A General Estimator for the Right Endpoint

Isabel Fraga Alves* and Cláudia Neves†

Abstract

In this paper, the right endpoint estimation problem is tackled via a recent estimator envisioned for distributions in the Gumbel domain, a domain of attraction induced by the extreme value theorem. The scope of this estimator is here lengthened to the case of the Weibull domain. This leads to a general estimator for the finite right endpoint that does not require the estimation of the (supposedly non-positive) extreme value index, thus unifying the problem of endpoint estimation in extreme value statistics. The asymptotic properties of the resulting general estimator are derived and its finite sample performance evaluated by means of simulations. These convey that the adopted endpoint estimator works remarkably well in case the true extreme value index stays above $-1/2$, embracing the most common cases in practical applications.

KEY WORDS AND PHRASES: Extreme value theory, semi-parametric estimation, tail estimation, theory of regular variation

*CEAUL and DEIO, FCUL, University of Lisbon
†CEAUL and University of Aveiro
1 INTRODUCTION

Extreme Value Theory (EVT) provides the adequate framework and asymptotic justification for modelling extreme values while extrapolating beyond the sample range (see e.g. de Haan and Ferreira [2006] Reiss and Thomas [2007]. Although the Central Limit Theorem prescribes the Normal distribution as the limiting distribution for the accumulation of many comparable events, this limit can be too slowly attained in the specific region of the tails of the underlying distribution function. In particular, if the underlying distribution exhibits a heavy right tail, then the sample maximum may well become (asymptotically) as important as the sum of tail-related observations (see Embrechts et al. 1997 p. 40).

The extreme value theorem (or extremal types theorem), with contributions from Fisher and Tippett (1928), Gnedenko (1943), de Haan (1970), Balkema and de Haan (1974), restricts the class of all possible limiting distribution functions to only three different types, while the induced domains of attraction embrace a great variety of distribution functions. Given that a non-degenerate limit is achieved for the maximum \( X_{n,n} \) of a sample of \( n \) independent and identically distributed random variables \( (X_1, X_2, \ldots, X_n) \), provided a suitable location-scale normalization, then the limit must be one of the following distributions: Gumbel, Fréchet or (negative) Weibull. In other words, if there exist constants \( a_n > 0, b_n \in \mathbb{R} \) such that

\[
\lim_{n \to \infty} F^n(a_n x + b_n) = G(x),
\] (1)
for all $x$, $G$ non-degenerate, then $G$ is one of the only possible three distribution functions:

$$
\Psi_\alpha(x) = \exp\{-(x)\alpha\}, \ x < 0, \ \alpha > 0,
$$

$$
\Lambda(x) = \exp\{-\exp(-x)\}, \ x \in \mathbb{R},
$$

$$
\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, \ x > 0, \ \alpha > 0.
$$

Redefining the constants $a_n > 0$ and $b_n \in \mathbb{R}$, then the above Weibull, Fréchet and Gumbel distribution functions, respectively, can be nested in a one-parameter family of distributions, the Generalized Extreme Value (GEV) distribution with distribution function

$$
G_\gamma(x) := \exp\{-(1 + \gamma x)^{-1/\gamma}\}, \ 1 + \gamma x > 0, \ \gamma \in \mathbb{R}.
$$

We then say that $F$ is in the (max-)domain of attraction of $G_\gamma$ and use the notation $F \in \mathcal{D}(G_\gamma)$. Observe that for $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$, the GEV distribution function reduces again to Weibull, Gumbel and Fréchet distribution functions, respectively. Hence, the GEV distribution is a unifying model that encompasses these three types of extreme value distributions. If $F \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$, then the distribution function $F$ is heavy-tailed, i.e., $F$ has a power-law decaying tail with infinite right endpoint. On the opposite end, $\gamma < 0$ refers to short tails which must detain a finite right endpoint. Uniform and Beta distributions are examples of distributions belonging to the Weibull domain of attraction.

The Gumbel domain of attraction $\mathcal{D}(G_0)$ renders itself a great variety of distributions, ranging from light-tailed distributions such as the Normal distribution, the exponential distribution, to moderately heavy distributions such as the Lognormal (see Embrechts et al., 1997, p. 144). Following a semi-parametric approach, the only assumption made is that the actual (and unknown) distribution function $F$ underlying the sampled data belongs to some domain of attraction, where the extreme value index $\gamma$ determines vary
INTRODUCTION

degrees of tail heaviness. Hence, under the semi-parametric setting, the main interest is on the magnitude of the parameter $\gamma$ instead of trying to fit an exact parametric model to the sampled data.

The theory of regular variation (Bingham et al., 1987; de Haan, 1970; de Haan and Ferreira, 2006), underpinning the EVT, provides necessary and sufficient conditions for $F \in D(G_\gamma)$. We shall concentrate on quantiles. Let $U$ be the tail quantile function, defined by the (generalized) inverse of $1/(1 - F)$,

$$U(t) := \left(\frac{1}{1 - F}\right)^{-1}(t) = F^{-1}\left(1 - \frac{1}{t}\right), \quad \text{for } t > 1.$$

$F \in D(G_\gamma)$ if and only if there exists a positive function $a(\cdot)$ such that the following condition of extended regular variation:

$$\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = x^\gamma - 1, \quad \text{for all } x > 0 \ (\text{notation: } U \in ERV_{\gamma}).$$

holds for all $x > 0$ (notation: $U \in ERV_{\gamma}$). From the above limiting relation, where the right hand-side coincides with the $U$-function of the Generalized Pareto distribution, we clearly see that the focus of our statistics of extremes is on the largest observations: given a sufficiently high threshold, the excesses above this threshold behave approximately as observations from the Generalized Pareto distribution (GPD), which enables us to discard the remainder lower part of the sample.

