Extreme Value Theory and Statistics of Univariate Extremes: A Review

M. Ivette Gomes
Universidade de Lisboa, FCUL, DEIO and CEAUL

Armelle Guillou
Université de Strasbourg et CNRS, IRMA UMR 7501, France

Abstract

Statistical issues arising in modeling univariate extremes of a random sample have been successfully used in the most diverse fields, such as biometrics, finance, insurance and risk theory. Statistics of univariate extremes (SUE), the subject to be dealt with in this review paper, has recently faced a huge development, partially due to the fact that rare events can have catastrophic consequences for human activities, through their impact on the natural and constructed environments. In the last decades there has been a shift from the area of parametric SUE, based on probabilistic asymptotic results in extreme value theory, towards semi-parametric approaches. After a brief reference to Gumbel's block methodology and more recent improvements in the parametric framework, we present an overview of the developments on the estimation of parameters of extreme events and on the testing of extreme value conditions under a semi-parametric framework. We further discuss a few challenging topics in the area of SUE.

Key words: statistics of univariate extremes; parametric and semi-parametric approaches; extreme value index and tail parameters; testing issues.

1 The role of extremes in society and scope of the paper

Statistics of univariate extremes (SUE) has been successfully used in the most diverse fields, such as finance, insurance and risk theory, where the *value-at-risk* (VaR) at any level p, the size of the loss that occurred with small probability p, and the adjustment coefficient, a rudimentary measure of risk in a collective of insurance risks, are important parameters of extreme or even rare events. Also, in fields like biology and environment, the Weibull tail coefficient, the regular variation coefficient of the inverse failure rate function, probabilities of exceedance of high levels and endpoints of underlying models (life-time of human beings, ultimate records in the field of athletics) are relevant extreme events' parameters or functionals. Statistical problems in all these areas have direct ethical, social, economic and environmental impact and this is one of the reasons why statistics of extremes, in general, and SUE, in particular, have faced a huge development in the last decades. Indeed, rare events can have catastrophic consequences for human activities, through

their impact on the natural and constructed environments. The recent development of sophisticated methodology for the estimation and prediction of functionals of rare events has contributed to save endangered natural resources, to model climate, earthquakes and other environmental phenomena, like precipitation, temperature and floods, situations where we have to deal with large risks or with very low probabilities of over (under)-passing a high (low) level. From a theoretical point of view, the key result obtained by Fisher & Tippett (1928) on the possible limiting laws of the sample maxima of a random sample $(X_1, \ldots, X_n)$, of *independent and identically distributed* (IID) *random variables* (RVs), formalized by Gnedenko (1943), and used by Gumbel (1958) for applications of *extreme value theory* (EVT) in engineering subjects, are some of the key tools that led to the way statistical EVT has been exploding in the last decades. The statistical applications of EVT gave emphasis to the relaxation of the independence condition, to the consideration of multivariate and spatial frameworks and to a deeper and deeper use of regular variation and point process approaches. These topics are well-documented in the books by David (1970), Galambos (1978), Leadbetter *et al.* (1983), Resnick (1987), Arnold *et al.* (1992), Falk *et al.* (1994), Embrechts *et al.* (1997), Reiss & Thomas (1997), Coles (2001), David & Nagaraja (2003), Beirlant *et al.* (2004), Castillo *et al.* (2005), de Haan & Ferreira (2006), Resnick (2007), Markovich (2007), and their posterior editions. For an overview of most of the topics in this field see the recent volumes of *Extremes* **11**:1 (2008) and *Revstat* **10**:1 (2012). Among review papers in the area of SUE, we mention Gomes *et al.* (2007b; 2008a), Neves & Fraga Alves (2008), Hüsler & Peng (2008), Beirlant *et al.* (2012) and Scarrot & MacDonald (2012).

In Section 2 of this review paper, we provide some details related to the non-degenerate limiting behaviour of the sequence of maximum values, other top order statistics (OSs) and excesses over high thresholds. Section 3 is dedicated to the most common parametric models in SUE and to a discussion of the more recent semi-parametric frameworks. Estimation procedures are discussed in sections 4, 5 and 6. In Section 7 we briefly discuss SUE for censored data. Section 8 is dedicated to a brief reference to testing issues either under parametric or semi-parametric frameworks. In Section 9 we make a brief reference to the estimation of the extremal index. Finally, in Section 10, we mention a few other relevant topics in the area and a few important open problems that we think useful to be dealt with in the nearby future.

2 Probabilistic backgrounds of EVT—limiting structure of maxima, top OSs and excesses over high thresholds

The field of EVT goes back to 1927, when Fréchet (1927) formulated the functional equation of stability for maxima, which later was solved with some restrictions by Fisher & Tippett (1928), and finally by Gnedenko (1943) and de Haan (1970). Let us assume that we have access to a sample, $(X_1, \ldots, X_n)$ of IID, or even stationary and weakly dependent RVs from an underlying model F, and let us denote by $(X_{1:n} \leq \cdots \leq X_{n:n})$ the sample of associated ascending OSs.

2.1 Limiting behaviour of the sequence of maximum values

Let us further assume that it is possible to linearly normalize the sequence of maximum values, $\{X_{n:n}\}_{n \geq 1}$, so that we get a non-degenerate limit for the sequence $(X_{n:n} - b_n)/a_n$, with $a_n > 0$ and

$b_n \in \mathbb{R}$. Then, Gnedenko's *extremal types theorem* (ETT) assures us that such a limiting RV has a cumulative distribution function (CDF) of the type of the *extreme value distribution* (EVD), given by

$$\mathrm{EV}_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right), \ 1+\xi x \geq 0, & \text{if} \quad \xi \neq 0, \\ \exp(-\exp(-x)), \ x \in \mathbb{R}, & \text{if} \quad \xi = 0, \end{cases} \tag{2.1}$$

and ξ is the so-called *extreme value index* (EVI), the primary parameter in SUE. We then say that F is in the max-domain of attraction (MDA) of EV_ξ, in (2.1), and use the notation $F \in \mathcal{D}_\mathcal{M}(\mathrm{EV}_\xi)$. The parameter ξ measures essentially the weight of the right tail-function (RTF),

$$\overline{F} := 1 - F.$$

If $\xi < 0$, the right-tail is short and light, i.e. $x^F := \sup\{x : F(x) < 1\}$, the *right endpoint* of F, is finite. This class is called the Weibull class and contains among others the uniform and reverse Burr CDFs. If $\xi > 0$, the right-tail is heavy, of a negative polynomial type, and F has an infinite *right endpoint*. Examples in this class (Fréchet class) are the Pareto, Burr, Student's, α-stable ($\alpha < 2$) and loggamma CDFs. If $\xi = 0$, the right-tail is of an exponential type, and the *right endpoint* can then be either finite or infinite. This class (Gumbel class) encompasses the exponential, normal, lognormal, gamma and classical Weibull CDFs, with an infinite right endpoint, but also models with a finite right endpoint, like $\overline{F}(x) = C \exp(-\delta/\{x^F - x\})$, for $x < x^F$, $\delta > 0$, and $C > 0$.

In Gnedenko's pioneering paper and also nowadays in many applications of EVT, the EVD, in (2.1), is often rewritten on the three domains of attractions as follows:

$$\begin{cases} \text{Weibull}: & \Psi_\xi(x) = \exp\left(-(-x)^{-1/\xi}\right), \ x \leq 0 & (\xi < 0), \\ \text{Fréchet}: & \Phi_\xi(x) = \exp\left(-x^{-1/\xi}\right), \ x \geq 0 & (\xi > 0), \\ \text{Gumbel}: & \Lambda(x) = \exp\left(-\exp(-x)\right), \ x \in \mathbb{R} & (\xi = 0). \end{cases} \tag{2.2}$$

The CDFs, either in (2.1) or in (2.2), are appropriate when the data consist of a set of maxima.

2.2 Joint limiting behaviour of top OSs

Apart from the ETT and the already mentioned EVD, in (2.1), it is also sensible mentioning the *multivariate* EVD (MEVD), related to the limiting distribution of the k largest values $X_{n-i+1:n}$, $1 \leq i \leq k$, also called the *extremal process* (Lamperti, 1964; Dwass, 1964), with associated probability density function (PDF)

$$h_\xi(x_1, x_2, \ldots, x_k) = \mathrm{ev}_\xi(x_k) \prod_{j=1}^{k-1} \frac{\mathrm{ev}_\xi(x_j)}{\mathrm{EV}_\xi(x_j)} \quad \text{if} \quad x_1 > \cdots > x_k, \tag{2.3}$$

where

$$\mathrm{ev}_\xi(x) = \frac{d\mathrm{EV}_\xi(x)}{dx} = \begin{cases} (1+\xi x)^{-1/\xi-1} \exp\left(-(1+\xi x)^{-1/\xi}\right), \ 1+\xi x \geq 0, & \text{if} \quad \xi \neq 0, \\ \exp(-x) \exp(-\exp(-x)), \ x \in \mathbb{R}, & \text{if} \quad \xi = 0. \end{cases}$$

2.3 Limiting structure of the excesses over a high threshold

Just as mentioned above, rather than the maxima, we can consider all values larger than a given threshold. The differences between these values and a given threshold are called *exceedances over the threshold*. These exceedances are typically assumed to have a $\mathrm{GP}_\xi(x/\sigma) =: \mathrm{GP}_{\xi,\sigma}(x)$ CDF, $\sigma > 0$, with GP_ξ the *generalized Pareto distribution* (GPD), defined by

$$\mathrm{GP}_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi}, \ 1 + \xi x > 0, \ x > 0, & \text{if } \ \xi \neq 0, \\ 1 - \exp(-x), \ x > 0, & \text{if } \ \xi = 0. \end{cases} \tag{2.4}$$

This distribution is generalized in the sense that it subsumes certain other distributions under a common parametric form. In the $\mathrm{GP}_{\xi,\sigma}(x)$ CDF, ξ is the important shape parameter of the distribution and σ is an additional scaling parameter.

The distribution of excesses over a high threshold u is defined to be

$$F_u(x) := \mathbb{P}\Big(X - u \leq x \Big| X > u\Big),$$

for $0 \leq x \leq x^F - u$. Depending on the field of applications, different names for F_u arise, for instance *excess-life* or *residual lifetime* in reliability or medical statistics, *excess-of-loss* in an insurance framework. According to Balkema & de Haan (1974) and Pickands (1975), the key result in EVT which explains the importance of the GPD is that $F \in \mathcal{D}_M(\mathrm{EV}_\xi)$ iff

$$\lim_{u \to x^F} \sup_{x \in [0, x^F - u)} \Big| F_u(x) - \mathrm{GP}_\xi(x/\sigma(u)) \Big| = 0$$

for some positive function $\sigma(u)$. Thus, the GPD is the natural model for the unknown excess distribution above sufficiently high thresholds. This approach based on the GPD approximation is called the *peaks-over-threshold* (POT) approach.

2.4 First, second and higher-order frameworks

We further use the notations

$$F^{\leftarrow}(y) := \inf \{x : F(x) \geq y\}, \quad y \in (0, 1),$$

for the *generalized inverse* function of F, $\mathcal{R}_a$ for the class of *regularly varying* functions at infinity with an index of regular variation a, i.e. positive Borel measurable functions $g(\cdot)$ such that $g(tx)/g(t) \to x^a$, as $t \to \infty$, for all $x > 0$ (see Bingham *et al.*, 1987, for details on regular variation), and

$$U(t) := F^{\leftarrow}\big(1 - 1/t\big), \quad t > 1,$$

for the tail quantile function (TQF).

A necessary and sufficient condition for $F \in \mathcal{D}_M(\mathrm{EV}_\xi)$, provided in de Haan (1984), is the following

$$F \in \mathcal{D}_M(\mathrm{EV}_\xi) \iff \lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \varphi_\xi(x) =: \begin{cases} \frac{x^\xi - 1}{\xi}, & \text{if } \ \xi \neq 0, \\ \ln x, & \text{if } \ \xi = 0, \end{cases} \tag{2.5}$$

for all $x > 0$, where a is a positive measurable function. Condition (2.5) is known as the first-order condition. For instance, in the Fréchet MDA, we can choose $a(t) = \xi U(t)$ and thus $F \in \mathcal{D}_M(\Phi_\xi)$ iff $\lim_{t \to \infty} U(tx)/U(t) = x^\xi$, $x > 0$, which means $U \in \mathcal{R}_\xi$ or equivalently $\overline{F} \in \mathcal{R}_{-1/\xi}$.

