RESAMPLING METHODOLOGIES AND RELIABLE TAIL ESTIMATION

M. Ivette Gomes [1]
Centro de Estatística e Aplicações (CEA), Faculdade de Ciências, Universidade de Lisboa
e-mail: *ivette.gomes@fc.ul.pt*
and
Fernanda Figueiredo
CEA, Universidade de Lisboa, and Faculdade de Economia, Universidade do Porto
and
M. João Martins
Centro de Estudos Florestais and Instituto Superior de Agronomia, Universidade de Lisboa
and
Manuela Neves
CEA, and Instituto Superior de Agronomia, Universidade de Lisboa

Key words: Bootstrap and jackknife methodologies, Semi-parametric estimation, Statistics of extremes.

Summary: Resampling methodologies, like the *generalised jackknife* and the *bootstrap* are important tools for a reliable semi-parametric estimation of parameters of extreme or even rare events. Among these parameters we mention the *extreme value index*, denoted ξ, the primary parameter in *statistics of extremes*, and the *extremal index*, denoted θ, a measure of clustering of extreme events. Most of the semi-parametric estimators of these parameters show the same type of behaviour: nice asymptotic properties, but a high variance for small k, the number of upper order statistics used in the estimation, a high bias for large k, and the need for an adequate choice of k. After a brief reference to some estimators of the aforementioned parameters and their asymptotic properties we present algorithms for an adaptive reliable estimation of ξ and θ.

1. Introduction

Resampling methodologies have recently been shown to be extremely fruitful in the field of *statistics of extremes*. Among others, we mention the importance of the *generalised jackknife* (GJ) (Gray and Schucany, 1972) and the *bootstrap* (Efron, 1979) for reliable semi-parametric estimation of any parameter of extreme or even rare events, like a *high quantile*, the *expected shortfall*, the *return period* of a high level or the two primary parameters of extreme events, the *extreme value index* (EVI) and the *extremal index* (EI).

[1] Corresponding author.

AMS: 62G32, 62G09, 62G05, 62E20, 65C05.

In order to illustrate such topics, we consider essentially GJ *minimum-variance reduced-bias* (MVRB) classes of estimators of a positive EVI and of a general EI. The MVRB EVI-estimators were introduced and studied in Caeiro, Gomes and Pestana (2005). The GJ-MVRB EVI-estimators were studied in Gomes, Martins and Neves (2013). We further consider a GJ Leadbetter-Nandagopalan EI-estimator, introduced and studied in Gomes, Hall and Miranda (2008c). In Section 2, we begin with a brief introduction to *extreme value theory* (EVT). Both the EVI and the EI are defined, and first, second and third-order conditions in EVT are made explicit. In Section 3, a set of classical and reduced-bias EVI and EI-estimators is presented. Section 4 is dedicated to a brief overview to resampling methodologies and its use in reliable EVI and EI-estimation. In Sections 5 and 6 we respectively present a few illustrative case-studies and some overall conclusions.

2. EVT—a brief introduction

2.1. The EVI

We shall use the notation ξ for the EVI for maxima, the shape parameter in the *extreme value* (EV) cumulative distribution function (CDF),

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

and we shall consider models with a heavy right-tail, i.e. an underlying right tail or survival function,

$$\overline{F} := 1 - F \in \mathscr{R}_{-1/\xi}, \quad \text{for some} \quad \xi > 0, \tag{2}$$

where the notation $\mathscr{R}_\alpha$ stands 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 all $x > 0$, $g(tx)/g(t) \to x^\alpha$, as $t \to \infty$ (see, Bingham, Goldie and Teugels, 1987).

2.2. First, second and third-order frameworks

For the above model, in (2), the CDF F is in the domain of attraction for maxima of a Fréchet-type *extreme value* CDF, i.e. an EV_ξ CDF with $\xi > 0$, in the sense that given a sequence of random samples, $(X_1, \ldots, X_n)$, it is possible to linearly normalise the sequence of maximum values $\{X_{n:n} := \max(X_1, \ldots, X_n)\}_{n\geq 1}$ and get convergence to a non-degenerate random variable (RV), with CDF EV_ξ, defined in (1), with $\xi > 0$. We then write

$$F \in \mathscr{D}_{\mathrm{M}}(\mathrm{EV}_{\xi>0}) =: \mathscr{D}_M{}^+$$

(Gnedenko, 1943).

In this same context of heavy right-tails, and with the notation

$$U(t) = F^{\leftarrow}(1-1/t),\ \ t \geq 1,$$

$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$ the *generalised inverse function* of the underlying model F, we can further say (de Haan, 1984) that

$$F \in \mathscr{D}_{\mathrm{M}}{}^+ \quad \Longleftrightarrow \quad \overline{F} \in \mathscr{R}_{-1/\xi} \quad \Longleftrightarrow \quad U \in \mathscr{R}_\xi, \tag{3}$$

the so-called *first-order conditions*.

