REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008, 83–100

TESTING EXTREME VALUE CONDITIONS — AN OVERVIEW AND RECENT APPROACHES

Authors: Cláudia Neves
– Department of Mathematics and UIMA,
University of Aveiro, Portugal
claudia.neves@ua.pt

M. Isabel Fraga Alves
– DEIO and CEAUL, Faculty of Sciences,
University of Lisbon, Portugal
isabel.alves@fc.ul.pt

Abstract:

- The aim of this paper is to give a brief overview about several tests published in the context of statistical choice of extreme value domains and for assessing extreme value conditions. Some of the most recent testing procedures encompassed in this framework will be illustrated using a teletraffic data set.

Key-Words:

- *extreme values; POT and PORT methodologies; location/scale invariance; statistical testing.*

AMS Subject Classification:

- Primary: 62G32; Secondary: 62F03, 62G10.

1. INTRODUCTION

Statistical inference about rare and damaging events can fairly be designed upon those observations which are considered extreme in some sense. There are different ways of mapping such observations yielding alternative approaches to statistical inference on extreme values: the classical Gumbel parametric method of block of *Annual Maxima*, *Peaks-Over-Threshold* (POT) parametric methods and the recently denominated *Peaks-Over-Random-Threshold* (PORT) semi-parametric methods, which is nothing more than a fairly small variant of POT for statistical inference conditionally on an intermediate random threshold.

However, regardless of the specific approach we intend to follow, statistical inference is clearly improved if one makes *a priori* assumptions about the most appropriate type of decay of the underlying tail distribution function $1-F$, i.e., about whether it decays exponentially fast, is polynomially decreasing or exhibits a light tail with finite right endpoint. This is supported by Extreme Value Theory, stemming from the fundamental Theorem of Fisher and Tippett (1928), which ascertains that all possible non-degenerate weak limit distributions of partial maxima of independent and identically distributed random variables $X_1, X_2, ...$ are (Generalized) Extreme Value distributions.

The Generalized Extreme Value distribution (GEVd) comprises Fréchet, Weibull and Gumbel distributions. A distribution function (d.f.) F that belongs to the Fréchet domain of attraction is called a heavy-tailed distribution, the Weibull domain encloses light-tailed distributions with finite right endpoint and the particularly interesting case of the Gumbel domain embraces a great variety of tail distribution functions ranging from light to moderately heavy, whether detaining finite right endpoint or not.

Hence, separating statistical inference procedures according to the most suitable domain of attraction for the underlying distribution has become a usual practice in the literature either by following a parametric or a semi-parametric approach. Following a semi-parametric approach, the only assumption made is that the underlying d.f. is in the domain of attraction of the GEVd. In this setup, any inference concerning the tail of the underlying distribution is based exclusively on those observations lying above an intermediate random threshold, giving rise to the PORT method. The latter compares with the alternative setup of restricting attention to a random number of observations exceeding a given high increasing deterministic level u, an approach engraved in the POT method.

Our aim here is to give a brief overview of several well-known testing procedures in the context of statistical choice of extreme value conditions, along with some recent proposals using location/scale invariant statistics that have been built on the k excesses above a random threshold. This random threshold is consensually an intermediate order statistic. The development of statistical procedures and techniques with the specific intention of dealing with extreme data in a more systematic and reliable way renders, to our best knowledge, a challenge that many applied fields such as environmetrics, climatology, telecommunications or finance hold in common.

The paper proceeds as follows. Section 2 contains some notation and sets general ground rules in the context of extreme value analysis. For analyzing extreme values there are different approaches, according to the underlying assumptions on F and the specific observations of the random sample available for statistical inference purposes. In this sequence, Sections 3 and 4 provide references and brief descriptions of several contributions in both parametric and semi-parametric setup. Finally, Section 5 brings the PORT-method into focus by means of an application to real data.

2. PRELIMINARIES AND SOME NOTATION

When we are interested in modeling large observations, we are usually confronted with two extreme value models:

- Generalized Extreme Value distribution (GEVd) with d.f.

$$G_\gamma(x) := \begin{cases} \exp\bigl(-(1+\gamma x)^{-1/\gamma}\bigr)\,, & 1+\gamma x>0 \quad \text{if } \gamma\neq 0\,,\\ \exp\bigl(-\exp(-x)\bigr)\,, & x\in\mathbb{R} \quad \text{if } \gamma=0\,. \end{cases} \tag{2.1}$$

- Generalized Pareto distribution (GPd) with d.f.

$$H_\gamma(x) := \begin{cases} 1-(1+\gamma x)^{-1/\gamma}\,, & 1+\gamma x>0 \text{ and } x\in\mathbb{R}^+ \quad \text{if } \gamma\neq 0\,,\\ 1-\exp(-x)\,, & x\in\mathbb{R}^+ \quad \text{if } \gamma=0\,. \end{cases} \tag{2.2}$$

The introduction of scale $\delta>0$ and location $\lambda\in\mathbb{R}$, results in the full GEV and GP families of distributions given by $G_\gamma(x;\lambda,\delta)=G_\gamma\bigl((x-\lambda)/\delta\bigr)$ and $H_\gamma(x;\lambda,\delta)=H_\gamma\bigl((x-\lambda)/\delta\bigr)$, respectively, which play a central role in statistical inference of extreme values.

