Adaptive Estimation of Heavy Right Tails: the Bootstrap Methodology in Action

M. Ivette Gomes

Universidade de Lisboa, FCUL, DEIO and CEAUL

Fernanda Figueiredo

Universidade do Porto, Faculdade de Economia and CEAUL

M. Manuela Neves

Universidade Técnica de Lisboa, Instituto Superior de Agronomia and CEAUL

November 30, 2010

Abstract

In this paper, we discuss an algorithm for the adaptive estimation of a positive *extreme value index*, γ, the primary parameter in *Statistics of Extremes*. Apart from the classical extreme value index estimators, we suggest the consideration of associated second-order corrected-bias estimators, and propose the use of bootstrap computer-intensive methods for an asymptotically consistent choice of the *thresholds* to use in the adaptive estimation of γ. The algorithm is described for a classical γ-estimator and associated corrected-bias estimator, but it can work similarly for the estimation of other parameters of extreme events, like a *high quantile*, the *probability of exceedance* or the *return period of a high level*.

AMS 2000 subject classification. Primary 62G32, 62E20; Secondary 65C05.

Keywords and phrases. *Statistics of extremes; semi-parametric estimation; bootstrap methodology.*

1 Introduction and outline of the paper

Heavy-tailed models appear often in practice in fields like Telecommunications, Insurance, Finance, Bibliometrics and Biostatistics. Power laws, such as the Pareto distribution and the Zipf's law, have been observed a few decades ago in some important phenomena in Economics and Biology, and have seriously attracted scientists in recent years.

We shall essentially deal with the estimation of a positive *extreme value index* (EVI), denoted γ, the primary parameter in *Statistics of Extremes*. Apart from the classical Hill, moment and generalized-Hill semi-parametric estimators of γ, we shall also consider associated classes of second-order reduced-bias estimators. These classes are based on the adequate estimation of a "scale" and a "shape" second-order parameters, β and ρ, respectively, are valid for a large class of heavy-tailed underlying parents and are appealing in the sense that we are able to reduce the asymptotic bias of a classical estimator without increasing its asymptotic variance. We shall call these estimators "*classical-variance reduced-bias*" (CVRB) estimators.

After the introduction, in Section 2, of a few technical details in the area of *Extreme Value Theory* (EVT), related with the EVI-estimators under consideration in this paper, we shall briefly discuss, in Section 3, the asymptotic properties of those estimators, and the kind of second-order parameters' estimation which enables the building of corrected-bias estimators with the same asymptotic variance of the associated classical estimators, i.e., the building of CVRB estimators. In Section 4, we propose an algorithm for the adaptive estimation of a positive EVI, through the use of bootstrap computer-intensive methods. The algorithm is described for a classical EVI estimator and associated CVRB estimator, but it can work similarly for the estimation of other parameters of extreme events, like a high quantile, the probability of exceedance or the return period of a high level. In Section 5, we present some of the results of a large-scale Monte-Carlo simulation related with the behaviour of the non-adaptive and adaptive estimators under consideration, emphasizing the low coverage probabilities of bootstrap confidence intervals. Finally, Section 6 is entirely dedicated to the application of the algorithm described in Section 4 to the analysis of environmental data related with the number of hectares burned during all wildfires that were recorded in Portugal in the period 1999-2003.

2 The EVI-estimators under consideration

In the area of EVT, and whenever dealing with large values, a model F is said to be *heavy-tailed* whenever the *right-tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation, denoted $-1/\gamma$, i.e., for all $x > 0$, there exists $\gamma > 0$, such that

$$\overline{F}(tx)/\overline{F}(t) \underset{t\to\infty}{\longrightarrow} x^{-1/\gamma}. \tag{2.1}$$

If (2.1) holds, we use the notation $\overline{F} \in RV_{-1/\gamma}$, with RV standing for *regular variation*, and we are then working in the whole domain of attraction (for maxima) of heavy-tailed models, denoted $\mathcal{D}_{\mathcal{M}}^{+} \equiv \mathcal{D}_{\mathcal{M}}\left(EV_\gamma\right)_{\gamma>0}$. Equivalently, with

$$U(t) := F^{\leftarrow}\big(1 - 1/t\big) = \inf\{x : F(x) \geq 1 - 1/t\}$$

denoting a reciprocal quantile function, we have the validity of the so-called *first-order condition*,

$$F \in \mathcal{D}_{\mathcal{M}}^{+} \quad\Longleftrightarrow\quad \overline{F} \in RV_{-1/\gamma} \quad\Longleftrightarrow\quad U \in RV_\gamma. \tag{2.2}$$

For these heavy-tailed parents, given a sample $\underline{X}_n := (X_1, X_2, \ldots, X_n)$ and the associated sample of ascending order statistics (o.s.'s), $(X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n})$, the classical EVI estimator is the Hill (H) estimator (Hill, 1975),

$$H(k) \equiv H_{k,n} := \frac{1}{k} \sum_{i=1}^{k} \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}, \tag{2.3}$$

the average of the k log-excesses over a high random threshold $X_{n-k:n}$, which needs to be an *intermediate* o.s., i.e., k needs to be such that

$$k = k_n \to \infty \quad \text{and} \quad k/n \to 0, \text{ as } n \to \infty. \tag{2.4}$$

But the Hill-estimator $H(k)$, in (2.3), reveals usually a high non-null asymptotic bias at optimal levels, i.e., levels k where the mean squared error (MSE) is minimum. 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, and a very sharp MSE pattern, as a function of k. Recently, several authors have been dealing with bias reduction in the field of *extremes* (for an overview, see Reiss and Thomas, 2007, Chapter 6, 189-204, as well as the more recent paper by Gomes *et al.*, 2008a). We then need to work in a region slightly more restrict than $\mathcal{D}_{\mathcal{M}}^{+}$. In this paper, we shall consider parents such that, as $t \to \infty$, the *third-order condition*,

$$U(t) = Ct^{\gamma}\big(1 + A(t)/\rho + O(A^2(t))\big), \quad A(t) =: \gamma\beta t^{\rho}, \tag{2.5}$$

holds, with $\gamma > 0$, $\rho < 0$, and $\beta \neq 0$. The most 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), with VaR_p standing for the Value-at-Risk at the level p, the size of the loss occurred with a small probability p. This class of EVI-estimators, here denoted $\overline{H} \equiv \overline{H}(k)$, is the CVRB-estimator associated with the Hill estimator $H = H(k)$, in (2.3), and depends upon the estimation of the second-order parameters (β, ρ), in (2.5). Its functional form is

$$\overline{H}(k) \equiv \overline{H}_{k,n;\hat\beta,\hat\rho} := H(k)\big(1 - \hat\beta(n/k)^{\hat\rho}/(1-\hat\rho)\big), \tag{2.6}$$

where $(\hat\beta, \hat\rho)$ is an adequate consistent estimator of (β, ρ). Algorithms for the estimation of (β, ρ) are provided, for instance, in Gomes and Pestana (2007a,b), and one of them will be reformulated in the **Algorithm** presented in Section 4.2 of this paper.

Apart from the *Hill* estimator, in (2.3), we suggest the consideration of two other classical estimators, valid for all $\gamma \in \mathbb{R}$, but considered here exclusively for heavy tails, the *moment* (Dekkers *et al.*, 1989) and the *generalized-Hill* (Beirlant *et al.*, 1996, 2005) estimators. The *moment* (M) estimator has the functional expression

$$M(k) \equiv M_{k,n} := M_{k,n}^{(1)} + \tfrac{1}{2}\Big\{1 - \big(M_{k,n}^{(2)}/(M_{k,n}^{(1)})^2 - 1\big)^{-1}\Big\}, \tag{2.7}$$

with

$$M_{k,n}^{(j)} := \frac{1}{k}\sum_{i=1}^{k}\big(\ln X_{n-i+1:n} - \ln X_{n-k:n}\big)^{j}, \ j \geq 1, \tag{2.8}$$

$M_{k,n}^{(1)} \equiv H(k)$, in (2.3). The *generalized Hill* (GH) estimator is defined for $k = 2, \ldots, n-1$, and it is given by

$$GH(k) \equiv GH_{k,n} := \frac{1}{k}\sum_{j=1}^{k}\ln UH_{j,n} - \ln UH_{k,n}, \tag{2.9}$$

$$UH_{j,n} := X_{n-j:n}H_{j,n}, \ 1 \leq j \leq k,$$

with $H_{k,n}$ defined in (2.3). To enhance the similarity between the moment estimator, in (2.7), and the generalized Hill estimator, in (2.10), we can also write an asymptotically equivalent expression for $GH(k)$, given by

$$GH^*(k) := H_{k,n} + \frac{1}{k}\sum_{i=1}^{k}\big\{\ln H_{i,n} - \ln H_{k,n}\big\}. \tag{2.10}$$

This means that $H_{k,n} \equiv M_{k,n}^{(1)}$ is estimating $\gamma^+ := \max(0,\gamma)$, both in (2.7) and (2.10), whereas $\gamma^- := \min(0,\gamma) = \gamma - \gamma^+$ is being estimated differently.

The associated bias-corrected moment and generalized Hill estimators have similar expressions, due to the same dominant component of asymptotic bias of the estimators in (2.7) and (2.10), whenever the EVI is positive (see Gomes and Neves, 2008, among others). Denoting generally $\overline{W}$, either $\overline{M}$ or $\overline{GH}$, and with the notation W for either M or GH, we get

$$\overline{W}(k) \equiv \overline{W}_{k,n;\hat\beta,\hat\rho} := W(k)\Big(1 - \hat\beta\,(n/k)^{\hat\rho}\,/(1-\hat\rho)\Big) - \hat\beta\;\hat\rho\,(n/k)^{\hat\rho}\,/(1-\hat\rho)^2. \tag{2.11}$$

In the sequel, we generally denote C any of the classical EVI-estimators, in (2.3), (2.7) and (2.10), and $\overline{C}$ the associated CVRB-estimator.

3 Asymptotic behaviour of the estimators

In order to obtain a non-degenerate behaviour for any EVI-estimator, under a semi-parametric framework, it is convenient to assume a second-order condition, measuring the rate of convergence in the first-order condition, given in (2.2). Such a condition involves a non-positive parameter ρ, and can be given by

$$\lim_{t\to\infty}\frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma\left(\frac{x^\rho - 1}{\rho}\right), \tag{3.1}$$

for all $x > 0$, where $A(\cdot)$ is a suitably chosen function of constant sign near infinity. Then, $|A| \in RV_\rho$ (Geluk and de Haan, 1987).

