Threshold Selection in Extreme Value Analysis*

Frederico Caeiro
D.M. and C.M.A., Universidade Nova de Lisboa
and
M. Ivette Gomes
D.E.I.O. and C.E.A.U.L., Universidade de Lisboa

September 1, 2014

Abstract. The main objective of *statistics of extremes* is the prediction of rare events, and its primary problem has been the estimation of the *extreme value index* (EVI). Whenever we are interested in large values, such estimation is usually performed on the basis of the largest $k+1$ order statistics in the sample or on the excesses over a high level u. The question that has been often addressed in practical applications of *extreme value theory* is the choice of either k or u, and an adaptive EVI-estimation. Such a choice can be either heuristic or based on sample paths stability or on the minimization of a mean squared error estimate as a function of k. Some of these procedures will be reviewed. Despite of the fact that the methods provided can be applied, with adequate modifications, to any real EVI and not only to the adaptive EVI-estimation but also to the adaptive estimation of other relevant right-tail parameters, we shall illustrate the methods essentially for the EVI and for heavy tails, i.e., for a positive EVI.

Keywords and phrases. *Bootstrap methodology, heuristic methods, optimal sample fraction, sample-paths stability, semi-parametric estimation, statistical theory of extremes*

1 Introduction

As usual, let us consider that we have access to a sample $(X_1, \ldots, X_n)$ of *independent, identically distributed* (IID) *random variables* (RVs), or possibly weakly dependent and stationary RVs from an underlying population with unknown *cumulative distribution function* (CDF) $F(\cdot)$. Let us further use the notation $(X_{1:n} \le \cdots \le X_{n:n})$ for the associated sample of ascending *order statistics* (OSs).

*Research partially supported by National Funds through **FCT** — Fundação para a Ciência e a Tecnologia, projects PEst-OE/MAT/UI0006/2014 (CEAUL) and PEst-OE/MAT/UI0297/2014 (CMA/UNL).

The main objective of *statistics of univariate extremes* (SUE) is the prediction of rare events, and thus the need for an adequate estimation of parameters such as *high quantiles*, *return periods*, the *extremal index*, and many other parameters or functionals related to natural 'disasters'. The primary question has however been for a long time the estimation of the *extreme value index* (EVI), either for large or small values. Anyway, as $\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n)$, results for small values can be easily derived from the analogue results for large events, and we shall deal with the *right tail function* (RTF),

$$\overline{F}(x) := 1 - F(x),$$

of the underlying CDF, F. We are thus interested in the parameter ξ in the extreme value (EV) CDF,

$$G_\xi(x) = \begin{cases} \exp(-(1+\xi x))^{-1/\xi}, & \text{if} \quad \xi \neq 0, \\ \exp(-\exp(-x)), & \text{if} \quad \xi = 0. \end{cases} \tag{1.1}$$

One of the most recent and general approaches to SUE is the semi-parametric one, where it is merely assumed that F belongs to the max-domain of attraction (MDA) of G_ξ, in (1.1). The estimation of ξ is then based on the largest $k+1$ OSs in the sample or on the excesses over a high level u, requiring thus the choice of a threshold at which the tail of the underlying distribution begins. The question that has been often addressed in practical applications of *extreme value theory* (EVT) is thus the choice of either k or u and an adaptive estimation of the parameter of interest. Despite of similar, we shall here address the choice of k. Anyway, we can always consider that u is chosen in the interval $[X_{n-k:n}, X_{n-k+1:n})$ or even that $u = X_{n-k:n}$. Then, our functionals of u can be transformed in functionals of an integer k in $[1, n)$, related to the number of top OSs to be used in the estimation.

Ideally, the estimates as a function of k, should not be highly sensitive to small changes in the threshold. Unfortunately, this is not always the case, and the most trivial example of this is the Hill estimator, $\hat{\xi}^H_{k,n}$, the classical EVI-estimator of a positive EVI (Hill, [41]). For the Hill EVI-estimator an inadequate choice of k can lead to large expected errors, essentially due to the fact that for small values of k, $\hat{\xi}^H_{k,n}$ has a large variance, whereas large values of k usually induce a high bias in $\hat{\xi}^H_{k,n}$. This leads to a very peaked *mean square error* (MSE), as a function of k. The problem of excessive oscillation of the Hill EVI-estimator has been discussed by Resnick and Stărică ([47]), who recommend the smoothing of the Hill-estimator by integrating it over a moving window (see also, Resnick, [49] and Martins, Gomes and Neves, [43]). Kernel EVI-estimators (Csörgő, Deheuvels, and Mason, [15]; Groeneboom, Lopuha, and de Wolf, [35]) and semi-parametric probability-weighted-moment (PWM) EVI-estimators, like the ones in Caeiro and Gomes ([9]) are also nice alternatives, usually providing much more less volatile sample paths as a function of k. And nowadays, at the current state-of-the-art, and particularly for heavy tails, there are several reduced-bias (RB) alternatives to the Hill EVI-estimator that are easy to use and much less sensitive to the choice of k. For recent overviews on RB estimation of parameters of extreme events, see Reiss and Thomas ([46]), Beirlant, Caeiro and Gomes ([2]) and Gomes and Guillou ([26]). Practitioners using this type of semi-parametric EVI-estimators can then get close to optimal EVI-estimates by using an estimator that is smooth in the tuning parameter k and that thus facilitates the threshold selection.

In any case, the problem of choosing the threshold k is still of high theoretical and practical interest. The existing methods rely either on heuristic and most of the times graphical procedures for threshold selection based on sample paths' stability as a function of k or minimization of MSE's estimates, also as functions of k. Most of these methods require the estimation of second-order parameters. Some of these procedures will be reviewed and applied to an adequate generated sample from a known model $F(\cdot)$. Although the methods provided may be applied, with adequate modifications, to the general MDA of the *extreme value* (EV) distribution, i.e. to a real EVI, and not only to the adaptive EVI-estimation but also to the adaptive estimation of any parameter of extreme events, like a high quantile, the return period and a probability of exceedance of a high level, among other relevant parameters, we shall illustrate the methods essentially for the EVI and for heavy tails, i.e., for a positive EVI.

The most common methods of adaptive choice of the threshold k are based on the minimization of some kind of MSE's estimates. We mention the pioneering papers by Hall ([37]) and Hall and Welsh ([39]), and discuss the role of the bootstrap methodology in the selection of $k_0(n) := \arg\min_k \mathrm{MSE}(\hat{\xi}_{k,n})$ (Hall, [38]; Draisma, de Haan, Peng and Themido Pereira [19]; Danielsson, de Haan, Peng and de Vries, [17]; Gomes and Oliveira, [32]; Gomes, Mendonça and Pestana, [31]; Gomes, Figueiredo and Neves, [25]). The regression methodologies developed in Beirlant Vynckier and Teugels ([8], [7]) and Beirlant, Dierckx, Goegebeur and Segers ([4]) have also the objective of minimization of the MSE. In [4], the exponential regression model (ERM), introduced in Beirlant, Dierckx, Goegebeur and Matthys ([3]), is considered and applications of the ERM to the selection of the optimal sample fraction, k/n, in EV estimation is discussed. A connection between the new choice strategy in that paper and the diagnostic proposed in Guillou and Hall ([36]), based on bias behaviour, is also provided. Drees and Kaufmann ([21]) proposal for the selection of the optimal sample fraction is also partially based on bias properties. Csörgő and Viharos ([16]) also provide a data-driven choice of k for the kernel class of estimators. Coles

([14]) outlines the common graphical diagnostics for threshold choice: mean residual life (or mean excess) plots, threshold stability plots, and all the usual distribution fit diagnostics, like probability plots, quantile plots and return level plots, among others. See also Beirlant, Goegebeur, Segers and Teugels ([5]), and references therein. For comparisons of different adaptive procedures on the basis of extensive small sample simulations, see Matthys and Beirlant ([44]), Gomes and Oliveira ([32]) and Beirlant, Dierckx, Guillou and Stărică ([4]). A recent overview of the topic can be found in Scarrot and MacDonald ([50]).