We shall base our statistical inference exclusively on a tiny top portion of the original sample, in the sense that we shall keep the number $k$ of sufficiently large observations, thus preserving the top sample fraction $k/n$, where $n$ (assumed a large number) denotes the original sample size. However, retaining only the number $k$ of upper observations means that we further need an increasing $k$ for the purpose of applying the large number
probabilistic theory. Therefore, on our quest for estimating the right endpoint of $x^F$, we shall assume the number $k$ is a sequence of positive integers $k = k_n$ such that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$. We notice that the last limit yields the above mentioned tiny top portion of the sample. We shall refer to $k = k_n$ as an intermediate sequence of positive integers, which is a common terminology in extreme value statistics. We are now in conditions to proceed by making $t = n/k$ in relation (3), so that we are actually restricting our attention to intermediate quantiles higher than $U(n/k)$ in a way that these might provide a satisfactory picture of the tail, approaching the GPD quantiles. This process for modelling the excesses above a high (random) threshold is ascertained by Balkema-de Haan-Pickands theorem (see e.g. Reiss and Thomas, 2007, p.27). The right endpoint (finite or infinite) of a distribution function $F$ is defined as

$$x^F := \sup \{ x : F(x) < 1 \} \in (0, \infty].$$

In terms of high quantiles, we notice that $x^F = \lim_{t \to \infty} U_t = U(\infty)$. This paper deals with semi-parametric estimation of the finite right endpoint of a distribution lying either in Gumbel or Weibull domains of attraction, i.e. the underlying distribution function $F$ is such that $F \in D(G_{\gamma})_{\gamma \leq 0}$ with $x^F < \infty$.

The outline of the paper is as follows. Section 2 expounds an equivalent extreme value condition to (3), devised for negative $\gamma < 0$, which is at the origin of a widely used class of endpoint estimators. Large sample results for this general endpoint estimator are presented in Section 3. Although this estimator is not asymptotically normal for all values of $\gamma < 0$, it proves to be a useful tool in terms of applications, for instance, in the case study addressed in Einmahl and Magnus (2008) aiming at the estimation of the ultimate records in several events in Athletics. We emphasise that different events or disciplines in Athletics carry different values of $\gamma \leq 0$. In this respect, de Haan and Ferreira (2006)
point out in their Remark 4.5.5 that using the sample maximum $X_{n,n}$ to estimate $x^F$ in case $\gamma < -1/2$ is approximately equivalent to using the estimator they advocate for the endpoint. We shall relate to this point later on in Section 4. Now, the proposed general estimator overcomes the nuisance of changing “tail estimation-goggles” each time we are dealing yet with another sample, aiming at extrapolating beyond the sample range on the basis of a few of the largest observed values. Section 4 contains simulation results which account for good performance of this more general endpoint estimator in many occasions. The simulation study also indicates that in the cited work by Einmahl and Magnus (2008), one could make well of use the same estimator at all times, for all the records in all events, free of any considerations about the estimated values of the extreme value index $\gamma$ and the domains of attraction endorsed by these estimates. An application of the general estimator to the long jump ultimate record has been provided by Fraga Alves et al. (2013). Finally, Section 5 encloses all the proofs pertaining to Section 3.

2 ENDPOINT ESTIMATORS

Several estimators for the right endpoint $x^F$ of a light-tailed distribution attached to an extreme value index $\gamma < 0$ are available in the literature (e.g. Hall, 1982; Cai et al., 2013; de Haan and Ferreira, 2006). These estimators often bear on the extreme value condition (3) with $x = x(t) \to \infty$, as $t \to \infty$. Since $\gamma < 0$ entails $\lim_{t \to \infty} U(t) = U(\infty)$ exists finite, then relation (3) rephrases as

$$\lim_{t \to \infty} \frac{U(\infty) - U(t)}{a(t)} = -\frac{1}{\gamma}.$$
A valid estimator for the right endpoint $x^F = U(\infty)$ thus arises by making $t = n/k$ in the approximate equality below

$$U(\infty) \approx U(t) - \frac{a(t)}{\gamma}$$

and then by replacing $U(n/k)$, $a(n/k)$ and $\gamma$ with suitable estimators, i.e.

$$\hat{x}^* = \hat{U}\left(\frac{n}{k}\right) - \frac{\hat{a}\left(\frac{n}{k}\right)}{\hat{\gamma}}$$

(cf. Section 4.5 of [de Haan and Ferreira 2006].)

There is however one estimator for the right endpoint $x^F$ that does not depend on the estimation of the extreme value index $\gamma$. This estimator, introduced in Fraga Alves and Neves (2014), has been tailored for light-tailed distributions in the Gumbel domain of attraction. The Gumbel domain of attraction assigns a null $\gamma$ thus making subsequent estimation of extreme characteristics free of any specific inference regarding the extreme value index. The study of consistency and asymptotic distribution of the announced endpoint estimator, presented in this paper, now completes the common scenario in extreme value theory, leading to a unified estimation procedure for the right endpoint in the broader sense of $\gamma \leq 0$.

Our methodology is as follows. Let $F$ be the distribution function of the random variable $X$. For simplicity we assume throughout that $X$ is non-negative and consider the $n$-th order statistics $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ resulting from the sample $X_1, \ldots, X_n$ of $n$ independent and identically distributed copies of $X$, after rearranging these by non-decreasing order. Hence the general right endpoint estimator:

$$\hat{x}^F := X_{n,n} + X_{n-k,n} - \frac{1}{\log 2} \sum_{i=0}^{k-1} \log\left(1 + \frac{1}{k+i}\right) X_{n-k-i,n}. \quad (5)$$
Despite the above estimator, recently introduced by Fraga Alves and Neves (2014), has been tailored for distributions in the Gumbel domain, we show in Section 3 that $\hat{x}_F$ is also a (strongly) consistent estimator with respect to distributions lying in the Weibull domain of attraction. Defining $a_{i,k} := \log \left( \frac{k+i+1}{k+i} \right) / \log 2$, the endpoint estimator $\hat{x}_F$ in (5) can be expressed in the equivalent form

$$\hat{x}_F := X_{n,n} + \sum_{i=0}^{k-1} a_{i,k} (X_{n-k,n} - X_{n-k-i,n}), \quad \text{with} \quad \sum_{i=0}^{k-1} a_{i,k} = 1. \quad (6)$$

From the non-negativeness of the weighted spacings in the sum in (6), it is clear that estimator $\hat{x}_F$ is greater than $X_{n,n}$, which constitutes a major advantage in the usual right endpoint estimation of a distribution belonging to the Weibull domain of attraction.