In this paper, as mentioned before and mainly because of the reduced-bias estimators in (2.6) and (2.11), generally denoted $\overline{C}(k) \equiv \overline{C}_{k,n;\hat\beta,\hat\rho}$, we shall slightly more restrictively assume that the third-order condition (2.5) holds. Then, (3.1) holds, with $A(t) = \gamma\beta t^\rho$, the parametrization used in (2.5). For the classical H, M and GH estimators, generally denoted C, we know that for any intermediate sequence k, as in (2.4), and even under the validity of the second-order condition in (3.1),

$$C(k) \stackrel{d}{=} \gamma + \frac{\sigma_{_C} Z_k^C}{\sqrt{k}} + b_{_{C,1}} A(n/k)\big(1 + o_p(1)\big), \tag{3.2}$$

where

$$\sigma_{_H} = \gamma, \qquad b_{_{H,1}} = \frac{1}{1-\rho}, \qquad \sigma_{_M} = \sigma_{_{GH}} = \sqrt{\gamma^2+1},$$
$$b_{_{M,1}} = b_{_{GH,1}} = \frac{\gamma(1-\rho)+\rho}{\gamma(1-\rho)^2} = \frac{1}{1-\rho} + \frac{\rho}{\gamma(1-\rho)^2}, \tag{3.3}$$

being Z_k^C ($C = H$ or M or GH) asymptotically standard normal r.v.'s (de Haan and Peng, 1998; Dekkers *et al.*, 1989; Beirlant *et al.*, 1996, 2005). See also de Haan and Ferreira, 2006.

The above mentioned properties, together with trivial adaptations of the results in Caeiro *et al.* (2005, 2009) and Gomes *et al.* (2008c), for $\overline{H}$, enable us to state, the following theorem, again for models with a positive EVI. We shall include in the statement of the theorem both the classical and the associated CVRB estimators.

Theorem 3.1. *Assume that condition* (3.1) *holds, and let* $k = k_n$ *be an intermediate sequence, i.e.,* (2.4) *holds. Then, there exist a sequence* Z_k^C *of asymptotically standard normal random variables, and for the real numbers* $\sigma_{_C} > 0$ *and* $b_{_{C,1}}$ *given in* (3.3), *the asymptotic distributional representation* (3.4) *holds. If we further assume that* (2.5) *holds, there exists an extra real number* $b_{_{C,2}}$, *such that we can write*

$$C(k) \stackrel{d}{=} \gamma + \frac{\sigma_{_C} Z_k^C}{\sqrt{k}} + b_{_{C,1}} A(n/k) + b_{_{C,2}} A^2(n/k)\big(1 + o_p(1)\big), \tag{3.4}$$

Under the validity of equation (2.5), *if we estimate* β *and* ρ *consistently through* $\hat\beta$ *and* $\hat\rho$, *in such a way that* $\hat\rho - \rho = o_p(1/\ln n)$, *we can guarantee that there exists a pair of real numbers* $(b_{_{\overline{C},1}}, b_{_{\overline{C},2}})$, *such that for adequate* k *values of an order up to* k *such that* $\sqrt{k}A^4(n/k) \to \lambda_{_A}$, *finite,*

$$\overline{C}(k) \stackrel{d}{=} \gamma + \frac{\sigma_{_C} Z_k^C}{\sqrt{k}} + b_{_{\overline{C},1}}\ A(n/k) + b_{_{\overline{C},2}}\ A^2(n/k)\ (1 + o_p(1)). \tag{3.5}$$

Moreover, $b_{_{\overline{C},1}} = 0$, $\forall \overline{C}$.

Proof. The proof of the theorem for $\overline{H}$, in (2.6), follows from the above mentioned papers. For the $\overline{W}$ estimators, in (2.11), the proof is also similar. If we estimate consistently β and ρ through the estimators $\hat{\beta}$ and $\hat{\rho}$ in the conditions of the theorem, we may use Cramer's delta-method, and write,

$$\begin{aligned}\overline{W}_{k,n,\hat{\beta},\hat{\rho}} \;=\; W_{k,n} \;\times \Big(1 &- \frac{\beta}{1-\rho}\left(\frac{n}{k}\right)^{\rho} - (\hat{\beta}-\beta)\frac{1}{1-\rho}\left(\frac{n}{k}\right)^{\rho}(1+o_p(1)) \\ &- \frac{\beta}{1-\rho}(\hat{\rho}-\rho)\left(\frac{n}{k}\right)^{\rho}\Big(\frac{1}{1-\rho}+\ln(n/k)\Big)(1+o_p(1))\Big) - \frac{\beta\,\rho}{(1-\rho)^2}\left(\frac{n}{k}\right)^{\rho} \\ &- \Big\{(\hat{\beta}-\beta)\frac{\rho}{(1-\rho)^2}\left(\frac{n}{k}\right)^{\rho} + \frac{\beta(\hat{\rho}-\rho)}{1-\rho}\left(\frac{n}{k}\right)^{\rho}\Big(\frac{\rho\ln(n/k)}{1-\rho}+3-\rho\Big)\Big\}(1+o_p(1)).\end{aligned}$$

We can then guarantee the existence of real values $u_{_W}$ and $v_{_W}$ such that

$$\overline{W}_{k,n,\hat{\beta},\hat{\rho}} \stackrel{d}{=} \overline{W}_{k,n,\beta,\rho} - \frac{A(n/k)}{1-\rho}\left(u_{_W}\left(\frac{\hat{\beta}-\beta}{\beta}\right) + v_{_W}(\hat{\rho}-\rho)\;\ln(n/k)\right)(1+o_p(1)).$$

The reasoning is then quite similar to the one used in Gomes *et al.* (2008c) for the $\overline{H}$-estimator. Since $\hat{\beta}$ and $\hat{\rho}$ are consistent for the estimation of β and ρ, respectively, and $(\hat{\rho}-\rho)\ln(n/k) = o_p(1)$, the last summand is obviously $o_p(A(n/k))$, and can even be $o_p(A^2(n/k))$. □

Remark 3.1. *Note that the values of* $b_{_{H,1}}$, $b_{_{M,1}}$ *and* $b_{_{GH,1}}$, *in* (3.3), *provide an easy heuristic justification for the CVRB estimators in* (2.6) *and* (2.11).

Remark 3.2. *Only the external estimation of both* β *and* ρ *at a level* k_1, *adequately chosen, and the estimation of* γ *at a level* $k = o(k_1)$, *or at a specific value* $k = O(k_1)$, *can lead to a CVRB estimator, with an asymptotic variance* $\sigma^2_{_C}$. *Such a choice of* k *and* k_1 *is theoretically possible, as shown in* Gomes *et al.* (2008c) *and in* Caeiro *et al.* (2009), *but under conditions difficult to guarantee in practice. As a compromise between theoretical and practical results, and with* $[x]$ *denoting, as usual, the integer part of* x, *we have so far advised any choice* $k_1 = [n^{1-\epsilon}]$, *with* ϵ *small* (*see* Caeiro *et al.*, 2005, 2009 *and* Gomes *et al.*, 2007, 2008b, *among others*). *Later on, in the algorithm described in* Section 4, *we shall consider* $\epsilon = 0.001$, *i.e.*, $k_1 = [n^{0.999}]$. *Then we get* $\sqrt{k_1}\;A(n/k_1) \to \infty$ *if and only if* $\rho > -499.5$, *an almost irrelevant restriction in the class* (2.5). *We can then guarantee that* $\hat{\rho}-\rho = o_p(1/\ln n)$, *and the above mentioned behaviour, described in* Theorem 3.1, *for the reduced-bias EVI-estimators. The estimation of* γ, β *and* ρ *at the same value* k *would* *lead to a high increase in the asymptotic variance of the RB-estimators* $\overline{C}_{k,n;\hat\beta,\hat\rho}$*, which would become* $\sigma^2_{_C}\,((1-\rho)/\rho)^4$ (*see* Feuerverger and Hall, 1999; Beirlant *et al.*, 1999; Peng and Qi, 2004, *also among others*). *The external estimation of* ρ *at* k_1*, but the estimation of* γ *and* β *at the same* $k = o(k_1)$*, enables a slight decreasing of the asymptotic variance to* $\sigma^2_{_C}\,((1-\rho)/\rho)^2$*, stilll greater than* $\sigma^2_{_C}$ (*see* Gomes and Martins, 2002, *again among others*). *However, even in such cases, the results in* Section 4 *are still valid.*

Remark 3.3. *Let* $k = k_n$ *be intermediate and such that* $\sqrt{k}\;A(n/k) \to \lambda$*, finite, the type of levels* k *where the MSE of* $C(k)$ *is minimum. Let* $\hat\gamma(k)$ *denote either* $C(k)$ *or* $\overline{C}(k)$*. Then*

$$\sqrt{k}\big(\hat\gamma(k)-\gamma\big) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal(\lambda\; b_{\hat\gamma,1},\ \sigma^2_{_C}),$$

even if we work with the CVRB EVI-estimators, and we thus get asymptotically a null mean value ($b_{_{\overline{C},1}} = 0$). *Since* $b_{_{C,1}} \neq 0$ *whereas* $b_{_{\overline{C},1}} = 0$*, the* $\overline{C}$*-estimators outperform the* C*-estimators for all* k*, as illustrated in* Figure 1.

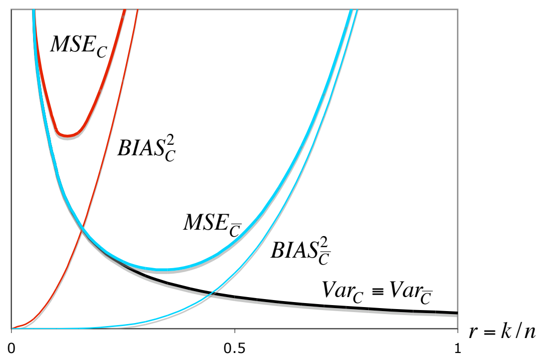

Figure 1: Patterns of asymptotic variances (Var), squared bias ($BIAS^2$) and MSE of a classical EVI-estimator, C, and associated CVRB estimator, $\overline{C}$.