The optimal sample fraction for second-order reduced-bias estimators is still giving its first steps. Possible heuristic choices are provided in Gomes and Pestana ([33]), Gomes, Henriques Rodrigues, Vandewalle and Viseu ([29]), Gomes, Henriques-Rodrigues, and Miranda ([28]), Figueiredo, Gomes, Henriques-Rodrigues and Miranda ([22]), Gomes, Henriques-Rodrigues, Fraga Alves, and Manjunath [27] and Neves, Gomes, Figueiredo and Prata-Gomes ([45]), among others.

In Section 2, after providing a few details on classical EVI-estimation for heavy RTFs, together with a direct threshold estimation, we refer the estimation of second-order parameters, providing an algorithm for the estimation of (β, ρ) in the RTF associated with the reciprocal tail quantile function (RTQF),

$$U(t) := F^{\leftarrow}(1 - 1/t) = Ct^{\xi}\left(1 + \xi\beta t^{\rho}/\rho + o(t^{\rho})\right), \tag{1.2}$$

as $t \to \infty$, with $C, \xi > 0$, $\beta \in \mathbb{R}$, $\rho < 0$ and $F^{\leftarrow}(x) := \inf\{y : F(y \geq x\}$ denoting the generalized inverse function of F. In Section 3, we review a few graphical tools for threshold selection. In Section 4, we make a review of some of the methods related to sample paths' stability, as functions of k. Section 5 is dedicated to methods that deal with the minimization of MSE's estimates, also as functions of k. In Section 6, we deal with a few other heuristic procedures devised for an adaptive EVI-estimation. Finally, in Section 7, we provide applications of the Algorithms to a simulated sample.

2 A few details on classical EVI-estimators for heavy right tails, direct estimation of the threshold and second-order parameters' estimation

Note first that a necessary and sufficient condition to have F in the MDA of G_ξ, in (1.1), with $\xi > 0$, is that $\overline{F} \in \mathcal{RV}_{-1/\xi}$ or equivalently that $U \in \mathcal{RV}_\xi$, where $\mathcal{RV}_{\mathrm{a}}$ denotes the class of regularly varying functions at infinity, with an index of regular variation a, i.e. positive measurable functions g such that $g(tx)/g(t) \to x^{\mathrm{a}}$, as $t \to \infty$. For such an MDA we shall use the notation

$\mathcal{D}_{\mathcal{M}}^{+} \equiv \mathcal{D}_{\mathcal{M}|1}^{+}$. We shall thus assume the validity of the so-called first-order condition,

$$F \in \mathcal{D}_{\mathcal{M}}^{+} \equiv \mathcal{D}_{\mathcal{M}|1}^{+} \quad \Longleftrightarrow \quad \overline{F} \in \mathcal{RV}_{-1/\xi} \quad \Longleftrightarrow \quad U \in \mathcal{RV}_{\xi}. \tag{2.1}$$

One of the pioneering classes of semi-parametric estimators of a positive EVI, usually known as Hill (H) EVI-estimators, was considered in [41]. Hill's estimators are based on the log-excesses over an OS $X_{n-k:n}$, $1 \leq k < n$, and have the functional form

$$\hat{\xi}_{k,n}^{\mathrm{H}} \equiv \hat{\xi}_{k,n}^{\mathrm{H}}(\underline{X}_n) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\}. \tag{2.2}$$

Consistency is achieved in the whole $\mathcal{D}_{\mathcal{M}|1}^{+}$, provided that $X_{n-k:n}$ is an *intermediate* OS, i.e. we need to have

$$k = k_n \to \infty \quad \text{and} \quad k/n \to 0, \ \text{as}\ n \to \infty.$$

In order to derive the asymptotic normality of the EVI-estimators in (2.2), among others, it is convenient to slightly restrict the class $\mathcal{D}_{\mathcal{M}|1}^{+}$, assuming the validity of a second-order condition either on $\overline{F}$ or on U. We now guarantee the existence of a function $A(t)$, going to zero as $t \to \infty$, such that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \xi \ln x}{A(t)} = \begin{cases} \frac{x^{\rho}-1}{\rho}, & \text{if} \quad \rho < 0, \\ \ln x, & \text{if} \quad \rho = 0, \end{cases} \tag{2.3}$$

where $\rho \leq 0$ is a second-order parameter, which measures the rate of convergence in the first-order condition, in (2.1). For such a class of models, we use the notation $\mathcal{D}_{\mathcal{M}|2}^{+}$. If the limit in the left hand side of Eq. (2.3) exists, it is necessarily of the above mentioned type and $|A| \in \mathcal{RV}_{\rho}$ (Geluk and de Haan, [24]). If we assume the validity of the second-order framework in (2.3), the aforementioned EVI-estimators are asymptotically normal, provided that $\sqrt{k}A(n/k) \to \lambda_{_A}$, finite, as $n \to \infty$, with A given in (2.3). Indeed, if we denote $\hat{\xi}_{k,n}^{\mathrm{C}}$, either the Hill estimator in (2.2) or any other classical (C) EVI-estimator, we have, with Z_k^{C} asymptotically standard normal and for adequate $(b_{_\mathrm{C}},\ \sigma_{_\mathrm{C}}) \in (\mathbb{R},\ \mathbb{R}^+)$, the validity of the asymptotic distributional representation

$$\hat{\xi}_{k,n}^{\mathrm{C}} \overset{d}{=} \xi + \xi\sigma_{_\mathrm{C}} Z_k^{\mathrm{C}}/\sqrt{k} + b_{_\mathrm{C}}\ A(n/k)(1 + o_p(1)), \quad \text{as} \quad n \to \infty. \tag{2.4}$$

In this paper, we shall often further assume that $\rho < 0$, in (2.3), and we shall use the following parameterization in $(\xi, \beta, \rho) \in (\mathbb{R}^+, \mathbb{R} - \{0\}, \mathbb{R}^-)$,

$$A(t) =: \xi\beta t^{\rho} \tag{2.5}$$

for the function A, in (2.3), possibly with $\beta = \beta(t)$, a slowly varying function. Then (1.2) holds, and in the lines of Caeiro, Gomes and Pestana ([12]), if

b_{C} depends only on (β, ρ), i.e. $b_{\mathrm{C}} = b_{\mathrm{C}}(\beta, \rho)$, as happens with the H EVI-estimators, we can built a RB classical (RBC) EVI-estimator associated with the C EVI-estimator, given by

$$\hat{\xi}^{\mathrm{RBC}}_{k,n} = \hat{\xi}^{\mathrm{C}}_{k,n} \left(1 - b_{\mathrm{C}}(\hat{\beta}, \hat{\rho})\ \hat{\beta}(n/k)^{\hat{\rho}}\right), \tag{2.6}$$

with $(\hat{\beta}, \hat{\rho})$ an adequate estimator of the vector of second-order parameters (β, ρ), in (1.2). And nowadays, we know how to estimate (β, ρ), externally and a bit more than consistently, in the sense the we need to have $\hat{\rho} - \rho = o_p(1/\ln n)$, so that we decrease the bias keeping the variance, i.e. (2.4) holds with $b_{\mathrm{C}} = 0$. Under an adequate third-order condition (see Caeiro *et al.* ([11]), we then get the validity of the asymptotic distributional representation

$$\hat{\xi}^{\mathrm{RBC}}_{k,n} \stackrel{d}{=} \xi + \xi \sigma_{\mathrm{C}} Z^{\mathrm{C}}_k/\sqrt{k} + b_{\mathrm{RBC}}\ A^2(n/k)(1 + o_p(1)), \quad \text{as} \quad n \to \infty. \tag{2.7}$$