However a first-order condition is in general not sufficient to study properties of tail parameters' estimators, in particular asymptotic normality. In that case a second-order condition is required specifying the rate of convergence in (2.5). Different types of conditions exist, some expressed in terms of either $\overline{F}$ or U or $\ln U$. The most common one is the following

$$\lim_{t \to \infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \varphi_\xi(x)}{A(t)} = \begin{cases} \frac{1}{\rho}\left(\frac{x^\rho-1}{\rho} - \ln x\right), & \text{if } \xi = 0, \ \rho \neq 0, \\ \frac{1}{\xi}\left(x^\xi \ln x - \frac{x^\xi-1}{\xi}\right), & \text{if } \xi \neq 0, \ \rho = 0, \\ \ln^2 x/2, & \text{if } \xi = \rho = 0, \\ \frac{1}{\rho}\left(\frac{x^{\xi+\rho}-1}{\xi+\rho} - \frac{x^\xi-1}{\xi}\right), & \text{otherwise,} \end{cases} \tag{2.6}$$

$x > 0$, where $\rho \leq 0$ is a second-order parameter, A is a function possibly not changing in sign and tending to 0 as $t \to \infty$, such that $|A| \in \mathcal{R}_\rho$ (de Haan & Stadtmüller, 1996). Note that both in (2.5) and (2.6) we could have provided a simpler limit by using the interpretation of the Box-Cox function as the logarithm when the power equals 0. Again the rate of convergence in (2.6) can also be specified in a third-order condition (see Gomes *et al.*, 2002; Fraga Alves *et al.*, 2003a,b, 2006). Other limiting results in EVT, important for the development of parametric approaches to SUE can be found in Gomes *et al.* (2008a).

3 Models in SUE

Statistical inference about rare events is clearly linked to observations which are extreme in some sense. There are different ways to define such observations, and such definitions lead to different alternative approaches to SUE.

3.1 Gumbel's approach or block maxima method

When the sample size $n \to \infty$, and due to the limiting result given before for the normalized sequence of maximum values, i.e. the ETT, we can write

$$\mathbb{P}[X_{n:n} \leq x] = F^n(x) \approx \text{EV}_\xi\left((x - \lambda_n)/\delta_n\right), \tag{3.1}$$

with $\text{EV}_\xi(x)$ given in (2.1) and $(\lambda_n, \delta_n) \in (\mathbb{R}, \mathbb{R}^+)$ an unknown vector of location and scale parameters, that replaces the attraction coefficients (b_n, a_n) in the normalized sequence of maximum values, $(X_{n:n} - b_n)/a_n$.

Remark 3.1. *Any result for maxima (top OSs) can be easily reformulated for minima (low OSs). Indeed, $\min_{1 \leq i \leq n} X_i = -\max_{1 \leq i \leq n}(-X_i)$, and consequently, when $n \to \infty$,*

$$\mathbb{P}[X_{1:n} \leq x] = 1 - (1 - F(x))^n \approx \text{EV}_\xi^*\left((x - \lambda_n')/\delta_n'\right),$$

with $\text{EV}_\xi^*(x) = 1 - \text{EV}_\xi(-x) = 1 - \exp\left\{-(1-\xi x)^{-1/\xi}\right\}$, for $1 - \xi x > 0$ and $(\lambda_n', \delta_n') \in (\mathbb{R}, \mathbb{R}^+)$.

The aforementioned ETT was used by Gumbel in several papers which culminated in his 1958 book, to give approximations of the type of the one provided in (3.1) but for any of the models in (2.2). He suggested the first model in SUE, usually called the *annual maxima* or *block maxima* (BM) *model* or the EV *univariate model* or merely *Gumbel's model*. Under *Gumbel's model*, the sample of size n is divided into k sub-samples of size r (usually associated to k years, with $n = r \times k$, and r reasonably large). Next, the maximum of the r observations in each of the k sub-samples is considered, and one of the extremal models in (2.2), obviously with extra unknown location and scale parameters, is fitted to the sample of those k maximum values. Nowadays, whenever using this approach, still quite popular in environmental sciences, it is more common to fit to the data an EVD, $\mathrm{EV}_\xi((x - \lambda_r)/\delta_r)$, with EV_ξ given in (2.1), $(\lambda_r, \delta_r, \xi) \in (\mathbb{R}, \mathbb{R}^+, \mathbb{R})$ unknown location, scale and 'shape' parameters. All statistical inference is then related to the above mentioned models.

3.2 Multivariate and multi-dimensional EV approaches: the method of largest observations

Although Gumbel's statistical procedure has proved to be fruitful in the most diverse situations, several criticisms have been made on Gumbel's technique, and one of them is the fact that we are wasting information when using only observed maxima and not further OSs, if available, because they certainly contain useful information about the RTF underlying the data. On the other hand, in most areas of application there is no natural seasonality of the data, and in such framework the method of sub-samples is subjective and artificial.

To infer on the right tail weight of the underlying model, it seems sensible to think of a small number k of top OSs from the original data. Indeed, if we have daily data, some years may have several values among those top OSs (that are for sure relevant to make inference upon the RTF), and other years may contain none of those top values. We can thus say that such an approach provides additional information, that has been disregarded in the traditional Gumbel's methodology.

This approach depends on the joint limiting distributional behaviour of those top OSs. When the sample size n is large and for fixed k, it is sensible to consider, on the basis of the PDF h_ξ defined in (2.3), the approximation,

$$\mathbb{P}[X_{n:n} \leq x_1, \cdots, X_{n-k+1:n} \leq x_k] \approx H_\xi\Big(\frac{x_1 - \lambda_n}{\delta_n}, \ldots, \frac{x_k - \lambda_n}{\delta_n}\Big),$$

where H_ξ is the MEV CDF associated with h_ξ, and λ_n and δ_n are unknown location and scale parameters, respectively, to be estimated on the basis of the k top OSs in the sample of size n. This approach to SUE is the so-called MEV model or *largest observations* (LOB) method or *extremal process*. Under this approach it is easier to increase the number k of observations, contrarily to what happens in Gumbel's approach, where a larger number n of original observations is usually needed. Such an approach has been introduced first, in a slightly different context, by Pickands (1975) and was used by Weissman (1978) and Gomes (1978, 1981).

Note finally that it is easy to combine both approaches. In each of the sub-samples associated to Gumbel's classical approach, we can collect a few top OSs modelled through a MEV model, and next consider the *multidimensional* EV *model*, based on the multivariate sample,

$$(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_k), \quad \text{where } \underline{X}_j = (X_{1j}, \ldots, X_{i_jj}), \ 1 \leq j \leq k,$$

are MEV vectors.

3.3 The POT approach

Another approach to SUE, in a certain sense parallel to the MEV model, is the one in which we restrict our attention only to the observations that exceed a certain high *threshold u*, fitting the appropriate statistical model to the *excesses* over such a high level u. From the results in Section 2, we get the approximation

$$\mathbb{P}\left[X - u \leq x | X > u\right] \approx \text{GP}_\xi(x/\sigma(u)),$$

with $\text{GP}_\xi(\cdot)$, the GPD in (2.4). We are then led to consider a deterministic high level u and work with the excesses. The adequate model is then the GPD. Such a model is the so-called *Paretian excesses model* or POT *model*, and was introduced in Smith (1987a). Here, all statistical inference is related to the GPD. Despite the existence of an extra scale parameter $\sigma = \sigma(u)$, the POT model can also be regarded as a semi-parametric model whenever we work in the MDA of EV_ξ.

3.4 Bayesian approaches

It has recently become more and more common to use Bayesian methods within EV analysis, as can be seen in the monographs by Reiss & Thomas (2001; 2007), by Coles (2001), by Beirlant *et al.* (2004), and references therein.

3.5 Summary of parametric approaches and a link to semi-parametric frameworks

More recently, the LOB and the POT methodologies have been considered under a semi-parametric framework. There is then no fitting of a specific parametric model, dependent upon a location parameter λ, a scale parameter δ and a 'shape' parameter ξ. It is merely assumed that $F \in \mathcal{D}_\mathcal{M}(\text{EV}_\xi)$, with EV_ξ given in (2.1), ξ being the unique primary parameter of extreme events to be estimated, on the basis of a few top observations, and according to adequate methodology, to be dealt with in Section 5. We now summarise the different approaches to SUE here discussed:

1. **Parametric approaches**:

   I The *univariate* EV *model* (for the k maximum values of sub-samples of size r, $n = r \times k$.) (Gumbel's classical approach or BM method).

   II The MEV model or LOB method (for the k top OSs associated with the original sample of size n).

   III The *multi-dimensional* EV *model* (MEV model for the i_j top observations, $j = 1, 2, \ldots, m$, in sub-samples of size r, $m \times r = n$).
   $m = k$ and $i_j = 1$ for $1 \leq j \leq k$ originates I;
   $m = 1$ ($r = n$) and $i_1 = k$ originates II.

   IV The *Paretian model* for the excesses, $X_j - u$, given $X_j > u$, $1 \leq j \leq k$, of a high deterministic threshold u, suitably chosen [POT approach].

   V *Bayesian* approaches.

2. **Semi-parametric approaches**:

VI Under these approaches we work with the k top OSs associated with all n observations or with the excesses over a high deterministic or random threshold, assuming only that the model F underlying the data is in $\mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi)$ or in specific sub-domains of $\mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi)$, with $\mathrm{EV}_\xi(\cdot)$ provided in (2.1). The POT approach can thus be considered under this framework.

4 Estimation under parametric frameworks

We have now several 'R-*Packages for Extreme Values*', such as **evd**, **evdbayes**, **evir**, **ismev**, **extRemes**, **extremevalues**, **fExtremes**, **lmom**, **lmomRFA**, **lmomco**, **POT** and **SpatialExtremes**, among others, that can help us in most of the inferential procedures given below.

4.1 Gumbel's approach or block maxima method

Computational details on maximum likelihood (ML) estimation of (λ, δ, ξ) in the EV model, $\mathrm{EV}_\xi((x - \lambda)/\delta)$, with $\mathrm{EV}_\xi(x)$ given in (2.1), can be found in Prescott & Walden (1980, 1983), Hosking (1985), Smith (1985) and Macleod (1989), among others.

As the ML estimators can be numerically difficult to handle, several alternative methods have been proposed for the estimation of (λ, δ, ξ). The *probability weighted moment* (PWM) method, introduced by Landwehr *et al.* (1979) and Greenwood *et al.* (1979) is an interesting alternative to the ML approach. The main idea of this method is to match the moments

$$\mathbb{E}\left[X^p(F(X))^r(1 - F(X))^s\right], \quad \text{with } p, r \text{ and } s \text{ real numbers,}$$

with their empirical versions, similarly to the classical method-of-moments. For the EVD, Hosking *et al.* (1985) show that $\mathbb{E}\left[X(F(X))^r\right]$ can be explicitly computed, which leads to the PWM estimation of the parameters under play and, just like the ML approach, implies some restriction on the value of ξ.

Several other estimation methods for the EVD can be found in the literature. Among them we mention: best linear unbiased estimation (Balakrishnan & Chan, 1992); method of moments (Christopeit, 1994); minimum distance estimation (Dietrich & Hüsler, 1996).

Robust methods for the EVD have been studied in Dupuis & Field (1998), who derived B-optimal robust M-estimators for the case that the observations follow an EVD. Modifications of the ML estimator are presented by Coles & Dixon (1999), who suggest penalised ML (PML) estimators, showing that PML estimation improves the small-sample properties of a likelihood-based analysis. The BM method has been recently revisited in Dombry (2013) and Ferreira & de Haan (2013).

4.2 Multivariate and multi-dimensional EV approaches: the method of largest observations

For estimation procedures under these MEV approaches, see Weissman (1984), Smith (1986), Tawn (1988) and references therein. ML estimators of the unknown parameters in the *multidimensional EV model* have been studied in Gomes (1981). See also Smith (1984) and references therein. The use of concomitants of OSs, to deal with statistical inference techniques in this model, appears

in Gomes (1984, 1985a). A comparison of the MEV model and the *multi-dimensional* EV *model* is performed in Gomes (1985b, 1989a). Discrimination among MEV models can also be found in Fraga Alves (1992).

4.3   The POT approach

ML estimates of ξ and σ in a GPD, non-regular for $\xi < -1/2$, have been studied in Smith (1987a). A survey of the POT methodology, together with several applications can be found in Davison & Smith (1990).

Just as for the EVD, as the ML estimators can be numerically hardly tractable (see Grimshaw, 1993), there have been several methods, other than ML, proposed for the estimation of ξ and σ. Hosking & Wallis (1987) suggest the use of PWM estimators (see also the comparative study of Singh & Guo, 1997). Castillo & Hadi (1997) have proposed estimators based on the elemental percentile method (EPM). PML estimators, containing a penalty function for the shape parameter, are presented by Coles & Dixon (1999) and Martins & Stedinger (2000). The PML estimator combines the flexibility of the ML estimator and the robustness of the PWM estimator. See also Resnick (1997), Crovella & Taqqu (1999) and references therein. The estimates depend significantly on the choice of the threshold and several authors, among whom we mention McNeil (1997) and Rootzén & Tajvidi (1997), explicitly state that the selection of an appropriate threshold u, above which the GPD assumption is appropriate, is a difficult task in practice.

Robust estimation for the GPD was first addressed by Dupuis (1998), who provides the optimally-biased robust estimator for the GPD and suggests a validation mechanism to guide the threshold selection. Peng & Welsh (2001) use the method of medians introduced by He & Fung (1999), and get estimators of the unknown parameters of the GPD with bounded influence functions. Juárez & Schucany (2004) implement the minimum density power divergence estimator (MDPDE), originally introduced by Basu *et al.* (1998), for the shape and scale parameters of the GPD. The MDPDE is indexed by a non-negative constant that controls the trade-off between robustness and efficiency. Frigessi *et al.* (2002) suggest an unsupervised alternative to the classical POT model, where a GPD is fitted beyond a threshold which is selected in a supervised way. They suggest to model the data with a dynamical mixture: one term of the mixture is a GPD and the other is a light-tailed density distribution. The weight of the GPD component is predominant for large values, and takes the role of threshold selection. A recent comparison between the BM approach (with several block-sizes, not only annual maximum values) and the POT approach has been performed by Engeland *et al.* (2004).