Under the conditions of Theorem 3.1*, if* $\sqrt{k}\;A(n/k) \to \infty$*, with* $\sqrt{k}\;A^2(n/k) \to \lambda_{_A}$*, finite, the type of levels* k *where the MSE of* $\overline{C}(k)$ *is minimum, then*

$$\sqrt{k}\;\;\big(\overline{C}(k)-\gamma\big) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\big(\lambda_{_A}\; b_{_{\overline{C},2}},\ \sigma^2_{_C}\big).$$

4 The bootstrap methodology and adaptive classical and CVRB EVI-estimation

With AMSE standing for "asymptotic MSE", $\hat\gamma$ denoting either C or $\overline{C}$, and with

$$k_0^{\widehat\gamma}(n) := \arg\min_k MSE(\widehat\gamma(k)), \tag{4.1}$$

we get, on the basis of (3.4) and (3.5),

$$\begin{aligned} k_{0|\widehat\gamma}(n) &:= \arg\min_k AMSE\big(\widehat\gamma(k)\big) \\ &= \arg\min_k \begin{cases} \big(\sigma_C^2/k + b_{C,1}^2\ A^2(n/k)\big) & \text{if} \quad \widehat\gamma = C \\ \big(\sigma_C^2/k + b_{\overline{C},2}^2\ A^4(n/k)\big) & \text{if} \quad \widehat\gamma = \overline{C} \end{cases} \\ &= k_0^{\widehat\gamma}(n)(1+o(1)). \end{aligned} \tag{4.2}$$

The bootstrap methodology can thus enable us to consistently estimate the optimal sample fraction (OSF), $k_0^{\widehat\gamma}(n)/n$, with $k_0^{\widehat\gamma}(n)$ defined in (4.1), on the basis of a consistent estimator of $k_{0|\widehat\gamma}(n)$, in (4.2), in a way similar to the one used for the classical EVI estimation in Draisma *et al.* (1999), Danielson *et al.* (2001) and Gomes and Oliveira (2001). We shall here use the auxiliary statistics

$$T_{k,n} \equiv T(k|\hat\gamma) \equiv T_{k,n|\hat\gamma} := \widehat\gamma([k/2]) - \widehat\gamma(k), \quad k = 2, \ldots, n-1, \tag{4.3}$$

which converge in probability to zero, for intermediate k, and have an asymptotic behaviour strongly related with the asymptotic behaviour of $\widehat\gamma(k)$. Indeed, under the above-mentioned third-order framework in (2.5), we easily get

$$T(k|\hat\gamma) \ \stackrel{d}{=}\ \frac{\sigma_{\hat\gamma}\ P_k^{\hat\gamma}}{\sqrt{k}} + \begin{cases} b_{\hat\gamma,1}(2^\rho - 1)\ A(n/k)(1+o_p(1)) & \text{if} \quad \widehat\gamma = C \\ b_{\hat\gamma,2}(2^{2\rho} - 1)\ A^2(n/k)(1+o_p(1)) & \text{if} \quad \widehat\gamma = \overline{C}, \end{cases}$$

with $P_k^{\hat\gamma}$ asymptotically standard normal.

Consequently, denoting $k_{0|T}(n) := \arg\min_k AMSE(T_{k,n})$, we have

$$k_{0|\widehat\gamma}(n) = k_{0|T}(n) \times \begin{cases} (1-2^\rho)^{\frac{2}{1-2\rho}}\ (1+o(1)) & \text{if} \quad \widehat\gamma = C \\ (1-2^{2\rho})^{\frac{2}{1-4\rho}}\ (1+o(1)) & \text{if} \quad \widehat\gamma = \overline{C}. \end{cases} \tag{4.4}$$

4.1 The bootstrap methodology in action

How does the bootstrap methodology then work? Given the sample $\underline{X}_n = (X_1, \ldots, X_n)$ from an unknown model F, and the functional in (4.3), $T_{k,n} =: \phi_k(\underline{X}_n)$, $1 < k < n$, consider for any $n_1 = O(n^{1-\epsilon})$, $0 < \epsilon < 1$, the bootstrap sample

$$\underline{X}^*_{n_1} = (X^*_1, \ldots, X^*_{n_1}),$$

from the model

$$F^*_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i \le x]},$$

the empirical d.f. associated with the available sample, $\underline{X}_n$.

Next, associate to the bootstrap sample the corresponding bootstrap auxiliary statistic, $T^*_{k_1,n_1} := \phi_{k_1}(\underline{X}^*_{n_1})$, $1 < k_1 < n_1$. Then, with $k^*_{0|T}(n_1) = \arg\min_{k_1} AMSE\big(T^*_{k_1,n_1}\big)$,

$$\frac{k^*_{0|T}(n_1)}{k_{0|T}(n)} = \left(\frac{n_1}{n}\right)^{-\frac{c\rho}{1-c\rho}}(1+o(1)), \qquad c = \begin{cases} 2 & \text{if} \quad \widehat{\gamma} = C \\ 4 & \text{if} \quad \widehat{\gamma} = \overline{C}. \end{cases}$$

Consequently, for another sample size n_2, and for every $\alpha > 1$,

$$\frac{\big(k^*_{0|T}(n_1)\big)^{\alpha}}{k^*_{0|T}(n_2)} = \left(\frac{n_1^{\alpha}}{n^{\alpha}}\frac{n}{n_2}\right)^{-\frac{c\rho}{1-c\rho}} \big\{k_{0|T}(n)\big\}^{\alpha-1}(1+o(1)).$$

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

$$\big(k^*_{0|T}(n_1)\big)^2 / k^*_{0|T}(n_2) = k_{0|T}(n)(1+o(1)), \text{ as } n \to \infty. \tag{4.5}$$

On the basis of (4.5), we are now able to consistently estimate $k_{0|T}$ and next $k_{0|\hat{\gamma}}$ through (4.4), on the basis of any estimate $\hat{\rho}$ of the second-order parameter ρ. Such an estimate is also a consistent estimate of $k_0^{\widehat{\gamma}}(n)$, in (4.1). With $\hat{k}^*_{0|T}$ denoting the sample counterpart of $k^*_{0|T}$, and $\hat{\rho}$ an adequate ρ-estimate, we have the k_0-estimate

$$\hat{k}^*_{0\hat{\gamma}} \equiv \hat{k}_0^{\widehat{\gamma}}(n; n_1) := \min\Big(n-1, \ \big[c_{\hat{\rho}} \ (\hat{k}^*_{0|T}(n_1))^2/\hat{k}^*_{0|T}([n_1^2/n]+1)\big]+1\Big), \tag{4.6}$$

with

$$c_{\hat{\rho}} = \begin{cases} \big(1-2^{\hat{\rho}}\big)^{\frac{2}{1-2\hat{\rho}}} & \text{if} \quad \widehat{\gamma} = C \\ \big(1-2^{2\hat{\rho}}\big)^{\frac{2}{1-4\hat{\rho}}} & \text{if} \quad \widehat{\gamma} = \overline{C}. \end{cases}$$

The adaptive estimate of γ is then given by

$$\widehat{\gamma}^* \equiv \widehat{\gamma}^*_{n,n_1|T} := \widehat{\gamma}(\hat{k}_0^{\widehat{\gamma}}(n; n_1)). \tag{4.7}$$

4.2 Algorithm for adaptive EVI-estimation through C and $\overline{C}$

Again, with $\widehat{\gamma}$ denoting any of the estimators C or $\overline{C}$, we proceed with the description of the algorithm for the adaptive bootstrap estimation of γ, where in **Steps 1.**, **2.** and **3.** we reproduce the algorithm provided in Gomes and Pestana (2007b) for the estimation of the second-order parameters β and ρ.

Algorithm

1. Given an observed 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 Fraga Alves *et al.* (2003). Such estimators depend on the statistics

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

where $M_{k,n}^{(j)}$, $j = 1, 2, 3$, are given in (2.8), and have the functional form

$$\hat{\rho}_\tau(k) := \min\left(0, \ \frac{3(V_{k,n}^{(\tau)} - 1)}{V_{k,n}^{(\tau)} - 3}\right). \tag{4.8}$$

2. Consider $\{\hat{\rho}_\tau(k)\}_{k \in \mathcal{K}}$, with $\mathcal{K} = \left([n^{0.995}], \ [n^{0.999}]\right)$, compute their median, 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 $\tau^* = 1$.

3. Work with $\hat{\rho} \equiv \hat{\rho}_{\tau^*} = \hat{\rho}_{\tau^*}(k_1)$ and $\hat{\beta} \equiv \hat{\beta}_{\tau^*} := \hat{\beta}_{\hat{\rho}_{\tau^*}}(k_1)$, with $k_1 = [n^{0.999}]$, being $\hat{\beta}_{\hat{\rho}}(k)$ the estimator in Gomes and Martins (2002), given by

$$\hat{\beta}_{\hat{\rho}}(k) := \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{d_k(\hat{\rho}) \ D_k(0) - D_k(\hat{\rho})}{d_k(\hat{\rho}) \ D_k(\hat{\rho}) - D_k(2\hat{\rho})}, \tag{4.9}$$

dependent on the estimator $\hat{\rho} = \hat{\rho}_{\tau^*}(k_1)$, and where, for any $\alpha \le 0$,

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

with

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

the *scaled log-spacings*.

4. Compute $\widehat{\gamma}(k)$, $k = 1, 2, \ldots, n-1$.

5. Next, consider the sub-sample size $n_1 = [n^{0.955}]$ and $n_2 = [n_1^2/n] + 1$.

6. For l from 1 till B, generate independently, from the empirical d.f. $F_n^*(x) = \frac{1}{n}\sum_{i=1}^n I_{[X_i \le x]}$, associated with the observed sample, B bootstrap samples

$$(x_1^*, \ldots, x_{n_2}^*) \quad \text{and} \quad (x_1^*, \ldots, x_{n_2}^*, x_{n_2+1}^*, \ldots, x_{n_1}^*),$$

of sizes n_2 and n_1, respectively.

7. Denoting $T_{k,n}^*$ the bootstrap counterpart of $T_{k,n}$, in (4.3), obtain, for $1 \le l \le B$, $t_{k,n_1,l}^*$, $1 < k < n_1$, $t_{k,n_2,l}^*$, $1 < k < n_2$, the observed values of the statistic T_{k,n_i}^*, $i = 1, 2$, and compute

$$MSE^*(n_i, k) = \frac{1}{B}\sum_{l=1}^{B}\left(t_{k,n_i,l}^*\right)^2, \quad k = 2, \ldots, n_i - 1.$$

8. Obtain $\hat{k}_{0T}^*(n_i) := \arg\min_{1<k<n_i} MSE^*(n_i, k)$, $i = 1, 2$.

9. Compute $\hat{k}_{0\hat{\gamma}}^* \equiv \hat{k}_0^{\widehat{\gamma}}(n; n_1)$, given in (4.6).

10. Compute $\hat{\gamma}^* \equiv \widehat{\gamma}_{n,n_1|T}^*$, given in (4.7).

In order to obtain a final adaptive estimate of γ on the basis of one of the estimators under consideration, we still suggest the estimation of the MSE of any of the EVI-estimators at the bootstrap k_0-estimate, in **Step 9.**, say the estimation of $MSE(\widehat{\gamma}(\hat{k}_{0\hat{\gamma}}^*))$, with $\widehat{\gamma} \in \{H, \overline{H}, M, \overline{M}, GH, \overline{GH}\}$, and the choice of the estimate $\widehat{\gamma}$ for which $MSE(\widehat{\gamma}(\hat{k}_{0\hat{\gamma}}^*))$ is minimum, i.e., the consideration of an extra step, after Step **7.**:

7'. For $k = 2, \ldots, n_2 - 1$, compute $Bias^*(n_i, k) = \frac{1}{B}\sum_{l=1}^{B} t_{k,n_i,l}^*$, $\; i = 1, 2$.