With $\mathbb{E}$ denoting the mean value operator, and for models in (1.2), a possible substitute for the MSE of any classical EVI-estimator $\hat{\xi}^{\mathrm{C}}_{k,n}$ is, according with the asymptotic distributional representation in (2.4),

$$\mathrm{AMSE}(\hat{\xi}^{\mathrm{C}}_{k,n}) = \xi^2 \big(\sigma^2_{\mathrm{C}}/k + b^2_{\mathrm{C}}\ \beta^2\ (n/k)^{2\rho}\big),$$

depending on n and k, and with AMSE standing for *asymptotic mean square error*. For the RBC EVI-estimator, in (2.6), and on the basis of (2.7), we obviously have

$$\mathrm{AMSE}(\hat{\xi}^{\mathrm{RBC}}_{k,n}) = \xi^2 \big(\sigma^2_{\mathrm{C}}/k + b^2_{\mathrm{RBC}}\ \beta^4\ (n/k)^{4\rho}\big).$$

Then, with $\bullet$ denoting either C or RBC, with $\sigma = \sigma_{\mathrm{C}}$, and considering

$$c_\bullet = \begin{cases} 2 & \text{if} \quad \bullet = \mathrm{C} \\ 4 & \text{if} \quad \bullet = \mathrm{RBC}. \end{cases} \tag{2.8}$$

we get

$$\begin{aligned} k_{0|\bullet}(n) &:= \arg\min_k \mathrm{AMSE}\big(\hat{\xi}^{\bullet}_{k,n}\big) \\ &= \big((-c_\bullet \rho)\ b^2_\bullet\ \beta^{c_\bullet}\ n^{c_\bullet \rho}/\sigma^2\big)^{-1/(1-c_\bullet \rho)} \\ &= k^\bullet_0(n)(1 + o(1)), \end{aligned}$$

with $k^\bullet_0(n) := \arg\min_k \mathrm{MSE}\big(\hat{\xi}^\bullet_{k,n}\big)$. This intuitively suggests the estimator

$$\hat{k}^\bullet_0 := \left\lfloor \left(\frac{\hat{\sigma}^2 n^{-c_\bullet \hat{\rho}}}{(-c_\bullet \hat{\rho}\ \hat{b}^2_\bullet\ \hat{\beta}^{c_\bullet})}\right)^{1/(1-c_\bullet \hat{\rho})} \right\rfloor, \tag{2.9}$$

with $\lfloor x \rfloor$ denoting the integer part of x.

For the Hill estimator, we have, in (2.4), $\sigma_{_H} = 1$ and $b_{_H} = 1/(1-\rho)$. Consequently, with $(\hat\beta, \hat\rho)$ any consistent estimator of the vector (β, ρ) of second-order parameters, Eq. (2.9) justifies asymptotically the estimator suggested in Hall ([37]), given by

$$\hat{k}_0^H := \left\lfloor \left(\frac{(1-\hat\rho)^2 n^{-2\hat\rho}}{-2\hat\rho\hat\beta^2} \right)^{1/(1-2\hat\rho)} \right\rfloor. \tag{2.10}$$

At the current state-of-the-art, the estimation of the second-order parameters β and ρ in (1.2), and also in (2.5), can be considered almost trivial. Despite of the great variety of recent classes of estimators now available in the literature, we suggest the use of the ρ–estimators in Fraga Alves *et al.* ([23]) and the β–estimators in Gomes and Martins ([30]). The second-order parameters' estimates can be obtained through algorithms already used in several articles, like Gomes and Pestana ([33]), among others. We can thus consider the following algorithm for the estimation of the threshold:

Algorithm 1.

Step 1. *Given the sample* $(x_1, \ldots, x_n)$*, compute for the tuning parameters* $\tau = 0$ *and* $\tau = 1$*, the observed values of* $\hat\rho_\tau(k)$*, the most simple class of estimators in* [23]*. Such estimators have the functional form*

$$\hat\rho_\tau(k) := -\left|3(W_{k,n}^{(\tau)} - 1)/(W_{k,n}^{(\tau)} - 3)\right|,$$

dependent on the statistics

$$W_{k,n}^{(\tau)} := \begin{cases} \dfrac{\left(M_{k,n}^{(1)}\right)^{\tau} - \left(M_{k,n}^{(2)}/2\right)^{\tau/2}}{\left(M_{k,n}^{(2)}/2\right)^{\tau/2} - \left(M_{k,n}^{(3)}/6\right)^{\tau/3}}, & \textit{if} \quad \tau \neq 0, \\[2ex] \dfrac{\ln M_{k,n}^{(1)} - \ln\left(M_{k,n}^{(2)}/2\right)/2}{\ln\left(M_{k,n}^{(2)}/2\right)/2 - \ln\left(M_{k,n}^{(3)}/6\right)/3}, & \textit{if} \quad \tau = 0, \end{cases}$$

where

$$M_{k,n}^{(j)} := \frac{1}{k} \sum_{i=1}^{k} \Big(\ln X_{n-i+1:n} - \ln X_{n-k:n} \Big)^j, \; j = 1, 2, 3.$$

Step 2. *Consider* $\mathcal{K} = \big(\lfloor n^{0.995} \rfloor, \; \lfloor n^{0.999} \rfloor\big)$*. Compute the median of* $\{\hat\rho_\tau(k)\}_{k\in\mathcal{K}}$*, denoted* χ_τ*, and compute* $I_\tau := \sum_{k\in\mathcal{K}} \left(\hat\rho_\tau(k) - \chi_\tau\right)^2$*,* $\tau = 0, 1$*. Next choose the tuning parameter* $\tau^* = 0$ *if* $I_0 \le I_1$*; otherwise, choose* $\tau^* = 1$*.*

Step 3. *Work with* $\hat\rho \equiv \hat\rho_{\tau^*} = \hat\rho_{\tau^*}(k_{01})$ *and* $\hat\beta \equiv \hat\beta_{\tau^*} := \hat\beta_{\hat\rho_{\tau^*}}(k_{01})$*, with* $k_{01} = \lfloor n^{0.999} \rfloor$*, being* $\hat\beta_{\hat\rho}(k)$ *the estimator in* [30]*, given by*

$$\hat\beta_{\hat\rho}(k) := \left(\frac{k}{n}\right)^{\hat\rho} \frac{d_k(\hat\rho)\; D_k(0) - D_k(\hat\rho)}{d_k(\hat\rho)\; D_k(\hat\rho) - D_k(2\hat\rho)},$$

dependent on the estimator $\hat\rho = \hat\rho_{\tau^*}(k_{01})$*, and where, for any* $\alpha \le 0$*,*

$$d_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k}(i/k)^{-\alpha} \quad \textit{and} \quad D_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k}(i/k)^{-\alpha}\,U_i,$$

with $U_i = i\,\left(\ln X_{n-i+1:n} - \ln X_{n-i:n}\right)$*,* $1 \le i \le k < n$*, the scaled log-spacings.*

Step 4. *On the basis of* $(\hat\beta, \hat\rho)$*, in* **Step 3.***, compute* $\hat k_0^{\mathrm{H}}$*, in* (2.10)*.*

Step 5. *Finally obtain* $\hat\xi^{\mathrm{H}} := \hat\xi^{\mathrm{H}}_{\hat k_0^{\mathrm{H}},n}$*.*

Note that if the second side of Eq. (2.9) depends on ξ, due to the fact that $(\sigma/b_\bullet)^2 = \varphi_\bullet(\xi)$, we can also try solving this equation numerically through the fixed point method, replacing **Steps 4.** and **5.**, in **Algorithm 1**, by:

Step 4' *Compute an initial estimate* $\hat\xi_0^\bullet = \hat\xi^\bullet_{\hat k^\bullet_{0,0},n}$*,* $\hat k^\bullet_{0,0} = \lfloor\sqrt{n}\rfloor$*.*

Step 5' *For* $i \ge 1$*, and with* $c_\bullet$ *defined in Eq.* (2.8)*, consider the iterative procedure*

$$\hat k^\bullet_{0,i} := \left\lfloor \left(\frac{\varphi^2_\bullet(\hat\xi^\bullet_{i-1})n^{-c_\bullet\hat\rho}}{-c_\bullet\hat\rho\;\hat\beta^{c_\bullet}}\right)^{1/(1-c_\bullet\hat\rho)} \right\rfloor, \qquad \hat\xi^\bullet_i = \hat\xi^\bullet_{\hat k^\bullet_{0,i},n},$$

and stop the procedure if we get $\hat k^\bullet_{0,i} = \hat k^\bullet_{0,i-1} =: \hat k^\bullet_0$*.*

Step 6' Finally obtain $\hat\xi^{\bullet|H} := \hat\xi^\bullet_{\hat k^\bullet_0,n}$.

