.

PICKANDS, J. (1975) Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119-131.

R DEVELOPMENT CORE TEAM (2009) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

REISS, R.D. and THOMAS, M. (2007) *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields* (Third ed.). Birkhäuser Verlag.

ROEDER, K. and WASSERMAN, L. (1997) Practical bayesian density estimation using mixtures of normals. *Journal of American Statistical Association* **92**, 894-902.

SMITH, R. (2003) *Extreme values in finance, telecommunications and the environment*, Chapter 1. Statistics of extremes, with applications in environment, insurance and finance. Chapman and Hall/CRC Press.

SMITH, R.L. (1984) *Statistical extremes and Applications*, Chapter Threshold methods for sample extremes, pp. 621-638. Reidel, Dordrecht.

SUN, D. and BERGER, J. (1994) Bayesian sequential reliability for weibull and related distributions. *Ann. Inst. Statist. Math.* **46(2)**, 221-249.

TANCREDI, A., ANDERSON, C. and O'HAGAN, A. (2006) Accounting for threshold uncertainty in extreme value estimation. *Extremes* **9**, 87-106.

VENTURA, L., CABRAS, S. and RACUGNO, W. (2009) Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *Journal of the American Statistical Association* **104(486)**, 768-774.

STEFANO CABRAS (Corresponding author)
Department of Mathematics, University of Cagliari (Italy).
Via Ospedale 72, 09124 Cagliari (Italy),
Tel.: +390706758536,
Fax: +390706758511.
E-mail: s.cabras@unica.it

MARÍA EUGENIA CASTELLANOS
Department of Statistics and Operations Research,
Rey Juan Carlos University (Spain).
C/ Tulipán, 28933, Móstoles (Spain).
E-mail: maria.castellanos@urjc.es