



HAL
open science

Detecting finiteness in the right endpoint of light tailed distributions

Cláudia Neves, António Pereira

► **To cite this version:**

Cláudia Neves, António Pereira. Detecting finiteness in the right endpoint of light tailed distributions. *Statistics and Probability Letters*, 2010, 80 (5-6), pp.437. 10.1016/j.spl.2009.11.021 . hal-00612018

HAL Id: hal-00612018

<https://hal.science/hal-00612018>

Submitted on 28 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Detecting finiteness in the right endpoint of light tailed distributions

Cláudia Neves, António Pereira

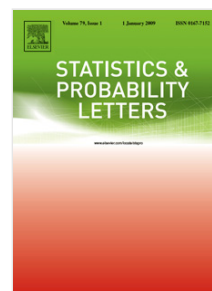
PII: S0167-7152(09)00445-3
DOI: 10.1016/j.spl.2009.11.021
Reference: STAPRO 5578

To appear in: *Statistics and Probability Letters*

Received date: 12 May 2009
Revised date: 29 September 2009
Accepted date: 28 November 2009

Please cite this article as: Neves, C., Pereira, A., Detecting finiteness in the right endpoint of light tailed distributions. *Statistics and Probability Letters* (2009), doi:10.1016/j.spl.2009.11.021

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Detecting finiteness in the right endpoint of light tailed distributions

Cláudia Neves

UIMA, Department of Mathematics
University of Aveiro, Portugal

António Pereira

CEOC, Department of Mathematics
University of Aveiro, Portugal

Abstract

In Extreme Value statistics we often encounter testing procedures for assessing the presence of the Gumbel domain, attached to the simple null hypothesis of shape parameter $\gamma = 0$, thus praising the selection of extreme domains of attraction. However, the problem of assessing for light tailed distributions with finite or infinite right endpoint is seldom referred. The latter is an impending problem of practical importance, particularly at the enrollment of subsequent estimation of extremal features such as small exceedance probabilities. In this paper, we present two test statistics whose asymptotic behavior, albeit under some restrictive yet reasonable conditions, enables to distinguish light tailed distribution functions with finite right endpoint from those with infinite endpoint lying in the Gumbel domain. An illustrative example is provided via application to significant wave height data recorded at Figueira da Foz, Portugal, from 1958 until 2001.

KEY WORDS AND PHRASES: Extreme Value Theory, right endpoint, test of hypothesis

1 Introduction

In statistical analysis of rare events, the Generalized Extreme Value distribution stems from the fundamental Fisher and Tippett theorem (Fisher and Tippett, 1928) as a unified version of the only three possible limits for the distribution of the maximum from a random sample with parent distribution function F , provided suitable normalization in scale and location, for a large enough sample size n . That is, assume there exist normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n \rightarrow \infty} P\{a_n^{-1}(\max(X_1, \dots, X_n) - b_n) \leq x\} = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad (1)$$

for all x , with some non-degenerate distribution function G . Then, G must be one of only three possible extreme value distributions - Fréchet, Gumbel or Weibull - while these, in turn, can be nested as a one parameter family of distributions (von Mises (1936)): the Generalized Extreme Value distribution,

$$G_\gamma(x) := \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & 1 + \gamma x > 0 \quad \text{if } \gamma \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbb{R} \quad \text{if } \gamma = 0. \end{cases} \quad (2)$$

We say that the distribution function F of the independent random variables $X_1, X_2, \dots, X_n, \dots$ is in the domain of attraction of G_γ if (1) holds with $G = G_\gamma$ (notation: $F \in \mathcal{D}(G_\gamma)$). The extreme value index $\gamma \in \mathbb{R}$ (featuring in (2)) is often regarded as a gauge of tail heaviness of the underlying distribution function F . For positive γ , the power-law behavior in the tail of F has important implications since it may suggest, for instance, the presence of infinite moments. The Fréchet domain of attraction $\mathcal{D}(G_\gamma)_{\gamma>0}$ encloses heavy-tailed distribution functions with polynomially decaying tails such as Pareto, Cauchy, Student's-t and the Fréchet distribution itself. All distribution functions in the Weibull domain of attraction are light-tailed with finite right endpoint. Beta distributions and, in particular, the Uniform distribution belong to the Weibull domain $\mathcal{D}(G_\gamma)_{\gamma<0}$. The intermediate case of the Gumbel domain $\mathcal{D}(G_0)$ is of particular interest in many applied sciences where extremes come into play, not only because of the simplicity of inference pertaining to $\gamma = 0$, but also for the great variety of distributions possessing an exponential tail, ranging from light-tailed distributions whether having finite or infinite right endpoint like the Normal and Gamma distributions, to moderately heavy tails such as the Lognormal distribution. In the context of extreme value analysis, where we wish to extrapolate beyond the sampled data, a screening process for the most likely underlying tail type may thus be considered an utter advantage. Statistical procedures on the selection of extreme domains are extensively used in order to have a grasp of the degree of tail heaviness. For instance, Neves and Fraga Alves (2007) and Hüsler and Peng (2008) offer a catalog of numerous procedures for assessing extreme value conditions, the first focusing different approaches within the univariate case and the second going through univariate and multivariate settings. Less attention has yet been paid to the problem of assessing the presence of a distribution function F with finite endpoint. Nevertheless, this is a striking problem of practical importance, for instance, when we wish to perform transformations to the sample observations that might hinder tail heaviness.