Therefore, the estimator $\hat{x}_F$ defined in (5) can be seen as a real asset in the context of semi-parametric estimation of the finite right endpoint, embracing all distributions connected by a non-positive extreme value index $\gamma$, which gains by far a broad spectrum of application to the usual alternatives.

## 3 RESULTS ON THE GENERAL ENDPOINT ESTIMATOR

This section contains a Proposition and the main Theorem of this paper, giving accounts of consistency and asymptotic distribution of the endpoint estimator defined in (5). All the proofs are postponed to Section 5.

**Proposition 1.** Suppose $x^F = U(\infty) := \lim_{t \to \infty} U(t)$ exists finite. Assume that the extended regular variation property (3) holds with $\gamma \leq 0$. If $k = kn \to \infty$, $kn/n \to 0$, as
RESULTS ON THE GENERAL ENDPOINT ESTIMATOR

\[ n \to \infty \text{ then the following almost sure convergence holds with respect to } \hat{x}_F \text{ defined in (5):} \]

\[ \hat{x}_F \xrightarrow{a.s.} x_F, \]

\[ \text{then } \hat{x}_F \text{ is a consistent estimator for } x_F < \infty, \text{ i.e. } \hat{x}_F \xrightarrow{p} x_F. \]

We now require a second order refinement of condition (3) and auxiliary second order conditions because we need to have a grasp at the speed of convergence in (3). In particular, we assume there exists a positive or negative function \( A_0 \) with \( \lim_{t \to \infty} A_0(t) = 0 \) such that for each \( x > 0 \)

\[ \lim_{t \to \infty} \frac{U(tx) - U(t)}{a_0(t)} - \frac{x^{\gamma - 1}}{\gamma} = \Psi_{\gamma, \rho}^*(x), \]

(7)

where \( \rho \) is a non-positive parameter and with

\[ \Psi_{\gamma, \rho}^*(x) := \begin{cases} \frac{x^\gamma + \rho - 1}{\gamma + \rho}, & \gamma + \rho \neq 0, \rho < 0, \\ \log x, & \gamma + \rho = 0, \rho < 0, \\ \frac{1}{\gamma} x^\gamma \log x, & \rho = 0 \neq \gamma, \\ \frac{1}{2} (\log x)^2, & \gamma = \rho = 0, \end{cases} \]

\[ a_0(t) := \begin{cases} a(t)(1 - A_0(t)), & \rho < 0, \\ a(t)(1 - A_0(t)/\gamma), & \rho = 0 \neq \gamma, \\ a(t), & \gamma = \rho = 0. \end{cases} \]

Moreover, \( |A_0| \in RV_\rho \) and

\[ \lim_{t \to \infty} \frac{a_0(tx)}{a_0(t)} - x^\gamma = x^\gamma \frac{x^\rho - 1}{\rho}, \]

(8)

for all \( x > 0 \) (cf. Theorem 2.3.3 and Corollary 2.3.5 of de Haan and Ferreira 2006). If
\( (7) \) holds with \( \gamma < 0 \) then, provided \( x = x(t) \to \infty \),

\[
\lim_{t \to \infty} \frac{U(\infty) - U(t)}{a_0(t)} + \frac{1}{\gamma} = \Psi_{\gamma,\rho}^*(\infty) := -\frac{1}{\gamma + \rho} I_{\{\rho < 0\}}
\]

(9)

by similar arguments of Lemma 4.5.4 of de Haan and Ferreira (2006), with \( I_A \) denoting the indicator function which is equal to 1 if \( A \) holds true and is equal to zero otherwise.

**Theorem 2.** Let \( F \) be a distribution function in the Weibull domain of attraction, i.e., \( F \in D(G_\gamma) \) with \( \gamma < 0 \). Suppose \( U \) satisfies condition \( (7) \) with \( \gamma < 0 \) and, in this sequence, assume that \( (9) \) holds. We define

\[
h(\gamma) := \frac{1}{\gamma} \left( \frac{2^\gamma - 1}{\gamma \log 2} + 1 \right).
\]

(10)

If the intermediate sequence \( k = k_n \) is such that \( \sqrt{k_n} A_0(n/k_n) \to \lambda \in \mathbb{R} \), then

\[
k^{\min(-\gamma,1/2)} \left( \frac{x^F - x^F}{a_0(\frac{x}{k})} - h(\gamma) \right) \xrightarrow{d} W I_{\{\gamma \geq -1/2\}} + \left( N - \lambda b_{\gamma,\rho} \right) I_{\{\gamma \leq -1/2\}},
\]

where \( W \) is a max-stable Weibull random variable, with distribution function \( \exp\{-(\gamma x)^{-1/\gamma}\} \) for \( x < 0 \), \( N \) is a normal random variable with zero mean and variance given by

\[
\text{Var}(N) = 1 + \frac{2}{\gamma \log^2 2} \left( \frac{2^{-(2\gamma + 1)}}{2\gamma + 1} + \frac{2^{-(\gamma + 1)}}{\gamma + 1} \right).
\]

(11)

and \( b_{\gamma,\rho} \) is defined as

\[
b_{\gamma,\rho} := \frac{1}{\log 2} \int_{1/2}^1 \Psi_{\gamma,\rho}^* \left( \frac{s}{2s} \right) ds = \begin{cases} \frac{1}{\gamma + \rho} \left( \frac{1}{\log 2} \frac{1 - 2^{-(\gamma + \rho)}}{\gamma + \rho} - 1 \right), & \rho < 0, \\ \frac{1}{\gamma \log 2} \left( 2^{-\gamma} (\log 2 + 1) - 1 \right), & \rho = 0. \end{cases}
\]

Moreover, the random variables \( W \) and \( N \) are independent.
Remark 3. The same normalization by \((a_0(n/k))^{-1}\), corresponding to \(\gamma = 0\), has been obtained by Fraga Alves and Neves (2014) in order to attain a Gumbel limiting distribution.