Finally, we add the extra step:

11. Compute $RMSE_{\hat{\gamma}}^* = \sqrt{\widehat{MSE}(\hat{k}_{0\hat{\gamma}}^*|\hat{\gamma}^*)}$, with $\widehat{MSE}(\hat{k}_{0\hat{\gamma}}^*|\hat{\gamma}^*)$ given by

$$\widehat{MSE}(\hat{k}_{0\hat{\gamma}}^*|\hat{\gamma}^*) := \begin{cases} \frac{(\hat{\gamma}^*)^2}{\hat{k}_{0\hat{\gamma}}^*} + \left(\frac{(Bias^*(n_1,\hat{k}_{0\hat{\gamma}}^*))^2}{(2^{\hat{\rho}}-1)Bias^*(n_2,\hat{k}_{0\hat{\gamma}}^*)}\right)^2 & \text{if } \widehat{\gamma} = H \\ \frac{(\hat{\gamma}^*)^2}{\hat{k}_{0\hat{\gamma}}^*} + \left(\frac{(Bias^*(n_1,\hat{k}_{0\hat{\gamma}}^*))^2}{(2^{2\hat{\rho}}-1)Bias^*(n_2,\hat{k}_{0\hat{\gamma}}^*)}\right)^2 & \text{if } \widehat{\gamma} = \overline{H} \\ \frac{(\hat{\gamma}^*)^2+1}{\hat{k}_{0\hat{\gamma}}^*} + \left(\frac{(Bias^*(n_1,\hat{k}_{0\hat{\gamma}}^*))^2}{(2^{\hat{\rho}}-1)Bias^*(n_2,\hat{k}_{0\hat{\gamma}}^*)}\right)^2 & \text{if } \widehat{\gamma} = M \text{ or } GH \\ \frac{(\hat{\gamma}^*)^2+1}{\hat{k}_{0\hat{\gamma}}^*} + \left(\frac{(Bias^*(n_1,\hat{k}_{0\hat{\gamma}}^*))^2}{(2^{2\hat{\rho}}-1)Bias^*(n_2,\hat{k}_{0\hat{\gamma}}^*)}\right)^2 & \text{if } \widehat{\gamma} = \overline{M} \text{ or } \overline{GH}, \end{cases}$$

and consider the final estimate, $\hat{\gamma}^{**} := \arg\min_{\hat{\gamma}^*} \widehat{MSE}(\hat{k}_{0\hat{\gamma}}^*|\hat{\gamma}^*)$.

4.3 Remarks on the adaptive classical or CVRB estimation

(i) If there are negative elements in the sample, the value of n, in the **Algorithm**, should be replaced by $n^+ := \sum_{i=1}^{n} I_{[X_i>0]}$, the number of positive elements in the sample.

(ii) In **Step 2.** of the **Algorithm** we are led in almost all situations to the *tuning parameter* $\tau = 0$ whenever $-1 \le \rho < 0$ and $\tau = 1$, otherwise. Due to the fact that bias reduction is really needed when $-1 \le \rho < 0$, we claim again for the relevance of the choice $\tau = 0$. Whenever we want to refer, in the estimation of γ through any of the reduced-bias estimators, the use of either $\tau = 0$ or $\tau = 1$ in the estimation of the second-order parameter ρ, we shall use the notation $\overline{C}_\tau$, for C equal to H or M or GH.

(iii) Regarding second-order parameters' estimation, attention should also be paid to the most recent classes of ρ-estimators proposed in Goegebeur *et al.* (2008, 2010) and in Ciuperca and Mercadier (2010), as well as to the estimators of β in Caeiro and Gomes (2006) and in Gomes *et al.* (2010).

(iv) As we shall see later on in Section 5.2, the method is only moderately dependent on the choice of the nuisance parameter n_1, in Step **5.** of the **Algorithm**, particularly for the MVRB estimators.

(v) The Monte-Carlo procedure in the **Steps 6.–10.** of the **Algorithm** can be replicated r_1 times if we want to associate standard errors to the OSF and the EVI-estimates. The value of B can also be adequately chosen.

(vi) We would like to stress again that the use of the random sample of size n_2, $(x_1^*, \ldots, x_{n_2}^*)$, and of the extended sample of size n_1, $(x_1^*, \ldots, x_{n_2}^*, x_{n_2+1}^*, \ldots, x_{n_1}^*)$, leads us to increase the precision of the result with a smaller B, the number of bootstrap samples generated.

(vi) We would like to notice again the "almost independence" on the choice of n_1, which enhances the practical value of the method. Consequently, although aware of the need of $n_1 = o(n)$, it seems that, once again, we get good results up till n.

5 Monte-Carlo simulations

5.1 Non-adaptive estimation

In this section, and comparatively with the behaviour of the classical estimators $H(k)$, $M(k)$ and $GH(k)$, in (2.3), (2.7) and (2.10), respectively, we are interested in the finite-sample behaviour, as functions of k, of the CVRB EVI-estimators, $\overline{H}(k)$ and $\overline{W}(k)$, in (2.6) and (2.11), respectively, with $\overline{W}$ denoting either $\overline{M}$, with M the estimator in (2.7), or $\overline{GH}$, with GH the estimator in (2.10). We have performed a multi-sample simulation with size 5000×10, i.e., 10 replicates with 5000 runs each. For details on multi-sample simulation refer to Gomes and Oliveira (2001). The patterns of mean values (E) and root mean squared errors (RMSE) are based on the first replicate and are considered as a function of $h = k/n$. We have considered in this article the following underlying parents:

I. the Fréchet(γ) model, $F(x) = \exp(-x^{-1/\gamma}),\ x > 0$, with $\gamma = 0.25$ ($\rho = -1$);

II. the Burr(γ, ρ) model, with d.f. $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho},\ x > 0$, for a few values of (γ, ρ), the pairs, $(0.25, -0.5)$, $(0.25, -1)$ and $(1, -1)$;

III. the Student's t_ν model with ν degrees of freedom, with a probability density function $f_{t_\nu}(t) = \Gamma((\nu+1)/2)\,[1 + t^2/\nu]^{-(\nu+1)/2} / (\sqrt{\pi\nu}\,\Gamma(\nu/2)),\ t \in \mathbb{R}\quad (\nu > 0)$, for which $\gamma = 1/\nu$ and $\rho = -2/\nu$. The illustration will be done for $\nu = 4$ degrees of freedom, i.e. $(\gamma, \rho) = (0.25, -0.5)$;

IV. the *extreme value* EV_γ model, with d.f. $EV_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma}),\ x > -1/\gamma$, for which $\rho = -\gamma$. We shall consider $\gamma = 0.25$ ($\rho = -0.25$) and $\gamma = 1$ ($\rho = -1$).

V. the *Generalized Pareto* GP_γ model, with d.f. $GP_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x > 0$, ($\rho = -\gamma,$ as in IV.), also for $\gamma = 0.25$ ($\rho = -0.25$) and $\gamma = 1$ ($\rho = -1$).

5.1.1 Mean values and root mean squared errors patterns

In Figure 2, as an illustration of the results obtained, we show the simulated patterns of mean values for all the estimators under study, as a function of the sample fraction $h = k/n$, for the underlying Fréchet parent, and sample sizes $n = 500$ and $n = 5000$. Figure 3 is equivalent to Figure 2, but with the root mean squared errors (RMSE) patterns of the estimators. Similar results have been obtained for all simulated models.

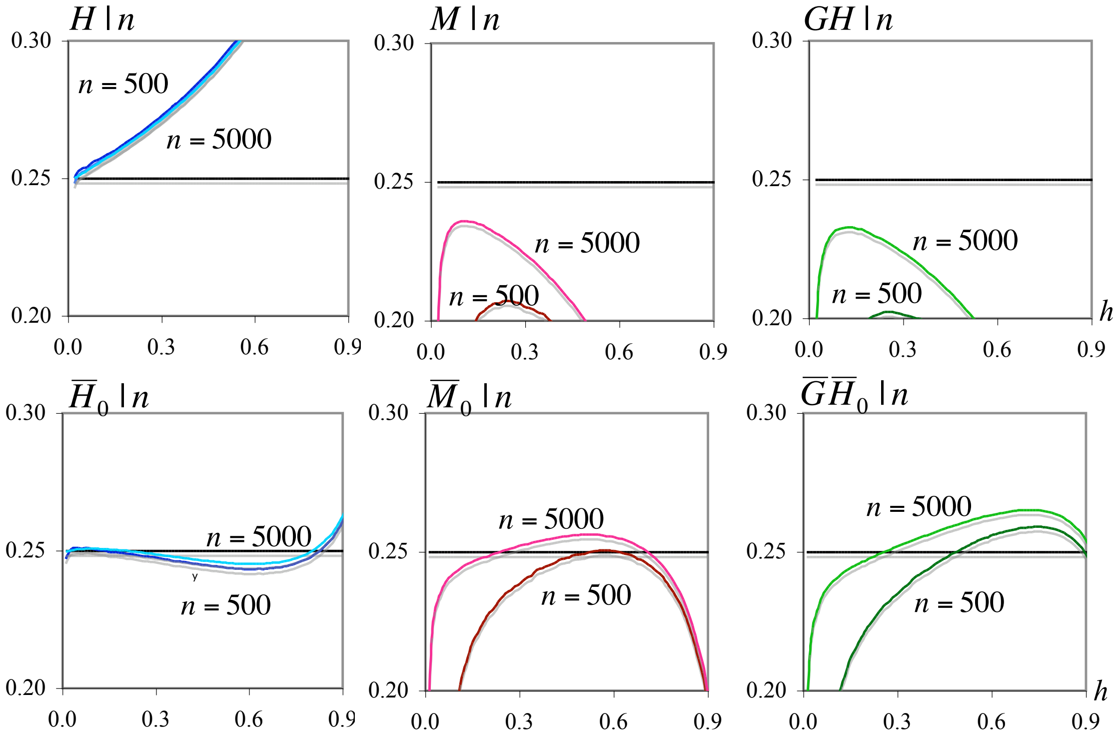

Figure 2: Patterns of mean values of the classical estimators H, M and GH, in (2.3), (2.7) and (2.10) (*top*) and the associated CVRB estimators (*bottom*), as functions of k/n, for an underlying Fréchet parent with $\gamma = 0.25$ ($\rho = -1$).