4.4   Bayesian approaches

We begin with some examples of the most direct use of Bayesian methodology within EV analysis: Smith & Naylor (1987) compare Bayesian and ML estimators for the Weibull distribution; Ashour & El-Adl (1980), Lingappaiah (1984), Achcar *et al.* (1987) and Engelund & Rackwitz (1992) consider estimation for specific extremal types for maxima; Lye *et al.* (1993) consider estimation of the EVD; Pickands (1994) and de Zea Bermudez & Amaral-Turkman (2003) consider estimation of the GPD.

Smith (1999) discusses predictive inference aspects of Bayesian and frequentist approaches, and

de Zea Bermudez *et al.* (2001) propose the use of a Bayesian predictive approach for the choice of the threshold, through a hierarchical Bayesian model. Stephenson & Tawn (2004) claim that in practice the appropriate asymptotically motivated extremal model, either EVD or GPD, fitted to data that can be regarded as maxima or exceedances of a high threshold, reduces the Gumbel (exponential) type to a single point in the parameter space, and consequently the Gumbel (exponential) model is never selected. They then decide to incorporate knowledge of the structure of the ETT into inference for the EVD and the GPD. To do this, they associate the probability p_ξ to the parameter subspace corresponding to the Gumbel (exponential) type. This approach requires an inference scheme which allows switching between the full EVD (GPD) and the Gumbel (exponential) sub-model. They then perform inference using reversible jump MCMC techniques. This Bayesian approach recognises the possibility that the data can come from any of the three extremal types. As a by-product of the analysis, posterior probabilities for $\mathbb{P}(\xi > 0)$, $\mathbb{P}(\xi = 0)$ and $\mathbb{P}(\xi < 0)$ are obtained.

Diebolt *et al.* (2005) propose a quasi-conjugate Bayesian inference approach for the GPD with $\xi > 0$, through the representation of a heavy-tailed GPD as a mixture of an exponential and a gamma distribution. For other papers on Bayesian approaches to, for instance, high quantile estimation, see, e.g., Coles & Powell (1996) and Coles & Tawn (1996), who provide a detailed review of Bayesian methods in EV modelling up to this date. See also Reiss & Thomas (1999), Walshaw (2000), Smith & Goodman (2000), Bottolo *et al.* (2003), Tancredi *et al.* (2006), among others.

5 Semi-parametric EVI-estimation

Just as mentioned above, under a semi-parametric framework, we do not need to fit a specific parametric model based on scale, shape and location parameters, but only to assume that F is in the MDA, $\mathcal{D}_M(\text{EV}_\xi)$, for a suitable index ξ. Then we construct an estimator for this index based on the $k-$largest observations in the sample $X_1, ..., X_n$, where k needs to be an intermediate sequence, i.e. such that $k \to \infty$, $k = o(n)$ as $n \to \infty$. We start by reviewing the most classical estimators proposed in the literature. Then we give their main asymptotic properties and briefly discuss reduced-bias estimation and the choice of k.

5.1 Classical Estimators

Hill estimator. The most famous estimator of $\xi > 0$ is the Hill (1975) estimator

$$\widehat{\xi}_{k,n}^H = \frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n}, \tag{5.1}$$

where $k = k(n) \to \infty$ in an appropriate way, so that an increasing sequence of upper OSs is used. One of the interesting facts concerning (5.1) is that various asymptotically equivalent versions of $\widehat{\xi}_{k,n}^H$ can be derived through essentially different methods (such as the ML method or the mean excess function approach), showing that the Hill estimator is very natural. This estimator is based on the assumption that the RTF is of Zipf or Pareto form for large x, i.e. $1 - F(x) \sim Cx^{-1/\xi}$ as $x \to \infty$, for some $\xi > 0$ and $C > 0$. Hence, Hill's estimator is only applicable in case the EVI is known to be positive, i.e., only in case the underlying CDF exhibits a heavy tail. Another reason

for the success of this estimator is the fact that it can be interpreted as an estimator of the slope of the Pareto quantile plot, which is a graphical tool for testing whether the data are Pareto distributed. Indeed as log-transformed Pareto distributed RVs are exponentially distributed, one can visually check the hypothesis of a strict Pareto behaviour by looking at the scatterplot with coordinates $(\log((n+1)/j), \log X_{n-j+1:n})$.

Kernel estimator. A general class of kernel estimators was given in Csörgő *et al.* (1985),

$$\widehat{\xi}_{k,n}^K = \frac{\sum_{j=1}^k \frac{j}{k} K(\frac{j}{k})(\log X_{n-j+1:n} - \log X_{n-j:n})}{\frac{1}{k}\sum_{j=1}^k K(\frac{j}{k})}.$$

This estimator is also restricted to the case $\xi > 0$. The Hill estimator is a member of this class, since it can be obtained by taking a uniform kernel $K(x)$. The kernel estimators can be interpreted as weighted least squares regression estimators of the slope of the Pareto quantile plot in case one considers regression lines passing through a fixed anchor point. Kernel EVI-estimators for a real ξ have been studied by Groeneboom *et al.* (2003).

Pickands estimator. A simple estimator for the general case, $\xi \in \mathbb{R}$, is the Pickands estimator (Pickands, 1975)

$$\widehat{\xi}_{k,n}^P = \frac{1}{\log 2} \frac{X_{n-\lfloor k/4 \rfloor:n} - X_{n-\lfloor k/2 \rfloor:n}}{X_{n-\lfloor k/2 \rfloor:n} - X_{n-k:n}},$$

where $\lfloor x \rfloor$ denotes the integer part of x.

Moment estimator. Dekkers *et al.* (1989) have proposed an alternative estimator that is not restricted to the case $\xi > 0$ and which has the following form

$$\widehat{\xi}_{k,n}^M = M_{k,n}^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_{k,n}^{(1)})^2}{M_{k,n}^{(2)}}\right)^{-1},$$

where

$$M_{k,n}^{(\ell)} = \frac{1}{k}\sum_{i=1}^k (\log X_{n-i+1:n} - \log X_{n-k:n})^\ell, \ (\ell = 1,2) \quad \left[M_{k,n}^{(1)} \equiv \widehat{\xi}_{k,n}^H\right].$$

Generalized Hill estimator. Beirlant *et al.* (1996b) proposed an estimator for $\xi \in \mathbb{R}$ by estimating the slope of the generalized quantile plot, $(\log((n+1)/j), \log UH_{j,n}), j = 1, ..., n$ where $UH_{j,n} = X_{n-j:n}\widehat{\xi}_{j,n}^H$. This leads to the generalized Hill (GH) estimator defined as

$$\widehat{\xi}_{k,n}^{GH} = \frac{1}{k}\sum_{j=1}^k \log UH_{j,n} - \log UH_{k+1,n}$$

further studied in Beirlant *et al.* (2005).

ML estimator. Conditionally on the OS, $X_{n-k:n}$, k intermediate, $X_{n-i+1:n} - X_{n-k:n}$, $1 \le i \le k$, are approximately the k top OSs associated with a sample of size k from $\mathrm{GP}_\xi(\alpha x/\xi)$, $\xi, \alpha \in \mathbb{R}$,

with $GP_\xi(x)$ given in (2.4). The solution of the ML equations associated with the above mentioned parameterization (Davison, 1984) gives rise to an explicit EVI-estimator, usually called the ML EVI-estimator, that can be named PORT-ML, with PORT standing for *peaks over random threshold*, after Araújo Santos *et al.* (2006). Such an EVI-estimator is given by

$$\widehat{\xi}^{\mathrm{ML}}_{n,k} := \tfrac{1}{k} \sum_{i=1}^{k} \ln(1 + \hat{\alpha}[X_{n-i+1:n} - X_{n-k:n}]),$$

where $\hat{\alpha}$ is the implicit ML estimator of the unknown 'scale' parameter α. A comprehensive study of the asymptotic properties of this ML estimator for $\xi > -1/2$ has been undertaken in Drees *et al.* (2004). As recently shown by Zhou (2009, 2010), such EVI–estimator is also valid for $\xi > -1$. We can also consider the random threshold $X_{n-k:n}$ replaced by a deterministic threshold u, working then under the POT methodology.

PWM estimators. The parametric PWM method, initially derived under parametric frameworks, and sketched in Section 4.1, can be devised under a semi-parametric framework. De Haan & Ferreira (2006) considered the semi-parametric *generalized Pareto* PWM (GPPWM) EVI-estimator, based on the sample of excesses over the high random level $X_{n-k:n}$, i.e.,

$$\widehat{\xi}^{GPPWM}_{k,n} := 1 - \frac{2\widehat{a}^{\star}_1(k)}{\widehat{a}^{\star}_0(k) - 2\widehat{a}^{\star}_1(k)},$$

with $\widehat{a}^{\star}_\ell(k) := \sum_{i=1}^{k} ((i-1)/(k-1))^\ell (X_{n-i+1:n} - X_{n-k:n})/k$, $\ell = 0,1$. See also Diebolt *et al.* (2007, 2008c). Caeiro & Gomes (2011) introduced and studied Pareto PWM (PPWM) EVI-estimators, given by

$$\widehat{\xi}^{PPWM}_{k,n} := 1 - \frac{\widehat{a}_1(k)}{\widehat{a}_0(k) - \widehat{a}_1(k)},$$

with $\widehat{a}_\ell(k) := \sum_{i=1}^{k} ((i-1)/(k-1))^\ell X_{n-i+1:n}/k$, $\ell = 0,1$.

Mixed-Moment estimator. We further refer the so-called *mixed moment* (MM) estimator (Fraga Alves *et al.*, 2009), valid for all $\xi \in \mathbb{R}$, and given by

$$\widehat{\xi}^{\mathrm{MM}}_n(k) := \frac{\widehat{\varphi}_{k,n} - 1}{1 + 2\min(\widehat{\varphi}_{k,n} - 1, 0)}, \quad \text{with} \quad \widehat{\varphi}_{k,n} := \frac{\widehat{\xi}^{H}_{k,n} - \mathrm{L}^{(1)}_{k,n}}{\left(\mathrm{L}^{(1)}_{k,n}\right)^2},$$

$\widehat{\xi}^{H}_{k,n}$ given in (5.1) and $\mathrm{L}^{(1)}_{k,n} := 1 - \tfrac{1}{k} \sum_{i=1}^{k} X_{n-k:n}/X_{n-i+1:n}$. The MM EVI-estimator has a very simple form and is asymptotically very close to the ML EVI-estimator for a wide class of heavy-tailed models.

The mean-of-order-p EVI-estimator. A competitive generalization of the Hill estimator has been recently introduced in the literature. Note that we can write

$$\widehat{\xi}^{H}_{k,n} = \sum_{i=1}^{k} \ln \left(\frac{X_{n-i+1:n}}{X_{n-k:n}} \right)^{1/k} = \ln \left(\prod_{i=1}^{k} \frac{X_{n-i+1:n}}{X_{n-k:n}} \right)^{1/k}, \quad 1 \le k < n.$$

The Hill estimator is thus the logarithm of the *geometric mean* (or *mean-of-order-*0) of $\underline{U} :=$ $\{U_{ik} := X_{n-i+1:n}/X_{n-k:n}, \ 1 \leq i \leq k < n\}$. More generally, Brilhante *et al.* (2013) considered as basic statistics the *mean-of-order-p* (MOP) of $\underline{U}$, with $p \geq 0$, i.e., the class of statistics

$$\mathrm{A}_p(k) = \begin{cases} \left(\frac{1}{k} \sum\limits_{i=1}^{k} U_{ik}^p \right)^{1/p}, & \text{if} \quad p > 0, \\ \left(\prod\limits_{i=1}^{k} U_{ik} \right)^{1/k}, & \text{if} \quad p = 0, \end{cases}$$

and the class of MOP EVI-estimators,

$$\widehat{\xi}_{k,n}^{H_p} := \begin{cases} \left(1 - \mathrm{A}_p^{-p}(k) \right)/p, & \text{if} \quad 0 < p < 1/\xi, \\ \ln \mathrm{A}_0(k), & \text{if} \quad p = 0, \end{cases}$$

with $\widehat{\xi}_{k,n}^{H_0} \equiv \widehat{\xi}_{k,n}^{H}$, given in (5.1). See also Beran *et al.* (2014).