GEVd and MAX-Domain: The Fisher–Tippett theorem of extreme values (Fisher and Tippett, 1928) states that all possible non-degenerate weak limit distributions of partial maxima of independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$ are (Generalized) Extreme Value distributions. That is, assume there exist normalizing constants $a_n>0$ and $b_n\in\mathbb{R}$ such that, for all x

$$\lim_{n\to\infty} P\Big\{a_n^{-1}\big(\max(X_1,...,X_n)-b_n\big)\le x\Big\} = G(x) , \tag{2.3}$$

where G is some non-degenerate distribution function, we can redefine the constants in such a way that the limit G is one of the GEV family of distributions given by (2.1) in the von Mises–Jenkinson form (von Mises, 1936; Jenkinson, 1955). We then say that $G = G_\gamma$ and the underlying d.f. F is in the domain of attraction of G_γ (notation: $F \in \mathcal{D}(G_\gamma)$). In case of $\gamma < 0$, $\gamma = 0$ or $\gamma > 0$, the G_γ reduces to Weibull, Gumbel or Fréchet distribution function, respectively.

GPd and POT-Domain: The use of GPd is suggested by the result of Balkema and de Haan (1974) and Pickands (1975), who proved that $F \in \mathcal{D}(G_\gamma)$ if and only if the upper tail of F is, in a certain sense, close to the upper tail of H_γ. While restricting attention to a top portion of the original sample, the GPd comes into play since it appears as the limiting distribution for the excesses $Y_i = X_i - u\,|X_i > u$, $i=1,...,k_u$ over a sufficiently high threshold u (POT method). For $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$, the H_γ d.f. in (2.2) reduces to Beta, Exponential and Pareto distribution functions, respectively. In both classes, the extreme value index γ is closely related to the tail heaviness of the distribution. In that sense, the value $\gamma = 0$ (*exponential tail*) can be regarded as a change point: $\gamma < 0$ refers to short tails with finite right endpoint $x^F := \sup\big\{x\colon F(x) < 1\big\}$, whereas for $\gamma > 0$ d.f.'s are heavy tailed. In many applied sciences where extremes come into play, it is assumed that the extreme value index γ of the underlying d.f. equals 0, and statistical inference procedures concerning rare events on the tail of F, such as the estimation of small exceedance probabilities or return periods, bear on this assumption. Moreover, Gumbel and exponential models are also preferred because of the greater simplicity of inference associated with Gumbel or exponential populations.

Here and throughout this paper, let us denote by $X_{1:n} \le ... \le X_{n:n}$ the order statistics pertaining to the i.i.d. random variable $X_1, X_2, ..., X_n$, after arranging these by nondecreasing order.

3. TESTING EXTREMES UNDER A PARAMETRIC APPROACH

In a parametric set-up, the main assumption regards the existence of a suitable class of models for describing the random variable attached to the process that is generating the data under study. These only three possible classes are motivated by Extreme Value Theory, and depend mainly on the shape parameter γ, and eventually on location and scale parameters.

Annual Maxima (AM): Suppose that the maximum of a random sample can be obtained in each of k equally spaced observation periods. The class of GEVd functions, G_γ, may be prescribed in order to model maxima of k subsamples taken from a given set of data of size $k \cdot n$, that is,

$$Z_i := X_{n:n}^{(i)} = \max\left\{X_1^{(i)}, ..., X_n^{(i)}\right\} , \qquad i = 1, ..., k \ . \tag{3.1}$$

A typical course of action lying in this classical Gumbel method is to take annual maxima. In this AM setup, the following testing problem has been treated extensively in the literature, with main emphasis on testing the Gumbel hypothesis for the d.f. of the $\{Z_i\}_{i=1}^k$ defined in (3.1):

$$H_0 \colon \gamma = 0 \qquad \text{vs.} \qquad H_1 \colon \gamma \neq 0 \ . \tag{3.2}$$

The testing problem (3.2) has received much attention in the literature; in fact, the hydrologists have long made use of extreme value distributions for estimating probabilities of flood events and the correct choice of the GEVd under approach is of crucial importance, since the three types differ considerably in their right tails. Among the papers concerned with the special testing problem G_0 against $\{G_\gamma \colon \gamma \neq 0\}$, or against one-sided alternatives $\{G_\gamma \colon \gamma > 0\}$, $\{G_\gamma \colon \gamma < 0\}$, we refer to Van Montfort (1970), Bardsley (1977), Otten and Van Montfort (1978), Tiago de Oliveira (1981), Gomes (1982), Tiago de Oliveira (1984), Tiago de Oliveira and Gomes (1984), Hosking (1984), Marohn (1994), Wang *et al.* (1996) and Marohn (1998a). Somewhat connected with the problem (3.2), there is the problem of goodness-of-fit tests for the Gumbel model, which has received the attention of Stephens (1976), Stephens (1977), Stephens (1986) and Kinnison (1989). The tests therein considered are mostly based on the well known goodness-of-fit statistics: Kolmogorov, Cramér–von Mises and Anderson–Darling statistics.

Largest Observations (LO): It may be that, when considering yearly data, some years contain several values that are larger then the maxima of other years. Although the requirement of only a simplified data summary carries reduction of possible dependencies in the sampled data, the loss of information provided by the largest observations in the sample can, by itself, motivate this alternative approach. Hence, suppose we take the k largest observations in the sample. If the underlying d.f. $F \in \mathcal{D}(G_\gamma)$, the non-degenerate joint limiting behavior of the k largest random variables determines the probability density function (p.d.f.)

$$f_\gamma(z_1, z_2, ..., z_k) \;=\; g_\gamma(z_k) \prod_{i=1}^{k-1} \frac{g_\gamma(z_i)}{G_\gamma(z_i)} \ , \qquad z_1 > z_2 > ... > z_k \ , \tag{3.3}$$

where $g_\gamma(z) = \partial G_\gamma(z)/\partial z$, in the sense that, after appropriately normalized with constants $a_n > 0$ and b_n,

$$\left(\frac{X_{n:n} - b_n}{a_n}, \frac{X_{n-1:n} - b_n}{a_n}, ..., \frac{X_{n-k+1:n} - b_n}{a_n}\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \left(Z_1, Z_2, ..., Z_k\right) .$$