In this paper we confine our attention to the Weibull and Gumbel domains, that is, we are addressing distribution functions in $\mathcal{D}(G_\gamma)_{\gamma \leq 0}$. Our proposal is to devise statistical tools, at the hypothesis testing level, with the ultimate aim of distinguishing between light-tailed distributions with infinite and finite right endpoint. The latter not only embraces the problem of statistical choice between Weibull and log-Normal distributions as recently treated by Kim and Yum (2008) although under a simulation perspective, but also accounts for ramifications on the possibility of applying data transforms, particularly in the case of an underlying distribution F with infinite right endpoint x_F . A noteworthy example corresponds to

changing data from the normal parent through an exponential function, which translates as a leap from an underlying light-tailed (normal) distribution to the moderately heavy lognormal parent.

Under a semi-parametric approach, we are only interested in the shape of the underlying distribution at high quantiles, meaning that the extreme value index γ is the only parameter to keep under scrutiny. The following *extended regular variation* property (de Haan (1984)) is a well known necessary and sufficient condition for F to belong to an extreme domain of attraction, i.e., $F \in \mathcal{D}(G_\gamma)$ if and only if

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \begin{cases} \frac{x^\gamma - 1}{\gamma}, & \gamma \neq 0 \\ \log x, & \gamma = 0, \end{cases} \quad (3)$$

for every $x > 0$ and some positive measurable function a , with U standing for a quantile type function pertaining to F defined by the generalized inverse

$$U(t) := \left(\frac{1}{1 - F} \right)^{\leftarrow}(t) = \inf \left\{ x : F(x) \geq 1 - \frac{1}{t} \right\}.$$

Observe that the limit function $(x^\gamma - 1)/\gamma$ is the tail quantile function of the Generalized Pareto distribution (GPD),

$$F_\gamma(x) := 1 + \log G_\gamma(x) = 1 - (1 + \gamma x)^{-\frac{1}{\gamma}} \quad \text{for } \begin{cases} x \geq 0 & \text{if } \gamma \geq 0 \\ 0 \leq x \leq -\frac{1}{\gamma} & \text{if } \gamma < 0. \end{cases} \quad (4)$$

This fact reflects its exceptional role in extreme value theory: while restricting attention to a top portion of the original sample, the GPD stems from the seminal works of Balkema and de Haan (1974) and Pickands (1975) as the limiting distribution for the excesses $W_i = X_i - u | X_i > u$, $i = 1, \dots, k_u$ over a sufficiently high threshold u . The GPD will be taken into account in Section 3 for the performance evaluation, via simulation, of the new test statistics proposed in Section 2, whereas the exponential distribution function $F_{\gamma=0}$ will be taken as a prototype of the subclass comprising light-tailed distributions with infinite right endpoint. The remainder of the paper regards an example of application in Section 4.

2 Two tests

Let X_1, X_2, \dots, X_n be independent random variables with the same distribution function F and let $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ denote their ascending order statistics. On account of simplicity, we assume throughout that the right endpoint $x_F := U(\infty) = \lim_{t \rightarrow \infty} U(t)$ is in $(0, \infty]$. In order to deal with the problem of casting out distributions with finite right endpoint lying in the Gumbel domain and therefore attached with $\gamma = 0$, we shall take k upper order statistics in a way that these might present a satisfactory picture of the tail of $1 - F$. Hence, let $k = k(n)$ be an intermediate sequence such that $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$ while $X_{n-k,n}$ thus refers to the k -th intermediate level which constitutes the designated eventually

high (random) threshold (see text below the GPD function (4)).

In this context, the main concern goes to casting out the presence of light-tailed distributions with infinite right endpoint $x_F := U(\infty)$, in the sense that we are particularly interested in testing the null hypothesis $H_0 : (F \in \mathcal{D}(G_0) \text{ and } x_F = \infty)$ against the alternative $H_1 : (F \in \mathcal{D}(G_\gamma)_{\gamma \leq 0} \text{ and } x_F < \infty)$. To tackle this testing problem, we introduce two test statistics:

$$T_{k,n}^{(1)} := \frac{1}{k} \sum_{i=1}^k \frac{X_{n-i,n} - X_{n-k,n} - \hat{a}(n/k)}{X_{n,n} - X_{n-k,n}} \quad (5)$$

$$T_{k,n}^{(2)} := \frac{1}{k} \sum_{i=1}^{k-1} i \frac{X_{n-i+1,n} - X_{n-i,n}}{X_{n-k,n}}, \quad (6)$$

where \hat{a} is a suitable estimator for the scale function a introduced in (3). These two tests are bound to be used as complement of each other under the general guidelines provided in Section 3 regarding finite samples. Convenient normalized versions of $T_{k,n}^{(1)}$ and $T_{k,n}^{(2)}$ are provided in the main Theorem below which expounds the asymptotic normality of the referred test statistics under suitable yet reasonable conditions involving the following second order behavior of U .

