Revisiting the Role of the Generalized Jackknife Methodology in the Field of Extremes*

M. Ivette Gomes

Universidade de Lisboa, FCUL, DEIO and CEAUL

M. João Martins

Universidade Técnica de Lisboa, ISA and Centro de Estudos Florestais

M. Manuela Neves

Universidade Técnica de Lisboa, ISA, and CEAUL

Abstract

In this article, we deal with the importance of the *generalized jackknife* methodology in the construction of a reliable semi-parametric estimate of any parameter of extreme or even rare events. In order to illustrate such a kind of methodology, we shall apply it to *corrected-bias* estimators of a positive *extreme value index*, the primary parameter in *statistics of extremes*.

Keywords. Statistics of extremes; extreme value index; semi-parametric estimation; bias reduction; generalized jackknife methodology.

*Research partially supported by National Funds through **FCT** — Fundação para a Ciência e a Tecnologia, project PEst-OE/MAT/UI0006/2011, and PTDC / FEDER.

1 Introduction and preliminaries

In the field of *statistics of extremes*, resampling methodologies, among which we mention the *jackknife* (Quenouille, 1949, 1956; Tukey, 1958) and the *bootstrap* (Efron, 1979; Efron and Tibshirani, 1993), have recently revealed to be of high relevance in the adequate estimation of any parameter of extreme events, like a *high quantile*, the *expected shortfall*, the *return period* of a high level or the primary parameter of extreme events, the *extreme value index* (EVI). In this article, we shall deal with the importance of the *generalized jackknife* (Gray and Schucany, 1972) in the obtention of reliable semi-parametric EVI estimates.

Given an underlying distribution function (d.f.) F, and with the notation $F^{\leftarrow}(y) := \inf\{x : F(x) \geq y\}$ for the generalized inverse function of F, let us denote $U(t) := F^{\leftarrow}(1 - 1/t)$, $t \geq 1$, the associated reciprocal quantile function. Let the notation RV_α stand for the class of regularly-varying functions with an index of regular variation equal to α, i.e., positive measurable functions $g(\cdot)$ such that for any $x > 0$, $g(tx)/g(t) \to x^\alpha$, as $t \to \infty$ (Bingham *et al.*, 1987). Under a semi-parametric framework, we shall deal with the estimation of a positive EVI, denoted γ, the primary parameter in *statistics of extremes*, i.e. we shall consider parents such that

$$U \in RV_\gamma \quad \Longleftrightarrow \quad \overline{F} := 1 - F \in RV_{-1/\gamma},$$

the usually called *heavy-tailed* parents characterized in Gnedenko (1943) and de Haan (1970). These heavy right-tails have revealed to be quite common in the most diversified areas of application, like finance, insurance, bibliometrics and environment.

We are then working in $\mathcal{D}^+_{\mathcal{M}} \equiv \mathcal{D}_{\mathcal{M}}(EV_\gamma)_{\gamma>0}$, the *domain of attraction* for maxima of a d.f. $EV_\gamma(x)$, $\gamma > 0$, the general *extreme value* (EV) d.f., given by

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

This means that, given a sample $\mathbf{X}_n = (X_1, \ldots, X_n)$, it is possible to linearly normalize the sequence of maximum values, $X_{n:n} := \max(X_1, X_2, \ldots, X_n)$, and to get convergence towards a non-degenerate random variable (r.v.), necessarily with the d.f. $EV_\gamma(x)$, in (1).

For these heavy-tailed parents, given the sample $\mathbf{X}_n$ and the associated sample of ascending order statistics (o.s.'s), $(X_{1:n} \leq \cdots \leq X_{n:n})$, the classical EVI-estimator is the Hill estimator (Hill, 1975), here denoted $H \equiv H(k)$, $k = 1, 2, \ldots, n-1$, and given by

$$H(k) \equiv H(k; \mathbf{X}_n) := \frac{1}{k} \sum_{i=1}^{k} \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}, \tag{2}$$

the average of the k log-excesses over a high random threshold $X_{n-k:n}$.

Consistency of the estimators in (2) is achieved if $X_{n-k:n}$ is an *intermediate* o.s., i.e., if

$$k \equiv k_n \to \infty \quad \text{and} \quad k/n \to 0, \ \text{as } n \to \infty. \tag{3}$$

Indeed, whenever working under a semi-parametric framework, functional EVI-estimators, like the Hill estimator, in (2), depend on the *tuning* or *nuisance* parameter k, related with the number of top o.s.'s involved in the estimation. Due to the high bias of the Hill estimator, in (2), for moderate up to large k, several authors have been dealing with bias reduction in the field of extremes, working usually in a slightly more restrict class than $\mathcal{D}_{\mathcal{M}}^+$, the class of models $U(\cdot)$ such that

$$U(t) = C\ t^\gamma \big(1 + A(t)/\rho + o(t^\rho)\big), \qquad A(t) = \gamma\ \beta\ t^\rho, \tag{4}$$

as $t \to \infty$, where $C > 0$, $\gamma > 0$, $\rho < 0$ and $\beta \neq 0$. This means that the slowly varying function $L(t)$ in $U(t) = t^\gamma L(t)$ is assumed to behave asymptotically as a constant C. Note that to assume (4) is equivalent to choose $A(t) = \gamma\, \beta\, t^\rho$, $\rho < 0$, in the more general second-order condition

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}. \tag{5}$$

As mentioned above, the Hill estimator $H(k)$, in (2), reveals usually a high asymptotic bias, i.e., as $n \to \infty$, and if (3) and (5) hold, $\sqrt{k}\,(H(k) - \gamma)$ is asymptotically normal with variance γ^2 and a non-null mean value, equal to $\lambda/(1-\rho)$, whenever $\sqrt{k}\ A(n/k) \ \to\ \lambda \neq 0$, finite, the type of k-values which lead to a minimum *mean squared error* (MSE). More specifically, it follows from the results of de Haan and Peng (1998) that for models in (5), with $H(k)$ given in (2), and with the notation Normal(μ, σ^2) for a normal r.v. with mean value μ and variance σ^2,

$$\sqrt{k}\,(H(k) - \gamma) \ \overset{d}{=}\ \text{Normal}\left(0, \gamma^2\right) + b_{_H}\sqrt{k}A(n/k) + o_p\left(\sqrt{k}A(n/k)\right),$$
$$b_{_H} = \frac{1}{1-\rho},$$

where the bias $b_{_H}\sqrt{k}A(n/k) = \gamma\ \beta\ \sqrt{k}\ (n/k)^\rho/(1-\rho)$, whenever (4) holds, can be very large, moderate or small (i.e. go to infinity, constant or zero) as $n \to \infty$. This non-null asymptotic bias, together with a rate of convergence of the order of $1/\sqrt{k}$, leads to sample paths with a high variance for small k, a high bias for large k, a very sharp MSE pattern, as a function of k, and hence a need for bias reduction.