This multivariate model, introduced in Weissman (1978), has received the general designation of *extremal process*. In light of this result, statistical procedures to discern between Gumbel and Fréchet or Weibull distributions have been considered, for instance, in Gomes and Alpuim (1986), Hasofer and Wang (1992), Wang (1995) and Wang *et al.* (1996). The key insight for the testing problem G_0 against $\{G_\gamma\colon \gamma \neq 0\}$, or against the one-sided alternatives $\{G_\gamma\colon \gamma > 0\}$, $\{G_\gamma\colon \gamma < 0\}$ is thus to assume that, with k fixed, the joint stochastic behavior of the largest k random variables tends to be properly described by the p.d.f (3.3) pertaining to $\gamma = 0$, i.e.,

$$f_0(z_1, z_2, ..., z_k) = \exp\left(-\exp(-z_k) - \sum_{i=1}^{k} z_i\right) , \qquad z_1 > z_2 > ... > z_k \ ,$$

which enables replacement of the normalized top order statistics with $(Z_1, Z_2, ..., Z_k)$. Under this assumption, Hasofer and Wang (1992) prove that the following test statistic

$$W(k) := \frac{1}{k-1} \, \frac{\left(\frac{1}{k} \sum\limits_{i=1}^{k} Z_i - Z_k\right)^2}{\frac{1}{k} \sum\limits_{i=1}^{k} \left(Z_i - Z_k\right)^2 - \left(\frac{1}{k} \sum\limits_{i=1}^{k} Z_i - Z_k\right)^2} \ ,$$

akin to the Shapiro–Wilk goodness-of-fit statistic (see Shapiro and Wilk, 1965), can be considered as approximately normal with mean $(k-1)^{-1}$ and variance $2^2 (k-2)(k-1)^{-2} \left((k+1)(k+2)\right)^{-1}$.

Despite the above results concern a fixed number k of top observations, we can find in Hasofer and Wang (1992) an attempt to make k to increase with n, but at a much slower rate, through the specification of $k = c_1 n^{c_2}$ in the simulation study. Wang (1995) also mention the case where $k \to \infty$ and $k = o(n)$, as $n \to \infty$. Furthermore, Wang (1995) relies on the Hasofer and Wang test to select the number k of largest order statistics for suitable statistical inference in the Gumbel domain. In general, if G is a goodness-of-fit statistic, then at a certain nominal level of the test α, say, choose $k+1 = \min\big\{i\colon\, g(i) \in \text{critical region of } G(i)\big\}$, provided the adopted statistic G is scale and location invariant and, of course, sensitive to small deviations from the null hypothesis.

Combination of AM and LO: In Gomes (1989), for instance, the testing problem specifying the Gumbel d.f. G_0 in the (simple) null hypothesis is handled with a combination of blocking split of the sample data and the k largest observations in each of the m blocks through what is called the *multidimensional* — GEV_γ *model*, as follows: a set of independent, identically distributed k-dimensional random vectors $\{\mathbf{X}_i\colon\, i=1,...,m\}$, and after suitable normalization, with common p.d.f of the vectors $\mathbf{Z}_i = (\mathbf{X}_i - \lambda)/\delta$ is given by $f_\gamma(\mathbf{z})$ defined in (3.3). Note that both AM and LO approaches can be particular cases of this multidimensional model, taking $k=1$ and $m=1$, respectively. In Gomes (1987) a truncated sample of the largest values of a sample, whose size increases to infinity and whose limiting distribution is in the class of GEVd is considered for goodness-of-fit purposes. Using the reduction to the exponentials of Gumbel distributions the author develops two-sided tests of exponentiality for the transformed variables, the tests being the Kolmogorov–Smirnov, Cramér–von Mises and Stephens goodness-of-fit test.

Peaks Over Threshold (POT): Suppose we pick up those observations exceeding a fixed high threshold u. As described in Section 2, given a random sample $(X_1, X_2, ..., X_n)$ from the d.f. F, the GPd is regarded as a good approximation for the distribution of the excesses $W_i := X_i - u$ over a sufficiently high threshold u if and only if $F \in \mathcal{D}(G_\gamma)$. A clear difference between the designated AM and POT setups is that the k yearly maxima do not necessarily carry over as to yield the k largest observations from the original sample.

In this POT setup the following testing problem has been frequently considered, rendering priority to testing the Exponential hypothesis for the d.f. of the excesses $\{W_i\}_{i=1}^{k_u}$, i.e., $H_0\colon \gamma=0$ *versus* $H_1\colon \gamma\neq 0$. The maximum likelihood method may then be applied under the assumption that those k_u observations over the threshold u follow exactly a GPd, provided a scale normalization σ_u, i.e.