Apart from the first order condition (3), we shall need a second order condition, specifying the inherent rate of convergence. Hence, we shall assume the existence of a function A , not changing in sign and tending to zero as $t \rightarrow \infty$ such that

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma, \rho}(x) := \frac{1}{\rho} \left(\frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right), \quad (7)$$

for all $x > 0$, where $\rho \leq 0$ is the second order parameter governing the speed of convergence in (3). We then say that the function U is of *second order extended regular variation* (notation: $U \in 2ERV(\gamma, \rho)$). In case $\rho = 0$ and/or $\gamma + \rho = 0$, note that the function $(x^a - 1)/a$ defined for all $x > 0$, $a \in \mathbb{R}$ reads as $\log x$ for $a = 0$. We also remark that $\lim_{t \rightarrow \infty} A(tx)/A(t) = x^\rho$, for every $x > 0$, i.e., $|A|$ is of regular variation at infinity with index ρ (cf. de Haan and Stadtmüller (1996)), hence the notation $|A| \in RV_\rho$.

Theorem 1 *Let X_1, X_2, \dots be i.i.d. random variables with the same distribution function F . Let $x_F = \infty$ and $F \in \mathcal{D}(G_0)$.*

1. *Furthermore, assume the second order condition (7) holds. If $k = k(n)$ is an intermediate sequence, i.e., $k \rightarrow \infty$ and $k = o(n)$, as $n \rightarrow \infty$, such that $\sqrt{k}A(n/k) \rightarrow 0$, as $n \rightarrow \infty$, and*

$$\sqrt{k} \left(\frac{1}{k} \sum_{i=1}^k \frac{X_{n-i,n} - X_{n-k,n}}{a(n/k)} - 1, \frac{\hat{a}(n/k)}{a(n/k)} - 1 \right) \xrightarrow{d} (B_1, B_2), \quad (8)$$

with (B_1, B_2) bivariate normal with zero mean vector and covariance matrix not depending on ρ , then

$$\sqrt{k} \log k T_{k,n}^{(1)} \xrightarrow{d} B_1 + B_2. \quad (9)$$

2. For any intermediate sequence $k = k(n)$,

$$\sqrt{k} \log(n/k) T_{k,n}^{(2)} - \sqrt{k} \xrightarrow{d} N(0, 1), \quad (10)$$

Before proceeding with the proof we call on Lemma 2 below.

Lemma 2 (Theorem B.2.2 and Corollary B.2.13 of de Haan and Ferreira (2006))

Suppose the function U is such that (3) holds with some $\gamma \leq 0$. Then, the auxiliary function a is in RV_γ and

- (i) if $\gamma = 0$ and $U(\infty) = \infty$, then U is of slow variation in the sense that $U \in RV_0$, i.e., $\lim_{t \rightarrow \infty} U(tx)/U(t) = 1$, for every $x > 0$, and $a(t) = o(U(t))$, as $t \rightarrow \infty$.
- (ii) If $\gamma = 0$ and $U(\infty) < \infty$, then $\lim_{t \rightarrow \infty} a(t)/(U(\infty) - U(t)) = 0$ and $U(\infty) - U(t) \in RV_0$.
- (iii) If $\gamma < 0$, then $U(\infty) < \infty$, $\lim_{t \rightarrow \infty} a(t)/(U(\infty) - U(t)) = -\gamma$ and $U(\infty) - U(t) \in RV_\gamma$.

Proof of Theorem 1:

We start by noting that given a random variable X with distribution function F and pertaining tail quantile function U , we have the equality in distribution $X \stackrel{d}{=} U(Y)$ with Y a standard Pareto random variable with distribution function $1 - (y \vee 1)^{-1}$. Naturally, if we consider a sample $\{X_i\}_{i=1}^n$ of independent and identically distributed random variables from the same parent F and let $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ be their ascending order statistics, we get the equality in distribution

$$\{X_{n-i,n}\}_{i=0}^{n-1} \stackrel{d}{=} \{U(Y_{n-i,n})\}_{i=0}^{n-1}. \quad (11)$$

In order to properly apply (7) and subsequent uniform inequalities in the sense of Drees (1998) (see Theorem 2.3.6 of de Haan and Ferreira, 2006), we shall bear in mind that if $k = k_n$ is an intermediate sequence of positive integers, i.e., $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$, then $(k/n)Y_{n-k,n} \xrightarrow{P} 1$ which entails in turn that $Y_{n-k,n} \xrightarrow{P} \infty$ (cf. Corollary 2.2.2 of de Haan and Ferreira, 2006). Hence, using (11), we have

$$\frac{a(n/k)}{X_{n,n} - X_{n-k,n}} \stackrel{d}{=} \frac{a(Y_{n-k,n})}{U(Y_{n,n}) - U(Y_{n-k,n})} \frac{a(n/k)}{a(Y_{n-k,n})}. \quad (12)$$