A simple class of second-order *minimum-variance reduced-bias* (MVRB) EVI-estimators is the one in Caeiro *et al.* (2005), used for a semi-parametric estimation of $\ln VaR_p$ in Gomes and Pestana (2007b). This class, here denoted $\overline{H} \equiv \overline{H}(k)$, depends upon the estimation of the second-order parameters (β, ρ) in (4). Its functional form is

$$\overline{H}(k) \equiv \overline{H}(k; \mathbf{X}_n) \equiv \overline{H}_{\hat\beta,\hat\rho}(k) := H(k)\left(1 - \hat\beta (n/k)^{\hat\rho}/(1-\hat\rho)\right), \tag{6}$$

with $H(k)$ the Hill estimator in (2), and where $(\hat{\beta}, \hat{\rho})$ needs to be an adequate consistent estimator of (β, ρ). Then, if (3) holds and $\sqrt{k}\ A(n/k) \ \rightarrow\ \lambda \neq 0$, finite, with $A(\cdot)$ given in (4), $\sqrt{k}\left(\overline{H}(k) - \gamma\right)$ is asymptotically normal with variance also equal to γ^2 but with a null mean value. Indeed, from the results in Caeiro *et al.* (2005), we know that it is possible to adequately estimate the second-order parameters β and ρ, so that we get

$$\sqrt{k}\left(\overline{H}(k) - \gamma\right) \ \stackrel{d}{=}\ \text{Normal}\left(0,\ \gamma^2\right) + o_p\left(\sqrt{k}A(n/k)\right),$$

for $\overline{H}(k)$ in (6), i.e., asymptotically, $\overline{H}(k)$ outperforms $H(k)$ for all k.

In order to obtain information on the order of the asymptotic bias of second-order reduced-bias EVI-estimators, like the class in (6), it is necessary to further assume a third-order condition, ruling the rate of convergence in (5), and which guarantees that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} - \frac{x^\rho - 1}{\rho}}{B(t)} = \frac{x^{\rho+\rho'} - 1}{\rho + \rho'}, \tag{7}$$

where $|B(t)|$ must then be of regular variation with index ρ'. There appears then this extra non-positive third-order parameter $\rho' \leq 0$. Such a condition has already been used in Gomes *et al.* (2002a) and Fraga Alves *et al.* (2003), for the full derivation of the asymptotic behaviour of ρ-estimators, in Gomes *et al.* (2004), for the study of a reduced-bias EVI-estimator and more recently in Caeiro *et al.* (2009), for a comparison of a few MVRB estimators.

It is often assumed that (7) holds with $\rho,\ \rho' < 0$ and that we can choose

$$A(t) = \alpha\ t^\rho = \gamma\ \beta\ t^\rho, \quad B(t) = \beta'\ t^{\rho'}, \quad \beta,\ \beta' \neq 0,$$

where β and β' are "scale" second and third-order parameters, respectively. More specifically, and only slightly more restrictively than (4), we sometimes assume that

$$U(t) = C\ t^\gamma\left(1 + A(t)/\rho + O(A^2(t))\right) \tag{8}$$

(see Caeiro and Gomes, 2011, for an asymptotic comparison of reduced-bias EVI-estimators, under condition (8)). Note that to assume (8) is equivalent to say that (7) holds with $\rho = \rho' < 0$ and that we can there choose

$$A(t) = \gamma\ \beta\ t^\rho, \qquad B(t) = \beta'\ t^\rho = \frac{\beta' A(t)}{\beta\gamma} =: \frac{\xi A(t)}{\gamma}, \qquad \xi = \beta'/\beta \neq 0. \quad (9)$$

We can then adequately estimate the vector of second-order parameters, (β, ρ), and write (Caeiro *et al.*, 2009; Caeiro and Gomes, 2011),

$$\sqrt{k}\left(\overline{H}(k) - \gamma\right) \overset{d}{=} \text{Normal}\left(0,\ \gamma^2\right) + b_{\overline{H}}\sqrt{k}A^2(n/k) + o_p(\sqrt{k}A^2(n/k)),$$
$$b_{\overline{H}} = \frac{1}{\gamma}\Big(\frac{\xi}{1-2\rho} - \frac{1}{(1-\rho)^2}\Big). \quad (10)$$

But even at the optimal level for the estimation of γ on the basis of $\overline{H}(k)$, in (6), i.e. levels k such that $\sqrt{k}A^2(n/k) \to \lambda$, finite and non-null, as $n \to \infty$, there exists a non-null asymptotic bias, as can be seen from (10). If we still want to remove such a bias, we can then make use of the *generalized jackknife* (GJ) methodology. It is then enough to consider an adequate pair of EVI-estimators, and to built a reduced-bias affine combination of them (see Gomes *et al.*, 2000, among others, for the application of this technique to the Hill estimator). We shall now consider the application of the GJ methodology to the MVRB estimators in Caeiro *et al.* (2005), i.e., the estimators $\overline{H}(k)$, given in (6).

Algorithms for an estimation of (β, ρ) which leads to MVRB EVI-estimators are provided in Gomes and Pestana (2007a,b), among others, and will be briefly reformulated in Section 2 of this paper. Section 3 is dedicated to the use of the jackknife methodology not only in the reduction of bias of Hill's estimator, in (2), but also in a further reduction of the bias of the MVRB EVI-estimator in (6). Finally, in Section 4, we provide a few results achieved through a Monte-Carlo simulation related with the finite-sample properties of the new reduced-bias EVI-estimators, denoted $\overline{\overline{H}}(k)$, comparatively with the MVRB EVI-estimator $\overline{H}(k)$, in (6).

2 Estimation of second-order parameters

All reduced-bias EVI-estimators, like the one in (6), require the estimation of shape and sometimes even scale second-order parameters, ρ and β, respectively. The estimation of $(\beta,\ \rho)$ will be briefly discussed next.