Remark 1.
Note that this procedure works essentially for classical EVI*-estimation. Indeed, since the estimation of the third-order parameters is still an almost open topic in* SUE *and* $\sigma/b_\bullet$ *depends usually on those parameters, we cannot yet directly estimate the optimal number of top* OSs *for* $\hat\xi^{\mathrm{RBH}}_{k,n}$

3 Graphical threshold selection—Zipf, Hill, empirical mean-excess and sum-plots

Zipf, Hill, empirical mean-excess and sum plots can be found in Beirlant, Goegebeur, Segers and Teugels ([5]), among others.

3.1 The Zipf-plot

In the Zipf-plot, the quantity $\ln((n+1)/i)$ is plotted against $X_{n-i+1:n}$, for $i = 1, \ldots, n$, and the estimate of ξ is given by the least squares estimate of the slope of the part of the plot that exhibits a linear behaviour. A nice feature of the Zipf-plot is that the lack of linearity at the left part of the graph suggests departure from a Pareto tail behaviour. Based on the Zipf-plot, Kratz and Resnick ([42]) considered the QQ-plot as an alternative to the Hill-plot, a plot of $\hat{\xi}^H_{k,n}$, in (2.2), versus k.

3.2 The Hill-plot

Just as mentioned above, the Hill-plot is based on the graph of $\hat{\xi}^H_{k,n}$, in (2.2), against k. The value of ξ is inferred by identifying a stable horizontal region in the graph. If data comes from a Pareto CDF, the ideal case, the Hill-plot, as well as the Zipf-plot, perform quite satisfactorily, almost independent of the threshold, allowing the data analyst to estimate correctly the EVI. But this is not always the case, and we can easily come to the so-called Hill "horror" plots (Resnick, [48]). Just as recommended by Drees, de Haan and Resnick ([20]), it is more informative to conceive the Hill-plot on a log-scale rather than on a linear scale. As illustrated in de Sousa and Michailidis ([18]), and despite of the fact that the QQ-plot tends to be smoother than the Hill-plot, problems are still easy to detect when departures from the Pareto CDF occur.

3.3 The mean-excess-plot

The mean-excess-plot is a quite common graphical tool in the field of insurance. Such a plot is related to the mean excess function (MEF)

$$e(t) := \mathbb{E}(X - t | X > t). \tag{3.1}$$

Given a sample $(x_1, \ldots, x_n)$, the MEF, in (3.1), is easily estimated by

$$\hat{e}_n(t) := \frac{\sum_{i=1}^n x_i \mathrm{I}_{\{t,\infty\}}(x_i)}{\sum_{i=1}^n \mathrm{I}_{\{t,\infty\}}(x_i)} - t,$$

where, as usual, $\mathrm{I}_{\mathcal{A}}(x) = \begin{cases} 1 & x \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$ is the indicator function of the set $\mathcal{A}$. The empirical function $\hat{e}_n$ is often plotted at the values $t = x_{n-k:n}$, $1 \le k < n$, the $(k+1)$-top OS. We then get

$$\hat{e}_{k,n} := \hat{e}_n(x_{n-k:n}) = \frac{1}{k}\sum_{j=1}^{k} x_{n-j+1:n} - x_{n-k:n}. \tag{3.2}$$

Note that $\hat{e}_{k,n}$ can also be interpreted as an estimate of the slope of the exponential QQ-plot to the right of a reference point with coordinates

$(-\ln((k+1)/(n+1)), x_{n-k:n})$. The MEF is constant for an exponential model. When the underlying CDF has a heavier tail than the exponential CDF, the MEF ultimately increases. For lighter tails the MFE ultimately decreases. To better understand the behaviour of the MEF, see Beirlant *et al.* ([5]). A mean-excess-plot is a plot of $\hat{e}_{k,n}$, in (3.2), either versus k or versus $x_{n-k:n}$.

3.4 Sum-plots

The sum plots were devised, among others, in de Sousa and Michailidis ([18]), for the Hill EVI-estimators, in Henry III ([40]), for the harmonic mean EVI-estimators and in Beirlant, Boniphace and Dierckx ([1]), for EVI-estimators that use a set of extreme OSs in the estimation of a real EVI. They are based on the plot $(k, \hat{S}^{\bullet}_{k,n} := k\hat{\xi}^{\bullet}_{k,n})$, which should be approximately linear for the k-values where $\hat{\xi}^{\bullet}_{k,n} \approx \xi$. The slope of the linear part of the graph $(k, \hat{S}^{\bullet}_{k,n})$ can then be used as an estimator of ξ. One can argue that the plots $(k, \hat{S}^{\bullet}_{k,n})$ and $(k, \hat{\xi}^{\bullet}_{k,n})$ are statistical equivalent. The sum-plot naturally leads to the estimation of the slope and the plot $(k, \hat{\xi}^{\bullet}_{k,n})$, which should be horizontal, leads to the estimation of the intercept. Taking the Hill as an example, whereas the Hill estimator can be seen as an estimator of the slope in a Pareto quantile plot (see, Beirlant, Teugels and Vynckier [6], and Kratz and Resnick, [42], among others), the Hill sum-plot can be viewed as a regression plot of $\hat{S}^{\mathrm{H}}_{k,n}$ on k. In [18] it is provided an algorithm for identifying k from the Hill sum-plot. Such algorithm is compared with two algorithms, in Beirlant, Dierck and Stărică ([4]) and Matthys and Beirlant ([44]), both based on AMSE minimization of the Hill EVI-estimators. The mispecification of the second-order parameter at $\rho = -1$ is considered, and the conclusion is that there does not exist a uniformly better procedure for all values of ξ and all possible underlying parents.

For other common graphical diagnostics for threshold choice, see Coles ([14]), de Sousa and Michailidis ([18]) and Wager ([51]), among others.

4 Threshold selection partially based on bias behaviour

4.1 Drees and Kaufmann's choice

Drees and Kaufmann ([21]) proposed a procedure partially based on bias and MSE minimization. The method was studied computationally in [32]. Apart from the need to estimate ρ, which relies on the choice of a tuning parameter, there are two other tuning parameters, and the method is quite sensitive to the choice of these parameters. However, the suggestions given by Drees and Kaufmann seem appropriate, and the method has an overall positive performance, as can be seen in the comparative simulation study performed in [32].