Since for $\gamma = 0$ condition (3) rephrases as

$$\frac{a(t)}{U(tx) - U(t)} = \frac{1}{\log x} - \frac{1}{\log^2 x} \left(\frac{U(tx) - U(t)}{a(t)} - \log x \right) (1 + o(1)), \quad (t \rightarrow \infty)$$

all $x > 0$, eq. (12) leads to

$$\begin{aligned} & \frac{a(n/k)}{X_{n,n} - X_{n-k,n}} \\ \stackrel{d}{=} & \frac{a(Y_{n-k,n})}{U\left(\frac{Y_{n,n}}{Y_{n-k,n}} Y_{n-k,n}\right) - U(Y_{n-k,n})} \frac{a(n/k)}{a(Y_{n-k,n})} \\ = & \left\{ \frac{1}{\log Y_{n,n} - \log Y_{n-k,n}} - A(Y_{n-k,n}) \frac{H_{0,\rho}\left(\frac{Y_{n,n}}{Y_{n-k,n}}\right)}{(\log Y_{n,n} - \log Y_{n-k,n})^2} (1 + o_p(1)) \right\} \frac{a(n/k)}{a(Y_{n-k,n})}, \quad (13) \end{aligned}$$

as $n \rightarrow \infty$. Define $Q_k = (\log Y_{k,k} - \log k)/\log k$. Owing to Rényi's representation (cf. Lemma 3.2.3 from de Haan and Ferreira, 2006), we see that with any measurable function g ,

$$\left\{ g\left(\frac{Y_{n-i+1,n}}{Y_{n-k,n}}\right) \right\}_{i=1}^k \stackrel{d}{=} \left\{ g(Y_{k-i+1,k}^*) \right\}_{i=1}^k, \quad (14)$$

where $Y_1^*, Y_2^*, \dots, Y_n^*$ are independent standard Pareto random variables with distribution function $1 - (y \vee 1)^{-1}$. Furthermore, $\lim_{k \rightarrow \infty} Q_k = 0$ a.s. (see page 194 of de Haan and Ferreira, 2006) and this takes care of the first term in the normalized version of $T_{k,n}^{(1)}$ written as follows:

$$\begin{aligned} \sqrt{k} \log k T_{k,n}^{(1)} &= \log k \frac{\hat{a}(n/k)}{X_{n,n} - X_{n-k,n}} \int_0^1 \sqrt{k} \left(\frac{X_{n-[ks],n} - X_{n-k,n}}{\hat{a}(n/k)} - \log s \right) ds \\ &= \log k \frac{a(n/k)}{X_{n,n} - X_{n-k,n}} \frac{a_0(n/k)}{a(n/k)} \left\{ \int_0^1 \sqrt{k} \left(\frac{X_{n-[ks],n} - X_{n-k,n}}{a_0(n/k)} - \log s \right) ds \right. \\ &\quad \left. - \sqrt{k} \left(\frac{\hat{a}(n/k)}{a_0(n/k)} - 1 \right) \int_0^1 \log s ds \right\} \\ &=: I \times (II - III). \end{aligned}$$

Here $[ks]$ denotes the highest integer not greater than ks . The new auxiliary function a_0 is such that $a_0(t) \sim a(t)$, as $t \rightarrow \infty$, and $a_0(t)/a(t) - 1 = o(A(t))$. It now intervenes because we wish to apply Corollary 2.4.6 of de Haan and Ferreira (2006) regarding the empirical tail quantile process. From (13) and (14) we have seen so far that

$$\begin{aligned} I &\stackrel{d}{=} \log k \frac{a(Y_{n-k,n})}{U(Y_{n,n}) - U(Y_{n-k,n})} \frac{a_0(n/k)}{a(Y_{n-k,n})} \\ &= \left\{ \frac{1}{Q_k + 1} - \frac{A(n/k)}{\log k} \frac{H_{0,\rho}(Y_{k,k})}{Q_k^2 + 2Q_k + \log k} \right\} \frac{a_0(n/k)}{a(n/k)} (1 + o_p(1)). \quad (15) \end{aligned}$$

Hence $I \xrightarrow{P} 1$, as $n \rightarrow \infty$. Note that $A(t)$ tends to zero as t goes to infinity even if $\rho = 0$. In (15) the random quantities $a(Y_{n-k,n})$ and $A(Y_{n-k,n})$ were replaced by their deterministic counterparts, sequences $a(n/k)$ and $A(n/k)$ respectively. This is sustained by the fact that both a and $|A|$ are of regular variation leading thereby to $a(Y_{n-k,n})/a(n/k) \xrightarrow{P} 1$ and $A(Y_{n-k,n})/A(n/k) \xrightarrow{P} 1$ as $n \rightarrow \infty$. Finally we have from

Corollary 2.4.6 of de Haan and Ferreira (2006) that

$$\begin{aligned} II - III &= \int_0^1 \left(\frac{W_n(s)}{s} - W_n(1) \right) ds - \alpha(W_n) \int_0^1 \log s ds \\ &\quad + \sqrt{k} A\left(\frac{n}{k}\right) \int_0^1 \rho\left(H_{0,\rho}\left(\frac{1}{s}\right) - H_{0,\rho}(1)\right) ds + o_p(1) \int_0^1 s^{-1/2-\varepsilon} ds, \end{aligned}$$

where $\{W_n(s)\}_{s>0}$ is a sequence of standard Brownian motions and α denotes a measurable functional (not depending on s) of this same sequence. Hence result in (9).