For models in (4), and after taking a decision of working in the region $|\rho| \le 1$, the region where bias reduction is indeed needed, as well as common in applications to real data, we consider a particular member of the class of estimators introduced in Fraga Alves *et al.* (2003), parameterized in a tuning parameter $\tau \in \mathbb{R}$. Here we take $\tau = 0$, the value suggested in previous papers whenever working in this region of ρ-values. Given a sample, $\mathbf{X}_n$, we shall thus work with

$$\hat{\rho}_0(k) \equiv \hat{\rho}_0(k; \mathbf{X}_n) := \min\left(0,\ \frac{3(T_n^{(0)}(k; \mathbf{X}_n) - 1)}{T_n^{(0)}(k; \mathbf{X}_n) - 3}\right), \tag{11}$$

dependent on the statistics

$$T_n^{(0)}(k; \mathbf{X}_n) := \frac{\ln\left(M_n^{(1)}(k; \mathbf{X}_n)\right) - \frac{1}{2}\ln\left(M_n^{(2)}(k; \mathbf{X}_n)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k; \mathbf{X}_n)/2\right) - \frac{1}{3}\ln\left(M_n^{(3)}(k; \mathbf{X}_n)/6\right)},$$

where

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\}^j, \quad j = 1, 2, 3.$$

Distributional properties of the estimators in (11) can be found in Fraga Alves *et al.* (2003). Consistency is achieved in the class of models in (4), for *intermediate* k-values, i.e., k-values such that (3) holds, and also such that $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$. As already suggested in previous papers, we have here decided for the computation of $\hat{\rho}_0(k)$ at $k = k_1$, given by

$$k_1 = \left[n^{1-\epsilon}\right], \quad \epsilon = 0.001, \tag{12}$$

the threshold used in Caeiro *et al.* (2005) and Gomes and Pestana (2007*a*; 2007*b*). With such a choice of k_1, and whenever $\sqrt{k_1}\ A(n/k_1) \to \infty$, we get $\hat{\rho} - \rho := \hat{\rho}_0(k_1) - \rho = o_p(1/\ln n)$, a condition needed, in order not to have any increase in the asymptotic variance of the new bias-corrected Hill estimator in equation (6). Note that with the choice of k_1 in (12), we get $\sqrt{k_1}\ A(n/k_1) \to \infty$ if and only if $\rho > (1 - 1/\epsilon)/2 = -499.5$, an almost irrelevant restriction, from a practical point of view. Interesting alternative classes of ρ-estimators have recently been introduced in Goegebeur *et al.* (2008, 2010) and Ciuperca and Mercadier (2010).

For the estimation of the scale second-order parameter β, in (4), and again on the basis of a sample $\mathbf{X}_n$, we shall here consider

$$\hat{\beta}_{\hat{\rho}}(k) \equiv \hat{\beta}_{\hat{\rho}}(k; \mathbf{X}_n) := \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{d_{\hat{\rho}}(k)\ D_0(k) - D_{\hat{\rho}}(k)}{d_{\hat{\rho}}(k)\ D_{\hat{\rho}}(k) - D_{2\hat{\rho}}(k)}, \tag{13}$$

dependent on the estimator $\hat{\rho} = \hat{\rho}_0(k_1; \mathbf{X}_n)$, with $\hat{\rho}_0(k; \mathbf{X}_n)$ given before, in (11), and where, for any $\alpha \leq 0$,

$$d_\alpha(k) := \frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{k}\right)^{-\alpha} \quad \text{and} \quad D_\alpha(k) := \frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{k}\right)^{-\alpha} U_i,$$

with $U_i,\ 1 \leq i \leq k$, the *scaled log-spacings* associated with $\mathbf{X}_n$ and given by

$$U_i := i \left(\ln \frac{X_{n-i+1:n}}{X_{n-i:n}}\right).$$

Details on the distributional behaviour of the estimator in (13) can be found in Gomes and Martins (2002) and more recently in Gomes *et al.* (2008) and Caeiro *et al.* (2009). Consistency is achieved for models in (4), k-values such that (3) holds and $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$, and estimators $\hat{\rho}$ of ρ such that $\hat{\rho} - \rho = o_p(1/\ln n)$. Alternative estimators of β can be found in Caeiro and Gomes (2006) and Gomes *et al.* (2010).

Remark 2.1. *In the simulation study, the* MVRB EVI-*estimators in* (6) *will be denoted* $\overline{H}_0(k)$, *to enhance the choice* $\tau = 0$ *in the class of* ρ-*estimators in* Fraga Alves *et al.* (2003).

3 Bias-reduction and the GJ methodology

The pioneering EVI reduced-bias estimators are, in a certain sense, GJ estimators, i.e., affine combinations of well-known estimators of γ. The GJ statistic was introduced by Gray and Shucany (1972). One of the main objectives of the method is related with *bias reduction*.

Definition 3.1. *Let $T_n^{(1)}$ and $T_n^{(2)}$ be two biased estimators of γ, with similar bias properties, i.e.,*

$$\mathrm{Bias}(T_n^{(i)}) = \gamma + \phi(\gamma) d_i(n), \quad i = 1, 2.$$

Then, the associated GJ *statistic is the affine combination*

$$T_n^{GJ} := \frac{T_n^{(1)} - q T_n^{(2)}}{1-q}, \quad \textit{with} \quad q = q_n = d_1(n)/d_2(n) \neq 1,$$

an unbiased estimator of γ.

Whenever we are dealing with semi-parametric estimators of the EVI or even of other parameters of extreme events, we have usually information about the asymptotic bias of those estimators. We can thus choose estimators with similar asymptotic properties, and build the associated *approximate* GJ (AGJ) r.v. or statistic. This methodology has been used in Gomes *et al.* (2000, 2002b), among others, who suggested several AGJ estimators of a positive EVI, $\gamma > 0$. Indeed, under the general second-order condition, in (5), and on the basis of the Hill estimator, in (2), it is easy to find two statistics $\hat{\gamma}_{n,k}^{(j)}$, such that, with $P_k^{(j)}$, $j = 1, 2$, asymptotically standard normal r.v.s, we have

$$\hat{\gamma}_{n,k}^{(j)} \stackrel{d}{=} \gamma + \frac{\sigma_j P_k^{(j)}}{\sqrt{k}} + b_j A(n/k) + o_p(A(n/k)), \quad b_j = b_j(\rho), \quad j = 1, 2.$$

The ratio between the dominant components of bias of $\hat{\gamma}_{n,k}^{(1)}$ and $\hat{\gamma}_{n,k}^{(2)}$ is thus $q = b_1/b_2 = q(\rho)$, and we thus get the AGJ r.v.,

$$\hat{\gamma}_{n,k}^{AGJ(\rho)} := \frac{\hat{\gamma}_{n,k}^{(1)} - q(\rho)\ \hat{\gamma}_{n,k}^{(2)}}{1 - q(\rho)},$$

where ρ must be replaced by an estimator $\hat{\rho}$. After dealing, in Section 3.1, with a natural AGJ EVI-estimator, and despite of the fact that in most applications the bias of $\overline{H}(k)$ is already small, this is not always the case and a similar argument will be used here to build, in Section 3.2, an AGJ EVI-estimator based on the MVRB EVI-estimators $\overline{H}(k)$, in (6). In Section 3.3, we deal with refined GJ corrected-bias EVI-estimators, and finally, in Section 3.4, we study the asymptotic bias and efficiency of an arbitrary affine combination of MVRB EVI-estimators.