4.2 Guillou and Hall's choice

Guillou and Hall ([36]) suggest an interesting approach for choosing the threshold when fitting the Hill estimator of a tail exponent to EV data. The argument is the following: for the Hill estimator, we have the validity of the distributional representation in (2.4), with $\sigma_H = 1$ and $b_H = 1/(1-\rho)$, and the AMSE of $\sqrt{k}\ \hat{\xi}^{\mathrm{H}}_{k,n}$ is equal to $\xi^2(1+\mu_k^2)$ where

$$\mu_k = \sqrt{k}\ A(n/k)/(\xi(1-\rho)).$$

Then, the value k_0 of k that is optimal in the sense of minimizing the AMSE of $\xi^{\mathrm{H}}_{k,n}$ may be taken as a solution of

$$\mu^2_{k_0} = 1/\sqrt{-2\rho}.$$

Similarly to what happens in the bootstrap methodology of Draisma, de Haan, Peng and Pereira ([19]) and Danielson, de Haan, Peng and de Vries ([17]), the idea underlying [36] is to replace the random variable $\sqrt{k}\{\hat{\xi}^{\mathrm{H}}_{k,n} - \xi\}/\xi$, by a statistic which also converges towards 0 and has similar bias properties. Such a statistic is merely a linear combination of the log-spacings

$$U_i = i\left\{\ln X_{n-i+1:n} - X_{n-i:n}\right\},\ 1 \le i \le k < n,$$

with null mean value. More precisely, the main role in this diagnostic technique is played by the auxiliary statistics

$$T_n(k) := \sqrt{\frac{3}{k^3}}\ \frac{\sum\limits_{i=1}^{k}(k-2i+1)\ U_i}{\frac{1}{k}\sum\limits_{i=1}^{k} U_i}, \tag{4.1}$$

for $1 \le k < n$, together with a moving average of its squares,

$$Q_n(k) := \left\{\frac{1}{2\lfloor k/2\rfloor + 1}\sum_{j=k-\lfloor k/2\rfloor}^{k+\lfloor k/2\rfloor} T_n^2(j)\right\}^{1/2}, \tag{4.2}$$

which dampens stochastic fluctuations of $T_n(k)$, in (4.1). On the basis of an ad hoc choice of c_{crit}, like for instance the value 1.25, or any value in [1.25, 1.5], as suggested in [36], together with the auxiliary statistic in (4.2), we can thus consider:

Algorithm 2.

Step 1. *Given the sample* $(x_1, \ldots, x_n)$*, compute the observed value of* $Q_n(k)$*, in* (4.2)*, for* $1 \leq k$ *and* $k + \lfloor k/2 \rfloor < n$.

Step 2. *Given* $c_{crit} = 1.25$*, consider the choice*

$$k_0^{\mathrm{GH}} := \inf\{k : |Q_n(j)| \geq c_{crit},\ \forall j \geq k\}.$$

Step 3. *Finally obtain* $\hat{\xi}^{\mathrm{GH}} := \hat{\xi}^{\mathrm{GH}}_{\hat{k}_0^{\mathrm{GH}},n}$.

For a more objective choice of the nuisance parametere c_{crit}, see Gomes and Pestana ([34]).

5 Threshold selection based on MSE behaviour

5.1 Hall and Welsh's threshold selection

Hall and Welsh's ([39]) provided a choice for the optimal threshold associated with Hill EVI-estimation, quite close to the one written in **Algorithm 1**, but with a different estimation of second-order parameters, dependent upon three tuning parameters, $\sigma < \tau_1 < \tau_2$. Again, the method is quite sensitive to changes in these parameters, but the choice suggested by Hall and Welsh seems to work well, as shown in the comparative study performed in [32].

5.2 Regression diagnostics' selections

The algorithm related to Beirlant *et al.*'s ([8], [7], [4]) regression diagnostics procedure has again been fully described and simulated in Gomes and Oliveira ([32]). The method provides interesting by-products: estimates of the MSE and of squared bias (SB), but, as detected in [32], there appear a few computational problems related to this method.

1. The procedure is quite time-consuming, particularly if we intend to obtain properties of the possible MSE and SB by-products.

2. The method is quite sensitive to the two main tuning parameters under play, but they may be used to overcome some non-convergence problems, like negative values for the MSE and non-admissible estimates of the second-order parameter, ρ. However, although time-consuming, the method provides nice results, particularly for small samples.

5.3 Hall's bootstrap methodology

The bootstrap methodology for the selection of the threshold k was first introduced in Hall ([38]). For a recent comparison between the simple-bootstrap and the double-bootstrap methodology, see Caeiro and Gomes ([10]), where an improved version of Hall's bootstrap methodology is introduced and compared with the double bootstrap methodology.

Given the sample $\underline{X}_n = (X_1, \ldots, X_n)$ from an unknown model F, and the functional $\hat{\xi}^{\mathrm{H}}_{k,n} = \phi_k(\underline{X}_n)$, $1 \leq k < n$, consider the bootstrap sample $\underline{X}^*_{n_1} = (X^*_1, \ldots, X^*_{n_1})$, $n_1 \leq n$, from the model $F^*_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{\{X_i \leq x\}}$, the empirical CDF associated with the original sample $\underline{X}_n$. Then associate with that bootstrap sample the corresponding bootstrap estimator $\hat{\xi}^{\mathrm{H}*}_{k_1,n_1} = \phi_{k_1}(\underline{X}^*_{n_1})$, $1 \leq k_1 < n_1$.

Given an initial value of k, say k_{aux}, such that $\hat{\xi}^{\mathrm{H}}_{k_{aux},n}$ is a consistent estimator of ξ, (i.e., we need to have $k_{aux} \to \infty$, $k_{aux}/n \to 0$, as $n \to \infty$), Hall suggests the minimization of the bootstrap estimate of the MSE of $\hat{\xi}^{\mathrm{H}}_{k_1,n_1}$,

$$\mathrm{MSE}^*(n_1, k_1) = \mathbb{E}\left[\left\{\hat{\xi}^{\mathrm{H}*}_{k_1,n_1} - \hat{\xi}^{\mathrm{H}}_{k_{aux},n}\right\}^2 |\underline{X}_n\right] \tag{5.1}$$

for a sub-sample size, n_1, together with the fact that for a special sub-class of Hall's class, for which $\rho = -1$, and for suitable $n_1 = o(n)$, as $n \to \infty$, the optimal k_1-choice is of the same type, $Dn_1^{\gamma}(1+o(1))$, with $\gamma = 2/3$, both for the original estimator $\hat{\xi}^{\mathrm{H}}_{k_1,n_1}$ and for the bootstrap statistic $\hat{\xi}^{\mathrm{H}*}_{k_1,n_1} - \hat{\xi}^{\mathrm{H}}_{k_{aux},n}|\underline{X}_n$.

More generally, under the validity of (2.5) and with $n_1 = O(n^{1-\epsilon})$, for some $0 < \epsilon < 1$, and k_1 intermediate, we can say that there exists a function $D(\rho)$ such that the optimal performance of $\hat{\xi}^{\mathrm{H}*}_{k_1,n_1}|\underline{X}_n$ (level k_1 where $\mathrm{MSE}\left[\hat{\xi}^*_{k_1,n_1}|\underline{X}_n\right]$ is minimal), is achieved at

$$k^*_0(n_1) = D(\rho) n_1^{-2\rho/(1-2\rho)}(1 + o_p(1)). \tag{5.2}$$

This follows in a way similar to the proof of Theorem 4 in [32]. Also from the results in Section 2, the AMSE of the H EVI-estimator is, for the same $D(\rho)$ as in (5.2), minimal for

$$k^{\mathrm{H}}_0(n) = D(\rho) n^{-2\rho/(1-2\rho)}(1 + o(1)). \tag{5.3}$$

On the basis of $\hat{\rho}$ any estimate of ρ, and an estimate, $\hat{k}^*_0(n_1)$, of $k^*_0(n_1)$, in (5.2), we can consider

$$\hat{k}^H_0 := \hat{k}^*_0(n_1)(n/n_1)^{-2\hat{\rho}/(1-2\hat{\rho})}. \tag{5.4}$$

As noticed in [32], and despite of an almost independence on n_1, there is a disturbing sensitivity of the method to the initial value of k_{aux}. These facts—strong dependence on k_{aux} and almost independence on n_1—arose strong confidence in the suitability of the alternative double-bootstrap methodology in [19], [17] and [32], investigated in the next section. The most common choices for k_{aux} and n_1 are $k_{aux} = 2\sqrt{n}$ and $n_1 = \lfloor n^{0.955} \rfloor$, respectively. As shown in [32] the bootstrap methodology is much more sensitive to the choice of k_{aux} than to the choice of n_1.