For consistent semi-parametric EVI-estimation, in the whole $\mathscr{D}_M^+$, we merely need to assume the validity of one of the *first-order conditions*, like $U \in \mathscr{R}_\xi$, and to work with adequate functionals, dependent on an *intermediate tuning* parameter k, the number of top order statistics involved in the estimation. This means that k needs to be such that

$$k = k_n \to \infty \quad \text{and} \quad k_n = o(n), \text{ as } n \to \infty. \tag{4}$$

To obtain information on the non-degenerate asymptotic behaviour of semi-parametric EVI-estimators, we need to further assume a *second-order condition*, ruling the rate of convergence in any of the *first-order conditions* in (3). The *second-order parameter*, ρ (≤ 0), rules such a rate of convergence, and it is the parameter appearing in the limiting result,

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

where we are using the interpretation of the Box-Cox transformation as the logarithm when the power equals zero. We often assume that (5) holds for every $x > 0$. Then $|A|$ must necessarily be in $\mathscr{R}_\rho$ (Geluk and de Haan, 1987). For technical simplicity we usually further assume that $\rho < 0$, writing

$$A(t) =: \xi \beta t^\rho, \tag{6}$$

dependent on the vector (β, ρ) of second-order parameters.

To obtain full information on the asymptotic bias of any corrected-bias EVI-estimator, it is often necessary to further assume a *general third-order condition*, ruling now the rate of convergence in the second-order condition in (5), which guarantees that, for all $x > 0$,

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

where $|B|$ must then be in $\mathscr{R}_{\rho'}$.

More restrictively, and equivalently to the aforementioned condition in (7) with $\rho = \rho' < 0$, we often consider a *Pareto third-order condition*, i.e., a Pareto-type class of models, with a tail function

$$1 - F(x) = Cx^{-1/\xi}\left(1 + D_1 x^{\rho/\xi} + D_2 x^{2\rho/\xi} + o\left(x^{2\rho/\xi}\right)\right), \tag{8}$$

as $x \to \infty$, with $C > 0$, $D_1,\ D_2 \neq 0$, $\rho < 0$.

Then we can choose in the aforementioned general *third-order condition*, in (7),

$$B(t) = \beta'\, t^\rho = \frac{\beta' A(t)}{\beta\xi} =: \frac{\zeta\, A(t)}{\xi}, \quad \beta,\ \beta' \neq 0, \quad \zeta = \frac{\beta'}{\beta}, \tag{9}$$

with β and β' 'scale' second and third-order parameters, respectively, and A the function in (6).

2.3. The EI

The EI is a parameter of extreme events related to the clustering of exceedances of high thresholds, a situation that occurs with stationary sequences (Leadbetter, 1973). We thus assume working with a strictly stationary sequence of RVs, $\{X_n\}_{n\geq 1}$, from F, under the long range dependence condition **D** (Leadbetter, Lindgren and Rootzén, 1983) and the local dependence condition **D''** (Leadbetter and Nandagopalan, 1989).

Definition 1 The stationary sequence $\{X_n\}_{n\geq 1}$ from an underlying model F is said to have an extremal index θ $(0 < \theta \leq 1)$ if, for all $\tau > 0$, we can find a sequence of levels $u_n = u_n(\tau)$ such that, with $\{Y_n\}_{n\geq 1}$ the associated *independent, identically distributed* (IID) sequence (i.e., an IID sequence from the same F),

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

Remark 1 **D** and **D''** are straightforwardly valid for IID data, and $\theta = 1$.

For dependent sequences there can thus appear a ‘shrinkage’ of maximum values, but the limiting CDF of $X_{n:n}$, linearly normalized, is still an EV CDF, i.e. from the CDF EV_ξ family, in (1).

The *extremal index* can also in most cases be defined as:

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

where u_n is a sequence of levels such that

$$F(u_n) = 1 - \tau/n + o(1/n), \text{ as } n \to \infty, \text{ with } \tau > 0, \text{ fixed.} \tag{10}$$

The ARMAX processes will be the ones used here for illustration. Such processes are based on an IID sequence of innovations $\{Z_i\}_{i\geq 1}$, with CDF H, and are defined through the relation,

$$X_i := \beta \max(X_{i-1}, Z_i), \quad i \geq 1, \quad 0 < \beta < 1.$$

The ARMAX sequences have a stationary distribution F, dependent on H through the stationarity equation $F(\beta x)/F(x) = H(x)$ (Alpuim, 1989). Conditions **D** and **D''** hold for these sequences and they can possess an extremal index $\theta < 1$.

For illustration, we shall consider ARMAX processes with Fréchet innovations, as illustrated in Figure 1. If $H(x) = \Phi_\xi^{\beta^{-1/\xi}-1}(x)$, then

$$F(x) = \Phi_\xi(x) = \exp\left(-x^{-1/\xi}\right), \; x \geq 0, \quad \text{and} \quad \theta = 1 - \beta^{1/\xi}.$$

Notice the richness of these processes regarding clustering of exceedances. Note also that for the same underlying model F there is a ‘shrinkage of maximum values’, together with the exhibition of larger and larger ‘clusters of exceedances’ of high values, as θ decreases.