3.1 The natural AGJ Hill EVI-estimator

In a context of heavy tails, Gomes *et al.* (2000) worked with a set of estimators of a positive extreme value index γ, which were real competitors to the well-known Hill estimator for γ, in (2), among which we refer the so-called *natural* AGJ (NAGJ) estimator associated to the Hill estimator, given by

$$H^{NAGJ}(k) := 2H([k/2]) - H(k). \tag{14}$$

The estimators in (14) were obtained, assuming a known value $\rho = -1$, possibly mispecified, in the NAGJ estimators studied later in Gomes and Martins (2002),

$$H^{GJ_1}(k) \equiv H^{GJ(\hat{\rho})}(k) := \frac{H(k) - 2^{-\hat{\rho}} H\big([k/2]\big)}{1 - 2^{-\hat{\rho}}}, \tag{15}$$

where $\hat{\rho}$ is an adequate consistent estimator of the second-order parameter ρ. More generally, we could have considered the affine combination

$$H^{GJ(\hat{\rho};\theta)}(k) := \frac{H(k) - \theta^{\hat{\rho}} H\big([\theta k]\big)}{1 - \theta^{\hat{\rho}}}, \qquad 0 < \theta < 1, \tag{16}$$

dependent on the *tuning parameter* $\theta \in (0, 1)$. The estimator in (15) is thus a particular case of the estimator in (16), whenever we consider $\theta = 1/2$.

In Gomes *et al.* (2000) the study of the estimators in (15) has been postponed due to practical reasons, and it has there been claimed that the known estimators of ρ, like the ones suggested, among other authors, by Hall (1982), Beirlant *et al.* (1996a,b), Drees and Kaufmann (1998) and Peng (1998) had a very high MSE for most of the common parent distributions, and that such erratic behaviour destroyed drastically the theoretical nice properties of the AGJ r.v.'s, should ρ be known. However it has also been argued there that these estimators would potentially be the "optimal" estimators, provided we were able to get a suitable way of estimating the second-order parameter ρ, as we think has been partially achieved by Fraga Alves *et al.* (2003), for heavy tails.

3.1.1 Asymptotic behaviour of the NAGJ Hill estimators

The class of estimators $H^{GJ_1}(k) \equiv H^{GJ(\hat{\rho})}(k)$, in (15), was built to remove the main component of bias of the original estimators $H(k)$, which is of the order of $A(n/k)$. For the associated r.v.'s, $H^{GJ(\rho)}(k)$, and again with the notation Normal(μ, σ^2) for a normal r.v. with mean value μ and variance σ^2, we now provide full information on the asymptotic bias of $H^{GJ(\rho)}$.

Theorem 3.1. *If the second-order condition* (5) *holds, if* $k = k_n$ *is a sequence of intermediate positive integers, i.e.,* (3) *holds, and if* $\sqrt{k}A(n/k) \underset{n\to\infty}{\longrightarrow} \lambda$, *finite, non necessarily null, then*

$$\sqrt{k}\left(H^{GJ(\rho)}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \textit{Normal}\left(0,\ \frac{\gamma^2(1+2^{1-2\rho}-2^{1-\rho})}{(1-2^{-\rho})^2}\right). \tag{17}$$

Under the third-order framework, in (7), *further assuming that* $A(\cdot)$ *and* $B(\cdot)$ *can be chosen as in* (9), *we have*

$$H^{GJ(\rho)}(k) \overset{d}{=} \gamma + \frac{\sigma_{GJ}}{\sqrt{k}}\,\overline{Z}_k^{GJ} + u_{GJ}A^2(n/k) + o_p(A^2(n/k)), \tag{18}$$

where

$$\sigma_{GJ} = \frac{\gamma\sqrt{1+2^{1-2\rho}-2^{1-\rho}}}{1-2^{-\rho}}, \qquad u_{GJ} = \frac{\xi(1-2^{\rho})}{\gamma(1-2\rho)}. \tag{19}$$

Consequently, $\sqrt{k}\ \left(H^{GJ(\rho)}(k)-\gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\big(b_{GJ} := \lambda_A u_{GJ}, \sigma^2_{GJ}\big)$, *provided that* $\sqrt{k}\ A^2(n/k) \to \lambda_A$, *finite, i.e. even when* $\sqrt{k}\ A(n/k) \to \infty$. *If* $\sqrt{k}\ A^2(n/k) \to \infty$, $\left(H^{GJ}(k)-\gamma\right)/A(n/k)$ *is* $O_p(A(n/k))$.

Proof. For the r.v. $H^{GJ(\rho)}$ we have a distributional representation of the type

$$\sqrt{k}\left(H^{GJ(\rho)}(k)-\gamma\right) \overset{d}{=} \gamma\, V_k^{GJ(\rho)} + o_p\left(\sqrt{k}\ A(n/k)\right), \tag{20}$$

where, for $r = 1, 2$,

$$V_k^{GJ(\rho)} := \frac{P_{k,2} - \sqrt{2}\ 2^{-\rho}\ P_{k,1}}{1-2^{-\rho}}, \qquad P_{k,r} := \sqrt{\frac{rk}{2}}\left\{\tfrac{2}{rk}\sum_{j=1}^{rk/2} E_j - 1\right\}, \tag{21}$$

with $\{E_i\}$ a sequence of independent, identically distributed (i.i.d.) standard exponential r.v.'s. Then $\mathrm{Var}(P_{k,1}) = \mathrm{Var}(P_{k,2}) = 1$, $\mathrm{Cov}(P_{k,1}, P_{k,2}) = \sqrt{2}/2$, and (17) follows.

Note next that under the third-order framework in (7), further assuming that $A(\cdot)$ and $B(\cdot)$ can be chosen as in (9), we can write

$$H(k) \overset{d}{=} \gamma + \frac{\gamma P_{k,2}}{\sqrt{k}} + \frac{A(n/k)}{1-\rho} + \frac{A(n/k)}{\gamma}\Big(\frac{\xi A(n/k)}{1-2\rho} + O_p\Big(\frac{1}{\sqrt{k}}\Big)\Big)(1+o_p(1)).$$

The results in (18) and (19) follow then straightforwardly. □

Corollary 3.1. *Under the conditions of* Theorem 3.1*, the same distributional results hold if we consider the tail index estimator* $H^{GJ_1}(k) \equiv H^{GJ(\hat\rho)}(k)$*, defined in* (15)*, for any second-order parameter estimator* $\hat\rho$ *such that* $\hat\rho - \rho = o_p(1/(\sqrt{k}A(n/k)))$.