From Figure 2 it is clear the reduction in bias achieved by any of the reduced-bias estimators. Such a bias reduction leads to much lower mean squared errors for the CVRB estimators, as can be seen from Figure 3.

5.1.2 Relative efficiencies and mean values at optimal levels

Given a sample $\underline{X}_n = (X_1, \ldots, X_n)$, let us denote $S(k) = S(k; \underline{X}_n)$ any statistic or r.v. dependent on k, the number of top o.s.'s to be used in an inferential procedure related with a parameter of extreme events. Just as mentioned before for the Hill estimator $H(k)$, in (2.3), the OSF for $S(k)$ is denoted k_0^S/n, with $k_0^S := \arg\min_k MSE\,(S(k))$. We have obtained, for $n = 100,\ 200,\ 500,\ 1000,\ 2000$ and 5000, and with $\bullet$ denoting H or M or GH or $\overline{H}_\tau$ or $\overline{M}_\tau$ or $\overline{GH}_\tau$, $\tau = 0$ and 1, the simulated OSF ($\text{OSF}_0^\bullet = k_0^\bullet/n$), bias ($B_0^\bullet = E_0^\bullet - \gamma$) and relative efficiencies ($REFF_0^\bullet$) of the EVI-estimators under study, at their optimal levels. The search of the minimum MSE has been performed over the region of k-values between 1 and $[0.95 \times n]$. For any EVI-estimator different from H,

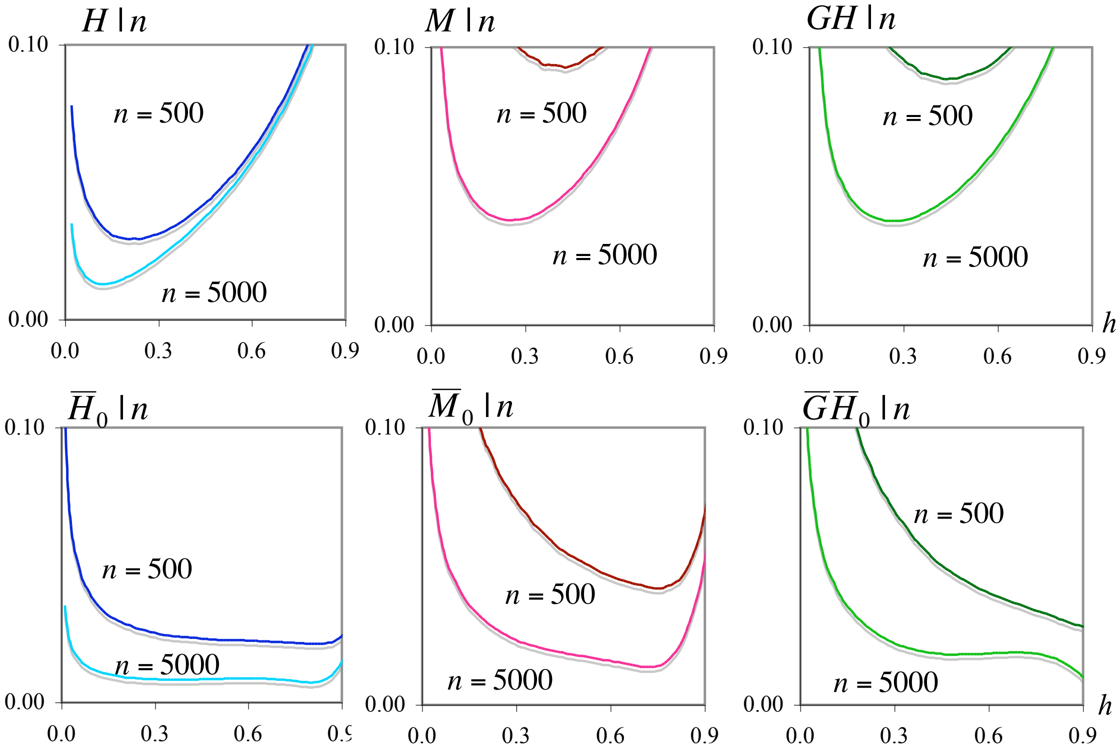

Figure 3: Patterns of RMSEs of the classical estimators H, M and GH, in (2.3), (2.7) and (2.10) (*top*) and the associated CVRB estimators (*bottom*), as functions of k/n, for an underlying Fréchet parent with $\gamma = 0.25$ ($\rho = -1$).

generally denoted S, and with the notation $S_0 = S(k_0^S)$, the $REFF_0^S$ indicator is

$$REFF_0^S := \sqrt{\frac{MSE\,(H_0)}{MSE\,(S_0)}} =: \frac{RMSE_0^H}{RMSE_0^S}.$$

As an illustration, we present in Tables 1, 2 and 3, the obtained simulated results for models with $|\rho| < 1$ (the generalized Pareto model with $\gamma = 0.25$, and consequently $\rho = -\gamma = -0.25$), $|\rho| = 1$ (the Burr parent with $\gamma = 0.25$ and $\rho = -1$) and $|\rho| > 1$ (the Student parent with $\nu = 1$, for which $\gamma = 1/\nu = 1$ and $\rho = -2/\nu = -2$), respectively. Among the estimators considered, and for all n, the one providing the smallest squared bias and the smallest MSE, i.e., the highest $REFF$ is underlined and in **bold**. The second highest $REFF$ indicators are written in *italic* and underlined. The MSE of H_0, Hill estimators at their simulated optimal level, is also provided so that it is possible to recover the MSE of any other EVI-estimator. Moreover, we present the B_0 and $REFF_0$ indicators of the r.v.'s $\overline{C}_{k,n;\beta,\rho}$ at their optimal levels, with $C = H$, M and GH, denoted $\overline{H}_{\beta,\rho}$, $\overline{M}_{\beta,\rho}$ and $\overline{GH}_{\beta,\rho}$, respectively, just to make it clear that in most situations some improvement is still possible with a better estimation of the second-order parameters. Note however, that some times the estimation of the parameters in the model (if nicely done) accommodates better the statistical fluctuations in the sample and produces better results than the use of the “true known values”. Extensive tables for all simulated models as well as 95% confidence intervals (CIs) associated with all the estimates are available from the authors.

Table 1: Simulated bias at optimal levels ($B_0^\bullet$), relative efficiency indicators ($REFF_0^\bullet$) and MSE of H_0, for a *generalized Pareto* GP_γ parent with $\gamma = 0.25$ ($\rho = -0.25$).

n	100	200	500	1000	2000	5000
			$B_0^\bullet$			
H	0.2352	0.1219	0.0869	0.0991	0.1292	0.0448
$\overline{H}_0$	0.1979	0.1339	0.1132	0.1007	0.1046	0.0414
$\overline{H}_1$	0.2347	0.1217	0.0908	0.0991	0.1292	0.0447
M	0.1290	0.0700	0.0801	0.0805	0.0644	0.0239
$\overline{M}_0$	0.1470	0.0761	0.0840	**0.0305**	**0.0124**	0.0082
$\overline{M}_1$	0.1512	0.0548	0.0801	0.0753	0.0658	0.0160
GH	0.1210	0.0301	0.0646	0.0543	0.0606	0.0097
$\overline{GH}_0$	**0.1169**	0.0337	**0.0650**	0.0567	0.0591	**0.0079**
$\overline{GH}_1$	0.1280	**0.0243**	0.0666	0.0576	0.0557	0.0079
$\overline{H}_{\beta,\rho}$	0.0325	0.0256	0.0212	0.0175	0.0102	0.0099
$\overline{M}_{\beta,\rho}$	0.0200	0.0098	0.0224	−0.0016	−0.0076	0.0010
$\overline{GH}_{\beta,\rho}$	0.0360	0.0177	0.0284	0.0236	0.0226	0.0086
			$REFF_0^\bullet$			
$\overline{H}_0$	1.1489	1.1170	1.0880	1.0688	1.0577	1.0421
$\overline{H}_1$	1.0015	1.0007	1.0003	1.0001	1.0001	1.0000
M	1.0585	1.1600	1.2506	1.3032	1.3479	1.3999
$\overline{M}_0$	1.3376	1.3194	1.3055	**1.4244**	**1.6963**	**2.1667**
$\overline{M}_1$	1.0832	1.1549	1.2346	1.2880	1.3334	1.4179
GH	1.3591	*1.3605*	*1.3798*	1.3946	*1.4166*	*1.4424*
$\overline{GH}_0$	**1.5844**	**1.4981**	**1.4313**	*1.3991*	1.3850	1.3767
$\overline{GH}_1$	*1.3654*	1.3563	1.3693	1.3851	1.4077	1.4364
$\overline{H}_{\beta,\rho}$	4.8574	5.3039	5.8883	6.4205	6.9599	7.7053
$\overline{M}_{\beta,\rho}$	4.1515	4.8043	5.2907	5.7535	6.7673	8.6844
$\overline{GH}_{\beta,\rho}$	4.6909	4.5833	4.4753	4.4443	4.4596	4.4941
MSE_{H_0}	0.0561	0.0382	0.0238	0.0172	0.0126	0.0084

In summary we may draw the following final conclusions:

Table 2: Simulated bias at optimal levels ($B_0^\bullet$), relative efficiency indicators ($REFF_0^\bullet$) and MSE of H_0, for a Burr parent with $(\gamma, \rho) = (0.25, -1)$.