The second part of the Theorem is supported on the extended regular variation of U for $\gamma = 0$ encapsulated in (3). Define $\tilde{A}(t) := a(t)/U(t)$. Using similar arguments as in the proof of 1., by virtue of the uniform inequalities in the sense of Drees (1998) (see also Proposition B.1.10 in de Haan and Ferreira, 2006) together with (i) of Lemma 2, we get

$$\begin{aligned} T_{k,n}^{(2)} &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^{k-1} i \left(\left(\frac{U(Y_{n-i+1,n})}{U(Y_{n-k,n})} - 1 \right) - \left(\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})} - 1 \right) \right) \\ &= \tilde{A}\left(\frac{k}{n}\right) \frac{1}{k} \sum_{i=1}^{k-1} i (\log Y_{k-i+1,k}^* - \log Y_{k-i,k}^*) + o_p\left(\tilde{A}\left(\frac{k}{n}\right)\right). \end{aligned} \quad (16)$$

Now, Theorem B.3.6 of de Haan and Ferreira (2006) ensures $\tilde{A}(t) \sim (\log t)^{-1}$, as $t \rightarrow \infty$, since the (slow) regular variation of second order we get for $\gamma \leq 0$ amounts to the extended regular variation statement in (3) (cf. Theorem 1 in Neves, 2009). Moreover, Rényi's representation theorem entails that $k \log(Y_{1,k}^*) \stackrel{d}{=} \log(Y_1^*)$, independent of $i (\log(Y_{i+1,k}^*) - \log(Y_{i,k}^*)) \stackrel{d}{=} \log(Y_{k-i+1}^*)$, $i \leq k-1$ (see e.g. Shorack and Wellner, 1986, p.336). Hence,

$$\sqrt{k} \log(n/k) T_{k,n}^{(2)} = \left(\sqrt{k} \left(\frac{1}{k} \sum_{i=1}^k \log Y_i^* - 1 \right) - \sqrt{k} \left(\frac{1}{k} \log Y_1^* - 1 \right) \right) (1 + o_p(1)).$$

The result thus follows from the Central Limit theorem. \square

Remark 3 Condition (8) holds for suitable estimators of $a(n/k)$. We shall use the estimator $\hat{a}(n/k)$ related with the Moment estimator for the extreme value index γ . In this respect we refer to Dekkers et al. (1989) and Section 2.1 of de Haan and Ferreira (2006). We define

$$\hat{a}\left(\frac{n}{k}\right) := X_{n-k,n} \frac{M_{k,n}^{(1)}}{2} \left(1 - \frac{(M_{k,n}^{(1)})^2}{M_{k,n}^{(2)}} \right)^{-1}, \quad (17)$$

pertaining to the moment statistics

$$M_{k,n}^{(j)} := \frac{1}{k} \sum_{i=1}^k \left(\log X_{n-i+1,n} - \log X_{n-k,n} \right)^j, \quad j = 1, 2. \quad (18)$$

Under the conditions in Theorem 1, it is not difficult to see that with $\hat{a}(n/k)$ defined above we have

$$T_1^* := \sqrt{k} \log k T_{k,n}^{(1)} \xrightarrow{d} N(0, 1).$$

Consistency of the tests is ensured by similar arguments to the ones in the proof of Theorem 1 coupled with Lemma 2. In particular, if $x_F < \infty$ and under the conditions of Theorem 1, $T_{k,n}^{(1)} \xrightarrow{P} \infty$ if $\gamma = 0$, and $T_{k,n}^{(1)} \xrightarrow{P} -\infty$ if $\gamma < 0$. With respect to $T_{k,n}^{(2)}$, we obtain the following representation, valid for $\gamma \leq 0$, $\rho < 0$ and $x_F = U(\infty) < \infty$, and akin to (16):

$$T_{k,n}^{(2)} = \tilde{A}\left(\frac{n}{k}\right) \left(\frac{1}{k} \sum_{i=1}^k \frac{(Y_i^*)^\gamma - 1}{\gamma} - \frac{e^{\gamma \log(Y_1^*)/k} - 1}{\gamma} \right) (1 + o_p(1)) = \tilde{A}\left(\frac{n}{k}\right) (I - II) (1 + o_p(1)). \quad (19)$$

The Central Limit theorem ascertains $\sqrt{k} (I - 1/(1 - \gamma)) = O_p(1)$. By straightforward calculations we can see that $II = o_p(1/\sqrt{k})$. Hence, we only have to take into account that $|\log t \tilde{A}(t)| \in RV_\gamma$, meaning that it converges to zero as long as $\gamma < 0$, while for $\gamma = 0$ and $x_F = U(\infty) \in (0, \infty)$ we have from Theorem B.3.6 of de Haan and Ferreira (2006) that $\tilde{A}(t)/(c/U(\infty)) \rightarrow 1$, $c > 0$. Altogether, we can ascertain that $\sqrt{k} \log(n/k) T_{k,n}^{(2)}$ shifts towards left for light-tailed distributions associated with negative γ and towards right for distributions in the Gumbel domain but with finite endpoint.