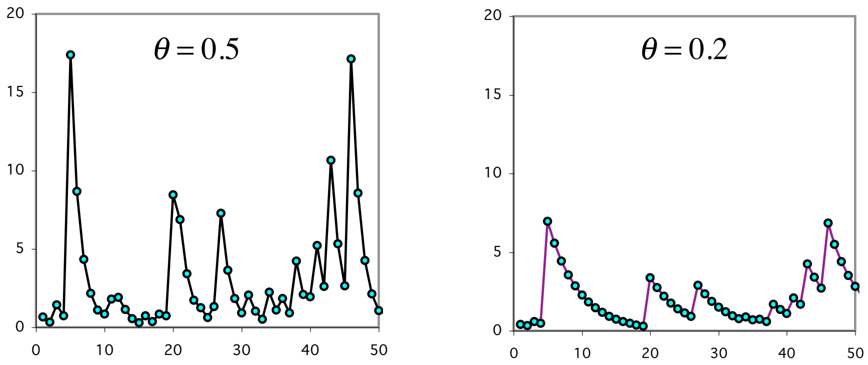

Figure 1: Sample paths of ARMAX processes with extremal index $\theta = 0.5$ (*left*) and 0.2 (*right*).

3. EVI and EI-ESTIMATORS

3.1. Classical EVI-estimators

For models in $\mathscr{D}^+_{\mathrm{M}}$, the classical EVI-estimators are the Hill estimators (Hill, 1975), averages of the *log-excesses*,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \qquad 1 \le i \le k < n,$$

i.e.,

$$\mathrm{H}_n(k) \equiv \mathrm{H}(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}, \quad 1 \le k < n. \tag{11}$$

But these EVI-estimators have often a strong asymptotic bias for moderate up to large values of k, of the order of $A(n/k)$, with A the function in (5), and the adequate accommodation of this bias has recently been extensively addressed in the literature.

3.2. Second-order reduced-bias (SORB) EVI-estimators

We mention the pioneering papers by Peng (1998), Beirlant, Dierckx, Goegebeur and Matthys (1999), Feuerverger and Hall (1999), Gomes, Martins and Neves (2000) and Gomes, Martins and Neves (2002), among others. In these papers, authors are led to SORB EVI-estimators, with asymptotic variances larger than or equal to $(\xi\ (1-\rho)/\rho)^2$, where $\rho(<0)$ is the aforementioned 'shape' second-order parameter, in (5). Note that $(\xi\ (1-\rho)/\rho)^2$ is the minimal asymptotic variance of an 'asymptotically unbiased' EVI-estimator in Drees' class of functionals (Drees, 1998).

3.3. MVRB EVI-estimators

Recently, Caeiro et al. (2005), Gomes, Martins and Neves (2007) and Gomes, de Haan and Henriques-Rodrigues (2008b) considered, in different ways, the problem of *corrected-bias* EVI-estimation, being able to *reduce the bias without increasing the asymptotic variance*, which was shown to be kept at ξ^2, the asymptotic variance of Hill's estimator, the maximum likelihood estimator of ξ for an underlying Pareto CDF, $F_P(x) = 1-(x/C)^{-1/\xi}$, $x \geq C$. Those estimators, called MVRB, from *minimum-variance reduced-bias*, are all based on an adequate 'external' consistent estimation of the pair of second-order parameters, $(\beta,\rho) \in (\mathbb{R},\mathbb{R}^-)$, in (6), done through estimators denoted $(\hat\beta,\hat\rho)$, and outperform the associated classical estimators for all k. We shall now consider the simplest class of MVRB EVI-estimators in Caeiro *et al.* (2005), a *corrected-Hill* (CH) EVI-estimator with the functional form

$$\overline{\mathrm{H}}(k) \equiv \overline{\mathrm{H}}_{\hat\beta,\hat\rho}(k) := \mathrm{H}(k)\Big(1-\frac{\hat\beta}{1-\hat\rho}\left(\frac{n}{k}\right)^{\hat\rho}\Big). \tag{12}$$

For the estimation of (β,ρ), and following Gomes and Pestana (2007) (see also, among others, Gomes, Caeiro, Henriques-Rodrigues and Manjunath, 2014), we consider the following:

Algorithm 1 (Second-order parameters' estimation).
Given $\underline{\mathbf{x}}_n := (x_1,\ldots,x_n)$*, an observed value of the random sample* $\underline{\mathbf{X}}_n := (X_1,\ldots,X_n)$*,*

S1 *Compute, for the tuning parameters* $\tau=0$ *and* $\tau=1$ *the observed values of the simplest* ρ*-estimator in* Fraga-Alves, Gomes and de Haan (2003)*,*

$$\hat\rho_\tau(k) \equiv \hat\rho_\tau(k;\underline{\mathbf{X}}_n) := -\left|\frac{3(V_\tau(k;\underline{\mathbf{X}}_n)-1)}{V_\tau(k;\underline{\mathbf{X}}_n)-3}\right|,$$