Corollary 4. Under the conditions of Theorem 2,

\[
\sqrt{k} \left( \frac{\hat{x}_F - x_F}{a_0(n/k)} - h(\gamma) \right) \xrightarrow{d} R, \quad n \to \infty
\]

where \(a^+ := \max(a,0)\) and \(R\) denotes a random variable with the following characterization:

1. Case \(-1/2 < \gamma < 0\): the random variable \(R\) is max-stable Weibull, with distribution function \(\exp\{- (\gamma x)^{-1/\gamma}\}\) for \(x < 0\), with mean \(\Gamma(1 - \gamma)/\gamma\) and variance equal to \(\gamma^{-2} (\Gamma(1 - 2\gamma) - \Gamma^2(1 - \gamma))\). Here and throughout, \(\Gamma(.)\) denotes the upper incomplete gamma function evaluated at zero, i.e. \(\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt, a > 0\).

2. Case \(\gamma < -1/2\): the random variable \(R\) has normal distribution with mean \(-\lambda b_{\gamma,\rho}\) and variance given in (11).

3. Case \(\gamma = -1/2\): the random variable \(R\) is the sum of the two cases above, taken as independent components, which yields a random part with mean \(\Gamma(1/2) - \lambda b_{-1/2,\rho} = \sqrt{\pi} - \lambda b_{-1/2,\rho}\) and variance equal to \(5 - \pi + 4(1 - (\sqrt{2} - 2)/\log 2)/\log 2\).

Remark 5. The function \(h(\gamma)\) is monotone decreasing for all \(\gamma < 0\). Taking into account the statement of Theorem 2, an adaptive reduced bias estimator based on the general estimator \(\hat{x}_F\) is given by \(\hat{x}_{RB}^F = \hat{x}_F + h(\gamma)\hat{a}(n/k)\). The dominant part comes from the scale function \(a(n/k)\) which, in case of \(\gamma = 0\), determines very slow convergence. We have conducted several simulations in this respect. These indicate that the bias correction has a very limited effect, in what can be consider as a residual improvement.
4 COMPARATIVE STUDY VIA SIMULATION

In this section, we shall study the exact properties of the general estimator $\hat{x}^F$, defined in [5], introduced by Fraga Alves and Neves (2014) and henceforth denoted by FAN. To this end, we consider four models already worked out in Girard et al. (2011, 2012), assigning different combinations of their parameters in order to obtain three distinct values of the extreme value index (EVI): $\gamma = -1/2, -1/3, -1/4$. This brief description is expounded in the following scheme:

- **Model 1**, with distribution function $F_1(x) = 1 - [1 + (-x)^{-\tau_1}]^{-\tau_2}, x < 0, \tau_1, \tau_2 > 0$. The EVI is $-1/(\tau_1\tau_2)$ and the endpoint $x_{F_1} = 0$.

- **Model 2**, with distribution function $F_2(x) = 1 - \int_{-\infty}^{\log(1-1/x)} \lambda^2 t e^{-\lambda t} dt, x < 0, \lambda > 0$. The EVI is $-1/\lambda$ and the endpoint $x_{F_2} = 0$. Moreover, $X_d = -1/(e^Z - 1)$, where $Z$ is Gamma($shape = 2, rate = \lambda$) distributed.

- **Model 3**, with distribution function $F_3(x) = 1 - [1 + (\frac{1}{x} - 1)^{-\tau_1}]^{-\tau_2}, x \in (0, 1), \tau_1, \tau_2 > 0$. The EVI is $-1/(\tau_1\tau_2)$ and the endpoint is $x_{F_3} = 1$.

- **Model 4**, with distribution function $F_4(x) = 1 - \int_{-\infty}^{\log(1-x)} \lambda^2 t e^{-\lambda t} dt, x \in (0, 1), \lambda > 0$. The EVI is $-1/\lambda$ and the endpoint is $x_{F_4} = 1$. Then $X_d = -1 - e^{-Z}$, where $Z$ is Gamma($shape = 2, rate = \lambda$) distributed.

At a first glance, the four models considered may suggest that a modest simulation study was undertaken. These are in fact taken as key examples from a comprehensive set of simulations we have conducted under the Weibull domain of attraction.

The finite-sample performance the general estimator FAN is here compared with naive maximum estimator $X_{n:n}$ (notation: MAX) and and the estimator $\hat{x}^*$ evolving from [4] by replacing $\hat{\gamma}$ and $\hat{a}(n/k)$ with moment related estimators (notation: MOM), which is
given by
\[ \hat{x}^* := X_{n-k,n} - \frac{X_{n-k,n}M_{n,k}^{(1)}(1 - \hat{\gamma}_{n,k})}{\hat{\gamma}_{n,k}}, \]
(12)
(see Section 4.5 of [de Haan and Ferreira 2006] and references therein) where the celebrated
Moment estimator for \( \gamma \in \mathbb{R} \) is defined as
\[ \hat{\gamma}_{n,k}^M := M_{n,k}^{(1)} + \hat{\gamma}_{n,k}^-, \quad \hat{\gamma}_{n,k}^- := 1 - \frac{1}{2} \left\{ 1 - \frac{\left( \frac{M_{n,k}^{(1)}}{M_{n,k}^{(2)}} \right)^2}{\left( \frac{M_{n,k}^{(1)}}{M_{n,k}^{(2)}} \right)^2} \right\}^{-1} \]
and the \( r \)-moment is
\[ M_{n,k}^{(r)} := \frac{1}{k} \sum_{i=0}^{k-1} \left\{ \log \left( \frac{X_{n-i,n}}{X_{n-k,n}} \right) \right\}^r, \quad r = 1, 2. \]