5.4 The use of auxiliary statistics—an alternative bootstrap method

Note now that on the basis of (5.2) and (5.3),

$$\frac{k_0^*(n_1)}{k_0^{\mathrm{H}}(n)} = \left(\frac{n_1}{n}\right)^{\frac{2\rho}{2\rho-1}} (1 + o_p(1)), \text{ as } n \to \infty,$$

and thus, for another sample size n_2, and for every $\alpha > 1$, we have

$$\frac{[k_0^*(n_1)]^\alpha}{k_0^*(n_2)} \left(\frac{n_1^\alpha}{n^\alpha} \frac{n}{n_2}\right)^{\frac{2\rho}{1-2\rho}} = \{k_0(n)\}^{\alpha-1} (1 + o_p(1)).$$

It is then enough to choose $n_2 = n\,(n_1/n)^\alpha$, in order to have independence of ρ. If we put $n_2 = n_1^2/n$, i.e. $\alpha = 2$, we have

$$\frac{[k_0^*(n_1)]^2}{k_0^*(n_2)} = k_0^{\mathrm{H}}(n)(1 + o_p(1)), \text{ as } n \to \infty.$$

This argument led to the so-called double-bootstrap method, first used in [19] for the general max-domain of attraction and in [17, 32] for heavy tailed models. More recently, [25] modified the double bootstrap algorithm for an adaptive choice of the thresholds for second-order corrected-bias estimators. See also [31] and [10]. **Algorithm 3**, provided in this section, follows closely [25] and [10]. We consider again an auxiliary statistic of the type of the one considered in [32], directly related to the EVI-estimator under consideration, but going to the known value zero,

$$T_{k,n}^\bullet := \hat{\xi}^\bullet_{\lfloor k/2\rfloor,n} - \hat{\xi}^\bullet_{k,n}, \quad k = 2, \ldots, n-1, \quad \bullet = \mathrm{H},\ \mathrm{RBH}. \tag{5.5}$$

Notice that if $\bullet = \mathrm{H}$ this approach is equivalent to replace the fixed threshold k_{aux} by $\lfloor k/2 \rfloor$ in (5.1). On the basis of the results similar to the ones in [32], and with $c_\bullet$ given in (2.8), we can get for $T_{k,n}^\bullet$, in (5.5), the asymptotic distributional representation,

$$T_{k,n}^\bullet \stackrel{d}{=} \frac{\xi\, P_k^\bullet}{\sqrt{k}} + b_\bullet (2^{c_\bullet \rho/2} - 1)\ A^{c_\bullet/2}(n/k)(1 + o_p(1)),$$

with $P_k^\bullet$ asymptotically standard normal. Then, the AMSE of $T_{k,n}^\bullet$ is minimal at a level $k_{0|T}^\bullet(n)$, such that

$$k_0^\bullet(n) = k_{0|T}^\bullet(n)(2^{c_\bullet \rho/2} - 1)^{\frac{2}{1-c_\bullet \rho}}.$$

Algorithm 3.
Let $\hat{\xi}^{\bullet}_{k,n}$ *denote either the* H *or the* RBH EVI-*estimators, in* (2.2) *and* (2.6), *respectively. We now proceed with the description of the algorithm for the adaptive estimation of the optimal threshold* $k^{\bullet}_0(n)$.

Step 1. *Given a sample* $(x_1, \ldots, x_n)$, *compute the estimates* $\hat{\rho}$ *and* $\hat{\beta}$ *of the second-order parameters* ρ *and* β *as described in* **Algorithm 1**.

Step 2. *Next, consider a sub-sample size* $n_1 = o(n)$ *and* $n_2 = \lfloor n_1^2/n \rfloor + 1$.

For l *from 1 until* B, *generate independently* B *bootstrap samples* $(x_1^*, \ldots, x_{n_2}^*)$ *and* $(x_1^*, \ldots, x_{n_2}^*, x_{n_2+1}^*, \ldots, x_{n_1}^*)$, *of sizes* n_2 *and* n_1, *respectively, from the empirical distribution* $F_n^*(x) = \frac{1}{n}\sum_{i=1}^n I_{\{X_i \le x\}}$ *associated with the observed sample* $(x_1, \ldots, x_n)$.

Step 3. *Denoting* $T^{\bullet *}_{k,n_i}$ *the bootstrap counterpart of* $T^{\bullet}_{k,n_i}$, *in* (5.5), *obtain* $t^{\bullet *}_{k,n_i,l}$, $1 \le l \le B$, $i = 1, 2$ *the observed values of* $T^{\bullet *}_{k,n}$. *For* $k = 2, \ldots, n_i - 1$, *and* $i = 1, 2$ *compute* $\mathrm{MSE}^{\bullet *}(n_i, k) = \frac{1}{B}\sum\limits_{l=1}^{B} \left(t^{\bullet *}_{k,n_1,l}\right)^2$, *and obtain* $\hat{k}^{\bullet *}_{0|T}(n_i) := \arg\min_{1<k<n_i} \mathrm{MSE}^{\bullet *}(n_i, k)$.

Step 4. *Compute the threshold estimate*

$$\hat{k}^{\bullet *}_0(n) \equiv \left\lfloor (1 - 2^{c_\bullet \hat{\rho}/2})^{\frac{2}{1-c_\bullet \hat{\rho}}} \, (\hat{k}^{\bullet *}_{0|T}(n_1))^2 / \hat{k}^{\bullet *}_{0|T}(n_2) \right\rfloor + 1.$$

If $\hat{k}^{\bullet *}_0(n) \notin [1, n)$ *go back to* Step 2, *generating different bootstrap samples.*

Step 5. *Obtain* $\hat{\xi}^{\bullet *} \equiv \hat{\xi}^{\bullet}_{\hat{k}^*_0(n),n}$.

Remarks:

- The use of the sample $(x_1^*, \ldots, x_{n_2}^*)$, and of the extended sample $(x_1^*, \ldots, x_{n_2}^*, \ldots, x_{n_1}^*)$, $n_2 < n_1$, lead us to an increased precision of the result with the same number B of bootstrap samples generated in **Step 2**. This is quite similar to the use of the 'Common Random Numbers' simulation technique.

- Bootstrap confidence intervals are easily obtained, through the replication of this algorithm r times. The replication can also provide us more precise estimates, if we consider the estimate given by the mean or the median of the r bootstrap estimates.

5.5 Semi parametric bootstrap

Caers et al. [13] proposed a simple alternative bootstrap method for the selection of the threshold u which does not require the choice of the sub-sample size, as in the previous two methods. Here, a Generalized Pareto (GP) distribution, with CDF

$$GP_{\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}x\right)^{-1/\xi}, & 0 < x < (\max(0, -\xi/\sigma))^{-1}, \quad \text{if} \quad \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\sigma}\right), & x > 0, \quad \text{if} \quad \xi = 0, \end{cases}$$

and $\sigma > 0$ a scale parameter, is fitted to the excesses above the threshold u and used to generate bootstrap tail observations. As mentioned at the very beginning of this Chapter, we shall consider $u = x_{n-k:n}$ and a slight modification of the method in [13], provided in the following algorithm.

Algorithm 4.
Let $\hat{\xi}_{k,n}$ and $\hat{\sigma}_{k,n}$ denote here any semi-parametric EVI *and scale estimators, respectively.*

Step 1. *Given a sample $(x_1, \ldots, x_n)$, consider a set of integer values $1 < k_1 < k_2 < \ldots < k_m < n$.*

Step 2. *For each threshold $x_{n-k_i:n}$, $1 \leq i \leq m$:*

- *Compute the estimates of the* GP *parameters, $\hat{\xi}_i \equiv \hat{\xi}_{k_i,n}$ and $\hat{\sigma}_i \equiv \hat{\sigma}_{k_i,n}$.*
- *For l from 1 until B, generate independently B bootstrap samples $(x_1^*, \ldots, x_n^*)$, of size n, from the semi-parametric model*

$$\hat{F}_{k_i,n}(x) = \begin{cases} 1 - \frac{k_i}{n} + \frac{k_i}{n} GP_{\hat{\xi}_i,\hat{\sigma}_i}(x - x_{n-k_i:n}), & x > x_{n-k_i:n} \\ \hat{F}_n(x), & x \leq x_{n-k:n}, \end{cases} \tag{5.6}$$

where $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{\{X_i \leq x\}}$ is the empirical CDF *associated with the original sample, and compute the estimates $\hat{\xi}^{*(l)}_{k_i,n}$, $l = 1, \ldots, B$.*

Step 3. *Compute*

$$\widehat{\mathrm{MSE}}^*(k_i) = \frac{1}{B}\sum_{l=1}^{B} \left(\hat{\xi}^{*(l)}_{k_i,n} - \hat{\xi}_{k_i,n}\right)^2, \quad 1 \leq i \leq m,$$

and obtain $\hat{k}_0^* := \arg\min_{1 \leq i \leq m} \widehat{\mathrm{MSE}}^*(k_i)$.