$$H_{\gamma,\sigma_u}(w) \;=\; 1-\bigl(1+\gamma\, w/\sigma_u\bigr)^{-1/\gamma} \;,$$

for all positive w such that $1+\gamma/\sigma_u\, w>0$. The parametrization $\tau=-\gamma/\sigma$ (Davison and Smith, 1990; Grimshaw, 1993) can be used for reducing dimensionality and therefore construct a likelihood ratio test based on the log-profile likelihood. In view of applications, the problem of detecting the presence of exponential distribution, under the POT approach, has received particular attention from hydrologists. Davison and Smith (1990) addresses this testing problem in the context of river-flow exceedances. Van Montfort and Witter (1986) illustrates the "lack"-of-fit statistic towards exponentiality $\hat\gamma/\sqrt{\widehat{\mathrm{var}}(\hat\gamma)}$, where $\hat\gamma$ denotes the Maximum Likelihood (ML) estimator of γ, in the sequence of a thorough application of the POT method to rainfall data. Among the numerous works connected with the special problem of testing exponential against other GPd upon the tail we mention, for instance, Van Montfort and Witter (1985), Gomes and Van Montfort (1986) and Brilhante (2004). Chaouche and Bacro (2004) introduce the test statistic $S=\overline{W}/(\overline{W}-W_{n:n})$, where again W_i are independent random variables with the same d.f. H_{γ,σ_u}, and obtain its empirical distribution via simulation. Moreover, when using Probability Weighted Moments of different orders to adapt S, a method to purge the influence of σ_u off these new test statistics is provided. Giving heed to the Local Asymptotic Normality theory, Falk (1995) followed by Marohn (1998b) and Marohn (2000), aim at asymptotically optimal tests for discriminating between different values of the extreme value index γ.

Goodness-of-fit tests for the Generalized Pareto distribution

Fitting the GPd function to data, which we expect to be lying far away in the tail, has been worked out in Castillo and Hadi (1997). The problem of goodness-of-fit tests for the GP model has been studied by Choulakian and Stephens (2001), with the following proposals for Cramér–von Mises and Anderson–Darling statistics:

$$W^2 = \sum_{i=1}^{k}\left(H_{\hat\gamma,\hat\sigma}(X_{n-i+1:n}) - \frac{2(k-i)+1}{2k}\right)^2 + \frac{1}{12\,k}\,,$$

$$A^2 = -k - \frac{1}{k}\sum_{i=1}^{k}\Big(2(k-i)+1\Big)\Big(\log H_{\hat\gamma,\hat\sigma}(X_{n-i+1:n}) + \log\big(1-H_{\hat\gamma,\hat\sigma}(X_{i:n})\big)\Big)\,,$$

where $(\hat\gamma,\hat\sigma)$ are ML estimators. A table of critical points is provided with good accuracy for $k \geq 25$. Konstantinides and Meintanis (2004) assess the presence of a GPd by means of a transformation of the data to reduce to exponential, then search for traces of exponentiality in the empirical Laplace transform. They also adapt the critical points leading to what promises to be a more accurate level of the test, pursuing the path of Davison and Smith (1990) claim that tables for testing the presence of a exponential distribution (see Van Montfort and Witter (1986) lack of fit statistic mentioned upstairs) give in general critical values which are too high, thus resulting in a very conservative test. Comparison with Choulakian and Stephens (2001) are also present by means of a simulation study. Luceño (2006) assigns more weight to the tails than the usual practice relating Cramér–von Mises and Aderson–Darling statistics goodness-of-fit test statistics and considers a maximum goodness-of-fit estimation method, which enables us to deal successfully with the estimation of GPd parameters, overcoming the occasional lack of convergence in ML estimation.

4. TESTING EXTREMES UNDER A SEMI-PARAMETRIC APPROACH

Following a semi-parametric approach, the only assumption made is that the extreme value condition (2.3) is satisfied, i.e., the underlying d.f. $F \in \mathcal{D}(G_\gamma)$. In this framework, the extreme value index γ is the parameter of prominent interest since, in both GEV and GP classes of distributions, it determines the shape of the tail of the underlying distribution function F.

To this extent, $\gamma = 0$ can be regarded as a benchmark value, since a negative γ is inevitably associated with short tails with finite right endpoint, while a positive (tail index) γ is connected with the presence of a heavy-tailed distribution. In many applied sciences where extremes are relevant, the case of simplest inference $\gamma = 0$ is assumed and bearing on this assumption, extreme characteristics such as exceedance probabilities or return periods are easily estimated.

As a matter of fact, separating statistical inference procedures according to the most suitable domain of attraction for the sampled distribution has become a usual practice. Methodologies for testing the Gumbel domain against Fréchet or Weibull max-domains have been of great usefulness. This fit-of-attraction problem, crafted from a semi-parametric setup, can be rephrased as a test for

$$H_0\colon\ F \in \mathcal{D}(G_0) \quad \textit{versus} \quad H_1\colon\ F \in \mathcal{D}(G_\gamma)_{\gamma\neq 0}\ . \tag{4.1}$$

or against one-sided alternatives $F \in \mathcal{D}(G_\gamma)_{\gamma<0}$ or $F \in \mathcal{D}(G_\gamma)_{\gamma>0}$.

Statistical tests that tackle the problem (4.1) can be traced back to the seminal papers by Galambos (1982) and Castillo *et al.* (1989). The latter presents a cunning procedure for fit of attraction diagnostics from the curvature of the graph of the sample distribution function hinged on the Gumbel probability paper. Predicated on this (so-called) curvature method, the authors introduce a test to assess whether the upper tail distribution function might be classified as convex, concave or a straight line.

Further testing procedures for (4.1) can be found in Fraga Alves and Gomes (1996), Fraga Alves (1999). Segers and Teugels (2000) have recently suggested a large sample test for the Gumbel domain with asymptotics deriving from the limiting distribution of Galton's ratio under the extreme value condition (2.3), which Rao's test statistic (see e.g. Serfling, 1980) for simple null hypothesis was applied to, with the ulterior aim of establishing a decision rule. In the process, the authors were confronted with the need of blocking the original sample of size n into m subsamples, each of size n_i, $i = 1, ..., m$ also under pledge of largeness.