We notice that these three estimators, MAX, FAN and MOM, are all based on a certain
number \( k^* \) of top order statistics (o.s.), and do not require the knowledge of the entire
sample, contrary to what happens with the endpoint estimators proposed in [Girard et al. 2011, 2012]. More precisely, MAX depends only on the first top extremal o.s. \( (k^* = 1) \),
MOM and FAN are functions of the \( k^* = k + 1 \) and \( k^* = 2k \) intermediate o.s., respectively.
It should be highlighted that in many practical applications there exist only records of
the largest observations, as it is the case of the records in sports (see e.g. [Einmahl and
Magnus 2008]), for which a very small amount of top observations is known. Inspired in the
numeric examples in [Girard et al. 2011, 2012], we have generated \( N = 500 \) replications
of a random sample with size \( n = 1000 \). Then the average \( L^1 \)-error
\[ E(k^*) := \frac{1}{N} \sum_{j=1}^{N} |\varepsilon(j, k^*)|, \quad \text{where} \quad \varepsilon(j, k^*) := \hat{x}_{k^*}^F(j) - x^F, \quad k^* \leq n \]
was obtained, where \( x_{k^*}^F(j) \) denotes the FAN endpoint estimator computed at the \( j \)-th
replicate. Afterwards, the “optimal” values of \( k^* \), \( k^*_0 := \text{argmin}\{E(k^*), \; k^* \leq n\} \), were
### Table 1: Average $L^1$-errors

| Model | MAX  | FAN  | MOM  |
|-------|------|------|------|
| Model 1 ($x^F = 0$) |
| $(\tau_1, \tau_2) = (2, 1)$ | 0.028 | 0.013 | 0.018 |
| $(\tau_1, \tau_2) = (3, 1)$ | 0.090 | 0.026 | 0.045 |
| $(\tau_1, \tau_2) = (4, 1)$ | 0.161 | 0.037 | 0.075 |
| Model 2 ($x^F = 0$) |
| $\lambda = 2$ | 0.009 | 0.005 | 0.006 |
| $\lambda = 3$ | 0.044 | 0.016 | 0.020 |
| $\lambda = 4$ | 0.101 | 0.029 | 0.034 |
| Model 3 ($x^F = 1$) |
| $(\tau_1, \tau_2) = (2, 1)$ | 0.027 | 0.011 | 0.039 |
| $(\tau_1, \tau_2) = (3, 1)$ | 0.082 | 0.022 | 0.064 |
| $(\tau_1, \tau_2) = (4, 1)$ | 0.136 | 0.066 | 0.089 |
| Model 4 ($x^F = 1$) |
| $\lambda = 2$ | 0.009 | 0.005 | 0.013 |
| $\lambda = 3$ | 0.042 | 0.014 | 0.042 |
| $\lambda = 4$ | 0.089 | 0.023 | 0.077 |

computed for each model with parameters set at convenient values. Similarly for the remainder estimators MAX and MOM. Since the estimator MAX does not depend on any parameter, the associated average $L^1$-error $E$ is constant. The resulting $E(k_0^*)$ is displayed in Table 1 for each one of the 12 cases generated. In all situations, the estimator FAN yields better results than the MAX or MOM estimators.

The comparison of the exact behavior of the adopted endpoint estimators (at the “optimal” threshold $k_0^*$ for MOM and FAN) is furthermore evaluated in terms the of boxplots of the associated errors $\varepsilon(j, k_0^*)$, $j = 1, \ldots, N$. These graphical tools, displayed in Figures 1 and 2 lead to the conclusion that the MAX estimator always underestimates the true value of the right endpoint $x^F$, as expected. MOM and FAN estimators present a different performance at the optimal level $k_0^*$, showing lower and higher values comparable with the true $\hat{x}^F$. Moreover, FAN estimates are not so spread out as MOM values, both
COMPARATIVE STUDY VIA SIMULATION

computed at level $k_0^*$, revealing greater variability.

Next, we draw comparison between the adopted endpoint estimators under the four models. In what follows, the EVI is set at the values $\gamma = -1/2, -1/4$. The main reason for excluding $\gamma = -1/3$ from the subsequent simulation study is that it delivers a snapshot of what happens in between, which is still in our grasp by screening the plots from top to bottom. The boxplots in Figures 1 and 2 already suggested this possibility. For instance, the outliers marked in these boxplots seems to move from lower to larger values of optimal bias $\varepsilon(\cdot, k_0^*)$ as we progress in $\gamma$. Now, Figures 3 to 4 display the average $L^1$-error as functions of the number $k^*$ of upper o.s. used in the estimation, while corresponding mean squared errors (MSE) are depicted in Figures 5 and 6. From these results it is clear that the MOM endpoint estimator presents large variability in upper part of the sample, contrasting with the small variance of the proposed FAN estimator. On the other hand, both MOM and FAN estimators show increasing $L^1$-error with increasing of $k^*$, a common feature in extreme semi-parametric estimators. Again, the more general endpoint estimator FAN returns values with lower $E(k^*)$ (i.e. average $L^1$-error) and estimated MSE for a top range of thresholds determined by $k^*$.

For all the distributions in the Weibull max-domain considered in the present simulation study, it is clear that the general endpoint estimator FAN tends to surpass the reminder two estimator involved in the study, both in terms of absolute bias and variability. This is particularly true for values of $\gamma$ closer to zero, as it could be expected by the motivation of the estimator in the Gumbel domain of attraction. Furthermore, the FAN estimator seems to work remarkably well under a fairly negative EVI, given that this is a general estimator which does not accommodate any specific information about the true value $\gamma$. In case of $\gamma = -1/2$, the $E(k^*)$ returned by the MOM estimator tends to approach the MAX-yields, which seems to verify the statement in Remark 4.5.5 of
Figure 1: Model 1 (left) and Model 2 (right) $x^{F_1} = x^{F_2} = x^F = 0$: Boxplots of $\varepsilon(j, k^*_0)$, $j = 1, \ldots, N$, for MAX, MOM and FAN estimates.
Figure 2: Model 3 (left) and Model 4 (right); $x^F_3 = x^F_4 = x^F = 1$: Boxplots of $\varepsilon(j,k_0^*), j = 1, \ldots, N$, for MAX, MOM and FAN estimates.
Figure 3: Model 1 (left) and Model 2 (right); $x^{F_1} = x^{F_2} = x^F = 0$: $E(k^*)$ plotted against $k^* \leq n/5$.

[Graphs showing $E(k^*)$ for different models and $k^*$ values]

De Haan and Ferreira (2006) that using the sample maximum $X_{n,n}$ to estimate $x^F$ in case $\gamma < -1/2$ is asymptotically equivalent to using the MOM estimator. The overall performance of the MOM endpoint estimator is clearly damaged by its variance.
Figure 4: Model 3 (left) and Model 4 (right); $x^F_3 = x^F_4 = x^F = 1$: $E(k^*)$ plotted against $k^* \leq n/5$.