n	100	200	500	1000	2000	5000
			$B_0^\bullet$			
H	0.0357	0.0112	0.0293	0.0280	0.0206	0.0115
$\overline{H}_0$	**0.0041**	**−0.0066**	**0.0035**	**−0.0006**	−0.0028	−0.0018
$\overline{H}_1$	0.0289	0.0110	0.0268	0.0278	0.0201	0.0107
M	−0.1557	−0.1408	−0.0481	−0.0305	−0.0132	−0.0305
$\overline{M}_0$	−0.0353	−0.0731	−0.0071	0.0043	−0.0014	−0.0026
$\overline{M}_1$	−0.1173	−0.1389	−0.0399	−0.0279	−0.0137	−0.0385
GH	−0.1581	−0.1364	−0.0514	−0.0395	−0.0200	−0.0403
$\overline{GH}_0$	−0.0065	−0.0210	0.0038	0.0028	**−0.0007**	**−0.0007**
$\overline{GH}_1$	−0.0223	−0.0354	0.0066	0.0045	−0.0003	−0.0008
$\overline{H}_{\beta,\rho}$	0.0206	0.0055	0.0068	0.0055	0.0029	0.0047
$\overline{M}_{\beta,\rho}$	−0.0438	−0.0760	−0.0078	−0.0023	−0.0037	−0.0042
$\overline{GH}_{\beta,\rho}$	−0.0026	−0.0199	0.0060	0.0028	−0.0006	−0.0007
			$REFF_0^\bullet$			
$\overline{H}_0$	**1.9835**	**2.1260**	**2.4273**	**2.6968**	**2.9520**	**3.3957**
$\overline{H}_1$	1.0432	1.0381	1.0243	1.0196	1.0159	1.0116
M	0.2648	0.2909	0.3148	0.3289	0.3431	0.3530
$\overline{M}_0$	0.5120	0.5659	0.6314	0.6826	0.7363	0.8146
$\overline{M}_1$	0.3168	0.3396	0.3553	0.3631	0.3710	0.3750
GH	0.3287	0.3330	0.3406	0.3463	0.3540	0.3593
$\overline{GH}_0$	*1.2003*	*1.3118*	*1.4779*	*1.6396*	*1.8180*	*2.0970*
$\overline{GH}_1$	0.7779	0.8633	0.9711	1.0776	1.2145	1.4126
$\overline{H}_{\beta,\rho}$	1.8925	1.9854	2.1210	2.2252	2.3388	2.4827
$\overline{M}_{\beta,\rho}$	0.4466	0.4976	0.5544	0.5957	0.6356	0.6875
$\overline{GH}_{\beta,\rho}$	1.0802	1.1947	1.3522	1.4988	1.6665	1.9248
MSE_{H_0}	0.0044	0.0026	0.0013	0.0008	0.0005	0.0003

1. For underlying parents with $|\rho| < 1$, the highest efficiency is generally achieved through $\overline{GH}_0$ for $n < 1000$ and through $\overline{M}_0$ otherwise. The highest bias reduction pattern is not so clear-cut, as can be seen in Table 1, but the results are not a long way from the ones related with efficiency.
2. Again at optimal levels, and for underlying parents with $\rho = -1$, the highest bias reduction as well as the highest efficiency is generally achieved through the use of $\overline{H}_0$, followed by $\overline{GH}_0$. Only for very large values of n, say $n \geq 2000$, did $\overline{GH}_0$ beat

Table 3: Simulated bias at optimal levels ($B_0^\bullet$), relative efficiency indicators ($REFF_0^\bullet$) and MSE of H_0, for a Student parent with $\nu = 1$ degrees of freedom ($\gamma = 1,\ \rho = -2$).

n	100	200	500	1000	2000	5000
			$B_0^\bullet$			
H	0.0794	0.1117	0.0936	0.0471	0.0573	0.0254
$\overline{H}_0$	−0.2632	−0.0608	0.0845	−0.0195	0.0250	**−0.0048**
$\overline{H}_1$	−0.1252	**0.0222**	0.0744	0.0116	0.0390	0.0275
M	**−0.0142**	0.0704	0.0915	0.0240	0.0539	0.0051
$\overline{M}_0$	−0.2392	−0.0999	**0.0018**	−0.0820	−0.0390	−0.0849
$\overline{M}_1$	−0.1865	−0.0235	0.0761	**0.0068**	0.0380	0.0073
GH	−0.0210	0.0503	0.0719	0.0190	0.0421	0.0334
$\overline{GH}_0$	−0.6775	−0.0997	0.0710	−0.0188	**0.0235**	0.0179
$\overline{GH}_1$	−0.6224	−0.0371	0.0667	0.0084	0.0328	0.0301
$\overline{H}_{\beta,\rho}$	0.0397	0.0488	0.0479	0.0144	0.0207	0.0152
$\overline{M}_{\beta,\rho}$	−0.0950	0.0021	0.0198	−0.0128	0.0039	−0.0073
$\overline{GH}_{\beta,\rho}$	−0.0805	0.0008	0.0413	0.0126	0.0274	0.0310
			$REFF_0^\bullet$			
$\overline{H}_0$	0.2359	0.5818	0.9376	0.9826	1.1084	**1.3939**
$\overline{H}_1$	0.2073	0.7966	**1.1590**	**1.1583**	**1.1593**	1.1641
M	0.7932	0.8668	0.9151	0.9234	0.9232	0.9273
$\overline{M}_0$	0.2145	0.5045	0.7283	0.7116	0.6741	0.6095
$\overline{M}_1$	0.1895	0.6698	0.9660	0.9962	1.0299	1.0676
GH	**1.0459**	**1.0683**	1.0600	1.0469	1.0266	1.0102
$\overline{GH}_0$	0.2969	0.5618	0.9363	0.9948	1.1002	<u>*1.3348*</u>
$\overline{GH}_1$	0.2705	0.7593	<u>*1.1151*</u>	<u>*1.1165*</u>	<u>*1.1260*</u>	1.1406
$\overline{H}_{\beta,\rho}$	1.3399	1.4229	1.5261	1.6359	1.7199	1.8236
$\overline{M}_{\beta,\rho}$	0.9507	1.1146	1.2612	1.3644	1.4499	1.5628
$\overline{GH}_{\beta,\rho}$	1.1897	1.3858	1.4854	1.5209	1.5228	1.5304
MSE_{H_0}	0.0693	0.0370	0.0166	0.0095	0.0053	0.0025

$\overline{H}_0$ regarding the bias-indicator (refer to Table 2).

3. Almost generally, and for models such that $|\rho| \le 1$, $\overline{M}_0$ ($\overline{GH}_0$) works better than M (GH). But, for models with $|\rho| = 1$, $\overline{M}_0$ never beats $\overline{H}_0$ regarding efficiency. Regarding bias reduction, $\overline{GH}_0$ beats $\overline{H}_0$ for large n, almost generally.

4. For the range of ρ-values close to zero ($-1 < \rho < 0$), the use of $\tau = 1$ in $\overline{H}_\tau$ provides results only slightly better than the ones associated with the classical estimator.

5. For underlying parents with $\rho < -1$, bias-reduced estimators work only for large n,

say $n \geq 500$. Then the highest efficiency is generally achieved through the use of $\overline{H}_1$, followed by $\overline{GH}_1$. Again, and even worse than for the case $|\rho| < 1$, the highest bias reduction pattern is not clear-cut, as can be seen in Table 3.

5.2 Adaptive estimation

In order to understand the performance of the adaptive bootstrap estimates as well as of bootstrap CIs, we have run Steps **6.–10.** of the **Algorithm** in Section 4.2, $r_1 = 100$ times, for the models considered in Section 5.1. The overall estimates of γ, denoted H^*, $\overline{H}^*_\tau$, M^*, $\overline{M}^*_\tau$, GH^* and $\overline{GH}^*_\tau$, $\tau = 0$ or 1, are the averages of the corresponding r_1 partial estimates.

5.2.1 Bootstrap EVI-estimates and CIs

In Figure 4, and as an illustration of the overall simulated behaviour of the bootstrap estimates, $\hat{\gamma}^*$, with $\hat{\gamma} = H, M, GH, \overline{H}_0, \overline{M}_0$ and $\overline{GH}_0$, we present, for a Fréchet model with $\gamma = 0.25$, the bootstrap adaptive EVI-estimates, as a function of the sample size n. The method works asymptotically, as can be seen from Figure 4. But is also works for small n, particularly if we take into account the estimate $\overline{H}^*_0$.

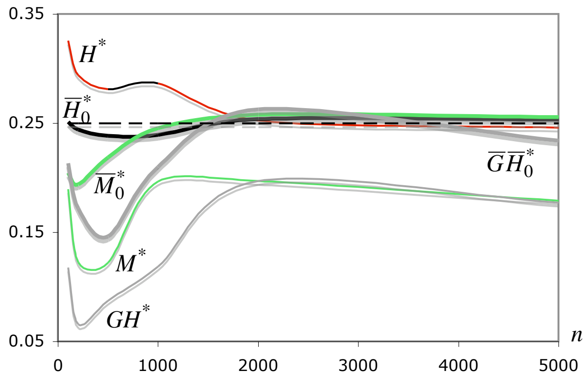

Figure 4: Bootstrap adaptive EVI-estimates as a function of the sample size n, for data simulated from a Fréchet parent with $\gamma = 0.25$ ($\rho = -1$).

Those estimates are also provided in Table 4, with associated standard errors provided between parenthesis, close to the estimates, at the first row of each entry. In the second row of each entry, we present the 99% bootstrap CIs. These bootstrap CIs are based on the quantiles of probability 0.005 and 0.995 of the $r_1 = 100$ partial bootstrap estimates, and are written in *italic* whenever they do not cover the true value of γ, with the upper limit smaller than γ (underestimation). They are written in *italic* and underlined, whenever they do not cover the true value of γ, with the lower limit larger than γ (overestimation).

H^*	$\overline{H}_0^*$	M^*	$\overline{M}_0^*$	GH^*	$\overline{GH}_0^*$
$n = 100$					
0.326 (0.0079)	0.251 (0.0044)	0.190 (0.0248)	0.205 (0.0139)	0.118 (0.1081)	0.213 (0.0815)
(*0.3003*, *0.3416*)	(0.2289, 0.2587)	(*0.1122, 0.2452*)	(0.1867, 0.2620)	(-0.0689, 0.2798)	(-0.0635, 0.2993)
$n = 200$					
0.294 (0.0136)	0.244 (0.0075)	0.123 (0.0201)	0.194 (0.0351)	0.066 (0.1083)	0.172 (0.1281)
(*0.2758*, *0.3181*)	(0.2106, 0.2521)	(*0.0804, 0.1668*)	(0.1548, 0.2936)	(*-0.3605, 0.1823*)	(-0.3409, 0.2759)
$n = 500$					
0.281 (0.0196)	0.239 (0.0031)	0.122 (0.0127)	0.219 (0.0068)	0.087 (0.0783)	0.147 (0.1205)
(0.2404, 0.3080)	(*0.2332, 0.2457*)	(*0.0854, 0.1561*)	(*0.2045, 0.2297*)	(*-0.1410, 0.1653*)	(-0.1853, 0.2566)
$n = 1000$					
0.286 (0.0021)	0.240 (0.0008)	0.197 (0.0061)	0.246 (0.0063)	0.119 (0.1442)	0.210 (0.1210)
(*0.2819*, *0.2905*)	(*0.2382, 0.2414*)	(*0.1841, 0.2181*)	(0.2370, 0.2626)	(*-0.4448, 0.2032*)	(-0.4258, 0.2627)
$n = 2000$					
0.253 (0.0092)	0.254 (0.0004)	0.197 (0.0073)	0.258 (0.0023)	0.197 (0.1327)	0.263 (0.0931)
(*0.2529*, *0.2548*)	(0.2233, 0.2626)	(*0.1885, 0.2172*)	(0.2436, 0.2645)	(-0.0766, 0.6833)	(0.1698, 0.7219)
$n = 5000$					
0.254 (0.0159)	0.246 (0.0023)	0.179 (0.0041)	0.256 (0.0082)	0.177 (0.1032)	0.234 (0.0422)
(*0.2501*, *0.2578*)	(0.2430, 0.2541)	(*0.1751, 0.1938*)	(0.2301, 0.2665)	(-0.0951, 0.4413)	(0.1593, 0.2788)

Table 4: Bootstrap adaptive estimates of γ through the classical C estimators, and the associated CVRB estimators $\overline{C}_\tau$, with $\tau = 0$, for $C = H, M$ and GH and for an underlying Fréchet parent, with $\gamma = 0.25$ ($\rho = -1$).