Proof. The result comes essentially from the fact that we have the distributional representation $\gamma_n^{\bullet(\hat\rho)}(k) \overset{d}{=} \gamma_n^{\bullet(\rho)}(k) + (\hat\rho-\rho)\ \nu_k^\bullet(\rho)\ (1+o_p(1))$, with $\nu_k^\bullet = O_p(1/\sqrt{k})$. □

Remark 3.1. *Note that, as it is well-known from the literature, we have a non-null asymptotic bias for the original estimators* $H(k)$*, if the threshold* k *is such that* $\sqrt{k}\ A(n/k) \to \lambda \neq 0$*. It is indeed a sequence* $k_0 = k_0(n)$ *such that* $\sqrt{k_0}\ A(n/k_0) \to \varphi(\rho, \gamma) \neq 0$*, the one which provides a minimum MSE for* $H(k)$*. From* Theorem 3.1*, even when* $\sqrt{k}\ A(n/k) \to \lambda \neq 0$*, finite, do we have a null dominant component of asymptotic bias for the r.v.'s,* $H^{GJ(\rho)}(k)$*, as can be seen from* (20)*. The same happens for the EVI-estimators,* $H^{GJ(\hat\rho)}$*, in a wide sub-class of* $\mathcal{D}_{\mathcal{M}}^{+}$*, whenever we consider the* ρ*-estimators in* (11)*, computed at an adequate fixed level* k_1*, like the one in* (12)*.*

3.2 An AGJ corrected-bias EVI-estimator

Let us now consider the GJ r.v. associated to the random pair $\big(\overline{H}(k),\ \overline{H}([\theta k])\big)$, $0<\theta<1$, i.e.,

$$\overline{H}^{GJ(q,\theta)}(k) := \frac{\overline{H}(k) - q_{\overline{H}}\ \overline{H}([\theta k])}{1-q_{\overline{H}}}, \qquad 0<\theta<1, \tag{22}$$

with

$$q \equiv q_{\overline{H}} = \frac{\mathrm{Bias}_\infty[\overline{H}(k)]}{\mathrm{Bias}_\infty[\overline{H}([\theta k])]}.$$

Since

$$\frac{\mathrm{Bias}_\infty[\overline{H}(k)]}{\mathrm{Bias}_\infty[\overline{H}(\theta k)]} = \frac{A^2(n/k)}{A^2(n/[\theta k])} \underset{n/k\to\infty}{\longrightarrow} \theta^{2\rho}$$

we shall consider, in (22), $q_{\overline{H}} = \theta^{2\rho}$, and for an adequate estimate $\hat\rho$ of ρ, we shall consider a class of AGJ EVI-estimators, parameterized in a *tuning parameter* $\theta \in (0,1)$, the class

$$\overline{H}^{GJ}(k;\theta,\hat\rho) \equiv \overline{H}^{GJ(\theta^{2\hat\rho},\theta)}(k) := \frac{\overline{H}(k) - \theta^{2\hat\rho}\ \overline{H}([\theta k])}{1-\theta^{2\hat\rho}}, \quad 0<\theta<1. \tag{23}$$

Again, and as before, for the sake of simplicity, we advise the choice $\theta = 1/2$ and the consideration of

$$\overline{H}^{GJ_1}(k) \equiv \overline{H}^{GJ}(k;1/2,\hat\rho) := \frac{\overline{H}(k) - 2^{-2\hat\rho}\ \overline{H}([k/2])}{1-2^{-2\hat\rho}}. \tag{24}$$

Recall the validity of the asymptotic distributional representation

$$\overline{H}(k) \stackrel{d}{=} \gamma + \frac{\sigma_{\overline{H}}}{\sqrt{k}}\, P_k + b_{\overline{H}} A^2(n/k)(1 + o_p(1)),$$

where $b_{\overline{H}} = b(\gamma, \rho, \xi) \in \mathbb{R}$, $\sigma_{\overline{H}} = \gamma > 0$, being $P_k \equiv P_{k,2}$, with $P_{k,2}$ given in (21), an asymptotically standard normal r.v. Indeed, we know that $P_k \stackrel{d}{=} \frac{1}{\sqrt{k}}\left(\sum_{i=1}^k E_i - k\right)$ with $\{E_i\}_{i\geq 1}$ a sequence of i.i.d. unit exponential r.v.s. For these estimators, and as $n \to \infty$, we know that

$$\sqrt{k}\left(\frac{\overline{H}(k) - \gamma}{\sigma_{\overline{H}}}\right) \stackrel{d}{\longrightarrow} N\left(\lambda b_{\overline{H}}, 1\right),$$

provided that we choose k such that $\sqrt{k}A^2(n/k) \to \lambda$, finite, as $n \to \infty$, i.e. $\overline{H}(k)$ is asymptotically normal, possibly with a non-null asymptotic bias. On the other hand, if $\sqrt{k}\ A^2(n/k) \to \infty$, then $\big(\overline{H}(k) - \gamma\big)/A^2(n/k) \stackrel{p}{\longrightarrow} b_{\overline{H}}$.

We can then straightforwardly prove the following theorem.

Theorem 3.2. *Let us further slightly restrict the class of models in* (8)*, working with models such that*

$$U(t) = C\ t^{\gamma}\left(1 + A(t)/\rho + O(A^2(t)) + O(A^3(t))\right).$$

Then, there exists b_{GJ_1}*, a function of first, second, third and fourth-order parameters, such that we get the validity of the asymptotic distributional representation*

$$\overline{H}^{GJ_1}(k) \stackrel{d}{=} \gamma + \frac{\gamma\sqrt{1 + 1/(2^{-2\rho} - 1)^2}}{\sqrt{k}} P_k + b_{GJ_1} A^3(n/k)) + o_p(A^3(n/k)),$$

again with P_k *a sequence of standard normal r.v.'s.*

3.3 Refined GJ corrected-bias EVI-estimators

Given the achieved theoretical results, the bias reduction we expected for the GJ EVI-estimators, $H^{GJ_1}(k)$ and $\overline{H}^{GJ_1}(k)$, in (15) and (24), respectively, was not fully achieved for finite sample sizes. This led us to turn back, to the AGJ estimator based on an affine combination of either H or $\overline{H}$-estimators computed at two different levels, considering new approximations for the bias quotient.