![Model 3 and Model 4 graphs](image1)

Figure 5: Models 1 (left) and 2 (right): mean squared error (MSE) as function of the upper o.s. $k^*$, $k^* \leq n$.

![Models 1 and 2 graphs](image2)
Figure 6: Models 3 (left) and 4 (right): mean squared error (MSE) as function of the upper o.s. $k^*, k^* \leq n$.

5 PROOFS

This section is entirely dedicated to the proofs of the results introduced in Section 3. In what follows we find more convenient to consider the estimator $\hat{x}^F$ in the functional form

$$\hat{x}^F = X_{n-k,n} - \frac{1}{\log 2} \int_0^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{ds}{s},$$

(13)

where $[a]$ denotes the integer part of $a \in \mathbb{R}$ (more details about the representation (13) can be obtained in Fraga Alves and Neves [2014]).

We note that if $s \in [0, 1/(2k)]$, then the integral in (13) is equal to zero. Bearing this in mind, we write

$$\hat{x}^F = X_{n-k,n} - \frac{1}{\log 2} \int_0^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{ds}{s}.$$
Moreover, if \( s \in [1/(2k), 1/k] \) then \([ks] = 0\) (not depending on \( s \)) and thus \( X_{n-[ks],n} = X_{n,n} \). Therefore, we have that

\[
\hat{x}^F = X_{n-k,n} - \frac{1}{\log 2} \left\{ \int_{\frac{1}{2k}}^{1} X_{n-[2ks],n} \frac{ds}{s} - X_{n,n} \int_{\frac{1}{2k}}^{1} \frac{ds}{s} + \int_{\frac{1}{2k}}^{1} (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{ds}{s} \right\}.
\]

(14)

With a suitable variable transform on the last integral, we can reassemble (14) in a tidy manner:

\[
\hat{x}^F = X_{n,n} + X_{n-k,n} - \frac{1}{\log 2} \int_{\frac{1}{2k}}^{1} X_{n-[2ks],n} \frac{ds}{s}.
\]

(15)

This is the main algebraic expression that will be used to derive the asymptotic distribution of \( \hat{x}^F \) in the proof of Theorem 2, which is a natural consequence of the three random contributions in (15).

**Proof of Proposition 1** We see that the integral in the functional form (15) satisfies the inequalities

\[
(\log 2) X_{n-k,n} \leq \int_{\frac{1}{2k}}^{1} X_{n-[2ks],n} \frac{ds}{s} \leq (\log 2) X_{n-2k,n}.
\]

Therefore, we obtain the following upper and lower bounds involving \( \hat{x}^F - x^F \),

\[
X_{n,n} - x^F \leq \hat{x}^F - x^F \leq (X_{n,n} - x^F) + X_{n-k,n} - X_{n-2k,n},
\]

and the result thus follows easily because the three order statistics \( X_{n,n}, X_{n-k,n} \) and \( X_{n-2k,n} \) all converge almost surely to \( x^F \), provided the intermediate nature of \( k = k_n \).

**Remark 6.** The result in Proposition 1 admits a general alternative proof based on the functional form (6) of the \( k^* := 2k \) top order statistics. For the estimator defined in (6),
strong consistence is assured by lower and upper bounds: on the one hand,

\[
\hat{x}^F - x^F = (X_{n,n} - x^F) + \left( X_{n-k,n} - \sum_{i=0}^{k-1} a_{i,k} X_{n-k-i,n} \right)
\geq (X_{n,n} - x^F) + \left( X_{n-k,n} - X_{n-k,n} \sum_{i=0}^{k-1} a_{i,k} \right) = X_{n,n} - x^F
\]

and on the other hand,

\[
\hat{x}^F - x^F \leq (X_{n,n} - x^F) + \left( X_{n-k,n} - X_{n-2k+1,n} \sum_{i=0}^{k-1} a_{i,k} \right) \
= (X_{n,n} - x^F) + (X_{n-k,n} - X_{n-2k+1,n});
\]

any intermediate \( k = k_n \) the order statistics \( X_{n,n}, X_{n-k,n}, X_{n-2k,n} \) converge almost surely to \( x^F \), the result follows.

Before getting under way to the proof of the main theorem, we need to lay down some ground results. These comprise a Proposition and a Lemma regarding the case \( \gamma < 0 \), along with a further development of the condition of regular variation introduced in Section 3.

**Proposition 7.** Suppose \( X_{n,n} \) is the maximum of a random sample whose parent distribution function \( F \) detains finite right endpoint of \( F \), i.e. \( x^F = U(\infty) < \infty \). Assume the second order condition (7) holds with \( \gamma < 0 \). If \( k = k_n \) is such that, as \( n \to \infty \), \( k \to \infty \), \( k/n \to 0 \) and \( \sqrt{kA_0(n/k)} \to \lambda \in \mathbb{R} \), then

1. for \( \gamma \geq -1/2 \), for each \( \varepsilon > 0 \),

\[
k^{-\gamma-\varepsilon} \left| \frac{X_{n,n} - x^F}{a_0(\frac{n}{k})} \right| \overset{p}{\to} 0. \tag{16}
\]
Moreover,
\[ k^{-\gamma} X_{n,n} - x^F \frac{d}{a_0 \left( \frac{n}{k} \right)} \xrightarrow{d} Z^\gamma, \]

where \( Z \) denotes a standard Fréchet with distribution function \( \Phi_1 \) as in (2).