A few comments on the bootstrap EVI-estimates and CIs:

- Almost generally, the bootstrap M^* and GH^*-estimates provide a systematic under-estimation of γ, which is compensated by the consideration of the associated CVRB-estimates.
- On another side, and except for EV_γ underlying parents, the bootstrap H^*-estimates provide a systematic over-estimation of γ, which is again compensated by the consideration of the associated CVRB-estimates, $\overline{H}^*$. Moreover, for several

values of n, the bootstrap 99% CIs, associated with the H^*-estimates, have lower limits above γ.

- The 99% bootstrap CIs associated with $\overline{GH}^*$ do always cover the true value of γ, but at expenses of very large sizes. Moreover, and despite of the volatility of the simulated GH^* estimates, some of the associated 99% bootstrap CIs have upper limits below γ.
- In general, and despite of a slight under-estimation for a few values of n and some of the simulated parents, the results are clearly in favour of the bootstrap $\overline{H}^*$-estimation procedure. However, the performance of $\overline{M}^*$ is also interesting for most of the simulated underlying parents, particularly for large sample size n.

5.2.2 Sensitivity of the algorithm to the subsample size n_1

In order to detect the sensitivity of the algorithm to changes of n_1, we have run it for values of $n_1 = [n^a]$, $a = 0.950(0.005)0.995$. In Figures 5 and 6, and again as an illustration, we present for the same Fréchet underlying parent, the bootstrap γ-estimates as a function of a, for $n = 200$ and $n = 2000$, respectively.

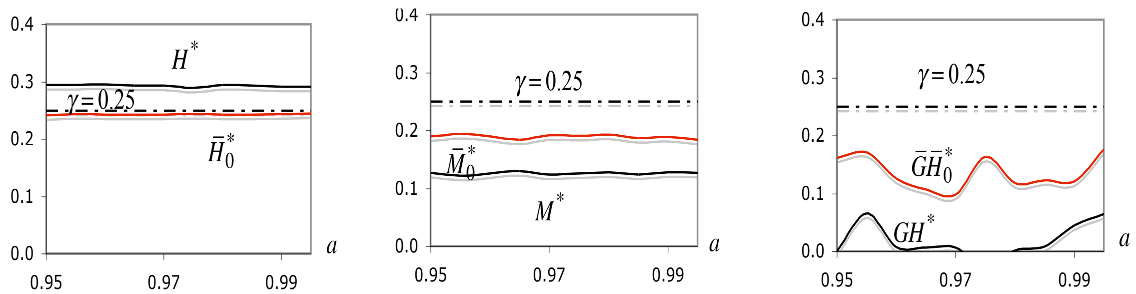

Figure 5: Bootstrap adaptive EVI-estimates, for samples of size $n = 200$ from a Fréchet parent with $\gamma = 0.25$ ($\rho = -1$).

A few comments on the results:

- As expected, and due to the fact that the method works asymptotically, there is a general improvement in the estimation as the sample size, n, increases.
- The sensitivity of the **Algorithm** in Section 4.2 to the nuisance parameter n_1 is weak for H^*, $\overline{H}^*$, M^* and $\overline{M}^*$, particularly if n is large. Such a dependency is however not so weak for both GH^* and $\overline{GH}^*$.

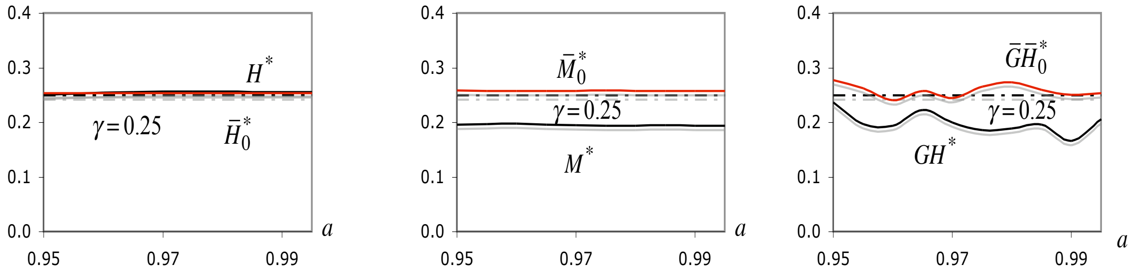

Figure 6: Bootstrap adaptive EVI-estimates, for samples of size $n = 2000$ from a Fréchet parent with $\gamma = 0.25$ ($\rho = -1$).

5.2.3 Bootstrap CIs' sizes and coverage probabilities

Due to the reasonably high number of bootstrap 99% CIs not covering the true value of γ (see Table 4), we felt the need and the curiosity of analyzing the performance of these bootstrap CIs, on the basis of a terribly time-consuming computer program. More specifically, in order to obtain information on the coverage probabilities and on the sizes of the bootstrap CIs, we have also run $r_2 = 100$ times, the whole algorithm in Section 4.2, after the $r_1 = 100$ replicates of **Steps 6.– 10.**, suggested in Section 5.2. This is a terrible time-consuming algorithm, and we have thus run it only for small values of n. Again as an illustration, we provide in Table 5, for a Student t_2 underlying parent and for $n = 100$, 200 and 1000, the overall EVI-estimates, the sizes of the 99% bootstrap CIs and the coverage probabilities of those CIs, in the first, second and third row, respectively. The overall EVI-estimate closer to the target, the minimum size and the maximum coverage probability are written in **bold**.

A few general comments:

- We need to be careful with the use of bootstrap CIs, due to the fact that we can indeed get very small coverage probabilities, comparatively with the target value, 0.99.
- Large coverage probabilities are attained only by the bootstrap GH^* and $\overline{GH}^*$ estimates, but at expenses of a very large size.
- As expected, and in general, there is a decreasing trend in the sizes, as n increases, and a slight increasing trend in the coverage probabilities. However, in some cases, the coverage probabilities decrease with n.
- As a compromise between size and coverage probability, we are inclined to the

H^*	$\overline{H}_0^*$	$\overline{H}_1^*$	M^*	$\overline{M}_0^*$	$\overline{M}_1^*$	GH	$\overline{GH}_0^*$	$\overline{GH}_1^*$
$n = 100$								
0.5751	**0.4882**	0.5885	0.4010	0.4238	0.4042	0.4753	0.4815	0.5427
0.1880	**0.1427**	0.2443	0.2431	0.2354	0.3850	0.6440	0.4842	0.5974
29%	31%	35%	27%	31%	36%	54%	44%	**55%**
$n = 200$								
0.5648	**0.4898**	0.5881	0.4362	0.4406	0.4704	0.4800	0.4818	0.5366
0.1573	**0.1154**	0.1899	0.17320	0.1722	0.2323	0.7195	0.5486	0.6544
29%	34%	31%	33%	36%	46%	64%	52%	**67%**
$n = 1000$								
0.5479	**0.4994**	0.5623	0.4877	0.4702	0.4965	0.5362	0.5220	0.5772
0.1139	**0.0862**	0.1327	0.1315	0.1396	0.2705	1.2217	0.7944	1.0453
28%	40%	27%	30%	31%	55%	82%	**84%**	81%

Table 5: Overall EVI-estimates (*first row*), sizes (*second row*) and percentage coverage probabilities (*third row*) of the 99% bootstrap CI's for γ, obtained on the basis of the classical C estimators, and the associated CVRB estimators $\overline{C}_\tau$, with $\tau = 0$ and $\tau = 1$, for $C = H, M$ and GH and for an underlying Student t_ν parent, with $\nu = 2$ ($\gamma = 1/\nu = 0.5$, $\rho = -2/\nu = -1$).

choice of $\overline{H}_0^*$ whenever $|\rho| \leq 1$ and $\overline{M}_1^*$ for models with $|\rho| > 1$. Indeed, these bootstrap EVI-estimates are quite close to the target γ.

6 An application to burned areas data

Most of the wildfires are extinguished within a short period of time, with almost negligible effects. However, some wildfires go out of control, burning hectares of land and causing significant and negative environmental and economical impacts. The data we analyse here consists of the number of hectares, exceeding 100 ha, burnt during wildfires recorded in Portugal during 14 years (1990-2003). The data (a sample of size $n = 2627$) do not seem to have a significant temporal structure, and we have used the data as a whole, although we think also sensible, to try avoiding spatial heterogeneity, the consideration of at least 3 different regions: the north, the centre and the south of Portugal (a study out of the scope of this paper).

The box-plot and the histogram of the available data, in Figure 7, provide evidence on the heaviness of the right tail.

In Figure 8, we present the sample path of the $\hat{\rho}_\tau(k)$ estimates in (4.8), as function of k, for $\tau = 0$, together with the sample paths of the associated β-estimators in (4.9),

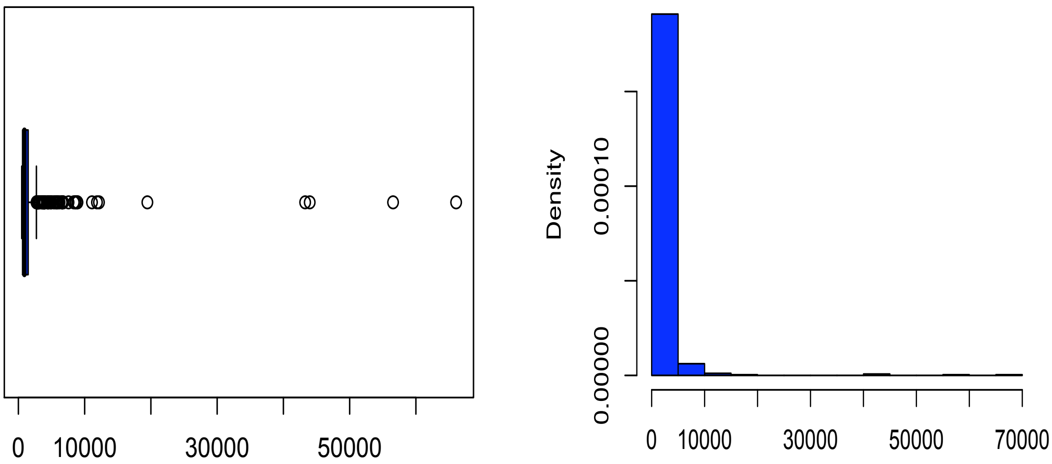

Figure 7: Box-and-whiskers plot (*left*) and Histogram (*right*) associated with burned areas in Portugal over 100 ha (1990-2003).

also for $\tau = 0$, the value obtained in the **Algorithm** of Section 4.2 for the tuning parameter τ. We have been led to the ρ-estimate, $\hat{\rho} \equiv \hat{\rho}_0 = -0.39$, obtained at the level $k_1 = [n^{0.999}] = 2606$, and to the associated β-estimate, $\hat{\beta} \equiv \hat{\beta}_0 = 0.47$, both recorded in Figure 8.