Recently, Neves *et al.* (2006) and Neves and Fraga Alves (2007) have introduced two testing procedures that are based on the sample observations lying above a random threshold. More specifically, in the last two references, the designed statistics for testing (4.1) are based on the k excesses over the $(n-k)$-th ascending intermediate order statistic $X_{n-k:n}$, where $k = k_n$ is such that $k \to \infty$ and $k = o(n)$ as $n \to \infty$. Clearly, the latter only differs from the POT approach on the absence of a parametric model and on the fact that the *intermediate* random threshold is now playing the role of the deterministic *sufficiently high* threshold u which, only by itself, we find relevant enough to motive the Peaks Over Random Threshold (PORT) methodology. Now following a semi-parametric approach supported on concepts from the theory of regularly varying functions, Neves and Fraga Alves (2007), reformulate the asymptotic properties of the Hasofer and Wang test statistic (denoted below with $W_n(k)$) in case $k = k_n$ behaves as an intermediate sequence rather than remaining fixed while the sample size n increases (which was case covered by Hasofer and Wang, 1992). In the process, a new Greenwood-type test statistic $G_n(k)$ (cf. Greenwood, 1946) proves to be useful in assessing the presence of heavy-tailed distributions.

Furthermore, motivated by eventual differences in the relative contribution of the maximum to the sum of the k excesses over the random threshold at different tail heaviness, a complementary test statistic $R_n(k)$ was introduced by Neves *et al.* (2006) in order to discern between max-domains of attraction.

Under the null hypothesis of Gumbel domain of attraction plus extra mild second order conditions on the upper tail of F and on the growth of the intermediate sequence k_n, we have that

$$\text{(4.2)} \qquad \text{[Ratio-test]} \qquad R_n(k) := \frac{X_{n:n} - X_{n-k:n}}{\frac{1}{k}\sum_{i=1}^{k}\big(X_{n-i+1:n} - X_{n-k:n}\big)} - \log k \;\underset{n\to\infty}{\overset{d}{\longrightarrow}}\; \Lambda\,,$$

$$\text{(4.3)} \qquad \text{[Gt-test]} \qquad G_n(k) := \frac{\frac{1}{k}\sum_{i=1}^{k}\big(X_{n-i+1:n} - X_{n-k:n}\big)^2}{\left(\frac{1}{k}\sum_{i=1}^{k} X_{n-i+1:n} - X_{n-k:n}\right)^2}\,,$$

$$\sqrt{k/4}\,\big(G_n(k) - 2\big) \;\underset{n\to\infty}{\overset{d}{\longrightarrow}}\; N(0,1)\,,$$

$$\text{(4.4)} \qquad \text{[HW-test]} \qquad W_n(k) := \frac{1}{k}\left[1 - \frac{G_n(k) - 2}{1 + \big(G_n(k) - 2\big)}\right],$$

$$\sqrt{k/4}\,\big(k\,W_n(k) - 1\big) \;\underset{n\to\infty}{\overset{d}{\longrightarrow}}\; N(0,1)\,,$$

where Λ stands for a Gumbel random variable. The critical regions for testing the two-sided alternative (4.1), at a nominal size α, are given by $V_n(k) < v_{\alpha/2}$ or $V_n(k) > v_{1-\alpha/2}$, where V has to be conveniently replaced by T, R, or W and v_ε denotes the ε-quantile of the corresponding limiting distribution. The limiting distribution of $G_n(k)$ [*resp.* $W_n(k)$] shifts towards the right [*resp.* left] for distributions in the Fréchet domain of attraction ($F \in \mathcal{D}(G_\gamma)_{\gamma>0}$) and towards the left [*resp.* right] for distributions lying in the Weibull domain ($F \in \mathcal{D}(G_\gamma)_{\gamma<0}$). Notice that the test statistic S in Chaouche and Bacro (2004) may be seen as the POT-counterpart of $\big(1 - R_n(k)\big)^{-1}$. An extensive simulation study involving Ratio, Gt and HW tests, let us to perceived the following guidelines:

- The test based on the G_n^* is shown to good advantage when testing the presence of heavy-tailed distributions is in demand.
- While the Gt-test barely detects small negative values of γ, the HW is the most powerful test under study with respect to alternatives in the Weibull domain of attraction.
- The simulations have emphasized the admonition for controlling the actual size of the test to apply, keeping low within acceptable bounds the probability of incorrect rejection of the null hypothesis. Since the test based on the very simple Ratio statistic tends to be a conservative test and yet detains a reasonable power, it proves to be a valuable complement to the remainder procedures.

Testing Extreme Value conditions

From its grounds, any inferential methodology considered in the field of Extreme Values is inextricably bound to the validity of an extreme value condition. Inevitably, the methods of the previous sections do not escape such a requirement. Hence, assessing whether the hypothesis that "$F \in \mathcal{D}(G_\gamma)$" is strongly supported by the data at hand, becomes an impending problem. On this matter, Dietrich *et al.* (2002) introduce the test statistic

$$E_n(k) := k \int_0^1 \left(\frac{\log X_{n-[kt]:n} - \log X_{n-k:n}}{\hat{\gamma}_+} - \frac{t^{-\hat{\gamma}_-} - 1}{\hat{\gamma}_-} (1 - \hat{\gamma}_-) \right)^2 t^\eta \, dt \,, \tag{4.5}$$

for some $\eta > 0$, where $\hat{\gamma}_+$ and $\hat{\gamma}_-$ are the same estimators of $\gamma_+ = \max(0, \gamma)$ and $\gamma_- = \min(\gamma, 0)$ as in Dekkers *et al.* (1989). Furthermore, in case we wish to test the null hypothesis that $F \in \mathcal{D}(G_\gamma)_{\gamma \geq 0}$, a simple version is available:

$$PE_n(k) := k \int_0^1 \left(\frac{\log X_{n-[kt]:n} - \log X_{n-k:n}}{\hat{\gamma}_+} + \log t \right)^2 t^\eta \, dt \,.$$