2. for \( \gamma < -1/2 \),
\[ \sqrt{k} \left| X_{n,n} - x^F \right| \frac{d}{a_0 \left( \frac{n}{k} \right)} \xrightarrow{p} 0. \]

Proof: Owing to the well-known equality in distribution that \( X_{i,n} \xrightarrow{d} U(Y_{i,n}) \), \( i = 1, 2, \ldots, n \), with \( \{Y_{i,n}\}_{i=1}^n \) the \( n \)-th order statistics from a sample of \( n \) independent random variables with common (standard) Pareto distribution function given by \( 1 - x^{-1} \), \( x \geq 1 \), then the following equality in distribution holds:
\[ \frac{X_{n,n} - x^F}{a_0 \left( \frac{n}{k} \right)} \xrightarrow{d} \left\{ \frac{U \left( \frac{k}{n} Y_{n,n} \right) - U \left( \frac{n}{k} \right)}{a_0 \left( \frac{n}{k} \right)} + \frac{1}{\gamma} \right\} - \left\{ \frac{U(\infty) - U \left( \frac{n}{k} \right)}{a_0 \left( \frac{n}{k} \right)} + \frac{1}{\gamma} \right\}. \]

Now we use conditions (7) and (9) with \( t \) replaced by \( n/k \) everywhere:
\[ \frac{X_{n,n} - x^F}{a_0 \left( \frac{n}{k} \right)} \xrightarrow{d} \left\{ \frac{k^\gamma \left( n^{-1} Y_{n,n} \right)^\gamma - 1}{\gamma} + A_0 \left( \frac{n}{k} \right) \Psi_{\gamma,\rho} \left( \frac{k}{n} Y_{n,n} \right) \left( 1 + o_p(1) \right) \right\}
- \left\{ A_0 \left( \frac{n}{k} \right) \Psi_{\gamma,\rho} (\infty) \left( 1 + o(1) \right) \right\}
= \frac{k^\gamma \left( n^{-1} Y_{n,n} \right)^\gamma}{\gamma} + A_0 \left( \frac{n}{k} \right) \left\{ \Psi_{\gamma,\rho} \left( \frac{k}{n} Y_{n,n} \right) + \frac{1}{\gamma + \rho} I_{\{\rho < 0\}} \right\} + o_p \left( A_0 \left( \frac{n}{k} \right) \right) \]

We note at this stage that \( n^{-1} Y_{n,n} \) is asymptotically a Fréchet random variable with distribution function given by \( \Phi_1 \) in (2). This non-degenerate limit yields \( (k/n) Y_{n,n} \) going to infinity with probability one, which implies in turn that \( \Psi_{\gamma,\rho} \left( k \left( n^{-1} Y_{n,n} \right) \right) \to - (\gamma + \rho)^{-1} I_{\{\rho < 0\}} \) as \( n \to \infty \). Therefore, we obtain for \( \gamma \geq -1/2 \),
\[ k^{-\gamma} \frac{X_{n,n} - x^F}{a_0 \left( \frac{n}{k} \right)} \xrightarrow{d} \frac{(n^{-1} Y_{n,n})^\gamma}{\gamma} + o_p \left( k^{-\gamma-1/2} \right), \quad (17) \]
by virtue of $\sqrt{k}A_0(n/k) = O(1)$, and (16) thus follows directly for each $\varepsilon > 0$. The second part in point 1. is ensured from (17) by the continuous mapping theorem. For $\gamma < -1/2$, we observe from (17) that

$$\sqrt{k} \frac{X_{n,n} - x_F}{a_0\left(\frac{n}{k}\right)} = k^{1/2 + \gamma} \left(\frac{n^{-1} Y_{n,n}}{\gamma}\right) + o_p(1).$$

Since we are addressing the case $\gamma + 1/2 < 0$, the fact that $n^{-1} Y_{n,n}$ converges in distribution to a Fréchet random variable suffices to conclude the proof.

Lemma 8. Given the conditions of Theorem 2, (18) converges in distribution to a bivariate normal $(P, Q)$ (of independent components) with zero mean and covariance matrix with entries

$$\text{Var}(P) = \frac{2}{\gamma} \left(\frac{2^{-(2\gamma+1)} - 1}{2\gamma + 1} + \frac{2^{-(\gamma+1)} - 1}{\gamma + 1}\right),$$

$$\text{Cov}(P, Q) = 0,$$

$$\text{Var}(Q) = 1.$$

Proof: The first component in (18) shall be tackled by Theorem 2.4.2 of de Haan and Ferreira (2006) with $k$ replaced by $2k$ therein, together with the second order conditions...
This yields for the finite integral:

\[
\sqrt{2k} \int_{1/2}^{1} \frac{X_{n-|2ks|,n} - U\left(\frac{n}{2ks}\right)}{a_0\left(\frac{n}{k}\right)} \, ds
= \frac{a_0\left(\frac{n}{2k}\right)}{a_0\left(\frac{n}{k}\right)} \sqrt{2k} \int_{1/2}^{1/2} \left\{ \frac{X_{n-|2ks|,n} - U\left(\frac{n}{2ks}\right)}{a_0\left(\frac{n}{2k}\right)} - U\left(\frac{n}{2ks}\right) \right\} \, ds
= \frac{1}{2} \int_{1/2}^{1} \left\{ s - \gamma^{-1} W_n(s) + o_p(1) s^{-\gamma - 1/2 - \varepsilon} + O\left(\frac{1}{s} \right) \right\} \, ds
+ O_p\left(A_0\left(\frac{n}{k}\right)\right),
\]

where \(\{W_n(s)\}_{n \geq 1}, s > 0\), denotes a sequence of Brownian motions. Under the assumption that \(\sqrt{k} A_0(n/(2k)) = O(1)\), we thus obtain

\[
\sqrt{k} \int_{1/2}^{1} \frac{X_{n-|2ks|,n} - U\left(\frac{n}{2ks}\right)}{a_0\left(\frac{n}{k}\right)} \, ds = \frac{1}{\sqrt{2}} \int_{1/2}^{1} (2s)^{-\gamma} W_n(s) \, ds + O_p\left(A_0\left(\frac{n}{k}\right)\right) + o_p(1) \quad (n \to \infty).
\]

We note that the integral above corresponds to the sum of asymptotically multivariate normal random variables. The second component of the random vector (18) is asymptotically standard normal (cf. Theorem 2.4.1 of de Haan and Ferreira, 2006). Finally, the independence between the two components of \((P, Q)\) comes from the equality in distribution \(X_{n-i,n} \stackrel{d}{=} U(Y_{n-i,n}), i = k, k + 1, \ldots, 2k\), upon which the argument of independence between \(\{Y_{n-i,n}/Y_{n-k,n}\}_{i=k+1}^{2k}\) and \(Y_{n-k,n}\) suffices (see p. 73 of Arnold et al., 2008). This completes the proof.