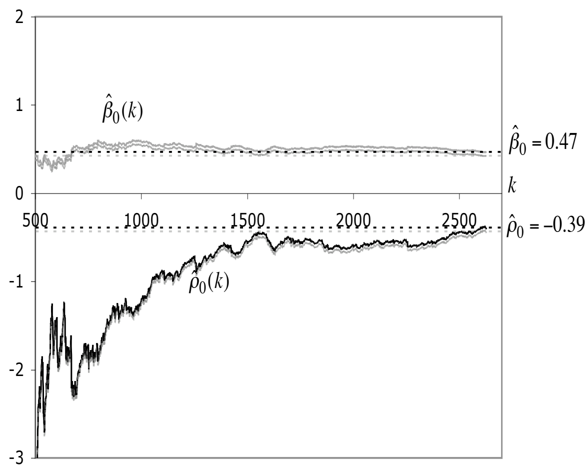

Figure 8: Estimates of the shape second-order parameter ρ and of the scale second-order parameter β for the burned areas data.

Next, in Figure 9, we present the adaptive and non-adaptive EVI-estimates provided by H and associated MVRB estimates $\overline{H}$ (*left*), as well as M, GH and the associated CVRB estimates, $\overline{M}$ and $\overline{GH}$ (*right*).

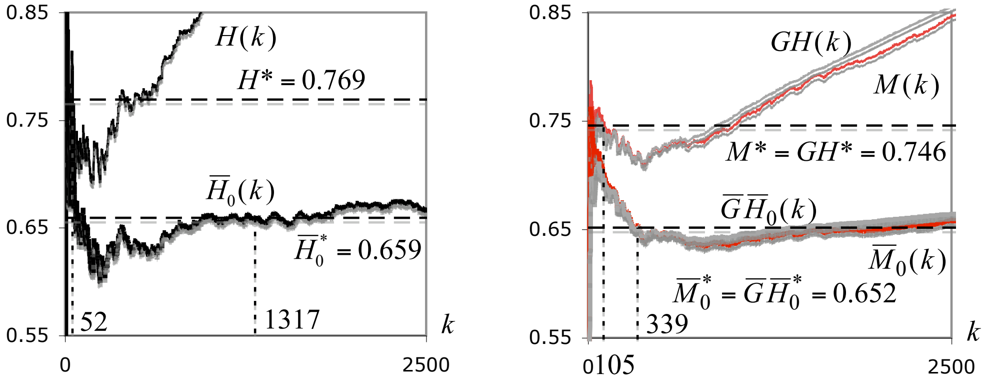

Figure 9: Estimates of the EVI, γ, through the EVI estimators under consideration, H, M and GH and associated CVRB estimators, $\overline{H}$, $\overline{M}$ and $\overline{GH}$, for the burned areas under analysis.

For the Hill estimator, we have simple techniques to estimate the OSF. Indeed, we get $\hat{k}_0^H(n) = \Big[\big((1-\hat\rho)^2 n^{-2\hat\rho}/\big(-2\ \hat\rho\ \hat\beta^2\big)\big)^{1/(1-2\hat\rho)}\Big] = 157$, and an associated γ-estimate equal to 0.73. The algorithm in this paper helps us to adaptively estimate the OSF associated not only with the classical EVI-estimates but also with the MVRB or even CVRB estimates. For a sub-sample size $n_1 = [n^{0.955}] = 1843$, and B=250 bootstrap generations, we have got $\hat{k}_0^{\overline{H}}(n;n_1) = 1317$ and the MVRB-EVI-estimate $\overline{H}^* = 0.659$, the value pictured in Figure 9, *left*, jointly with the bootstrap adaptive Hill estimate, H^*, equal to 0.769, due to the fact that we were led to $\hat{k}_0^H(n;n_1) = 52$. The estimated RMSEs, in **Step 11.** of the **Algorithm**, were $RMSE_H^* = 0.164$ and $RMSE_{\overline{H}_0}^* = 0.049$. Again with W denoting either M or GH, we were led to $\hat{k}_0^W(n;n_1) = 105$, $\hat{k}_0^{\overline{W}}(n;n_1) = 339$, $W^* = 0.746$ and $\overline{W}^* = 0.652$, the values pictured at Figure 9 (*right*). The estimated RMSEs were $RMSE_W^* = 0.173$ and $RMSE_{\overline{W}_0}^* = 0.276$. Note the fact that the MVRB EVI-estimators look practically "unbiased" for the data under analysis and the associated adaptive estimator $\overline{H}_0^*$, was the chosen one, due to the smallest estimated RMSE, the value $RMSE_{\overline{H}_0}^* = 0.049$.

Regarding the dependency of the bootstrap methodology on the subsample size n_1, we refer to Figure 10, where apart from the adaptive bootstrap estimates we also picture the medians of the values obtained for n_1 from 1750 until 2600, with step 1.

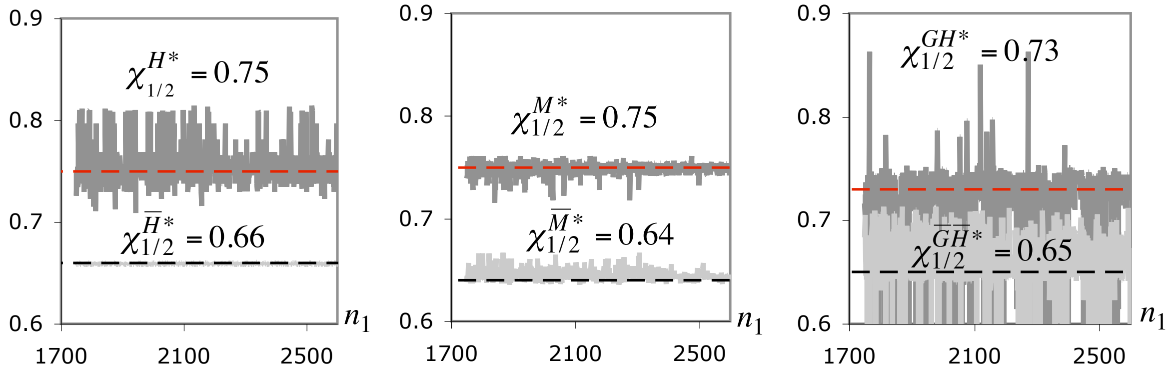

Figure 10: Bootstrap adaptive estimates of the EVI, γ, as a function of the subsample size n_1, done through $(H^*, \overline{H}_0^*)$ (*left*), $(M^*, \overline{M}_0^*)$ (*center*) and $(GH^*, \overline{GH}_0^*)$ (*right*), for the burned areas data under analysis.

It is clear the small sensitivity, to changes in n_1, of $\overline{H}_0^*(n; n_1)$, contrarily to the high sensitivity of $GH^*(n; n_1)$. We consider this to be another point in favour of $\overline{H}_0^*$. The consideration of all the above mentioned values of n_1, i.e., $n_1 = 1750, 1751, \ldots 2600$, led us to a minimum RMSE given by $RMSE^*_{\overline{H}_0} = 0.018$, attained at $n_1 = 1320$, with an associated EVI-estimate given by $\overline{H}_0^* = 0.66$, just as before. We finally exhibit, in Figure 11, not only a zoom of the adaptive bootstrap estimates $\overline{H}_0^*$ (*right*) but also of $\hat{k}_{0\overline{H}_0^*}$ (*left*), again as a function of n_1, as well as the medians of the values obtained for n_1 from 1750 until 2600, with step 1.

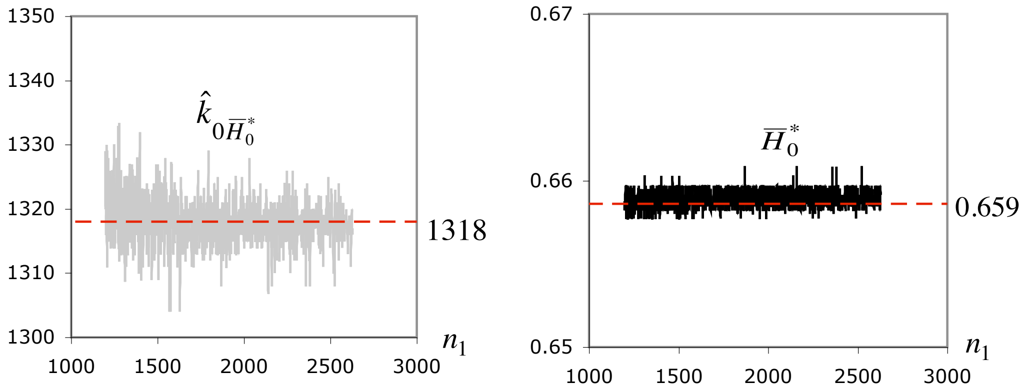

Figure 11: Bootstrap adaptive estimates of the optimal level (*left*) and of the EVI (*right*), done through the adaptive MVRB estimator, $\overline{H}_0^*$, for the burned areas data under analysis.

References

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

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

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

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

[6] 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.

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

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

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

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

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

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

[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, 51–69.

[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, 2632-2652.

[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 Neves, C. (2008). Asymptotic comparison of the mixed moment and classical extreme value index estimators. *Statistics and Probability Letters* **78**:6, 643-653.

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

[19] 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.

[20] Gomes, M.I. and Pestana, D. (2007b). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. American Statistical Association* **102**, no. 477, 280–292.

[21] Gomes, M.I., Martins, M.J. and Neves, M.M. (2007). Improving second order reduced-bias tail index estimation. *Revstat* **5**:2, 177-207.

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

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

[24] Gomes, M.I., Pestana, D. and Caeiro, F. (2008c). A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Statist. and Probab. Letters* **79**, 295-303.

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

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

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

[29] Peng, L. and Qi, Y. (2004). Estimating the first and second order parameters of a heavy tailed distribution. *Australian and New Zealand J. of Statist.* **46**, 305-312.

[30] Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd edition, Birkhäuser Verlag, Basel-Boston-Berlin, 2007.