where

$$V_\tau(k;\underline{\mathbf{X}}_n) := \begin{cases} \dfrac{\left(M_n^{(1)}(k;\underline{\mathbf{X}}_n)\right)^{\tau}-\left(M_n^{(2)}(k;\underline{\mathbf{X}}_n)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k;\underline{\mathbf{X}}_n)/2\right)^{\tau/2}-\left(M_n^{(3)}(k;\underline{\mathbf{X}}_n)/6\right)^{\tau/3}}, & \textit{if} \quad \tau \neq 0, \\[2ex] \dfrac{\ln M_n^{(1)}(k;\underline{\mathbf{X}}_n)-\ln\left(M_n^{(2)}(k;\underline{\mathbf{X}}_n)/2\right)/2}{\ln\left(M_n^{(2)}(k;\underline{\mathbf{X}}_n)/2\right)/2-\ln\left(M_n^{(3)}(k;\underline{\mathbf{X}}_n)/6\right)/3}, & \textit{if} \quad \tau = 0, \end{cases}$$

is defined for any tuning parameter $\tau\in\mathbb{R}$*, with*

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

With $\lfloor x\rfloor$ *denoting the integer part of* x*, consider* $\{\hat\rho_\tau(k)\}_{k\in\mathscr{K}}$*, with* $\mathscr{K} = \left(\lfloor n^{0.995}\rfloor,\ \lfloor n^{0.999}\rfloor\right)$*, compute their median, denoted* χ_τ*, and further compute* $I_\tau := \sum_{k\in\mathscr{K}}\left(\hat\rho_\tau(k)-\chi_\tau\right)^2$*,* $\tau=0,1$*. Next choose the tuning parameter* $\tau^*=0$ *if* $I_0\leq I_1$*; otherwise, choose* $\tau^*=1$*.*

S2 *Work with* $\hat\rho \equiv \hat\rho_{\tau^*}(k_1)$ *and* $\hat\beta = \hat\beta_{\hat\rho}(k_1)$*, where*

$$k_1 = \left\lfloor n^{1-\varepsilon}\right\rfloor, \quad \varepsilon = 0.001,$$

and with

$$d_\alpha(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\alpha},\ D_\alpha(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\alpha} i\ln\frac{X_{n-i+1:n}}{X_{n-i:n}},\ \alpha\in\mathbb{R},$$

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

is the β*-estimator introduced and studied in* Gomes and Martins (2002).

For recent overviews on reduced-bias EVI-estimation see (Chapter 6, Reiss and Thomas, 2007), Gomes, Canto-Castro, Fraga-Alves and Pestana (2008a), Beirlant, Caeiro and Gomes (2012) and Gomes and Guillou (2014).

3.4. Asymptotic comparison of classical and MVRB EVI-estimators

The Hill estimator reveals usually a high asymptotic bias. Indeed, with $\mathcal{N}(\mu,\sigma^2)$ denoting a normal RV with mean value μ and variance σ^2, it follows from the results of de Haan and Peng (1998) that under the *general second-order condition*, in (5),

$$\sqrt{k}\big(\mathrm{H}(k)-\xi\big) \stackrel{d}{=} \mathcal{N}(0,\sigma_H^2)+b_H\sqrt{k}A(n/k)+o_p\Big(\sqrt{k}A(n/k)\Big),$$

where $\sigma_H^2=\xi^2$, and for $\rho<0$ and $A(t)=\xi\beta t^\rho$, already defined in (6), the bias $b_H\sqrt{k}A(n/k)=\xi\;\beta\;\sqrt{k}\;(n/k)^\rho/(1-\rho)$ can be very large (going to infinity), moderate (going to a constant) or small (going to zero) as $n\to\infty$, depending on the rate of increase of the sequence k_n with n.

This non-null asymptotic bias of the order of $A(n/k)$, 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 mean square error (MSE) pattern, as a function of k.

Under the same conditions as before, $\sqrt{k}\,\big(\overline{\mathrm{H}}(k)-\xi\big)$ is asymptotically normal with variance also equal to ξ^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, through for instance *Algorithm* 1, so that we get

$$\sqrt{k}\,\big(\overline{\mathrm{H}}(k)-\xi\big)\stackrel{d}{=}\mathcal{N}(0,\xi^2)+o_p\Big(\sqrt{k}A(n/k)\Big).$$

Consequently, $\overline{\mathrm{H}}(k)$ outperforms $\mathrm{H}(k)$ for all k. Under the validity of the aforementioned third-order condition related to Pareto-type class of models, i.e. the condition in (8), and with ζ defined in (9), we can then adequately estimate the vector of second-order parameters, (β,ρ), and write (Caeiro, Gomes and Henriques-Rodrigues, 2009)

$$\sqrt{k}\,\big(\overline{\mathrm{H}}(k)-\xi\big)\stackrel{d}{=}\mathcal{N}(0,\xi^2)+b_{\overline{\mathrm{H}}}\sqrt{k}A^2(n/k)+o_p\Big(\sqrt{k}A^2(n/k)\Big),$$

$$b_{\overline{\mathrm{H}}}=\frac{1}{\xi}\left(\frac{\zeta}{1-2\rho}-\frac{1}{(1-2\rho)^2}\right),$$

i.e. the bias is now of the order of $A^2(n/k)$.