**Proof of Theorem 2** Let \(h(\gamma) = (\log 2)^{-1} \int_{1/2}^{1} \left\{ (2s)^{-\gamma} - 1 \right\} / (-\gamma) \, ds/s\), which has been defined in (10). Taking the auxiliary function \(a_0\) from the second order condition (7) we write the following normalization of \(\hat{x}^F\) (cf. (15)):

\[
\frac{\hat{x}^F - x^F}{a_0\left(\frac{n}{k}\right)} - h(\gamma) = R_1 - \frac{1}{\log 2} R_2 + R_3 - \frac{1}{\log 2} \int_{1/2}^{1} \left( \frac{U\left(\frac{n}{2ks}\right) - U\left(\frac{n}{2k}\right)}{a_0\left(\frac{n}{2k}\right)} - \frac{(2s)^{-\gamma} - 1}{\gamma} \right) \, ds.
\]
with
\[
R_1 := \frac{X_{n,n} - x^F}{a_0 \left( \frac{n}{k} \right)},
\]
\[
R_2 := \int_{1/2}^{1} \frac{X_{n-[2ks],n} - U \left( \frac{n}{2ks} \right)}{a_0 \left( \frac{n}{k} \right)} \frac{ds}{s},
\]
\[
R_3 := \frac{X_{n-k,n} - U \left( \frac{n}{k} \right)}{a_0 \left( \frac{n}{k} \right)}.
\]

Lemma 8 entails that $\sqrt{k}(R_2, R_3)$ is asymptotically bivariate normal, i.e. for the remainder proof, we shall bear in mind that $R_2$ and $R_3$ are of order $k^{-1/2}$. Proposition 7 expounds the limiting distribution of $R_1$ provided suitable normalization, possibly different than $\sqrt{k}$. Hence, the crux of the proof is in the following distributional expansion, under the second order condition (7), for large enough $n$:

\[
k^{-\gamma} \left( \frac{x^F - x^F}{a_0 \left( \frac{n}{k} \right)} - h(\gamma) \right) = k^{-\gamma} R_1 + k^{-(\gamma+1/2)} \left\{ \sqrt{k} R_3 - \frac{\sqrt{k}}{\log 2} \left( R_2 + A_0 \left( \frac{n}{k} \right) \int_{1/2}^{1} \Psi_{\gamma, \varphi} \left( \frac{1}{2s} \right) \frac{ds}{s} \right) \right\}.
\]

(19)

We shall consider the cases $\gamma > -1/2$, $\gamma = -1/2$ and $\gamma < -1/2$ separately.

**Case** $\gamma > -1/2$: Proposition (7) and Lemma 8 upon (19) ascertain the result, by virtue that $W = Z^\gamma / \gamma$ with $Z$ a standard Fréchet random variable.

**Case** $\gamma = -1/2$: The random components $R_1$ and $R_2$ are asymptotic independent. This claim is supported on Lemma 21.19 of van der Vaart (1998). Again, the combination of Proposition 7 with Lemma 8 ascertains the result.

**Case** $\gamma < -1/2$: It is more convenient to rephrase (19) with a suitable normalization in view of Proposition 7 and the precise statement thus follows:

\[
\sqrt{k} \left( \frac{x^F - x^F}{a_0 \left( \frac{n}{k} \right)} - h(\gamma) \right) = \sqrt{k} \left\{ R_3 - \frac{R_2}{\log 2} - A_0 \left( \frac{n}{k} \right) \frac{1}{\log 2} \int_{1/2}^{1} \Psi_{\gamma, \varphi} \left( \frac{1}{2s} \right) \frac{ds}{s} \right\} + O_p(k^{\gamma+1/2}).
\]
Proof of Corollary 4. The result follows immediately from Theorem 2 given that the random variables $W$ and $N$ are independent.

Acknowledgment

Cláudia Neves’ research is partially supported by DEXTE–Development of Extremes in Time and Space, FCT–project EXPL/MAT-STA/0622/2013 and by national funds through the Fundação Nacional para a Ciência e Tecnologia, Portugal–FCT under the project PEst-OE/MAT/UI0006/2014.

Isabel Fraga Alves’ research is partially supported by national funds through the Fundação Nacional para a Ciência e Tecnologia, Portugal–FCT under the project PEst-OE/MAT/UI0006/2014.

References

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (2008). A First Course in Order Statistics. SIAM - Society for Industrial and Applied Mathematics, Philadelphia, second edition.

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

Bingham, N. H., Goldie, C., and Teugels, J. L. (1987). Regular Variation. Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press.
Cai, J. J., de Haan, L., and Zhou, C. (2013). Bias correction in extreme value statistics with index around zero. *Extremes*, 16:173–201.

de Haan, L. (1970). On regular variation and its application to the weak convergence of sample extremes. Mathematisch Centrum Amsterdam.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.

Einmahl, J. H. J. and Magnus, J. R. (2008). Records in athletics through extreme-value theory. *JASA*, 103:1382–1391.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Math. Proc. Cambrige Philos. Soc.*, 24:180–190.

Fraga Alves, M. I., de Haan, L., and Neves, C. (2013). How far can Man go? In Torelli, N., Pesarin, F., and Bar-Hen, A., editors, *Advances in Theoretical and Applied Statistics*. Springer, pages 185–195.

Fraga Alves, M. I. and Neves, C. (2014). Estimation of the finite right endpoint in the Gumbel domain. *Statist. Sinica*, 24:1811–1835.

Girard, S., Guillou, A., and Stupfler, G. (2011). Estimating an endpoint with high order moments. *Test*, 21:697–729.

Girard, S., Guillou, A., and Stupfler, G. (2012). Estimating an endpoint with high order moments in the weibull domain of attraction. *Statist. Probab. Lett.*, 82:2136–2144.
Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Ann. of Math.*, 44:423–453.

Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.*, 10:556–568.

Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser Verlag.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.