3 Simulation Results

According to Theorem 1 and subsequent Remark 3, the testing procedures that statistics $T_{k,n}^{(j)}$, $j = 1, 2$, yield unfold as follows. With view towards testing the hypothesis

$$H_0 : F \in \mathcal{D}(G_0) \text{ and } x_F = \infty \quad \text{versus} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \leq 0} \text{ and } x_F < \infty,$$

take the normalized versions

$$T_1^* = \sqrt{k} \log k T_{k,n}^{(1)} \quad \text{and} \quad T_2^* = \sqrt{k} \left(\log\left(\frac{n}{k}\right) T_{k,n}^{(2)} - 1 \right).$$

For the estimation of the scale required in T_1^* we settle with the one related to the Moment estimator (see Remark 3). Under the null hypothesis, provided mild restrictions upon the growth of the intermediate sequence $k = k(n)$, Theorem 1 ascertains that T_j^* behave as standard normal random variables, eventually. Corresponding rejection regions at a significance level $\alpha \in (0, 1)$ are thus allocated by fulfilling $|T_j^*| > z_{1-\alpha/2}$, for $j = 1, 2$, where $z_\alpha = \Phi^\leftarrow(\alpha)$ denotes the α -quantile pertaining to the standard normal distribution function Φ .

The present section aims at illustrating the finite sample behavior of the proposed asymptotic test procedure by conducting Monte-Carlo simulations. In order to estimate the empirical rejection probabilities (type I error and power) we generated 5000 samples (replicates) of size $n = 1000$ from several distributions satisfying condition (7). Hence we shall be interested in the relative frequency of replicates which the null hypothesis was rejected for. Inevitably, the extreme value index γ is the most important design factor here but the second order parameter ρ has also a relevant part since it determines the speed at which T_1^* and T_2^* converge to their normal limits. The following models are taken as key examples:

- GPD ($\gamma \leq 0$), with distribution function F_γ given in (4).
- Negative Fréchet distribution ($\gamma = \rho = 0$): we say that a random variable X has a negative Fréchet distribution if $X = -Y$ with Y a Fréchet random variable, that is, the distribution function of X is given by

$$F_X(x) = 1 - \exp\{-(x_0 - x)^{-\beta}\}, \quad x < x_0, \beta > 0.$$

- Gumbel distribution ($\gamma = 0, \rho = -1$).
- Logistic distribution ($\gamma = 0, \rho = -1$), with distribution function

$$F(x) = 1 - \frac{2}{1 + \exp(x)}, \quad x > 0.$$

- Lognormal distribution ($\gamma = 0 = \rho$). A random variable Y is said to have a Lognormal distribution if $\log Y$ is a normal random variable.
- Fréchet ($\gamma > 0, \rho = -1$), with distribution function

$$F(x) = \exp\{-x^{-\frac{1}{\gamma}}\}, \quad x > 0.$$

Figure 1 displays the empirical power of the aforementioned tests, at the nominal size $\alpha = 0.05$, for the underlying GPD by retaining the $k+1 = 101$ most extreme observations. The GPD, including the particular case of the exponential distribution for $\gamma = 0$, satisfies the extreme value condition (3) but not the second order relation (7). The GPD has tail quantile function $(t^\gamma - 1)/\gamma$ and we may read the convergence in (7) as being so fast that $\rho = -\infty$ can be assumed. On the light of the information shared by Figure 1, we can ascertain superiority of the T_1^* -test as far as testing exponential-type against Beta-type models is concerned mainly because T_2^* yields too many rejections for underlying distributions other than the (exact) exponential. The latter is verified by Figure 2.

In Figure 2, we have depicted the estimated type I error of the two tests under study for the designated light tailed distributions with infinite right endpoint. The somewhat conservative behavior expounded by the test statistic T_1^* can be seen as a valuable complement to the T_2^* -test since this test unleashes high number of wrong rejections for distributions with infinite endpoint that belong to the Gumbel domain of

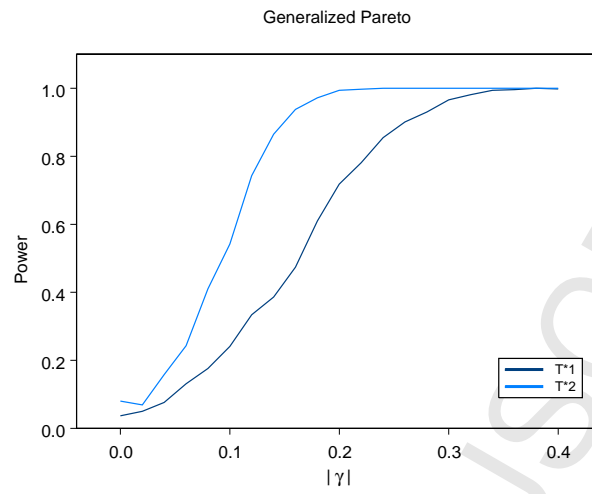


Figure 1: Empirical power of T_1^* and T_2^* with $k = 100$, at a significance level $\alpha = 0.05$, on the basis of 5000 samples of size $n = 1000$ from the $\text{GPD}(\gamma)$, plotted against $|\gamma|$, $\gamma \leq 0$.