Under extra mild condition upon the growth of k, the limit distributions of $E_n(k)$ and $PE_n(k)$ are attainable, with their specific forms being established by using an asymptotic expansion for the tail empirical quantile function due to Drees (1998). A table of critical points several values of γ is provided, although some corrections have become available in Hüsler and Li (2006). Aside from the latter, Drees *et al.* (2006) deal with the testing of extreme value conditions pertaining to $\gamma > -1/2$, via the statistic

$$T_n(k) := k \int_0^1 \left(\frac{n}{k} \, \overline{F}_n \left(\hat{a}\Big(\frac{n}{k}\Big) \frac{x^{-\hat{\gamma}} - 1}{\hat{\gamma}} + \hat{b}\Big(\frac{n}{k}\Big) \right) - x \right)^2 x^{\eta - 2} \, dx \,, \tag{4.6}$$

for some $\eta > 0$, with $\overline{F}_n = 1 - F_n$. The use ML estimators for γ and a as in Drees *et al.* (2004) is recommend, while $\hat{b}(n/k) := X_{n-k:n}$. Similarly as before, under mild restrictions upon the growth of k, the limit distribution of $T_n(k)$ is attainable and its specific form can be established using a tail approximation to the empirical distribution function. Again, tables of critical points at quite good accuracy are provided in Hüsler and Li (2006), where an exhaustive simulation study is carried out in order to draw general guidelines for the adequate specification of η in the most suitable test for the problem at hand.

Notwithstanding, if we strongly suspect we are dealing with heavy tailed phenomena, Beirlant *et al.* (2006) provide a goodness-of-fit procedure for testing the inherent Pareto-type behavior upon the tail of the underlying distribution function F.

5. AN ILLUSTRATIVE EXAMPLE

The potential of Extreme Value theory in assessing statistical models for tail-related values has gained widespread recognition in fields ranging from hydrology to insurance, finance and, more recently, in telecommunications and engineering.

As an illustrative example of methodologies embraced in the previous section, consider the 36 699 file lengths, in bytes, extracted from the Internet Traffic Archive (`http://ita.ee.lbl.gov/index.html`). In the light of extreme value analysis, the main concern here is not towards the accumulation of many file lengths, none of these being dominant (in which case the normal assumption would reasonably follow from the Central Limit Theorem), but the interest goes instead to the transmission of such huge batches of data that could possibly compromise the capacity of the system, thus making the normal distribution inadequate to describe the small set of data arising with such individual large and, therefore, dominant contributors. This same data set is analyzed in a paper by Tsourti and Panaretos (2004). Their exploratory analysis for independence seems to ascertain that an application of the testing procedures mentioned in this paper, to the available data set, will not be hindered by the pernicious effects of seasonality and clustering.

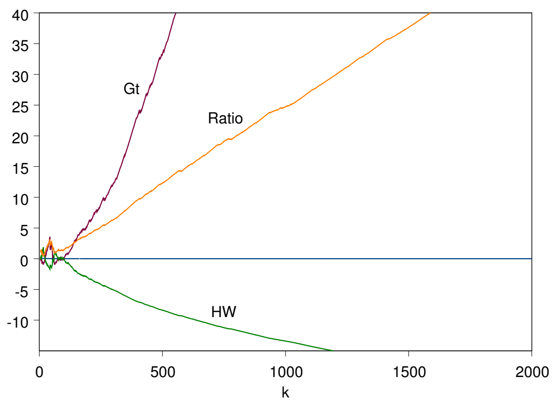

Figure 1: Plot of the sample paths returned by the three test statistics for the Gumbel domain.

Hence, we have found it reasonable to proceed with three tests in (4.2)–(4.4). The results are depicted in Figure 1. All the tests point towards a definite rejection of the null hypothesis that the underlying distribution function F belongs to the Gumbel domain. Nevertheless, the validity of condition (2.3) is still questionable. So far, we have only found evidences in the data of that F can be in any domain except for the Gumbel domain, but the question "does the underlying d.f. F belongs to any domain of attraction at all?" remains unanswered.