3.5. Classical EI-estimators

Given a sample $(X_1,\ldots,X_n)$ and a suitably chosen threshold u, with I_A the indicator function of A, a possible estimator of θ (Leadbetter and Nandagopalan, 1989) is given by

$$\hat\theta_n^{\mathrm{N}}=\hat\theta_n^{\mathrm{N}}(u):=\frac{\sum_{j=1}^{n-1}I_{[X_j>u,X_{j+1}\le u]}}{\sum_{j=1}^{n}I_{[X_j>u]}}=\frac{\sum_{j=1}^{n-1}I_{[X_j\le u<X_{j+1}]}}{\sum_{j=1}^{n}I_{[X_j>u]}}.$$

To have consistency, the high level u must be such that $n(1-F(u_n)) = c_n\tau = \tau_n$, $\tau_n \to \infty$ and $\tau_n/n \to 0$ (Nandagopalan, 1990). Indeed, the intermediate sequence k_n, in (4), in an EVI-estimation is replaced, in an EI-estimation, by the sequence $\tau_n = c_n\tau$ with $c_n \to \infty$ as $n \to \infty$.

To make the semi-parametric EI-estimation closer to the semi-parametric EVI-estimation, it is sensible to consider any level $u \in [X_{n-k:n}, X_{n-k+1:n})$ (Gomes et al., 2008c), and the estimator

$$\hat{\theta}_n^{\mathrm{N}}(k) \equiv \hat{\theta}_n^{\mathrm{N}}(u) := \frac{1}{k}\sum_{j=1}^{n-1} I_{[X_j \le X_{n-k:n} < X_{j+1}]}. \tag{13}$$

Bias assumption on the data structures. For IID data ($\theta = 1$):

$$\mathbb{E}\left\{\hat{\theta}_n^{\mathrm{N}}(k)\right\} = 1 + \left(\frac{1}{2\,k} - \frac{k}{n}\right)(1+o(1)).$$

Moreover, for ARMAX processes, we get

$$\mathbb{E}\left\{\hat{\theta}_n^{\mathrm{N}}(k)\right\} = \theta - \left(\frac{\theta(\theta+1)}{2}\left(\frac{k}{n}\right) - \frac{3-2\,\theta}{2\,k}\right)(1+o(1)).$$

We shall thus consider the EI-estimator as a function of k, the number of order statistics larger than the chosen threshold, as given in (13). Moreover, we shall further assume a sensible structure for the asymptotic bias, given by

$$\mathrm{Bias}\left\{\hat{\theta}_n^{\mathrm{N}}(k)\right\} = \varphi_1(\theta)\left(\frac{k}{n}\right) + \varphi_2(\theta)\left(\frac{1}{k}\right) + o\left(\frac{1}{k}\right) + o\left(\frac{k}{n}\right), \tag{14}$$

as $n \to \infty$, and for any intermediate k (see Gomes et al., 2008c).

In the semi-parametric EI-estimation we have thus to cope with problems similar to the ones appearing in the EVI-estimation: *increasing bias, as the threshold decreases and a high variance for high thresholds*, and it is sensible to ask whether it is possible to improve the performance of estimators through the use of resampling methods.

We are next interested in the use of the GJ methodology, in order to reduce the bias of the MVRB EVI-estimators, in (12), and the classical EI-estimators, in (13). In statistics we often put the question, *"Will the combination of information improve the quality of estimators of a certain parameter or functional?"* The *jackknife* or GJ are resampling methodologies, which usually give a positive answer to such a question. Indeed, the main objectives of the *jackknife methodology* are:

1. Bias and variance estimation of a statistic, only through manipulation of observed data $\underline{x}$.
2. The building of estimators with bias and MSE smaller than those of an initial set of estimators.

4. Resampling methodologies

As mentioned in the very beginning of this article, the use of resampling methodologies was shown to be promising in the estimation of the nuisance parameter k, and in the reduction of bias of any estimator of a parameter of extreme events. If we ask how to choose the tuning parameter k in the EVI-estimation, either through $\mathrm{H}(k)$ or through $\overline{\mathrm{H}}(k)$, given respectively in (11) and (12), we usually consider the estimation of $k_0^{\mathrm{H}} := \arg\min_k \mathrm{MSE}(\mathrm{H}(k))$ or $k_0^{\overline{\mathrm{H}}} = \arg\min_k \mathrm{MSE}(\overline{\mathrm{H}}(k))$. To obtain estimates of k_0^{H} and $k_0^{\overline{\mathrm{H}}}$ one can then use a *double-bootstrap* method applied to an adequate *auxiliary statistic* which tends to **zero** and has an asymptotic behaviour similar to the one of either $\mathrm{H}(k)$ (Draisma, de Haan, Peng and Pereira, 1999; Danielsson, de Haan, Peng and de Vries, 2001; Gomes and Oliveira, 2001, among others) or $\overline{\mathrm{H}}(k)$ (Gomes, Mendonça and Pestana, 2011; Gomes, Figueiredo and Neves, 2012). See also, Caeiro and Gomes (2014) and Gomes et al. (2014), for short reviews on the role of bootstrap in statistics of extremes.