Step 4. *Obtain $\hat{\xi}^{*|GP} \equiv \hat{\xi}_{\hat{k}_0^*,n}$.*

6 Other heuristic or data-driven methods for adaptive EVI-estimation

A heuristic algorithm in the line of the ones in [28], [27] and [45], essentially based on sample path stability (PS), and known to work in a quite reliable way for RB EVI-estimators, but that can also be used for C EVI-estimators and estimators of other parameters of extreme events, is now sketched:

Algorithm 5.
Let $\hat{\xi}^{\bullet}_{k,n} =: \mathrm{T}(k)$ *be any consistent* EVI-*estimator.*

Step 1. *Given an observed sample* $(x_1, \ldots, x_n)$, *compute, for* $k = 1, \ldots, n-1$, *the observed values of* $\mathrm{T}(k)$.

Step 2. *Obtain* j_0, *the minimum value of* j, *a non-negative integer, such that the rounded values, to* j *decimal places, of the estimates in* **Step 1** *are distinct. Define* $a_k^{(\mathrm{T})}(j) = round(\mathrm{T}(k), j)$, $k = 1, 2, \ldots, n-1$, *the rounded values of* $\mathrm{T}(k)$ *to* j *decimal places.*

Step 3. *Consider the sets of* k *values associated to equal consecutive values of* $a_k^{(\mathrm{T})}(j_0)$, *obtained in Step* **2.** *Set* $k_{min}^{(\mathrm{T})}$ *and* $k_{max}^{(\mathrm{T})}$ *the minimum and maximum values, respectively, of the set with the largest range. The largest run size is then* $l_{\mathrm{T}} := k_{max}^{(\mathrm{T})} - k_{min}^{(\mathrm{T})}$.

Step 4. *Consider all those estimates,* $\mathrm{T}(k)$, $k_{min}^{(\mathrm{T})} \le k \le k_{max}^{(\mathrm{T})}$, *now with two extra decimal places, i.e. compute* $\mathrm{T}(k) = a_k^{(\mathrm{T})}(j_0+2)$. *Obtain the mode of* $\mathrm{T}(k)$ *and denote* $\mathcal{K}_{\mathrm{T}}$ *the set of* k*-values associated with this mode.*

Step 5. *Take* $\hat{k}_{\mathrm{T}}$ *as the maximum value of* $\mathcal{K}_{\mathrm{T}}$.

Step 6. *Compute* $\hat{\xi}^{\bullet|PS} = \hat{\xi}^{\bullet}_{\hat{k}_{\mathrm{T}},n}$.

Other heuristic algorithms, essentially devised for RB estimators of parameters of extreme events, can be found in [33], [29] and [22]. Other heuristic methods for classical positive EVI-estimators can be found in Csörgo and Viharos ([16]). See also Scarrot and MacDonald ([50]) for a review of this topic.

7 Application to a simulated sample

We shall now consider an illustration of the performance of the aforementioned algorithms, when applied to the analysis of a simulated sample with a size $n = 500$ from a Burr model with $\xi = 0.3$ and $\rho = -0.7$. The CDF of a $\mathrm{Burr}_{\xi,\rho}$ model is given by $F(x) = 1-(1+x^{-\rho/\xi})^{1/\rho}$, $x > 0$, $\xi > 0$ and $\rho < 0$.

We are interested in the selection of the threshold and in the EVI-estimation provided by the the Hill estimator in (2.2) and the corresponding RB Hill (RBH) EVI-estimator in (2.6). Figure 1 shows the estimates of the EVI, ξ, provided by the selected EVI-estimators, as a function of the threshold k.

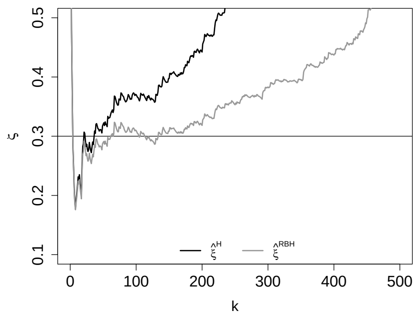

Figure 1: H and RBH estimates of ξ, as a function of k, for a $\text{Burr}_{0.3,-0.7}$ sample

If we consider **Algorithm 1**, the estimates of the second-order parameters, provided in **Step 3.**, are $\hat{\rho} = -0.773$ and $\hat{\beta} = 1.008$. The estimated optimal level is then $\hat{k}_0^H = 57$ and we obtain $\hat{\xi}^{\text{H}}_{57,500} = 0.333$. Note that since the estimation of the third-order parameters is still an almost open topic in SUE, we cannot yet directly estimate the optimal number of top OSs for $\hat{\xi}^{\text{RBH}}_{k,n}$.

If we use now Guillou and Hall choice of the threshold for the Hill estimator, presented in section 4.2, using **Algorithm 2**, we obtain the estimates $\hat{k}_0^H = 51$ and the EVI-estimate $\hat{\xi}^{GH} = \hat{\xi}^{\text{H}}_{51,500} = 0.323$. A plot with all values of the statistic $Q_n(k)$, in (4.2), is provided in Figure 2.

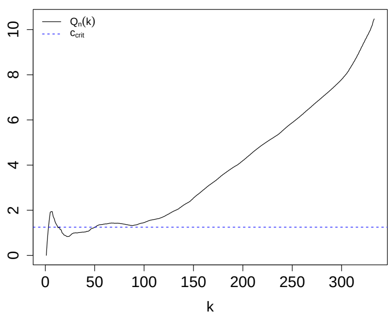

Figure 2: Values of the statistic $Q_n(k)$, in (4.2), for the simulated $\text{Burr}_{0.3,-0.7}$ sample

Another possibility is to estimate the optimal threshold through the algorithms based on the bootstrap methodology. We have applied for both EVI-estimators $\hat{\xi}^{\mathrm{H}}_{k,n}$ and $\hat{\xi}^{\mathrm{RBH}}_{k,n}$, **Algorithm 3** with $n_1 = 366$ and $B = 5000$ and **Algorithm 4** with $B = 400$ and the scale parameter σ estimated through maximum likelihood.

Finally, if we use the heuristic method in **Algorithm 5** we obtain $\hat{k}_{\mathrm{H}} = 291$ and $\hat{\xi}^{H|PS} = 0.579$ for the H estimator in (2.2). If we use the RBH estimator, we get $\hat{k}_{\mathrm{RBH}} = 180$ and $\hat{\xi}^{RBH|PS} = 0.315$. The obtained results are summarized in Table 1.

Table 1: Estimates of the optimal threshold and the EVI for the simulated $\mathrm{Burr}_{0.3,-0.7}$ sample

	EVI-estimator	$\hat{k}_0$	$\hat{\xi}$
Algorithm 1	H	57	0.333
Algorithm 2	H	51	0.332
Algorithm 3	H	61	0.335
	RBH	130	0.292
Algorithm 4	H	122	0.361
	RBH	128	0.283
Algorithm 5	H	291	0.579
	RBH	180	0.315

As a final remark, we can say that most of the methods provide accurate estimates of the true EVI-value, $\xi = 0.3$, associated with the underlying model. The smallest bias was attained by the RBH EVI-estimator and **Algorithm 3**, followed by the same EVI-estimator and **Algorithm 5**. But we cannot forget that the above results are taken from just one simulated sample and therefore we cannot compare the performance of the different methods without making first an extended Monte-Carlo comparison study.