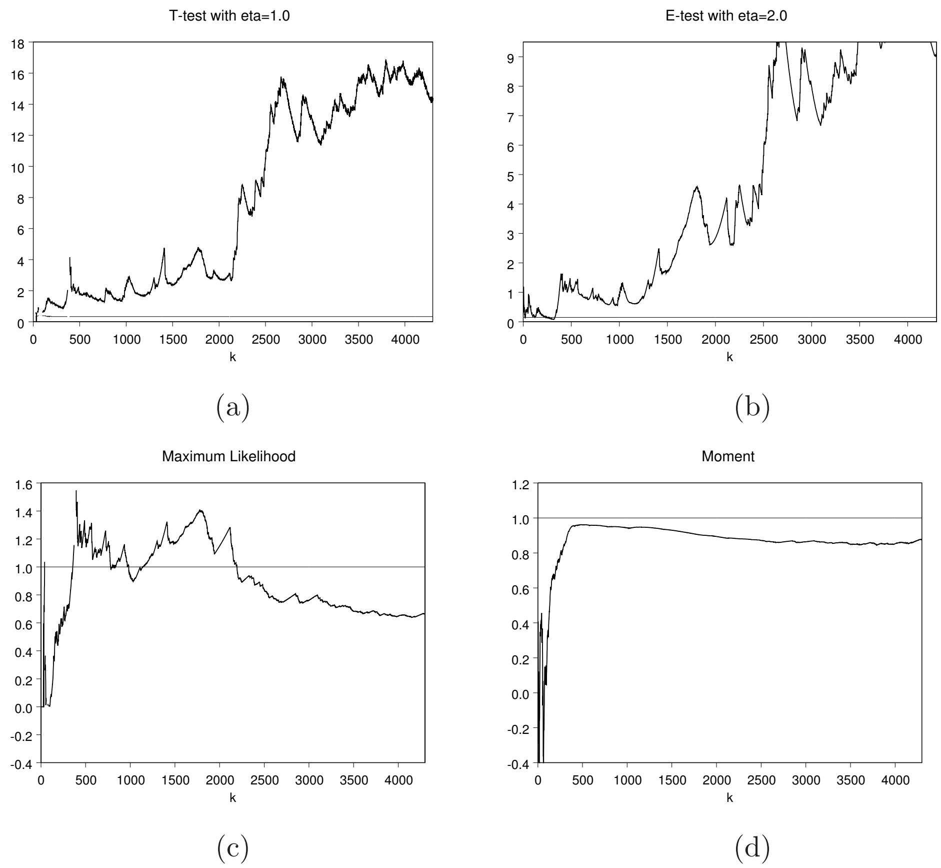

Figure 2: (a)–(b) Plot of the sample paths for the T-test and E-test statistics with accompanying critical points; (c) Plot of the ML estimates; (d) Plot of the Moment estimates.

Owing to these last remarks and following practical recommendations of Hüsler and Li (2006), we have furthermore considered application of the T-test and the E-test, given in (4.6) and (4.5), with $\eta = 1$ and $\eta = 2$, respectively. Figure 2 displays the results with respect to a significance level $\alpha = 0.05$. Although the moment estimator yields a stable plateau near $\gamma = 1$ for quite long, the conjunction of the two testing procedures seems to advise rejection of the null hypothesis on that the tail of F obeys the dictates of an extreme value law.

ACKNOWLEDGMENTS

Research partially supported by FCT/POCI 2010 and POCI — ERAS Project MAT/58876/2004.

Cláudia Neves is grateful to the organization of the Invited Paper Meeting Session on Extremes, Risk and the Environment, from the ISI 2007 — 56^{th} Session of the International Statistical Institute, which took place in Lisbon, for the opportunity to amplify her contribution to the Session in this paper.

REFERENCES

[1] Balkema, A.A. and de Haan, L. (1974). Residual life time at great age, *Ann. Probab.*, **2**, 792–804.

[2] Bardsley, W.E. (1977). A test for distinguishing between extreme value distributions, *J. Hydrology*, **34**, 377–381.

[3] Beirlant, J.; de Wet, T. and Goegebeur, Y. (2006). A goodness-of-fit statistic for Pareto-type behaviour, *J. Comput. Appl. Math.*, **186**, 99–116.

[4] Brilhante, M.F. (2004). Exponentiality versus generalized Pareto — a resistant and robust test, *RevStat*, **2**, 1–13.

[5] Castillo, E.; Galambos, J. and Sarabia, J.M. (1989). *The selection of the domain of attraction of an extreme value distribution from a set of data*. In "Extreme value theory (Oberwolfach, 1987) — Lecture Notes in Statistics 51" (J. Hüsler and R.-D. Reiss, Eds.), Springer, Berlin–Heidelberg, pp. 181–190.

[6] Castillo, E. and Hadi, A.S. (1997). Fitting the Generalized Pareto distribution to data, *J. Amer. Statist. Assoc.*, **92**, 1609–1620.

[7] Chaouche, A. and Bacro, J.-N. (2004). Fitting the Generalized Pareto distribution to data, *Comput. Statist. Data Anal.*, **45**, 787–803.

[8] Choulakian, V. and Stephens, M.A. (2001). Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, **43**, 478–484.

[9] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *J. Roy. Statist. Soc. Ser. B*, **52**, 393–442.

[10] Dekkers, A.L.M.; Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Ann. Statist.*, **17**, 1833–1855.

[11] Dietrich, D.; de Haan, L. and Husler, J. (2002). Testing extreme value conditions, *Extremes*, **5**, 71–85.

[12] Drees, H. (1998). On smooth statistical tail functionals, *Scand. J. Statist.*, **25**, 187–210.

[13] Drees, H.; de Haan, L. and Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions, *J. Statist. Plann. Inference*, **136**, 3498–3538.

[14] Drees, H.; Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Ann. Probab.*, **14**, 1179–1201.

[15] Falk, M. (1995). On testing the extreme value index via the POT-method, *Ann. Statist.*, **23**, 2013–2035.

[16] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample, *Proc. Camb. Phil. Soc.*, **24**, 180–190.

[17] Fraga Alves, M.I. (1999). Asymptotic distribution of Gumbel statistic in a semi-parametric approach, *Port. Math.*, **56**, 282–298.

[18] Fraga Alves, M.I. and Gomes, M.I. (1996). Statistical choice of extreme value domains of attraction — a comparative analysis, *Commun. Statist. — Theory Meth.*, **25**, 789–811.

[19] Galambos, J. (1982). *A statistical test for extreme value distributions.* In "Non-parametric Statistical Inference" (B.W. Gnedenko *et al.*, Eds.), North Holland, Amsterdam, pp. 221–230.

[20] Gomes, M. (1982). *A note on statistical choice of extremal models.* In "Proc. IX Jornadas Mat. Hispano-Lusas", Salamanca, pp. 653–655.

[21] Gomes, M. and Alpuim, M. (1986). Inference in a multivariate generalized extreme value model — asymptotic properties of two test statistics, *Scand. J. Statist.*, **13**, 291–300.