As we have seen before, when we consider a weight ($q_\bullet$) equal to the quotient between the bias of two estimators, we get an unbiased estimator of γ. Both the H^{GJ_1} and the $\overline{H}^{GJ_1}$ EVI-estimators were developed with this objective in mind, and using the relationships

(*i*) $\text{Bias}_\infty\,(H(k)) = b_{_H} A(n/k)$, $\text{Bias}_\infty\,\big(\overline{H}(k)\big) = b_{_{\overline{H}}}\, A^2(n/k)$,

(*ii*) $A(n/(\theta k)) \sim \theta^{-\rho} A(n/k)$, as $n \to \infty$.

This has led us to $q_{_H} = \theta^\rho$, and the GJ estimator in (16), and $q_{_{\overline{H}}} = \theta^{2\rho}$, associated with the GJ estimator in (23).

But we can consider a better approximation for $A(n/k)/A(n/(\theta k))$, valid for a wide class of models. Indeed, we can add a second term to such a quotient, assuming the validity of an adequate second-order condition on $A(\cdot)$. In a sub-class of Hall's class of models, to which belong models like the *generalized Pareto* and the *Burr*, we have, as $n/k \to \infty$,

$$\frac{A(n/k)}{A(n/(\theta k))} \sim \theta^\rho \Big(1 - \Big(\frac{k}{n}\Big)^{-\rho} (\theta^{-\rho} - 1)\Big).$$

Working in a way similar to the one used for the construction of either $H^{GJ_1}(k)$ or $\overline{H}^{GJ_1}(k)$, i.e. working with $\theta = 1/2$, we get

$$q_{_H} = 2^{-\rho}\Big(1 - \Big(\frac{k}{n}\Big)^{-\rho} (2^\rho - 1)\Big), \qquad q_{_{\overline{H}}} = 2^{-2\rho}\Big(1 - \Big(\frac{k}{n}\Big)^{-\rho} (2^\rho - 1)\Big)^2,$$

and the refined GJ EVI-estimators

$$\begin{aligned} &H^{GJ_2}(k) := a_{_H}(\hat\rho) H([k/2]) + (1 - a_{_H}(\hat\rho)) H(k), \\ &\quad \overline{H}^{GJ_2}(k) := a_{_{\overline{H}}}(\hat\rho) \overline{H}([k/2]) + (1 - a_{_{\overline{H}}}(\hat\rho)) \overline{H}(k), \quad a_\bullet := q_\bullet/(1 - q_\bullet), \quad (25) \end{aligned}$$

with the same asymptotic properties of $H^{GJ_1}(k)$ and $\overline{H}^{GJ_1}(k)$, respectively.

The improvement achieved for finite samples is usually significant, particularly for the GJ Hill estimator, as can be seen from the results in Section 4. It is thus worth finding better approximations for the asymptotic bias, or even estimate the bias through the bootstrap methodology.

3.4 Asymptotic bias and efficiency of an arbitrary affine combination of MVRB EVI-estimators

Note that an AGJ EVI-estimator can be regarded as an affine combination of two consistent EVI-estimators, with related asymptotic properties. Just as done before in Gomes *et al.* (2000) for the classical Hill estimator, in (2), the most obvious affine combination associated to the MVRB EVI-estimator $\overline{H}(k)$, in (6), is

$$\overline{H}^{GJ(a)}(k;\theta) := a\overline{H}([k\theta]) + (1-a)\overline{H}(k). \tag{26}$$

or the particular case associated to $\theta = 1/2$,

$$\overline{H}^{GJ(a)}(k) \equiv \overline{H}^{GJ(a)}(k;1/2) := a\overline{H}([k/2]) + (1-a)\overline{H}(k). \tag{27}$$

We have thus a class of estimators parameterized in the *tuning* parameter a, to be chosen in the most adequate way.

Remark 3.2. *Note that the r.v.* $\overline{H}^{GJ(a)}(k;\theta)$*, in* (26)*, has the functional form of* $\overline{H}^{GJ(q,\theta)}(k)$*, in* (22)*, with* $a = q/(q-1) = \theta^{2\rho}/(\theta^{2\rho}-1)$*. Since* $a = q/(q-1) > 1$*, we shall thus consider only* $a \geq 1$*, in* (26)*, the type of affine combinations that can lead to a bias reduction. Note however that* $\overline{H}(k)$ *corresponds to* $a = 0$*, in* (26)*.*

We shall next analyze the behaviour of the asymptotic bias and the efficiency of the particular affine combination, $\overline{H}^{GJ(a)}(k)$, in (27), studying first their asymptotic properties for a possibly non-optimal choice of $a \geq 1$.

i) For a fixed level k, the reduction in the asymptotic bias of $\overline{H}^{GJ(a)}(k)$ comparatively with the original estimator $\overline{H}(k)$ is measured through the *asymptotic bias reduction* indicator

$$ABR_a := \lim_{n\to\infty}\left(\left|\frac{\text{Bias}_\infty[\overline{H}(k)]}{\text{Bias}_\infty[\overline{H}^{(GJ(a))}(k)]}\right|\right) = \frac{1}{|1-a(1-2^{2\rho})|}.$$

In Figure 1, we present, in the (a,ρ)-plane, the values of this quotient (independent of the tail index γ).

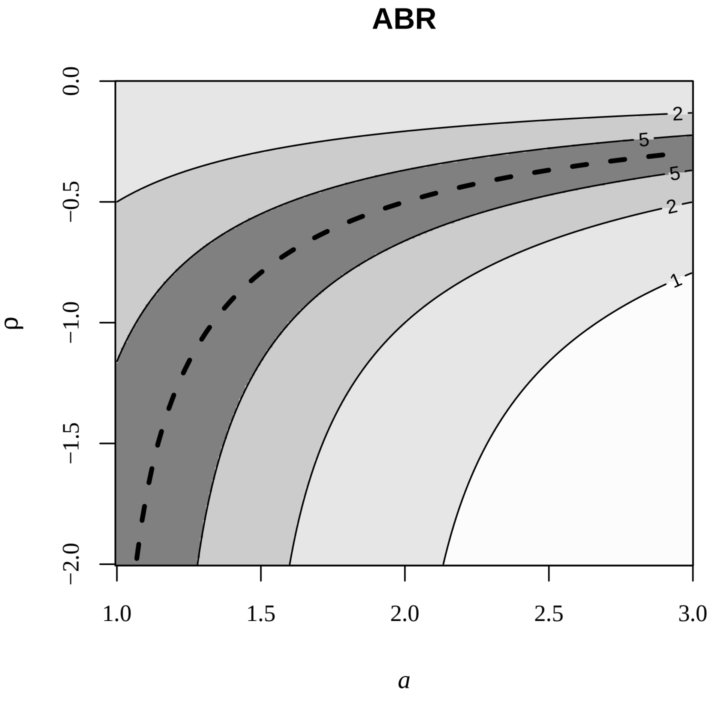

Figure 1: Asymptotic bias reduction indicator.

ii) Let us define the asymptotic efficiency of $\overline{H}^{GJ(a)}(k)$ relatively to $\overline{H}(k)$ as the quotient between the two asymptotic MSEs, computed at optimal levels. Provided that $a \neq 1/(1-2^{2\rho})$, we have

$$AREFF_a = \left(\frac{(a^2+1)^{2\rho}}{1-a(1-2^{2\rho})} \right)^{\frac{2}{1-4\rho}}.$$

We next show, in Figure 2, and again in the (a, ρ)-plane, the values of this asymptotic efficiency.