But at such optimal levels, we still have a non-null asymptotic bias even when we work with the CH EVI-estimator $\overline{\mathrm{H}}$, in (12). If we still want to remove such a bias, we can still make use of the GJ methodology. It is then enough to consider an adequate pair of estimators of the parameter of extreme events under consideration, and to build a *reduced-bias affine combination* of them. In Gomes et al. (2000) and Gomes et al. (2002), among others, we can find an application of this technique to the Hill estimator and in Gomes et al. (2013) an application to the CH EVI-estimators. To illustrate here the use of these methodologies in EVT, we begin with an application of the GJ methodology to the aforementioned MVRB EVI-estimators $\overline{\mathrm{H}}(k)$ in Caeiro et al. (2005), just as performed in Gomes et al. (2013).

4.1. The jackknife methodology and bias reduction

The pioneering EVI reduced-bias estimators are, in a certain sense, GJ estimators, i.e., affine combinations of well-known estimators of ξ. The GJ statistic was introduced by Gray and Schucany (1972): Let $T_n^{(1)}$ and $T_n^{(2)}$ be two biased estimators of ξ, with similar bias properties, i.e.,

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

Then, if $q = q_n = d_1(n)/d_2(n) \neq 1$, the affine combination

$$T_n^{\mathrm{G}} := \left(T_n^{(1)} - qT_n^{(2)}\right)/(1-q)$$

is an unbiased estimator of ξ.

4.2. A GJ corrected-bias EVI-estimator

Given $\overline{\mathrm{H}}$, defined in (12), the most natural GJ RV is the one associated with the random pair $\left(\overline{\mathrm{H}}(k), \overline{\mathrm{H}}(\lfloor \delta k \rfloor)\right)$, $0 < \delta < 1$, i.e.

$$\overline{\mathrm{H}}^{\mathrm{GJ}(q,\delta)}(k) := \frac{\overline{\mathrm{H}}(k) - q\,\overline{\mathrm{H}}(\lfloor \delta k \rfloor)}{1-q}, \quad 0 < \delta < 1,$$

with

$$q = q_n = \frac{\mathrm{Bias}_\infty\left\{\overline{\mathrm{H}}(k)\right\}}{\mathrm{Bias}_\infty\left\{\overline{\mathrm{H}}(\lfloor \delta k \rfloor)\right\}} = \frac{A^2(n/k)}{A^2(n/\lfloor \delta k \rfloor)} \underset{n/k \to \infty}{\longrightarrow} \delta^{2\rho}.$$

It is thus sensible to consider $q = \delta^{2\rho}$, $\delta = 1/2$, and, with $\hat{\rho}$ a consistent estimator of ρ, the GJ EVI-estimator,

$$\overline{\mathrm{H}}^{\mathrm{GJ}}(k) := \frac{2^{2\hat{\rho}}\,\overline{\mathrm{H}}(k) - \overline{\mathrm{H}}(\lfloor k/2 \rfloor)}{2^{2\hat{\rho}} - 1}. \tag{15}$$

Then (Gomes et al., 2013), provided that the condition $\hat{\rho}-\rho=o_p(1)$ holds,

$$\sqrt{k}\left(\overline{\mathrm{H}}^{\mathrm{GJ}}(k)-\xi\right) \stackrel{d}{=} \mathcal{N}(0,\sigma^2_{GJ})+o_p\left(\sqrt{k}A^2(n/k)\right),$$

with

$$\sigma^2_{\mathrm{GJ}}=\xi^2\left(1+1/(2^{-2\rho}-1)^2\right).$$

We have thus the 'old' trade-off between variance and bias, as happened with all the aforementioned SORB EVI-estimators associated with classical EVI-estimators. The bias decreases, but the variance increases. However, at optimal levels in the sense of minimal MSE, these third-order reduced-bias GJ EVI-estimators can often beat the MVRB EVI-estimators, as seen in Section 4.3.

4.3. Asymptotic bias and efficiency of an affine combination of corrected-bias EVI-estimators

Just as mentioned before, the most obvious affine combination associated with the CH EVI-estimator $\overline{\mathrm{H}}(k)$, is

$$\overline{\mathrm{H}}^{\mathrm{GJ}(a)}(k):=a\overline{\mathrm{H}}(\lfloor k/2\rfloor)+(1-a)\overline{\mathrm{H}}(k). \tag{16}$$

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

We next discuss the behaviour of the asymptotic bias and the efficiency of the affine combination in (16), discussing first their asymptotic properties for a possibly non-optimal choice of $a>1$. For a fixed level k, the reduction in the asymptotic bias of $\overline{\mathrm{H}}^{\mathrm{GJ}(a)}(k)$ compared to $\overline{\mathrm{H}}(k)$ can be measured by the indicator

$$\mathrm{ABR}_a:=\lim_{n\to\infty}\left(\left|\frac{\mathrm{Bias}_\infty\{\overline{\mathrm{H}}(k)\}}{\mathrm{Bias}_\infty\{\overline{\mathrm{H}}^{\mathrm{GJ}(a)}(k)\}}\right|\right)=\frac{1}{|1-a(1-2^{2\rho})|}. \tag{17}$$

In Figure 2 we present, in the (a,ρ)-plane, the values of the indicator ABR_a, in (17), independent of the tail index ξ.