The PORT EVI-estimators. Apart from Pickands and ML's estimators, all aforementioned EVI-estimators are scale invariant but not location-invariant. And particularly the Hill estimator can suffer drastic changes when we induce a shift in the data, given rise to the so-called 'Hill horror-plots', a terminology used in Resnick (1997). This led Araújo Santos *et al.* (2006) to introduce the so-called PORT methodology. The estimators are then functionals of a sample of excesses over a random level $X_{n_q:n}$, $n_q := \lfloor nq \rfloor + 1$, i.e. functionals of the sample,

$$\underline{\mathbf{X}}_n^{(q)} := \left(X_{n:n} - X_{n_q:n}, \dots, X_{n_q+1:n} - X_{n_q:n} \right). \tag{5.2}$$

Generally, we can have $0 < q < 1$, for any $F \in \mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi)$ (*the random level is an empirical quantile*). If the underlying model F has a finite *left endpoint*, $x_F := \inf\{x : F(x) \geq 0\}$, we can also use $q = 0$ (*the random level can then be the minimum*). If we think, for instance, of Hill EVI-estimators, the new classes of PORT-Hill EVI-estimators, theoretically studied in Araújo Santos *et al.* (2006), and for finite samples in Gomes *et al.* (2008b), are given by

$$\widehat{\xi}_{k,n}^{(q)} := \frac{1}{k} \sum_{i=1}^{k} \left\{ \ln \frac{X_{n-i+1:n} - X_{n_q:n}}{X_{n-k:n} - X_{n_q:n}} \right\}, \quad 0 \leq q < 1. \tag{5.3}$$

And a similar dependence on this extra *tuning* parameter q can be conceived for any other EVI-estimator. See also Gomes *et al.* (2011, 2013). Other interesting location invariant EVI-estimators for Pareto-type tails can be found in Fraga Alves (2001) and Ling *et al.* (2012).

5.2 Main asymptotic properties

In order to obtain weak consistency of all these estimators, we only need to assume that F belongs to the MDA, $\mathcal{D}_M(\mathrm{EV}_\xi)$ and that k is an intermediate sequence. However, if we want to obtain asymptotic normality, we need to strengthen this condition into a second-order one, of the form (2.6). More precisely, assuming that $\sqrt{k}A\left(n/k\right) \to \lambda < \infty$, we can expand each estimator $\widehat{\xi}_{k,n}^{\star}$ as follows

$$\widehat{\xi}_{k,n}^{\star} - \xi \stackrel{d}{=} \sigma^{\star} \frac{N_k}{\sqrt{k}} + b^{\star} A\left(\frac{n}{k} \right) + o_{\mathbb{P}}\left(A\left(\frac{n}{k} \right) \right),$$

13

where N_k is a standard normal RV and $(\sigma^\star, b^\star) \in \mathbb{R}^+ \times \mathbb{R}$. Thus, with $\mathcal{N}(\mu, \sigma^2)$ denoting a normal RV with mean value μ and variance σ^2,

$$\sqrt{k}(\widehat{\xi}_{k,n}^\star - \xi) \xrightarrow{d} \mathcal{N}(\lambda b^\star, \sigma^{\star 2}). \tag{5.4}$$

5.3 Reduced-bias estimators

The adequate accommodation of the high asymptotic bias of some of the above mentioned EVI-estimators has recently been extensively addressed. We mention the pioneering papers by Peng (1998), Beirlant *et al.* (1999), Feuerverger & Hall (1999) and Gomes *et al.* (2000), among others, based on log-excesses and on scaled log-spacings between subsequent extreme OSs from a Pareto-type distribution. In these papers, authors are led to propose *second-order reduced-bias* (SORB) EVI-estimators, with asymptotic variances larger than or equal to $(\xi (1 - \rho)/\rho)^2$, where $\rho(< 0)$ is the aforementioned 'shape' second-order parameter, ruling the rate of convergence of the distribution of the normalized sequence of maximum values towards the limiting law EV_ξ, in (2.1). Similar results in the general MDA, $\mathcal{D}_M(\mathrm{EV}_\xi)$, can be found in Beirlant *et al.* (2005), and more recently in Cai *et al.* (2013), who introduced a SORB EVI-estimator for ξ around zero, based on the PWM methodology. For those estimators, the asymptotic mean is 0, instead of the value $\lambda b^\star$, in (5.4), whatever the value of λ. However, the asymptotic variance increases compared to $\sigma^{\star 2}$.

For Pareto-type models, Caeiro *et al.* (2005; 2009) and Gomes *et al.* (2007a; 2008c) have been able to reduce the bias without increasing the asymptotic variance, kept at ξ^2, just as happens with the Hill EVI-estimators. Those estimators, called *minimum-variance reduced-bias* (MVRB) EVI-estimators, are all based on an adequate 'external' and a bit more than consistent estimation of the pair of second-order parameters, $(\beta, \rho) \in (\mathbb{R}, \mathbb{R}^-)$ in $A(t) = \xi \beta t^\rho$, done through consistent estimators, denoted by $(\widehat{\beta}, \widehat{\rho})$, such that $\hat{\rho} - \rho = o_p(1/\ln n)$, and outperform the classical estimators for all k. Different algorithms for the estimation of (β, ρ) can be found in Gomes & Pestana (2007), among others. Among the most common MVRB EVI-estimators, we just mention the simplest class in Caeiro *et al.* (2005). Such a class has the functional form

$$\overline{\widehat{\xi}_{k,n}^H}(\widehat{\beta}, \widehat{\rho}) := \widehat{\xi}_{k,n}^H \left(1 - \widehat{\beta}(n/k)^{\widehat{\rho}}/(1 - \widehat{\rho})\right), \tag{5.5}$$

with $\widehat{\xi}_{k,n}^H$ the Hill EVI-estimator, in (5.1), and where $(\widehat{\beta}, \widehat{\rho})$ is an adequate consistent estimator of the aforementioned vector of second-order parameters (β, ρ). Note that the MVRB EVI-estimator in (5.5) is easily justified by the fact that in (5.4), $b^H = 1/(1 - \rho)$. For recent overviews on reduced-bias EVI-estimation, see Gomes *et al.* (2007b), Chapter 6 in Reiss & Thomas (2007), Gomes *et al.* (2008a) and Beirlant *et al.* (2012).

5.4 Choice of k

An important problem has often hampered a wide practical use of these estimators: the number k used in the implementation of these estimators depends strongly on the tail itself and needs to be estimated adaptively from the data. The choice of this number is clearly a question of trade-off between bias and variance: as k increases, the bias will grow since the tail satisfies less the convergence criterion, while if fewer data are used, the variance increases. It is therefore typically suggested that the optimal value of k should coincide with the value that minimizes the

mean-squared error (MSE, the sum of the bias-squared and the variance). However it has been theoretically shown that this optimum depends on both the sample size and the unknown values of ξ and ρ (see Hall & Welsh, 1985, among others). Therefore, some authors have suggested to plot the value of the ξ-estimator as a function of k and to judgementally choose a 'stable' point (see Drees *et al.*, 2000, and de Sousa & Michailids, 2004, among others). Nevertheless, this stable part is actually lacking in many cases and the estimation problem turns into a guessing practice. Other approaches based on bootstrap methods (Hall, 1990; Draisma *et al.*, 1999; Danielsson *et al.*, 2001; Gomes & Oliveira, 2001) or on regression diagnostics on a Pareto quantile plot (Beirlant *et al.*, 1996c; 2002) have been proposed. We further mention interesting selection procedures based on bias, like the ones in Drees & Kaufmann (1998) and Guillou & Hall (2001). Possible heuristic choices are provided in Gomes & Pestana (2007), Gomes *et al.* (2008e) and Beirlant *et al.* (2011). The adaptive SORB and MVRB EVI-estimation is still giving its first steps. We can however mention the recent papers by Gomes *et al.* (2012, 2013).

6 Semi-parametric estimation of other parameters

6.1 Classical semi-parametric estimation

Despite the fact that the estimation of quantiles, return periods of high levels and exceedance probabilities, among other parameters of extreme events, are at least as important in applications as the estimation of the EVI, we shall only briefly refer this topic.

High quantile or VaR-estimation. In a semi-parametric framework, the most usual estimators of a quantile $\mathrm{VaR}_p = \chi_{1-p} := U(1/p)$, with p small, can be easily derived from (2.5), through the approximation

$$U(tx) \approx U(t) + a(t)(x^\xi - 1)/\xi.$$

The fact that $X_{n-k+1:n} \overset{\mathbb{P}}{\sim} U(n/k)$ enables us to estimate χ_{1-p} on the basis of this approximation and adequate estimates of ξ and $a(n/k)$. For the simpler case of heavy tails, the approximation turns out to be

$$U(tx) \approx U(t)x^\xi,$$

and we get

$$\widehat{\chi}_{1-p,k,n} := X_{n-k:n} \ (k/(np))^{\widehat{\xi}_{k,n}},$$

where $\widehat{\xi}_{k,n}$ is any consistent semi-parametric EVI-estimator. This estimator is of the type of the one introduced by Weissman (1978). Details on semi-parametric estimation of extremely high quantiles for $\xi \in \mathbb{R}$, can be found in Dekkers & de Haan (1989), de Haan & Rootzén (1993), Ferreira *et al.* (2003), and more recently in de Haan & Ferreira (2006). Fraga Alves *et al.* (2009) also provide, jointly with the MM-estimator, accompanying shift and scale estimators that make high quantile estimation almost straightforward. Other approaches to high quantile estimation can be found in Matthys & Beirlant (2003). None of the above mentioned quantile estimators reacts adequately to a shift of the data. Araújo Santos *et al.* (2006) provide a class of semi-parametric VaR_p estimators which enjoy such a feature, the empirical counterpart of the theoretical linearity of any quantile χ_α, $\chi_\alpha(\delta X + \lambda) = \delta\chi_\alpha(X) + \lambda$, for any real λ and positive δ. This class of estimators is based

on the PORT methodology, providing exact properties for risk measures in finance: translation-equivariance and positive homogeneity. For Pareto-type models and a Hill EVI-estimation, they are given by

$$\widehat{\chi}^{(q)}_{1-p,k,n} := (X_{n-k:n} - X_{n_q:n})\big(k/(np)\big)^{\widehat{\xi}^{(q)}_{k,n}} + X_{n_q:n},$$

with $\widehat{\xi}^{(q)}_{k,n}$ the PORT-Hill EVI-estimator in (5.3).

Probability of exceedance estimation. The estimation of the probability of exceedance of a fixed high level is the dual problem of estimation of a high quantile. It has been dealt with by Dijk & de Haan (1992) and Ferreira (2002), among others. Again for Pareto-type underlying models, we get for the probability of exceedance of a high level $x = x_n$,

$$\widehat{p}_n = \frac{k}{n}\left\{\frac{x_n}{X_{n-k:n}}\right\}^{-1/\widehat{\xi}_{k,n}}.$$

See also Caeiro *et al.* (2012).

Estimation of other parameters. The estimation of the endpoint of an underlying CDF has been studied by Hall (1982), Csörgő & Mason (1989), Aarssen & de Haan (1994), among others. We further refer to the recent article by Fraga Alves & Neves (2013), dealing with the endpoint semi-parametric estimation of a model in $\mathcal{D}_{\mathcal{M}}(\mathrm{EV}_0)$. Estimation of the mean of a heavy-tailed distribution has been undertaken by Peng (2001) and Johansson (2003). Estimation of the conditional tail expectation can be found in Deme *et al.* (2014). Estimation of the Weibull tail coefficient dates back to Girard (2004). See also Goegebeur *et al.* (2010), among others.

6.2 SORB semi-parametric estimation

Reduced-bias quantile estimators have been studied in Matthys *et al.* (2004) and Gomes & Figueiredo (2006), who consider the classical SORB EVI-estimators. Gomes & Pestana (2007) and Beirlant *et al.* (2008) incorporate the MVRB EVI-estimators in Caeiro *et al.* (2005) and Gomes *et al.* (2007b) in high quantile semi-parametric estimation. See also Diebolt *et al.* (2008b), Beirlant *et al.* (2009), Caeiro & Gomes (2009) and Li *et al.* (2010). For a SORB estimation of the Weibull-tail coefficient, we mention Diebolt *et al.* (2008a). Finally, for a SORB endpoint estimation, we mention Li & Peng (2009).

7 SUE for censored data

Censoring and truncation are two of the most relevant concepts in Statistics. Although conceptually different, with the first one related to a cutting in the support of the underlying model and the second one with a cutting of the sample, they are formally quite similar. Whenever working in the area of SUE, and if interested in inference for large values, we are dealing under a kind of censoring, left censoring, working only with the k top OSs. But apart from this, we can also have a random right censoring. Censoring occurs both in industrial life-testing (i.e. investigation of the distribution of the lifetime of manufactured components) and in medical trials and biological experiments. So terms synonymous to a "censored observation" are "withdrawal", a "loss", or a "death due to a

competing risk", while an "uncensored observation" might be a "failure", a "relapse", or a "death from the cause under study". Statistical techniques for analyzing censored data sets are quite well studied now, but they mostly concern central characteristics of the underlying distribution. The framework of extreme-value analysis under censoring has not been extensively studied in the literature. To the best of our knowledge, the first who mentioned the topic are Beirlant *et al.* (1996a, Section 2.7) and Reiss & Thomas (1997, Section 6.1). Then some estimators of tail parameters have been proposed in Beirlant & Guillou (2001) for truncated data and extended to the random-right censoring by Beirlant *et al.* (2007) and Einmahl *et al.* (2008).