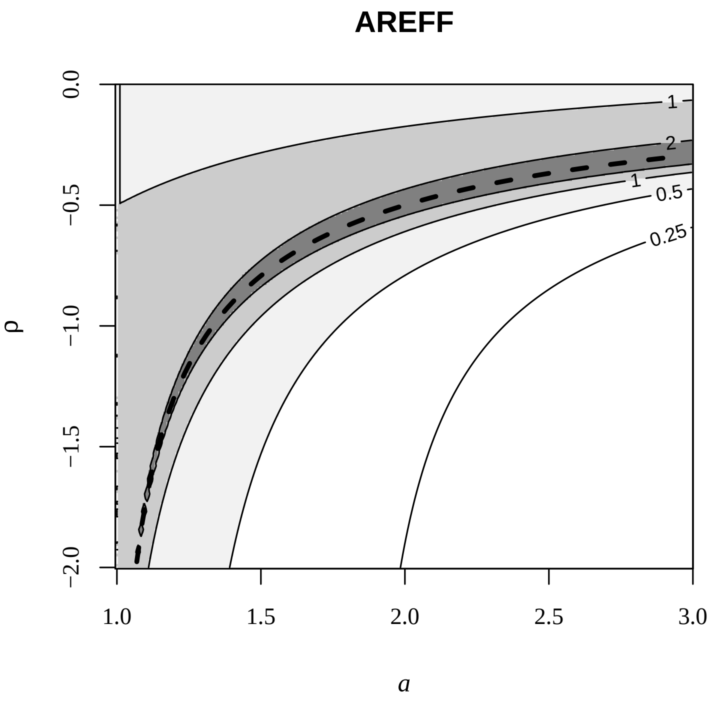

Figure 2: Asymptotic relative efficiency indicator.

It is clear from Figure 2 that for a reduction in MSE we indeed need to work in a line close to $a = 1/(1-2^{2\rho})$. However, the reduction in bias holds in a wide region of the (a, ρ)-plane, as can be seen in Figure 1.

4 Finite sample behaviour: a small-scale Monte-Carlo simulation

The Monte-Carlo simulation of mean values (Es) and root MSEs (RMSEs) of the estimators under study are based on 5000 runs. The relative efficiency of an estimator is defined as the quotient between the simulated RMSE of the H-estimator and the one of any of the estimators under study, both computed at their optimal levels, i.e. for any T-statistic,

$$REFF_{T|H} := \frac{RMSE(H_0)}{RMSE(T_0)},$$

with $T_0 := T(k_0^T)$, with $k_0^T := \arg\min_k MSE(T(k))$. The simulation of those efficiencies is based on 20×5000 replicates. Details on multi-sample Monte Carlo simulation can be found in Gomes and Oliveira (2001).

In this section we will denote the MVRB EVI-estimator $\overline{H}_{\hat\beta,\hat\rho}$, defined in (6), by $\overline{H}_0$, due to the reason referred in Remark 2.1. The AGJ EVI-estimators based on the Hill and the MVRB estimators, defined in (15) and (24), will be denoted $\overline{H}_{01}$ and $\overline{\overline{H}}_{01}$, respectively. Finally, for the refined GJ corrected-bias EVI-estimators based on the Hill and the MVRB estimators, and defined in (25), we shall use the notations $\overline{H}_{02}$ and $\overline{\overline{H}}_{02}$, respectively.

We have considered different parents with a heavy right-tail, like the Fréchet, with d.f., $F(x) = \exp(-x^{-1/\gamma}),\ x > 0$, the EV d.f., in (1) and the *generalized Pareto* (GP) d.f., $GP_\gamma(x) = 1 + \ln EV_\gamma(x),\ x > 0$, with $EV_\gamma(x)$ given in (1). As an illustration of the results obtained whenever $-1 < \rho < 0$, we present in Figure 3 the simulated mean values and RMSEs of the different estimators under study, for a sample of size $n = 1000$ from an EV d.f. with $\gamma = 0.5$ ($\rho = -0.5$).

The situations are diversified, and sometimes, particularly when ρ approaches -1, the GJ MVRB EVI-estimators are not able to overpass the $\overline{H}$

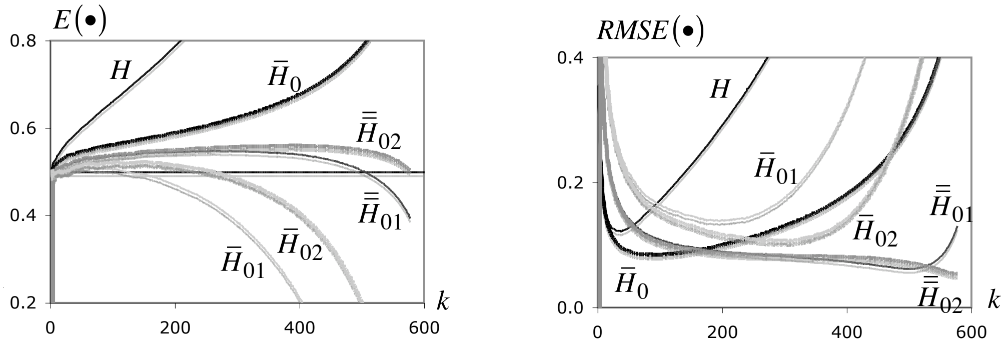

Figure 3: Simulated mean values (*left*) and RMSEs (*right*) of the estimators under study, for a sample of size $n = 1000$ from an $EV_{0.5}$-model.

estimator at optimal levels, in the sense of RMSE minimization, like happens with a Fréchet underlying parent. This can be seen from Figure 4, similar to Figure 3, but for a Fréchet underlying parent with $\gamma = 1$ ($\rho = -1$).