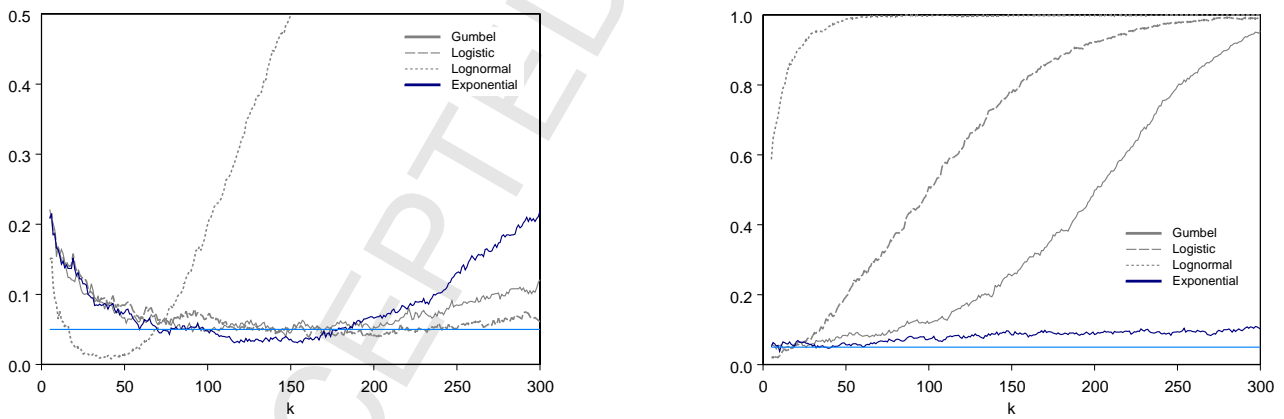


Figure 2: Estimated type I error probability of T_1^* (left) and T_2^* (right) plotted against $k = 5, 6, \dots, 300$. The solid straight line stands for the significance level $\alpha = 0.05$.

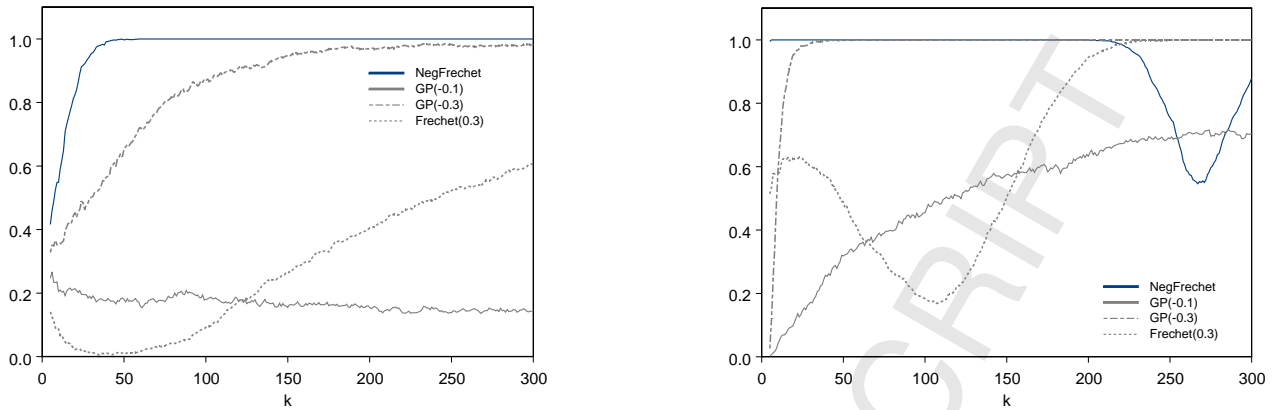


Figure 3: Estimated power of T_1^* (left) and T_2^* (right) at a significance level $\alpha = 0.05$ for GPD with $\gamma = -0.05, -0.1, -0.3$, plotted against $k = 5, 6, \dots, 300$.

attraction corresponding to small deviations from the exponential distribution. In the particular case of underlying lognormal distribution, the T_2^* -test performs rather poorly for almost all values of k , mainly because this is already a moderately heavy tailed distribution, further away from the ideal exponential model.