Under random censoring, there is a RV Y such that only $Z = X \wedge Y$ and $\delta = I_{\{X \leq Y\}}$ are observed, with I_A denoting the indicator function of the event A. The indicator variable δ determines whether X has been censored or not. Consequently, we have access to the random sample (Z_i, δ_i), $1 \leq i \leq n$, of independent copies of (Z, δ), but our goal is to make inference on the RTF of the unknown lifetime distribution, i.e. on $\overline{F}_X(x) := \mathbb{P}(X > x) = 1 - F_X(x)$, while F_Y, the CDF of Y, is considered to be a non-parametric nuisance parameter. As mentioned in Einmahl *et al.* (2008), all the EVI-estimators need to be slightly modified in order to be consistent for the estimation of ξ. The following possible and simple modification is suggested in Einmahl *et al.* (2008) (see also Beirlant *et al.*, 2007)

$$\widehat{\xi}_{k,n}^{c,\star} = \widehat{\xi}_{k,n}^{\star}/\widehat{p}, \ \text{with} \ \widehat{p} = \frac{1}{k}\sum_{j=1}^{k} \delta_{[n-j+1,n]},$$

where $\widehat{\xi}_{k,n}^{\star}$ can be one of the EVI-estimators defined in Section 5, but based on the observed sample $(Z_1, \ldots, Z_n)$, $\delta_{[1,n]}, ..., \delta_{[n,n]}$ are the δ's corresponding to the $Z_{1:n}, \ldots, Z_{n:n}$, $\widehat{p}$ being the proportion of non-censored observations in the $k-$largest Z's. Another approach to derive estimators in the case of censoring is to adapt the likelihood to this context (see Beirlant *et al.*, 2010). For applications of this methodology to different sets of survival data as well as simulated data, with hints for the adequate EVI, high quantiles and right endpoint estimation of X, see also Gomes & Neves (2011).

8 Testing issues

8.1 Statistical choice of EV models under parametric frameworks

The Gumbel type CDF, $\Lambda = \text{EV}_0$, or the exponential type CDF, GP_0, with EV_ξ and GP_ξ given in (2.1) and (2.4), respectively, are favorites in SUE, essentially because of the simplicity of inference associated with these populations. Additionally, $\xi = 0$ can be regarded as a change-point, because for $\xi < 0$ the data come from a CDF with a finite right endpoint and for $\xi > 0$ the right endpoint is infinite. Thus, any separation between EV models, with Λ playing a central and prominent position (called 'trilemma' by Tiago de Oliveira), turns out to be an important statistical problem, that has been recently considered under a semi-parametric framework. From a parametric point of view, empirical tests of the hypothesis $H_0 : \xi = 0$ versus a sensible one-sided or two-sided alternative, either for the EVD or the GPD, date back to Jenkinson (1955) and Gumbel (1965). Next, we can find in the literature:

1. Quick tests, suggested by heuristic reasons (van Montfort, 1970, 1973; Bardsley, 1977; van Montfort & Otten, 1978; Otten & van Montfort, 1978; Galambos, 1982; Gomes, 1982, 1984;

Tiago de Oliveira & Gomes, 1984; van Montfort & Gomes, 1985; van Monfort & Witter, 1985; Gomes & van Montfort, 1987; Brilhante, 2004).

2. Modified locally most powerful tests (Tiago de Oliveira, 1981, 1984; Tiago de Oliveira & Gomes, 1984; Gomes & van Montfort, 1987).

3. Locally asymptotically normal (LAN) tests (Falk, 1995a,b; Marohn, 1994, 1998a,b, 2000; Falk *et al.*, 2008).

4. Goodness-of-fit (GOF) tests for the Gumbel model (Stephens, 1976, 1977, 1986; Kinnison, 1989). The fitting of the GPD to data has been worked out in Castillo & Hadi, (1997) and Chaouche & Bacro (2004). The problem of GOF tests for the GPD has been studied by Choulakian & Stephens (2001) and Luceño (2006), among others. Further non-parametric tests appear in Jurečková & Picek (2001).

5. Tests from large sample theory, like the likelihood ratio test and Wald test, among others (Hosking, 1984; Gomes, 1989b).

Statistical choice in the MEV model is developed in Gomes (1984; 1987; 1989a,b), Gomes & Alpuim (1986), Hasofer & Wang (1992), Wang (1995), Fraga Alves & Gomes (1996) and Wang *et al.* (1996). Some of these authors already go beyond the extremal process, working under a semi-parametric framework.

8.2 Semi-parametric framework

Under a semi-parametric framework, it is naturally sensible to test the hypothesis

$$H_0 : F \in \mathcal{D}_\mathcal{M}(G_0) \quad \text{against} \quad H_1 : F \in \mathcal{D}_\mathcal{M}(G_\xi), \ \xi \neq 0$$

or the corresponding one-sided alternatives. Tests of this nature can already be found in several papers prior to 2000, among which we mention Galambos (1982), Castillo *et al.* (1989), Hasofer & Wang (1992), Fraga Alves & Gomes (1996), Wang *et al.* (1996), Marohn (1998a,b) and Fraga Alves (1999). More recently, further testing procedures of this type can be found in Segers & Teugels (2000), Neves *et al.* (2006) and Neves & Fraga Alves (2007), among others. The testing of first-order EV conditions can be dated back to Dietrich *et al.* (2002), who propose a test statistic to check whether the hypothesis $F \in \mathcal{D}_\mathcal{M}(G_\xi)$ is supported by the data, together with a simpler version devised to test whether $F \in \mathcal{D}_\mathcal{M}(G_{\xi \geq 0})$. Further results of this last nature can be found in Beirlant *et al.* (2006). Drees *et al.* (2006) deal with the testing of $F \in \mathcal{D}_\mathcal{M}(G_{\xi > -1/2})$. Accurate tables of critical points of this statistic are provided in Hüsler & Li (2006). See also Canto e Castro *et al.* (2011). Reviews of the topic up to the date can be found in Neves & Fraga Alves (2008) and Hüsler & Peng (2008).

9 Dependent frameworks—a brief reference to the extremal index estimation

The same EV_ξ CDF, in (2.1), appears as the limiting CDF of the maximum for a large class of stationary sequences, $\{X_n\}_{n \geq 1}$, like, for instance, the ones for which the mixing condition **D**,

introduced in Leadbetter *et al.* (1983), holds. Let us assume we have data from a stationary process with an underlying CDF, F, and let $\{Y_n\}_{n\geq 1}$ be the associated IID sequence (from the same model F). Under adequate local dependence conditions, the limiting CDF of the maximum $X_{n:n}$ of the stationary sequence may be directly related to the maximum, $Y_{n:n}$, of the IID associated sequence, through a new parameter, the so-called *extremal index* (EI).

More specifically, the stationary sequence $\{X_n\}_{n\geq 1}$ has an extremal index θ $(0 < \theta \leq 1)$ if, for every $\tau > 0$, we may find a sequence of levels $u_n = u_n(\tau)$ such that

$$\mathbb{P}\left(Y_{n:n} \leq u_n\right) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\tau} \quad \text{and} \quad \mathbb{P}\left(X_{n:n} \leq u_n\right) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\theta\tau}.$$

The *extremal index* θ may thus be informally defined by the approximation

$$\mathbb{P}[X_{n:n} \leq x] \approx F^{n\theta}(x) \approx \mathrm{EV}_\xi\left(\frac{x - b'_n}{a'_n}\right), \quad \begin{cases} a'_n = a_n\, \theta^\xi \\ b'_n = b_n + a_n\, \frac{\theta^\xi - 1}{\xi} \end{cases},$$

where $F(.)$ is the marginal CDF of a strictly stationary sequence $\{X_n\}_{n\geq 1}$, satisfying adequate local and asymptotic dependence conditions. One of the local dependence conditions which enable us to guarantee the existence of an *extremal index* is the $\mathbf{D}$" condition, introduced by Leadbetter & Nandagopalan (1989). Under the validity of such a condition, the EI can also be defined as the reciprocal of the 'mean time of duration of extreme events', being directly related to the exceedances of high levels. Indeed, we have

$$\theta = \frac{1}{\text{limiting mean size of clusters}} = \lim_{n\to\infty} \mathbb{P}(X_2 \leq u_n | X_1 > u_n),$$

where u_n is a sequence of values such that $F(u_n) = 1 - \tau/n + o(1/n)$, as $n \to \infty$. Then, given a sample $(X_1, \ldots, X_n)$, an obvious non-parametric estimator of θ is immediately suggested: once a suitable threshold u is chosen, put

$$\theta_n = \theta_{n,u} := \frac{\sum_{j=1}^{n-1} I_{[X_j > u, X_{j+1} \leq u]}}{\sum_{j=1}^{n} I_{[X_j > u]}} = \frac{\sum_{j=1}^{n-1} I_{[X_j \leq u < X_{j+1}]}}{\sum_{j=1}^{n} I_{[X_j > u]}}.$$

In order to have consistency of this estimator the high level $u = u_n$ must be such that $n(1 - F(u_n)) = c_n\tau = \tau_n$, $\tau_n \to \infty$ and $\tau_n/n \to 0$ (Nandagopalan, 1990). And it is sensible to replace the deterministic level u by the stochastic level $X_{n-k:n}$, and to consider the EI-estimator as a function of k, the number of OSs higher than the chosen threshold, getting the upcrossing (UC) EI-estimator

$$\widehat{\theta}_{k,n}^{UC} := \frac{1}{k} \sum_{j=1}^{n-1} I_{[X_j \leq X_{n-k:n} < X_{j+1}]}. \tag{9.1}$$

Like this, we are thus placing ourselves in a situation similar to the one we have with the semi-parametric EVI-estimation, being consistency attained only if k is intermediate. Reduced-bias versions of the UC EI–estimator, in (9.1), have been obtained in Gomes *et al.* (2008d). Limit theorems for empirical processes of cluster functionals have been obtained by Drees & Rootzén (2010).

10   Other related topics and open problems

We shall next almost reproduce the discussion in Beirlant *et al.* (2012), since most of the topics considered there still deserve attention. Indeed, SUE is a quite lively topic of research. Important developments have appeared recently in the area of *spatial extremes*, where *parametric models* became again relevant. And now, that we have access to highly sophisticated computational techniques, a great variety of *parametric models* can further be considered. And in a semi-parametric framework, topics like *robustness and extremes*, *threshold selection*, *trends and change points* in the *tail behaviour*, and *clustering*, among others, are still quite challenging.

10.1   Penultimate approximations

An important problem in EVT concerns the rate of convergence of $F^n(a_n x + b_n)$ towards $\mathrm{EV}_\xi(x)$, in (2.1), or, equivalently, the search for estimates of the difference

$$d_n(F, \mathrm{EV}_\xi, x) := F^n(a_n x + b_n) - \mathrm{EV}_\xi(x).$$

Indeed, parametric inference on the RTF, usually unknown, is done on the basis of the identification of $F^n(a_n x + b_n)$ and of $\mathrm{EV}_\xi(x)$. And the rate of convergence can validate or not the most usual models in SUE. As noted by Fisher & Tippett (1928), despite the fact that the normal CDF, $\Phi \in \mathcal{D}_\mathcal{M}(G_0)$, the convergence of $\Phi^n(a_n x + b_n)$ towards $G_0(x)$ is extremely slow. They then show that $\Phi^n(x)$ is 'closer' to a suitable penultimate Weibull rather than the Gumbel CDF. Such an approximation is the so-called *penultimate approximation* and several penultimate models have been advanced. Dated overviews of the modern theory of rates of convergence in EVT, introduced in Anderson (1971), can be seen in Galambos (1984) and Gomes (1994). More recently, Gomes & de Haan (1999) derived, for all $\xi \in \mathbb{R}$, exact penultimate approximation rates with respect to the variational distance, under adequate differentiability assumptions. Kaufmann (2000) proved, under weaker conditions, a result related to the one in Gomes & de Haan (1999). This penultimate or pre-asymptotic behaviour has further been studied by Raoult & Worms (2003) and Diebolt & Guillou (2005), among others. Quite recently, the role of MS penultimate approximations in reliability has been considered in Reis *et al.* (2013). Dealing with regular and homogeneous parallel-series systems, these authors assess the gain in accuracy when a penultimate approximation is used instead of the ultimate one. Other type of penultimate approximations have been considered in the unpublished paper by Smith (1987b). Among them, we mention a *penultimate* EV parametric model of the type,

$$\mathrm{PEV}_\xi(x; r) = \exp\left(-(1 + \xi x)^{-1/\xi}\left(1 + r(1 + \xi x)^{-1/\xi}\right)\right)$$

or the associated *penultimate* GP CDF,

$$\mathrm{PGP}_\xi(x; r) = 1 - (1 + \xi x)^{-1/\xi}\left(1 + r(1 + \xi x)^{-1/\xi}\right).$$

This type of model surely deserves a deeper consideration under statistical backgrounds. *Penultimate models* seem to be possible and interesting alternatives to the classical models but have never been deeply used in the literature. Shouldn't we invest more on this type of models from an inferential point of view?