[22] Gomes, M.I. (1987). *Extreme value theory — statistical choice.* In "Colloq. Math. Soc. János Bolyai 45", Debrecen, pp. 195–210.

[23] Gomes, M.I. (1989). *Comparison of extremal models through statistical choice in multidimensional backgrounds.* In "Extreme value theory (Oberwolfach, 1987) — Lecture Notes in Statistics 51" (J. Hüsler and R.-D. Reiss, Eds.), Springer, Berlin–Heidelberg, pp. 191–203.

[24] Gomes, M.I. and Van Montfort, M.A.J. (1986). *Exponentiality versus generalized* P*areto-quick tests.* In "Proc. III Internat. Conf. Statistical Climatology", pp. 185–195.

[25] Greenwood, M. (1946). The statistical study of infectious diseases, *J. Roy. Statist. Soc. Ser. A*, **109**, 85–109.

[26] Grimshaw, S.D. (1993). Computing the maximum likelihood estimates for the Generalized Pareto distribution, *Technometrics*, **35**, 185–191.

[27] Hasofer, A. and Wang, J.Z. (1992). A test for extreme value domain of attraction, *J. Amer. Statist. Assoc.*, **87**, 171–177.

[28] Hosking, J. (1984). Testing whether the shape parameter is zero in the generalized extreme value distribution, *Biometrika*, **71**, 367–374.

[29] Hüsler, J. and Li, D. (2006). On testing extreme value conditions, *Extremes*, **9**, 69–86.

[30] Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quart. J. Roy. Meteo. Soc.*, **81**, 158–171.

[31] Kinnison, R. (1989). Correlation coefficient goodness-of-fit test for the Extreme-Value distribution, *Amer. Statist.*, **43**, 98–100.

[32] Konstantinides, D. and Meintanis, S.G. (2004). *A test of fit for the generalized* P*areto distribution based on transforms.* In "Proc. 3rd Conf. in Actuarial Science and Finance in Samos".

[33] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators, *Comput. Statist. Data Anal.*, **51**, 904–917.

[34] Marohn, F. (1994). *On testing the exponential and* G*umbel distribution.* In "Extreme Value Theory" (J. Galambos, Ed.), Kluwer, Dordrecht, pp. 159–174.

[35] Marohn, F. (1998a). An adaptive efficient test for Gumbel domain of attraction, *Scand. J. Statist.*, **25**, 311–324.

[36] Marohn, F. (1998b). Testing the Gumbel hypothesis via the POT-method, *Extremes*, **1**(2), 191–213.

[37] Marohn, F. (2000). Testing extreme value models, *Extremes*, **3**(4), 363–384.

[38] Neves, C. and Fraga Alves, M.I. (2007). Semi-parametric approach to Hasofer-Wang and Greenwood statistics in extremes, *Test*, **16**, 297–313.

[39] Neves, C.; Picek, J. and Fraga Alves, M.I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction, *J. Statist. Plann. Inference*, **136**(4), 1281–1301.

[40] Otten, A. and Van Montfort, M. (1978). The power of two tests on the type of distribution of extremes, *J. Hydrology*, **37**, 195–199.

[41] Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

[42] Segers, J. and Teugels, J. (2000). Testing the Gumbel hypothesis by Galton's ratio, *Extremes*, **3**(3), 291–303.

[43] Serfling, R.J. (1980). *Approximations Theorems of Mathematical Statistics*, Wiley, New York.

[44] Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591–611.

[45] Stephens, M.A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters, *Ann. Statist.*, **4**, 357–369.

[46] Stephens, M.A. (1977). Goodness-of-fit for the extreme value distribution, *Biometrika*, **64**, 583–588.

[47] Stephens, M.A. (1986). *Tests for the exponential distribution*, In "Goodness-of-Fit Techniques" (R.B. D'Agostinho and M.A. Stephens, Eds.), Marcel Dekker, New York, pp. 421–459.

[48] Tiago de Oliveira, J. (1981). *The selection of the domain of attraction of an extreme value distribution from a set of data*. In "Statistical Distributions in Scientific Work", Vol. 6 (C. Taillie, Ed.), D. Reidel, Dordrecht, pp. 367–387.

[49] Tiago de Oliveira, J. (1984). *Univariate extremes: Statistical choice*. In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, Dordrecht, pp. 91–107.

[50] Tiago de Oliveira, J. and Gomes, M.I. (1984). *Two statistics for choice of univariate extreme value models*. In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, Dordrecht, pp. 651–668.

[51] Tsourti, Z. and Panaretos, J. (2004). Extreme-value analysis of teletraffic data, *Comput. Statist. Data Anal.*, **45**, 85–103.

[52] Van Montfort, M.A.J. (1970). On testing that the distribution of extremes is of type I when type II is the alternative, *J. Hydrology*, **11**, 421–427.

[53] Van Montfort, M.A.J. and Witter, J.V. (1985). Testing exponentiality against Generalized Pareto distribution, *J. Hydrology*, **78**, 305–315.

[54] Van Montfort, M.A.J. and Witter, J.V. (1986). The Generalized Pareto distribution applied to ranfall depths, *J. Hydrology*, **31**, 151–162.

[55] von Mises, R. (1936). *La distribution de la plus grande de n valeurs*. Reprinted in Selected Papers Volumen II, American Mathematical Society, Providence, R.I., 1954, pp. 271–294.

[56] Wang, J.Z. (1995). Selection of the k largest order statistics for the domain of attraction of the gumbel distribution, *J. Amer. Statist. Assoc.*, **90**, 1055–1061.

[57] Wang, J.Z.; Cooke, P. and Li, S. (1996). Determination of domains of attraction based on a sequence of maxima, *Austral. J. Statist.*, **38**, 173–181.

[58] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.