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

$$\mathrm{AREFF}_a:=\frac{\mathrm{MSE}_\infty\left\{\overline{\mathrm{H}}(k_0^{\overline{\mathrm{H}}})\right\}}{\mathrm{MSE}_\infty\{\overline{\mathrm{H}}^{\mathrm{GJ}(a)}(k_0^{\overline{\mathrm{H}}^{\mathrm{GJ}(a)}})\}}=\left(\frac{(a^2+1)^{2\rho}}{1-a(1-2^{2\rho})}\right)^{\frac{2}{1-4\rho}}. \tag{18}$$

We next show, in Figure 3, and again in the (a,ρ)-plane, the values of the asymptotic relative efficiency indicator, in (18), which becomes infinity for $a=1/(1-2^{2\rho})$, the value of a associated with the GJ RV.

ρ \ a	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00	2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-0.10	1.15	1.17	1.18	1.20	1.22	1.24	1.26	1.28	1.30	1.33	1.35	1.37	1.40	1.42	1.45	1.48	1.51	1.54	1.57	1.60	1.63
-0.20	1.32	1.36	1.41	1.46	1.51	1.57	1.63	1.70	1.77	1.85	1.94	2.03	2.14	2.26	2.39	2.53	2.70	2.89	3.11	3.36	3.66
-0.30	1.52	1.60	1.69	1.79	1.91	2.04	2.19	2.37	2.58	2.83	3.13	3.50	3.98	4.60	5.45	6.69	8.67	12.3	21.1	75.3	48.2
-0.40	1.74	1.88	2.04	2.24	2.47	2.77	3.14	3.62	4.28	5.23	6.73	9.42	15.7	47.6	46.4	16	10.0	6.70	5.21	4.27	3.61
-0.50	2.00	2.22	2.50	2.86	3.33	4.00	5.00	6.67	10.0	20.0		20.0	10.0	6.67	5.00	4.00	3.33	2.86	2.50	2.22	2.00
-0.60	2.30	2.64	3.10	3.76	4.78	6.54	10	25.0	60.6	14	7.73	5.38	4.13	3.35	2.81	2.43	2.14	1.91	1.72	1.57	1.44
-0.70	2.64	3.16	3.93	5.19	7.66	14.6	159	17.9	8.48	5.55	4.13	3.29	2.73	2.33	2.04	1.81	1.63	1.48	1.35	1.25	1.16
-0.80	3.03	3.80	5.11	7.76	16.2	193	13.9	7.18	4.85	3.66	2.94	2.46	2.11	1.85	1.64	1.48	1.35	1.24	1.14	1.06	0.99
-0.90	3.48	4.63	6.92	13.6	489	14.4	7.12	4.72	3.53	2.82	2.35	2.01	1.76	1.56	1.41	1.28	1.17	1.08	1.00	0.94	0.88
-1.00	4.00	5.71	10.0	40.0	20.0	8.00	5.00	3.64	2.86	2.35	2.00	1.74	1.54	1.38	1.25	1.14	1.05	0.98	0.91	0.85	0.80
-1.10	4.59	7.17	16.3	58.6	10.5	5.76	3.97	3.03	2.45	2.06	1.77	1.56	1.39	1.25	1.14	1.05	0.97	0.90	0.84	0.79	0.74
-1.20	5.28	9.22	36.6	18.6	7.42	4.63	3.37	2.65	2.18	1.85	1.61	1.42	1.28	1.16	1.06	0.97	0.90	0.84	0.79	0.74	0.70
-1.30	6.06	12.3	482	11.7	5.91	3.96	2.98	2.38	1.99	1.70	1.49	1.33	1.19	1.09	1.00	0.92	0.85	0.80	0.75	0.70	0.66
-1.40	6.96	17.3	36.1	8.82	5.03	3.51	2.70	2.19	1.85	1.59	1.40	1.25	1.13	1.03	0.95	0.88	0.82	0.76	0.72	0.67	0.64
-1.50	8.00	26.7	20.0	7.27	4.44	3.20	2.50	2.05	1.74	1.51	1.33	1.19	1.08	0.99	0.91	0.84	0.78	0.73	0.69	0.65	0.62
-1.60	9.19	50.8	14.4	6.31	4.04	2.97	2.35	1.94	1.66	1.44	1.28	1.15	1.04	0.95	0.88	0.81	0.76	0.71	0.67	0.63	0.60
-1.70	10.6	238	11.6	5.65	3.74	2.79	2.23	1.86	1.59	1.39	1.23	1.11	1.01	0.92	0.85	0.79	0.74	0.69	0.65	0.62	0.58
-1.80	12.1	108	9.90	5.19	3.51	2.66	2.14	1.79	1.53	1.35	1.20	1.08	0.98	0.90	0.83	0.77	0.72	0.68	0.64	0.60	0.57
-1.90	13.9	48	8.78	4.84	3.34	2.55	2.06	1.73	1.49	1.31	1.17	1.05	0.96	0.88	0.81	0.76	0.71	0.66	0.63	0.59	0.56
-2.00	16.0	32	8.00	4.57	3.20	2.46	2.00	1.68	1.45	1.28	1.14	1.03	0.94	0.86	0.80	0.74	0.70	0.65	0.62	0.58	0.55

$ABR_a = 1$ | $1 \le ABR_a < 2$ | $2 \le ABR_a < 10$ | $ABR_a \ge 10$

Figure 2: Asymptotic bias reduction (ABR) indicator.