10.2 Max-semistable laws as alternative parametric models

We also refer the class of *max-semistable* (MSS) laws, introduced by Grinevigh (1992a, 1992b), Pancheva (1992), and further studied in Canto e Castro *et al.* (2002) and in Temido & Canto e Castro (2003). Such a class is more general than the class of *max-stable* (MS) laws, given in (2.1). The possible MSS laws are

$$G_{\xi,\nu}(x) = \begin{cases} \exp\left(-\nu(\ln(1+\xi x))(1+\xi x)^{-1/\xi}\right), \ 1+\xi x > 0 & \text{if} \ \xi \neq 0 \\ \exp\left(-\nu(x)\exp(-x)\right), \ x \in \mathbb{R} & \text{if} \ \xi = 0, \end{cases}$$

where $\nu(\cdot)$ is a positive, limited and periodic function. A unit ν-function enables us to get the MS laws in (2.1). Discrete models like the geometric and negative binomial, and some multimodal continuous models, are in $\mathcal{D}_{\mathcal{MSS}}(G_{\xi,\nu})$ but not in $\mathcal{D}_{\mathcal{M}}(\text{EV}_\xi)$. A survey of the topic can be found in Pancheva (2010). Generalized Pickands-statistics have been used in Canto e Castro & Dias (2011), to develop methods of estimation in the MSS context. See also Canto e Castro *et al.* (2011). Such a diversity of models, if duly exploited from a statistical point of view, can surely provide fruitful topics of research, both in parametric and semi-parametric setups.

10.3 Robustness and extremes

In most statistical applications, outliers often occur and thus can have a disproportional effect on the estimation procedures. Some robust algorithms, replacing the classical statistical methodologies, can be useful, leading to alternative outlier-resistant estimators. In the context of EVT, this notion of robustness can appear at first sight as a contradiction since the aim is to reduce the influence of extreme observations whereas EVT mainly focus on these data points (see e.g. Dell'Aquila & Embrechts, 2006). However tail parameters can be seriously affected by these suspicious observations and thus robust methods are required. Such a topic has been recently studied in the literature. We can mention among others Brazauskas & Serfling (2000) and Vandewalle *et al.* (2007) for strict Pareto-type distributions, Dupuis & Field (1998), Peng & Welsh (2001), and Juárez & Schucany (2004) for EVDs or GPDs. Different approaches can be used, for instance the density power divergence criterion introduced by Basu *et al.* (1998) (see Dierckx *et al.*, 2013). Other methods based on an adaptation of classical tail parameters, such as the Hill's one, can be proposed (see e.g. Beran *et al.*, 2014). In all the cases, these estimators depend on a single parameter that controls the trade-off between robustness and efficiency. The choice of this parameter is delicate and its adaptive selection, depending on the data, could be a challenging open problem. From a practical point of view, it is also difficult to be convinced of the presence or not of outliers in a data set. Thus it is recommended to compare robust and non-robust estimators. In case of a strong discrepancy between the estimates, we can suspect the presence of outliers in the tails.

10.4 The PORT methodology

In SUE most of the methods of estimation are dependent on the log-excesses or scaled log-spacings and do not react adequately to changes in the location of the model underlying the data. The steps given in PORT-estimation of parameters of extreme events are promising but still quite incomplete at the current state-of-the-art. For PORT quantile estimation, we mention Henriques-Rodrigues & Gomes (2009). The shift invariant versions, dependent on the tuning parameter q in (5.2), have

properties similar to the ones of the original estimator T, provided we keep to adequate k-values and choose an adequate *tuning parameter q*. Recent research on this topic can be seen in Gomes *et al.* (2011, 2013), but further theoretical research is welcome. The PORT estimation of a shape second-order parameter has been dealt with in Henriques Rodrigues *et al.* (2013).

10.5 Adaptive selection of sample fraction or threshold

A *threshold* is often set "almost arbitrarily" (for instance at the 90% or the 95% sample quantile)! However, the choice of the threshold, or equivalently of the number k of top OSs to be used, is crucial for a reliable estimation of any parameter of extreme events. The topic has already been extensively studied for classical and even reduced-bias EVI-estimators, as mentioned in Section 5.4. Is it sensible to use bootstrap computational intensive procedures for threshold selection or there will be simpler techniques possibly related with bias pattern? Is it possible to apply a similar methodology for the estimation of other parameters of extreme events?

10.6 Other possible topics of research in SUE

Recent parametric models, like the *extreme value Birnbaum-Saunders model* in Ferreira *et al.* (2012), can also become relevant in the area of SUE. Testing whether $F \in \mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi)$, for a certain ξ, is a crucial topic, already dealt with in several articles referred in Section 8. And what about testing second-order and even third-order conditions? Change-points detection is also a challenging topic of research. And SUE for weakly dependent data, with all problems related with clustering of extreme values still deserves further research. SUE for randomly censored data and estimation of the endpoint are also still relevant topics in SUE. We mention the recent papers by Einmahl & Magnus (2008), Li & Peng (2009), Einmahl & Smeets (2011), Henriques-Rodrigues *et al.* (2011), Li *et al.* (2011) and Fraga Alves & Neves (2013). Moreover, the estimation of second and higher-order parameters still deserves further attention, particularly due to the importance of such estimation in SORB estimators of parameters of extreme events.

References

[1] Aarssen, K. & de Haan, L. (1994). On the maximal life span of humans. *Math. Popul. Stud.*, **4**:4, 259–281.

[2] Achcar, A.J., Bolfarine, H. & Pericchi, L.R. (1987). Transformation of a survival data to an extreme value distribution. *The Statistician*, **36**, 229–234.

[3] Anderson, C.W. (1971). *Contributions to the Asymptotic Theory of Extreme Values*, Ph.D. Thesis, University of London.

[4] Araújo Santos, P., Fraga Alves, M.I. & Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and high quantile estimation. *REVSTAT*, **4**:3, 227–247.

[5] Arnold, B., Balakrishnan, N. & Nagaraja, H.N. (1992). *A First Course in Order Statistics*, Wiley, New York.

[6] Ashour, S.K. & El-Adl, Y.M. (1980). Bayesian estimation of the parameters of the extreme value distribution. *Egypt. Statist. J.*, **24**, 140–152.

[7] Balakrishnan, N. & Chan, P.S. (1992). Order statistics from extreme value distribution, II: best linear unbiased estimates and some other uses. *Comm. Statist. Simulation Comput.*, **21**, 1219–1246.

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

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

[10] Basu, A., Harris, I.R., Hjort, N.L. & Jones, M.C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, **85**, 549–559.

[11] Beirlant, J., Boniphace, E. & Dierckx, G. (2011). Generalized sum plots. *REVSTAT*, **9**:2, 181–198.

[12] Beirlant, J., Caeiro, F. & Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes. *REVSTAT*, **10**:1, 1–31.

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

[14] Beirlant, J., Dierckx, G., Goegebeur, Y. & Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, **2**, 177–200.

[15] Beirlant, J., Dierckx, G. & Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots. *Bernoulli*, **11**:6, 949–970.

[16] Beirlant, J., Dierckx, G., Guillou, A. & Starica, C. (2002). On exponential representations of log-spacings of extreme order statistics. *Extremes*, **5**:2, 157–180.

[17] Beirlant, J., Figueiredo, F., Gomes, M.I. & Vandewalle, B. (2008). Improved reduced bias tail index and quantile estimators. *J. Statist. Plann. Inference*, **138**:6, 1851–1870.

[18] Beirlant, J., Goegebeur, Y., Teugels, J. & Segers, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, England.

[19] Beirlant, J. & Guillou, A. (2001). Pareto index estimation under moderate right censoring. *Scand. Actuar. J.*, **2**, 111–125.

[20] Beirlant, J., Guillou, A., Dierckx, G. & Fils-Viletard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, **10**:3, 151–174.

[21] Beirlant, J., Guillou, A. & Toulemonde, G. (2010). Peaks-Over-Threshold modeling under random censoring. *Comm. Statist. Theory Methods*, **39**, 1158–1179.

[22] Beirlant, J., Joossens, E. & Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *J. Statist. Plann. Inference*, **139**, 2800–2815.

[23] Beirlant, J., Teugels, J. & Vynckier, P. (1996a). *Practical Analysis of Extreme Values*, Leuven University Press, Leuven, Belgium.

[24] Beirlant, J., Vynckier, P. & Teugels, J. (1996b). Excess functions and estimation of the extreme-value index. *Bernoulli*, **2**, 293–318.

[25] Beirlant, J., Vynckier, P. & Teugels, J. (1996c). Tail index estimation, Pareto quantile plots and regression diagnostics. *J. Amer. Statist. Assoc.*, **91**, 1659–1667.

[26] Beran, J., Schell, D. & Stehlik, M. (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Ann. Inst. Statist. Math.*, **66**, 193–220.

[27] Bingham, N., Goldie, C.M. & Teugels, J.L. (1987). *Regular Variation*, Cambridge Univ. Press, Cambridge.

[28] Bottolo, L., Consonni, G., Dellaportas, P. & Lijoi, A. (2003). Bayesian analysis of extreme values by mixture modeling. *Extremes*, **6**:1, 25–47.

[29] Brazauskas, V. & Serfling, R. (2000). Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes*, **3**, 231–249.

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

[31] Brilhante, F., Gomes, M.I. & Pestana, D. (2013). A simple generalization of the Hill estimator. *Comput. Statist. Data Anal.*, **57**:1, 518–535.

[32] Caeiro, F. & Gomes, M.I. (2009). Semi-parametric second-order reduced-bias high quantile estimation. *TEST*, **18**:2, 392–413.

[33] Caeiro, F. & Gomes, M.I. (2011). Semi-parametric tail inference through probability-weighted moments. *J. Statist. Plann. Inference*, **141**:2, 937–950.

[34] Caeiro, F., Gomes, M.I. & Henriques-Rodrigues, L. (2009). Reduced-bias tail index estimators under a third order framework. *Comm. Statist. Theory Methods*, **38**:7, 1019–1040.

[35] Caeiro, F., Gomes, M.I. & Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *REVSTAT*, **3**:2, 113–136.

[36] Caeiro, F., Gomes, M.I. & Vandewalle, B. (2012). Semi-parametric probability-weighted moments estimation revisited. *Methodol. Comput. Appl. Probab.*, DOI: 10.1007/s11009-012-9295-6.

[37] Cai, J., de Haan, L & Zhou, C. (2013). Bias correction in extreme value statistics with index around zero. *Extremes*, **16**, 173–201.

[38] Canto e Castro, L. & Dias, S. (2011). Generalized Pickands' estimators for the tail index parameter and max-semistability. *Extremes*, **14**:4, 429–449.

[39] Canto e Castro, L. Dias, S. & Temido, M.G. (2011). Looking for max-semistability: a new test for the extreme value condition. *J. Statist. Plann. Inference*, **141**, 3005-3020.

[40] Canto e Castro, L., de Haan, L. & Temido, M.G. (2002). Rarely observed maxima. *Theory Probab. Appl.*, **45**, 658–661.

[41] Castillo, E., Galambos, J. & Sarabia, J.M. (1989). The selection of the domain of attraction of an extreme value distribution from a set of data, In Hüsler, J. & Reiss, R.-D. (eds.), *Extreme Value Theory, Proc. Oberwolfach* 1987, 181–190, Springer, Berlin-Heidelberg.

[42] Castillo, E. & Hadi, A. (1997). Fitting the generalized Pareto distribution to data. *J. Amer. Statist. Assoc.*, **92**, 1609–1620.

[43] Castillo, E., Hadi, A., Balakrishnan, N. & Sarabia, J.M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley, Hoboke, New Jersey.

[44] Chaouche, A. & Bacro, J.-N. (2004). A statistical test procedure for the shape parameter of a generalized Pareto distribution. *Comput. Statist. Data Anal.*, **45**, 787–803.

[45] Choulakian, V. & Stephens, M.A. (2001). Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, **43**:4, 478–484.

[46] Christopeit, N. (1994). Estimating parameters of an extreme value distribution by the method of moments. *J. Statist. Plann. Inference*, **41**, 173–186.

[47] Coles, S.G. (2001). *An Introduction to Statistical Modelling of Extreme Values*, Springer Series in Statistics.

[48] Coles, S.G. & Dixon, M.J. (1999). Likelihood-based inference for extreme value models. *Extremes*, **2**:1, 5–23.

[49] Coles, S. & Powell, E. (1996). Bayesian methods in extreme value modelling: a review and new developments. *Int. Stat. Rev.*, **64**:1, 119–136.

[50] Coles, S. & Tawn, J. (1996). A Bayesian analysis of extreme rainfall data. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, **45**:4, 463–478.

[51] Crovella, M. & Taqqu, M. (1999). Estimating the heavy tail index from scaling properties. *Methodol. Comput. Appl. Probab.*, **1**, 55–79.

[52] Csörgő, S., Deheuvels, P. & Mason, D.M. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.*, **13**, 1050–1077.

[53] Csörgő, S. & Mason, D.M. (1989). Simple estimators of the endpoint of a distribution, In Hüsler, J. & Reiss, R.-D. (1989). *Extreme Value Theory, Proceedings Oberwolfach* 1987, 132–147, Springer-Verlag, Berlin, Heidelberg.

[54] Danielsson, J., de Haan, L., Peng, L. & de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.*, **76**, 226–248.

[55] David, H.A. (1970; 1981). *Order Statistics*, 1st ed.; 2nd ed. Wiley, New York.

[56] David, H.A. & Nagaraja, H.N. (2003). *Order Statistics*, 3rd ed. Wiley, Hoboken, New Jersey.