References

[1] Beirlant, J., Boniphace, E., and Dierckx, G. (2011), "Generalized sum plots," *Revstat*, 9, 181–198.

[2] Beirlant, J., Caeiro, F., and Gomes, M. I. (2012), "An overview and open research topics in the field of statistics of univariate extremes," *Revstat*, 10, 1–31.

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

[4] Beirlant, J., Dierckx, G., Guillou, A., and Stărică, C. (2002), "On exponential representations of log-spacings of extreme order statistics," *Extremes*, 5, 157–180.

[5] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004), *Statistics of Extremes: Theory and Applications*, John Wiley & Sons.

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

[7] Beirlant, J., Vynckier, P., and Teugels, J. L. (1996), "Tail index estimation, Pareto quantile plots regression diagnostics," *Journal of the American statistical Association*, 91, 1659–1667.

[8] Beirlant, J., Vynckier, P., Teugels, J. L., et al. (1996), "Excess functions and estimation of the extreme-value index," *Bernoulli*, 2, 293–318.

[9] Caeiro, F. and Gomes, M. I. (2011), "Semi-parametric tail inference through probability-weighted moments," *Journal of Statistical Planning and Inference*, 141, 937–950.

[10] — (2014), "On the bootstrap methodology for the estimation of the tail sample fraction," in *Proceedings of COMPSTAT 2014*, eds. Gilli, M., Gonzalez-Rodriguez, G., and Nieto-Reyes, A., pp. 545–552.

[11] Caeiro, F., Gomes, M. I., and Henriques-Rodrigues, L. (2009), "Reduced-bias tail index estimators under a third order framework," *Communications in Statistics-Theory and Methods*, 38, 1019–1040.

[12] Caeiro, F., Gomes, M. I., and Pestana, D. (2005), "Direct reduction of bias of the classical Hill estimator," *Revstat*, 3, 113–136.

[13] Caers, J., Beirlant, J., and Maes, M. A. (1999), "Statistics for modeling heavy tailed distributions in geology: Part I. Methodology," *Mathematical geology*, 31, 391–410.

[14] Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer.

[15] Csörgő, S., Deheuvels, P., and Mason, D. (1985), "Kernel estimates of the tail index of a distribution," *The Annals of Statistics*, 1050–1077.

[16] Csörgő, S. and Viharos, L. (1998), "Estimating the tail index," in *Asymptotic Methods in Probability and Statistics'*, ed. Szyszkowicz, B., Elsevier, pp. 833–881.

[17] Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001), "Using a bootstrap method to choose the sample fraction in tail index estimation," *Journal of Multivariate Analysis*, 76, 226–248.

[18] de Sousa, B. and Michailidis, G. (2004), “A diagnostic plot for estimating the tail index of a distribution,” *Journal of Computational and Graphical Statistics*, 13.

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

[20] Drees, H., de Haan, L., and Resnick, S. (2000), “How to make a Hill plot,” *Annals of Statistics*, 254–274.

[21] Drees, H. and Kaufmann, E. (1998), “Selecting the optimal sample fraction in univariate extreme value estimation,” *Stochastic Processes and their Applications*, 75, 149–172.

[22] Figueiredo, F., Gomes, M. I., Henriques-Rodrigues, L., and Miranda, M. C. (2012), “A computational study of a quasi-PORT methodology for VaR based on second-order reduced-bias estimation,” *Journal of Statistical Computation and Simulation*, 82, 587–602.

[23] Fraga Alves, M. I., Gomes, M. I., and de Haan, L. (2003), “A new class of semi-parametric estimators of the second order parameter,” *Portugaliae Mathematica*, 60, 193–214.

[24] Geluk, J. and de Haan, L. (1987), *Regular Variation, Extensions and Tauberian Theorems*, CWI Tract 40, Amsterdam, The Netherlands: Centre for Mathematics and Computer Science.

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

[26] Gomes, M. I. and Guillou, A. (2014), “Extreme value theory and statistics of univariate extremes: a review,” *International Statistical Review*, doi:10.1111/insr.12058.

[27] Gomes, M. I., Henriques-Rodrigues, L., Fraga Alves, M. I., and Manjunath, B. (2013), “Adaptive PORT–MVRB estimation: an empirical comparison of two heuristic algorithms,” *Journal of Statistical Computation and Simulation*, 83, 1129–1144.

[28] Gomes, M. I., Henriques-Rodrigues, L., and Miranda, M. C. (2011), “Reduced-bias location-invariant extreme value index estimation: a simulation study,” *Communications in Statistics–Simulation and Computation*, 40, 424–447.

[29] Gomes, M. I., Henriques-Rodrigues, L., Vandewalle, B., and Viseu, C. (2008), “A heuristic adaptive choice of the threshold for bias-corrected

Hill estimators," *Journal of Statistical Computation and Simulation*, 78, 133–150.

[30] Gomes, M. I. and Martins, M. J. (2002), "Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter," *Extremes*, 5, 5–31.

[31] Gomes, M. I., Mendonça, S., and Pestana, D. (2011), "Adaptive reduced-bias tail index and VaR estimation via the bootstrap methodology," *Communications in Statistics-Theory and Methods*, 40, 2946–2968.

[32] Gomes, M. I. and Oliveira, O. (2001), "The bootstrap methodology in statistics of extremes–choice of the optimal sample fraction," *Extremes*, 4, 331–358.

[33] Gomes, M. I. and Pestana, D. (2007), "A sturdy reduced-bias extreme quantile (VaR) estimator," *Journal of the American Statistical Association*, 102, 280–292.

[34] — (2011), "A note on the adaptive choice of the optimal threshold in extreme value analysis," in *Proc. 58th World Statistics Congress.*

[35] Groeneboom, P., Lopuhaä, H., and De Wolf, P. (2003), "Kernel-type estimators for the extreme value index," *The Annals of Statistics*, 31, 1956–1995.

[36] Guillou, A. and Hall, P. (2001), "A diagnostic for selecting the threshold in extreme value analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 293–305.

[37] Hall, P. (1982), "On some simple estimates of an exponent of regular variation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 37–42.

[38] — (1990), "Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems," *Journal of multivariate analysis*, 32, 177–203.

[39] Hall, P. and Welsh, A. (1985), "Adaptive estimates of parameters of regular variation," *The Annals of Statistics*, 13, 331–341.

[40] Henry III, J. (2009), "A harmonic moment tail index estimator," *Journal of Statistical Theory and Applications*, 8, 141–162.

[41] Hill, B. M. (1975), "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, 3, 1163–1174.

[42] Kratz, M. and Resnick, S. I. (1996), "The QQ-estimator and heavy tails," *Stochastic Models*, 12, 699–724.

[43] Martins, M. J. a., Gomes, M. I., and Neves, M. M. (2004), "Averages of Hill estimators," *Test*, 13, 1–16.

[44] Matthys, G. and Beirlant, J. (2000), "Adaptive threshold selection in tail index estimation," in *Extremes and Integrated Risk Management*, ed. Embrechts, P., London, UK: Risk Books, pp. 37–49.

[45] Neves, M. M., Gomes, M. I., Figueiredo, F., and Gomes, D. P. (2014), "Modeling extreme events: Sample fraction adaptive choice in parameter estimation," *Journal of Statistical Theory and Practice.*

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

[47] Resnick, S. and Stărică, C. (1997), "Smoothing the Hill estimator," *Advances in Applied Probability*, 271–293.

[48] Resnick, S. I. (1997), "Heavy tail modeling and teletraffic data: special invited paper," *The Annals of Statistics*, 25, 1805–1869.

[49] — (2007), *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*, Springer.

[50] Scarrott, C. and MacDonald, A. (2012), "A review of extreme value threshold estimation and uncertainty quantification," *REVSTAT–Statistical Journal*, 10, 33–60.

[51] Wager, S. (2012), "Subsampling Extremes: From Block Maxima to Smooth Tail Estimation," *arXiv preprint arXiv:1204.0316.*