Figure 3 displays empirical power obtained by generating from the GPD with small negative values of γ thus closer to the null hypothesis. According to both tests, the rejection of the null hypothesis is a highly frequent event in case of the negative Fréchet (with $x_0 = \beta = 1$), just like a distribution with finite endpoint belonging to the Gumbel domain should determine. The Fréchet distribution is also included for the purpose of illustration embracing the possibility of an underlying heavy-tailed distribution. We have considered $\gamma = 0.3$ in this case which corresponds to a not so heavy-tailed distribution with finite variance. For small values of k , the T_2^* -test leads to a prompt rejection of the hypothesis of infinite right endpoint for all distributions but offers an overall erratic behavior for the Fréchet(0.3) distribution. On the other hand, T_1^* places the Fréchet distribution where it should be in the graph: below all the others by yielding the smallest relative number of rejections. However, the T_1^* -test is not very sharp in detecting very small negative values of γ while the statistic T_2^* seems to perform relatively well in this respect.

4 An illustrative example

From the enormous amount of data available, comprising 3-hourly significant wave heights (Hs), collected at a particular location in the coast off Portugal (Figueira da Foz) from 1958 up to 2001, arises the perception that the assumption of independent data may be compromised by the latent tendency for extreme conditions to linger over several observations. However, in the situation at hand, we may circumvent such

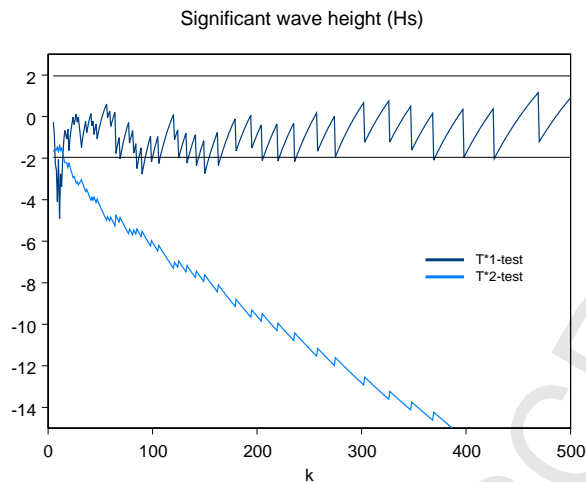


Figure 4: Sample paths of the statistics T_1^* and T_2^* with respect to the H_s data, in companion critical lines $z_{0.025} = -1.96$ and $z_{0.975} = 1.96$, all plotted against k .

difficulty in application of the tests by retaining only the k largest daily maxima above the threshold 5. This leaves us with a sample comprising $n = 1272$ independent observations. Application of the testing procedures under study yield the sample paths presented in Figure 4 which suggest the presence of a distribution detaining finite right endpoint. The test supported on T_2^* returns values in the negative direction leaning fast towards a light-tailed distribution in the Weibull domain ($\gamma < 0$). The latter comes as no surprise from the high frequency of rejections that T_2^* has accounted for in the simulations presented in Section 3: at a significance level $\alpha = 0.05$ T_1^* -test is usually a bit more reluctant in rejection of the null hypothesis in favor of the one-sided alternative pertaining to the Weibull domain. Now it yields a sample path near the lower the critical level $z_{0.025} = -1.96$ for several values of k lying in the most stable region of the graph that can be taken as intermediate and thus we can lean on our findings from the T_2^* -test .

Acknowledgments

Cláudia Neves' research was partially supported by the FCT Projects EXES-PTDC/MAT/64924/2006 and EXTREMA-PTDC/MAT/101736/2008. António Pereira's research was supported by *Center for Research on Optimization and Control (CEOC), FCT/POCI 2010*.

References

- Balkema, A.A., de Haan, L. (1974). *Residual life time at great age*. Ann. Probab., **2**, 792–804.
- 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.
- Drees, H. (1998). *On smooth statistical tail functionals*, Scand. J. Statist., **25**, 187-210.
- 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. Cambridge Philos. Soc., Mathematical Proceedings, **24**, 180-190.
- de Haan, L. (1984). *Slow variation and characterization of domains of attraction*, In: Statistical Extremes and Applications, Reidel Publishing, Dordrecht, J. Tiago de Oliveira (Editor), 31-48.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering.
- de Haan, L. and Stadtmüller, U. (1996). *Generalized regular variation of second order*, J. Australian Math. Soc., **61**, 381-395.
- Hüsler, J. and Peng, L. (2008). *Review of testing issues in extremes: in honor of Professor Laurens de Haan*, Extremes, **11**, 99-111.
- Kim, J.S. and Yum, B.-J. (2008). *Selection between Weibull and lognormal distributions: a comparative simulation study*, Comput. Statist. Data Anal., **53**, 477-485.
- Neves, C. (2009). *From extended regular variation to regular variation with application in extreme value statistics*, J. Math. Anal. Appl., **335**, 216-230.
- Neves, C. and Fraga Alves, I. (2007). *Testing extreme value conditions – an overview and recent approaches*, REVSTAT Statistical Journal, **6**, 83-100.
- Pickands III, J. (1975). *Statistical inference using extreme order statistics*. Ann. Statist., **3**, 119–131.
- Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- 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, 271-294.