[57] Davison, A. (1984). Modeling excesses over high threshold with an application, In J. Tiago de Oliveira ed., *Statistical Extremes and Applications*, D. Reidel, 461–482.

[58] Davison, A.C. & Smith, R.L. (1990). Models for exceedances over high thresholds. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **52**, 393–442.

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

[60] Dekkers, A.L.M. & de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.*, **17**, 1795–1832.

[61] Dell'Aquila, R. & Embrechts, P. (2006). Extremes and robustness: a contradiction? *Fin. Mkts. Portfolio Mgmt.*, **20**, 103–118.

[62] Deme, E., Girard, S. & Guillou, A. (2014). Reduced-bias estimator of the conditional tail expectation of heavy-tailed distributions. In Mason, D. (ed.), *Mathematical Statistics and Limit Theorems*, to appear, Springer.

[63] Diebolt, J., El-Aroui, M.A., Garrido, M. & Girard, S. (2005). Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling. *Extremes*, **8**, 57–78.

[64] Diebolt, J., Gardes, L., Girard, S. & Guillou, A. (2008a). Bias-reduced estimators of the Weibull tail-coefficient. *TEST*, **17**, 311–331.

[65] Diebolt, J., Gardes, L., Girard, S. & Guillou, A. (2008b). Bias-reduced extreme quantiles estimators of Weibull tail-distributions. *J. Statist. Plann. Inference*, **138**, 1389–1401.

[66] Diebolt, J. & Guillou, A. (2005). Asymptotic behaviour of regular estimators. *REVSTAT*, **3**:1, 19–44.

[67] Diebolt, J., Guillou, A., Naveau, P. & Ribereau, P. (2008c). Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT*, **6**:1, 33–50.

[68] Diebolt, J., Guillou, A. & Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method. *J. Statist. Plann. Inference*, **137**, 841–857.

[69] Dierckx, G., Goegebeur, Y. & Guillou, A. (2013). An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *J. Multivariate Anal.*, **121**, 70–86.

[70] Dietrich, D., de Haan, L. & Hüsler, J. (2002). Testing extreme value conditions. *Extremes*, **5**, 71–85.

[71] Dietrich, D. & Hüsler, J. (1996). Minimum distance estimators in extreme value distributions. *Comm. Statist. Theory Methods*, **25**, 695–703.

[72] Dijk, V. & de Haan, L. (1992). On the estimation of the exceedance probability of a high level. In Sen, P.K., and Salama, I.A. (eds.), *Order Statistics and Nonparametrics: Theory and Applications*, 79–92, Elsevier, Amsterdam.

[73] Dombry, C. (2013). *Maximum likelihood estimators for the extreme value index based on the block maxima method*, arXiv:1301.5611.

[74] Draisma, G., de Haan, L., Peng, L. & Pereira, T.T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme value index. *Extremes*, **2**:4, 367–404.

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

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

[77] Drees, H., de Haan, L. & Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.*, **28**, 254–274.

[78] Drees, H. & Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.*, **75**, 149–172.

[79] Drees, H. & Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.*, **38**, 2145–2186.

[80] Dupuis, D.J. (1998). Exceedances over high thresholds: a guide to threshold selection. *Extremes*, **1**:3, 251–261.

[81] Dupuis, D.J. & Field, C.A. (1998). A comparison of confidence intervals for generalized extreme-value distributions. *J. Statist. Comput. Simulation*, **61**, 341–360.

[82] Dwass, M. (1964). Extremal processes. *Ann. Math. Statist.*, **35**, 1718–1725.

[83] Einmahl, J.H.J., Fils-Villetard, A. & Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli*, **14**:1, 207–227.

[84] Einmahl, J. & Magnus, J.R. (2008). Records in athletics through extreme-value theory. *J. Amer. Statist. Assoc.*, **103**, 1382–1391.

[85] Einmahl, J. & Smeets, S.G.W.R. (2011). Ultimate 100-m world records through extreme-value theory. *Stat. Neerl.*, **65**:1, 32–42.

[86] Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, Heidelberg.

[87] Engeland, K., Hisdal, H. & Frigessi, A. (2004). Practical extreme value modelling of hydrological floods and droughts: a case study. *Extremes*, **7**:1, 5–30.

[88] Engelund, S. & Rackwitz, R. (1992). On predictive distribution functions for the three asymptotic extreme value distributions. *Structural Safety*, **11**, 255–258.

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

[90] Falk, M. (1995b). LAN of extreme order statistics. *Ann. Inst. Statist. Math.*, **47**, 693–717.

[91] Falk, M., Guillou, A. & Toulemonde, G. (2008). A LAN based Neyman smooth test for the generalized Pareto distribution. *J. Statist. Plann. Inference*, **138**, 2867–2886.

[92] Falk, M., Hüsler, J. & Reiss, R.-D. (1994; 2004; 2011). *Laws of Small Numbers: Extremes and Rare Events*, 1st ed.; 2nd ed.; 3rd ed., Birkhäuser, Basel.

[93] Ferreira, A. (2002). Optimal asymptotic estimation of small exceedance probabilities. *J. Statist. Plann. Inference*, **104**, 83–102.

[94] Ferreira, M., Gomes, M.I. & Leiva, V. (2012). On an extreme value version of the Birnbaum-Saunders distribution. *REVSTAT*, **10**:2, 181–210.

[95] Ferreira, A. & de Haan, L. (2013). *On the block maxima method in extreme value theory*, arxiv.org/pdf/1310.3222.

[96] Ferreira, A., de Haan, L. & Peng, L. (2003). On optimizing the estimation of high quantiles of a probability distribution. *Statistics*, **37**:5, 401–434.

[97] Feuerverger, A. & Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, **27**, 760–781.

[98] Fisher, R.A. & Tippett, L.H.C. (1928). Limiting forms of the frequency of the largest or smallest member of a sample. *Math. Proc. Cambridge Philos. Soc.*, **24**, 180–190.

[99] Fraga Alves, M.I. (1992). The influence of central observations on discrimination among multivariate extremal models. *Theory Probab. Appl.*, **37**, 334–337.

[100] Fraga Alves, M.I. (1999). Asymptotic distribution of Gumbel statistic in a semi-parametric approach. *Port. Math.*, **56** (3), 282–298.

[101] Fraga Alves, M.I. (2001). A location invariant Hill-type estimator. *Extremes*, **4**:3, 199–217.

[102] Fraga Alves, M.I. & Gomes, M.I. (1996). Statistical choice of extreme value domains of attraction — a comparative analysis. *Comm. Statist. Theory Methods*, **25**:4, 789–811.

[103] Fraga Alves, M.I., Gomes M.I. & de Haan, L. (2003a). A new class of semi-parametric estimators of the second order parameter. *Port. Math.*, **60**:2: 193–213.

[104] Fraga Alves, M.I., Gomes, M.I., de Haan, L. & Neves, C. (2009). Mixed moment estimator and location invariant alternatives. *Extremes*, **12**, 149–185.

[105] Fraga Alves, M. I., de Haan, L. & Lin, T. (2003b). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods Statist.*, **12**, 155–176.

[106] Fraga Alves, M.I., de Haan, L. & Lin, T. (2006). Third order extended regular variation. *Publ. Inst. Math. (Beograd)*, **80** (94), 109–120.

[107] Fraga Alves, M.I. & Neves, C. (2013). Estimation of the finite right endpoint in the Gumbel domain. *Statist. Sinica*, accepted.

[108] Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polon. Math.*, **6**, 93–116.

[109] Frigessi, A., Haug, O. & Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, **5**:3, 219–235.

[110] Galambos, J. (1978; 1987). *The Asymptotic Theory of Extreme Order Statistics*, 1st ed.: Wiley, New York; 2nd ed.: Krieger, Malabar, Florida.

[111] Galambos, J. (1982). A statistical test for extreme value distributions. In Gnedenko, B.V. *et al.* (eds.), *Non-Parametric Statistical Inference*, 221–230, North-Holland, Amsterdam.

[112] Galambos, J. (1984). Rates of convergence in extreme value theory. In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*, 347–352, D. Reidel, Dordrecht.

[113] Girard, S. (2004). A Hill type estimator of the Weibull tail-coefficient. *Comm. Statist. Theory Methods*, **33**:2, 205–234.

[114] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.*, **44**, 423–453.

[115] Goegebeur, Y., Beirlant, J. & de Wet, T. (2010). Generalized kernel estimators for the Weibull-tail coefficient. *Comm. Statist. Theory Methods*, **39**, 3695–3716.

[116] Gomes, M.I. (1978). *Some Probabilistic and Statistical Problems in Extreme Value Theory*, Ph. D. Thesis, Univ. Sheffield.

[117] Gomes, M.I. (1981). An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes, In C. Taillie *et al.* (eds.), *Statistical Distributions in Scientific Work*, Vol. **6**, 389–410, D. Reidel.

[118] Gomes, M.I. (1982). A note on statistical choice of extremal models. *Actas IX Jorn. Hispano-Lusas*, 653–655, Salamanca.

[119] Gomes, M.I. (1984). Concomitants in a multidimensional extreme model. In J. Tiago de Oliveira ed., *Statistical Extremes and Applications*, 353–364, D. Reidel, Dordrecht.

[120] Gomes, M.I. (1985a). Concomitants and linear estimators in an i-dimensional extremal model. *Trabajos Estadist. Investigación Oper.*, **36**, 129–140.

[121] Gomes, M.I. (1985b). Statistical theory of extremes — comparison of two approaches. *Statist. Decision*, **2**, 33–37.

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

[123] Gomes, M.I. (1989a). Comparison of extremal models through statistical choice in multi-dimensional backgrounds. In Hüsler, J. & R.-D. Reiss (eds.), *Extreme Value Theory, Proc. Oberwolfach* 1987, Lecture Notes in Statistics 51, 191–203, Springer-Verlag.

[124] Gomes, M.I. (1989b). Generalized Gumbel and likelihood ratio test statistics in the multivariate GEV model. *Comput. Statist. Data Anal.*, **7**, 259–267.

[125] Gomes, M.I. (1994). Penultimate behaviour of the extremes, In Galambos, J. *et al.* (eds.), *Extreme Value Theory and Applications*, 403–418, Kluwer, The Netherlands.

[126] Gomes, M.I. & Alpuim, M.T. (1986). Inference in multivariate generalized extreme value models–Asymptotic properties of two test statistics. *Scand. J. Stat.*, **13**, 291–300.

[127] Gomes, M.I., Canto e Castro, L., Fraga Alves, M.I & Pestana, D. (2008a). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions. *Extremes*, **11**:1, 3–34.

[128] Gomes, M.I. & Figueiredo, F. (2006). Bias reduction in risk modelling: semi-parametric quantile estimation. *TEST*, **15**:2, 375–396.

[129] Gomes, M.I., Figueiredo, F. & Neves, M.M. (2012). Adaptive estimation of heavy right tails: resampling-based methods in action. *Extremes*, **15**, 463–489.

[130] Gomes, M.I., Fraga Alves, M.I. & Araújo Santos, P. (2008b). PORT Hill and moment estimators for heavy-tailed models. *Comm. Statist. Simulation Comput.*, **37**, 1281–1306.

[131] Gomes, M.I. & de Haan, L. (1999). Approximation by penultimate extreme-value distributions. *Extremes*, **2**:1, 71–85.

[132] Gomes, M. I., de Haan, L. & Henriques Rodrigues, L. (2008c). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **70**:1, 31–52.

[133] Gomes, M.I., de Haan, L. & Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, **5**:4, 387–414.

[134] Gomes, M.I., Hall, A. & Miranda, C. (2008d). Subsampling techniques and the Jackknife methodology in the estimation of the extremal index. *Comput. Statist. Data Anal.*, **52**:4, 2022–2041.

[135] Gomes, M.I., Henriques-Rodrigues, L., Fraga Alves, M.I. & Manjunath, B.G. (2013). Adaptive PORT-MVRB estimation: an empirical comparison of two heuristic algorithms. *J. Stat. Comput. Simul.*, **83**:6, 1129–1144.

[136] Gomes, M.I., Henriques-Rodrigues, L. & Miranda, C. (2011). Reduced-bias location-invariant extreme value index estimation: a simulation study. *Comm. Statist. Simulation Comput.*, **40**:3, 424–447.

[137] Gomes, M.I., Henriques Rodrigues, L., Vandewalle, B. & Viseu, C. (2008e). A heuristic adaptive choice of the threshold for bias-corrected Hill estimators. *J. Stat. Comput. Simul.*, **78**:2, 133–150.

[138] Gomes, M.I., Martins, M.J. & Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events: the Jackknife methodology. *Extremes*, **3**, 207–229.

[139] Gomes, M.I., Martins, M.J. & Neves, M.M. (2007a). Improving second order reduced bias extreme value index estimation. *REVSTAT*, **5**:2, 177—207.

[140] Gomes, M.I. & van Montfort, M.A.J. (1987). Exponentiality versus Generalized Pareto—quick tests. *Proc. third Int. Conf. Statistical Climatology*, 185–195.

[141] Gomes, M.I. & Neves, M.M. (2011). Estimation of the extreme value index for randomly censored data. *Biometrical Letters*, **48**:1, 1–22.