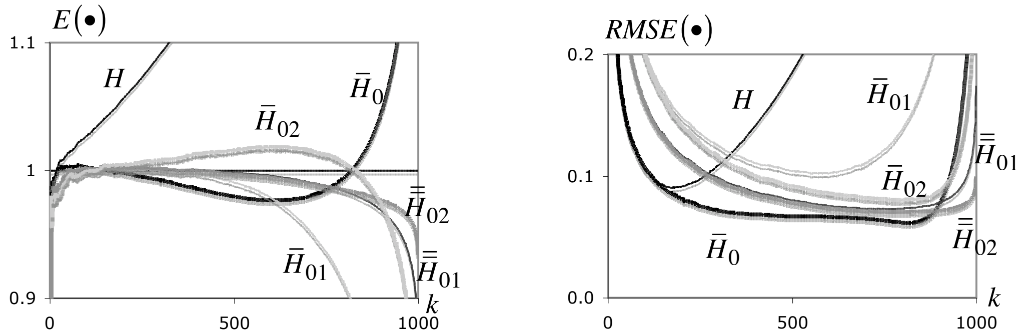

Figure 4: Simulated mean values (*left*) and RMSEs (*right*) of the estimators under study, for a sample of size $n = 1000$ from a Fréchet parent with $\gamma = 1$.

Indeed, as shown, in Figure 5, the REFF-indicators as a function of n, can be smaller than one, for the GJ Hill EVI-estimators and small n, both for the $EV_{0.5}$ and the Fréchet($\gamma = 1$) models. Moreover, whereas $\overline{\overline{H}}_{02}$, computed at its optimal level, in the sense of minimum MSE, just as mentioned above, attains the highest REFF for $EV_{0.5}$ underlying parents, as well as for other simulated parents with $\rho > -1$, it is only the second-best for underlying Fréchet parents.

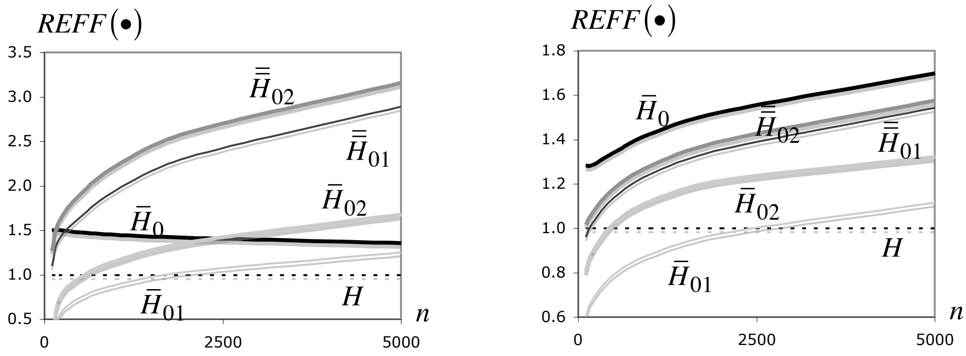

Figure 5: Simulated REFF indicators, as a function of the sample size n, for $EV_{0.5}$ (*left*) and Fréchet parents with $\gamma = 1$ (*right*).

Some general comments:

- All GJ-estimators have a bias smaller than the one of the original estimator.
- The reduction is higher when the weight q depends on the level k, as happens with the GJ_2 EVI-estimators. Regarding MSE, we are able to go below the MSE of the MVRB $\overline{H}$-estimator for a large variety of underlying parents and small values of $|\rho|$.
- Apart from what happens for very small values of ρ, there is a high reduction in the MSE of the GJ-estimator, at optimal levels, and comparatively with the original $\overline{H}$-estimator.

References

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

[2] Beirlant, J., Vynckier, P. and J. L. Teugels (1996b). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.* **91**, 1659-1667.

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

[4] Caeiro, F. and Gomes, M.I. (2006). A new class of estimators of a "scale" second order parameter. *Extremes* **9**, 193-211, 2007.

[5] Caeiro, F. and Gomes, M.I. (2011). Asymptotic comparison at optimal levels of reduced-bias extreme value index estimators. *Statistical Neerlandica* **65**:4, 462-488.

[6] Caeiro, F., Gomes, M.I. and Pestana, D.D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3**:2, 111-136.

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

[8] Ciuperca, G. and Mercadier, C. (2010). Semi-parametric estimation for heavy tailed distributions. *Extremes* **13**:1, 55-87.

[9] Drees, H. and E. Kaufmann (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Proc. and Appl.* **75**, 149-172.

[10] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**:1, 1-26.

[11] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

[12] 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**:2, 194-213.

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

[14] Goegebeur, Y., Beirlant, J. and de Wet, T. (2008). Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation. *Revstat* **6**:1, 5169.

[15] Goegebeur, Y., Beirlant, J. and de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics. *J. Statist. Plann. Inference* **140**:9, 2632-2654.

[16] 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**:1, 5-31.

[17] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes: choice of the optimal sample fraction. *Extremes* **4**:4, 331-358, 2002.

[18] Gomes, M. I. and Pestana, D. (2007a). A simple second order reduced-bias' tail index estimator. *J. Statist. Comp. and Simul.* **77**:6, 487-504.

[19] Gomes, M. I. and Pestana, D. (2007b). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.* **102**:477, 280–292.

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

[21] Gomes, M.I., de Haan, L. and Peng, L. (2002a). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour. *Extremes* **5**:4, 387–414.

[22] Gomes, M.I., Martins, M.J. and Neves, M.M. (2002b). Generalized jackknife semi-parametric estimators of the tail index. *Portugaliae Mathematica* **59**:4, 393–408.

[23] Gomes, M.I., Caeiro, F. and Figueiredo, F. (2004). Bias reduction of a tail index estimator through an external estimation of the second order parameter. *Statistics* **38**:6, 497-510.

[24] Gomes, M. I., de Haan, L. and Henriques-Rodrigues, L. (2008). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. Royal Statist. Soc.* **B70**:1, 31-52.

[25] Gomes, M.I., Henriques-Rodrigues, L., Pereira, H. and Pestana, D. (2010). Tail index and second order parameters' semi-parametric estimation based on the log-excesses. *J. Statist. Comput. and Simul.* **80**:6, 653-666.

[26] Gray, H.L. and Schucany, W.R. (1972). *The Generalized Jackknife Statistic.* Marcel Dekker.

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

[28] Haan, L. de and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52**, 60–70.

[29] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc.* **B44**, 37-42.

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

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

[32] Quenouille, M.H. (1949). Approximate tests of correlation in time series. *J. Royal Statist. Soc.* **B11**, 68-84.

[33] Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353–360.

[34] Tukey, J.W. (1958). Bias and confidence in not quite large samples (Abstract). *Ann. Math. Statist.* **29**, 614.