[142] Gomes, M.I. & Oliveira, O. (2001). The bootstrap methodology in statistics of extremes — Choice of the optimal sample fraction. *Extremes*, **4**:4, 331–358.

[143] Gomes, M.I. & Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.*, **102**, No. 477, 280–292.

[144] Gomes, M.I., Reiss, R.-D. & Thomas, M. (2007b). Reduced-bias estimation, In Reiss, R.-D. and Thomas, M., *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd Ed., Chapter 6, 189–204, Birkhäuser Verlag, Basel-Boston-Berlin.

[145] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resour. Res.*, **15**, 1049–1054.

[146] Grimshaw, S. (1993). Computing maximum likelihood estimates for the genaralized Pareto distribution. *Technometrics*, **35**:2, 185–191.

[147] Grinevigh, I.V. (1992a). Max-semistable laws corresponding to linear and power normalizations. *Theory Probab. Appl.*, **37**, 720–721.

[148] Grinevigh, I.V. (1992b). Domains of attraction of max-semistable laws under linear and power normalizations. *Theory Probab. Appl.*, **38**, 640–650.

[149] Groeneboom, P., Lopuha, H.P. & de Wolf, P.P. (2003). Kernel-type estimators for the extreme value index. *Ann. Statist.*, **31**, 1956–1995.

[150] Guillou, A. & Hall, P. (2001). A diagnostic for selecting the threshold in extreme-value analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **63**, 293–305.

[151] Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia Univ. Press, New York.

[152] Gumbel, E.J. (1965). A quick estimation of the parameters in Fréchet's distribution. *Rev. Inst. Internat. Statist.*, **33**, 349–363.

[153] de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Amsterdam.

[154] de Haan, L. (1984). Slow variation and characterization of domains of attraction. In Tiago de Oliveira, ed., *Statistical Extremes and Applications*, 31–48, D. Reidel, Dordrecht, Holland.

[155] de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: an Introduction*, Springer Science+Business Media, LLC, New York.

[156] de Haan, L. & Rootzén, H. (1993). On the estimation of high quantiles. *J. Statist. Plann. Inference*, **35**, 1–13.

[157] de Haan, L. & Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A*, **61**, 381–395.

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

[159] Hall, P. (1990). Using the bootstrap to estimate mean-squared error and select smoothing parameter in non-parametric problems. *J. Multivariate Anal.*, **32**, 177–203.

[160] Hall, P. & Welsh, A.W. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.*, **13**, 331–341.

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

[162] He, X. & Fung, W.K. (1999). Method of medians for lifetime data with Weibull models. *Stat. Med.*, **18**, 1993–2009.

[163] Henriques-Rodrigues, L. & Gomes, M.I. (2009). High quantile estimation and the PORT methodology. *REVSTAT*, **7**:3, 245–264.

[164] Henriques-Rodrigues, L., Gomes, M.I., Fraga Alves, M.I. & Neves, C. (2013). PORT estimation of a shape second-order parameter. *REVSTAT*, accepted.

[165] Henriques-Rodrigues, L., Gomes, M.I. & Pestana, D. (2011). Statistics of extremes in athletics. *REVSTAT*, **9**:2, 127–153.

[166] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.

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

[168] Hosking, J.R.M. (1985). Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, **34**, 301–310.

[169] Hosking, J.R.M. & Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339–349.

[170] Hosking, J.R.M., Wallis, J.R. & Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251–261.

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

[172] Hüsler, J. & Peng, L. (2008). Review of testing issues in extremes: in honor of Professor Laurens de Haan. *Extremes*, **11**:1, 99–111.

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

[174] Johansson, J. (2003). Estimating the mean of heavy-tailed distributions. *Extremes*, **6**, 91–109.

[175] Juárez, S.F. & Schucany, W.R. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes*, **7**:3, 237–251.

[176] Jurečková, J. & Picek, J. (2001). A class of tests on the tail index. *Extremes*, **4**:2, 165–183.

[177] Kaufmann, E. (2000). Penultimate approximations in extreme value theory. *Extremes*, **3**:1, 39–55.

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

[179] Lamperti, J. (1964). On extreme order statistics. *Ann. Math. Statist.*, **35**, 1726–1737.

[180] Landwher, J., Matalas, N. & Wallis, J. (1979). Probability Weighted Moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.*, **15**, 1055–1064.

[181] Leadbetter, M.R., Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer.

[182] Leadbetter, M.R. & Nandagopalan, S. (1989). On exceedance point processes for stationary sequences under mild oscillation restrictions. In Hüsler, J. & R.-D. Reiss (eds.), *Extreme Value Theory*, Springer-Verlag, 69–80.

[183] Li, D. & Peng, L. (2009). Does Bias Reduction with External Estimator of Second Order Parameter Work for Endpoint? *J. Statist. Plann. Inference*, **139**, 1937–1952.

[184] Li, D., Peng, L. & Qi, Y. (2011). Empirical likelihood confidence intervals for the endpoint of a distribution function. *TEST*, **20**, 353–366.

[185] Li, D., Peng, L. & Yang, J. (2010). Bias reduction for high quantiles. *J. Statist. Plann. Inference*, **140**, 2433–2441.

[186] Ling, C., Peng, Z. & Nadarahja, S. (2012). Location invariant Weiss-Hill estimator. *Extremes*, **15**, 197–230.

[187] Lingappaiah, G.S. (1984). Bayesian prediction regions for the extreme order statistics. *Biometrical J.*, **26**, 49–56.

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

[189] Lye, L.M., Hapuarachchi, K.P. & Ryan, S. (1993). Bayes estimation of the extreme-value reliability function. *IEEE Trans. Reliab.*, **42**, 641–644.

[190] Macleod, A.J. (1989). A remark on the algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Applied Statistics*, **38**, 198–199.

[191] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*, Wiley.

[192] Marohn, F. (1994). On testing the exponential and Gumbel distribution. In Galambos, J. et al. (eds.), *Extreme Value Theory*, 159–174, Kluwer, Dordrecht.

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

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

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

[196] Martins, E.S. & Stedinger, J.R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.*, **36**:3, 737–744.

[197] Matthys, G. & Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statist. Sinica*, **13**, 853–880.

[198] Matthys, G., Delafosse, M., Guillou, A. & Beirlant, J. (2004). Estimating catastrophic quantile levels for heavy-tailed distributions. *Insurance Math. Econom.*, **34**, 517–537.

[199] McNeil, A.J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *Astin Bull.*, **27**, 117–137.

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

[201] van Montfort, M.A.J. (1973). *An asymmetric test on the type of distribution of extremes*, Tech. Rep. 73-18, Medelelugen Landbbouwhoge-School Wageningen.

[202] van Montfort, M.A.J. & Gomes, M.I. (1985). Statistical choice of extremal models for complete and censored data. *J. Hydro.*, **77**, 77–87.

[203] van Montfort, M.A.J. & Otten, A. (1978). On testing a shape parameter in the presence of a location and scale parameter. *Math. Operationsforsch. Statist. Ser. Statist.*, **9**, 91–104.

[204] van Montfort, M.A.J. & Witter, J.V. (1985). Testing exponentiality against generalized Pareto distribution. *J. Hydro.*, **78**, 305–315.

[205] Nandagopalan, S. (1990). *Multivariate Extremes and Estimation of the Extremal Index*, Ph.D. Thesis, Univ. North Carolina at Chapel Hill.

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

[207] Neves, C. & Fraga Alves, M.I. (2008). Testing extreme value conditions — an overview and recent approaches. *REVSTAT*, **6**(1), 83–100.

[208] Neves, C., Picek, J. & 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.

[209] Otten, A. & van Montfort, M.A.J. (1978). The power of two tests on the type of distributions of extremes. *J. Hydro.*, **37**, 195–199.

[210] Pancheva, E. (1992). Multivariate max-semistable distributions. *Theory Probab. Appl.*, **37**, 794–795.

[211] Pancheva, E. (2010). Max-semistability: a survey. *ProbStat Forum*, **3**, 11–24.

[212] Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statist. Probab. Lett.*, **38**:2, 107–115.

[213] Peng, L. (2001). Estimating the mean of a heavy-tailed distribution. *Statist. Probab. Lett.*, **52**, 255–264.

[214] Peng, L. & Welsh, A.W. (2001). Robust estimation of the generalized Pareto distribution. *Extremes*, **4**:1, 53–65.

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

[216] Pickands III, J. (1994). Bayes quantile estimation and threshold selection for the generalized Pareto family. In Galambos, J., Leigh, S., & Simiu, E. (eds.), *Extreme Value Theory and Applications*, 123–138, Kluwer, Amsterdam.

[217] Prescott, P. & Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, **67**, 723–724.

[218] Prescott, P. & Walden, A.T. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. *J. Statist. Comput. Simulation*, **16**, 241–250.

[219] Raoult, J.P. & Worms, R. (2003). Rate of convergence for the generalized Pareto approximation of the excesses. *Adv. in Appl. Probab.*, **35**:4, 1007–1027.

[220] Reis, P., Canto e Castro, L., Dias, S. & Gomes, M.I. (2013). Penultimate approximations in Statistics of Extremes and reliability of large coherent systems. *Methodol. Comput. Appl. Probab.*, DOI: 10.1007/s11009-013-9338-7.

[221] Reiss, R.-D. & Thomas, M. (1997; 2001; 2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*, 1st ed.; 2nd ed.; 3rd ed., Birkhäuser Verlag.

[222] Reiss, R.-D. & Thomas, M. (1999). A new class of Bayesian estimators in paretian excess-of-loss reinsurance. *Astin Bull.*, **29**:2, 339–349.

[223] Resnick, S.I. (1987). *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, New York.

[224] Resnick, S.I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.*, **25**:5, 1805–1869.

[225] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer-Verlag.

[226] Rootzén, H. & Tajvidi, N. (1997). Extreme value statistics and wind storm losses: a case study. *Scand. Actuar. J.*, **1**, 70–94.

[227] Scarrot, C. & McDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT*, **10**:2, 33–60.

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

[229] Singh, V.P. & Guo, H. (1997). Parameter estimation for 2-parameter generalized Pareto distribution by POME. *Stochastic Hydrology and Hydraulics*, **11**:2, 211–227.

[230] Smith, R.L. (1984). Threshold methods for sample extremes. In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*, 621–638, D. Reidel, Dordrecht.

[231] Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.

[232] Smith, R.L. (1986). Extreme value theory based on the r largest annual events. *J. Hydro.*, **86**, 27–43.

[233] Smith, R.L. (1987a). Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174–1207.

[234] Smith, R.L. (1987b). *Approximations in extreme value theory*, Preprint, Univ. North-Carolina.

[235] Smith, R.L. (1999). Bayesian and frequentist approaches to parametric predictive inference. In Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F. (eds.), 589–612, Oxford Press, U.K.

[236] Smith, R.L. & Goodman, D.J. (2000). Bayesian risk analysis. In Embrechts, P. (ed.), *Extremes and Integrated Risk Management*, 235–251, Risk Books, London.

[237] Smith, R.L. & Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, **36**, 358–369.

[238] de Sousa, B. & Michailids, G. (2004). A Diagnostic Plot for Estimating the Tail Index of a Distribution. *J. Comput. Graph. Statist.*, **13**:4, 974–995.

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

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

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

[242] Stephenson, A. & Tawn, J. (2004). Bayesian inference for extremes: accounting for the three extremal types. *Extremes*, **7**:4, 291–307.

[243] Tancredi, A., Anderson, C. & O'Hagan, T. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, **9**, 87–106.

[244] Tawn, J. (1988). An extreme value theory model for dependent observations. *J. Hydro.*, **101**, 227–250.

[245] Temido, M.G. & Canto e Castro, L. (2003). Max-semistable laws in extremes of stationary random sequences. *Theory Probab. Appl.*, **47**:2, 365–374.

[246] Tiago de Oliveira, J. (1981). Statistical choice of univariate extreme models. In C. Taillie et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. **6**, 367–387.

[247] Tiago de Oliveira, J. (1984). Univariate extremes: statistical choice. In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Application*, 91–107, D. Reidel, Dordrecht.

[248] Tiago de Oliveira, J. & Gomes, M.I. (1984). Two test statistics for choice of univariate extreme models. In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Application*, 651–668, D. Reidel, Dordrecht.

[249] Vandewalle, B., Beirlant, J., Christmann, A. & Hubert, M. (2007). A robust estimator for the tail index of Pareto-type distributions. *Comput. Statist. Data Anal.*, **51**, 6252–6268.

[250] Walshaw, D. (2000). Modelling extreme wind speeds in regions prone to hurricanes. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, **49**, 51–62.

[251] 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.

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

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

[254] Weissman, I. (1984). Statistical estimation in extreme value theory. In Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*, 109–115, D. Reidel.

[255] de Zea Bermudez, P. & Amaral-Turkman, M.A. (2003). Bayesian approach to parameter estimation of the generalized pareto distribution. *TEST*, **12**, 259–277.

[256] de Zea Bermudez, P., Amaral-Turkman, M.A. & Turkman, K.F. (2001). A predictive approach to tail probability estimation. *Extremes*, **4**:4, 295–314.

[257] Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivariate Anal.*, **100**:4, 794–815.

[258] Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *J. Multivariate Anal.*, **101**:4, 971–983.