Some general comments:

- The reduction in bias is achieved in a wide region of the (a,ρ)-plane, making the choice of the *tuning parameter a* almost irrelevant. See Figure 2.

- However, it is clear from Figure 3 that for a reduction in MSE we indeed need to work close to the line $a = 1/(1-2^{2\rho})$. This justifies the introduction of the GJ estimator $\overline{\mathrm{H}}^{\mathrm{GJ}}(k)$, in (15).

- There is then a high reduction in the MSE of the GJ $\overline{\mathrm{H}}$ EVI-estimators, at optimal level, in the sense of minimal MSE as a function of k, compared to the original $\overline{\mathrm{H}}$ EVI-estimators, also at optimal level.

- The sample paths of these corrected-bias estimators are usually quite stable. The choice of the optimal level is thus of lesser importance.

- But, even so, we can use the *bootstrap methodology* for the choice of such an optimal level, as already mentioned in this article.

ρ \ a	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00	2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-0.10	1.00	0.99	0.99	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.99	0.99	1.00	1.01	1.02	1.03	1.05
-0.20	1.00	0.99	0.99	0.98	0.98	0.98	0.98	0.99	0.99	1.01	1.02	1.04	1.06	1.09	1.12	1.16	1.21	1.27	1.34	1.42	1.52
-0.30	1.00	0.99	0.99	0.99	1.00	1.01	1.02	1.05	1.08	1.12	1.17	1.24	1.34	1.47	1.65	1.91	2.33	3.09	4.88	14.96	9.66
-0.40	1.00	1.00	1.00	1.01	1.03	1.06	1.10	1.17	1.26	1.39	1.61	1.99	2.81	6.30	5.90	2.44	1.58	1.18	0.93	0.77	0.65
-0.50	1.00	1.00	1.02	1.04	1.08	1.15	1.25	1.43	1.77	2.66	####	2.39	1.43	1.04	0.82	0.67	0.57	0.49	0.43	0.38	0.34
-0.60	1.00	1.01	1.04	1.08	1.17	1.31	1.62	2.55	4.03	1.59	1.07	0.82	0.66	0.56	0.48	0.42	0.37	0.33	0.30	0.27	0.24
-0.70	1.00	1.02	1.06	1.15	1.31	1.72	5.65	1.68	1.06	0.80	0.64	0.54	0.46	0.40	0.36	0.32	0.29	0.26	0.24	0.22	0.20
-0.80	1.00	1.03	1.10	1.25	1.65	4.99	1.33	0.91	0.71	0.58	0.49	0.42	0.37	0.33	0.30	0.27	0.24	0.22	0.20	0.19	0.17
-0.90	1.00	1.05	1.15	1.44	6.32	1.27	0.87	0.68	0.56	0.47	0.41	0.36	0.32	0.29	0.26	0.24	0.22	0.20	0.18	0.17	0.16
-1.00	1.00	1.06	1.23	1.98	1.39	0.89	0.69	0.57	0.48	0.41	0.36	0.32	0.29	0.26	0.24	0.22	0.20	0.18	0.17	0.16	0.14
-1.10	1.00	1.09	1.36	2.02	0.99	0.73	0.59	0.50	0.43	0.38	0.33	0.30	0.27	0.24	0.22	0.20	0.19	0.17	0.16	0.15	0.14
-1.20	1.00	1.12	1.65	1.21	0.81	0.64	0.53	0.45	0.40	0.35	0.31	0.28	0.25	0.23	0.21	0.19	0.18	0.16	0.15	0.14	0.13
-1.30	1.00	1.15	3.47	0.96	0.71	0.58	0.49	0.42	0.37	0.33	0.30	0.27	0.24	0.22	0.20	0.18	0.17	0.16	0.15	0.14	0.13
-1.40	1.00	1.21	1.39	0.84	0.65	0.54	0.46	0.40	0.35	0.31	0.28	0.26	0.23	0.21	0.19	0.18	0.17	0.15	0.14	0.13	0.12
-1.50	1.00	1.29	1.10	0.75	0.60	0.51	0.44	0.38	0.34	0.30	0.27	0.25	0.23	0.21	0.19	0.17	0.16	0.15	0.14	0.13	0.12
-1.60	1.00	1.46	0.95	0.70	0.57	0.48	0.42	0.37	0.33	0.29	0.27	0.24	0.22	0.20	0.18	0.17	0.16	0.15	0.14	0.13	0.12
-1.70	1.00	2.04	0.86	0.66	0.54	0.47	0.41	0.36	0.32	0.29	0.26	0.24	0.22	0.20	0.18	0.17	0.16	0.14	0.13	0.13	0.12
-1.80	1.00	1.56	0.80	0.63	0.52	0.45	0.39	0.35	0.31	0.28	0.25	0.23	0.21	0.19	0.18	0.16	0.15	0.14	0.13	0.12	0.12
-1.90	1.00	1.22	0.75	0.60	0.51	0.44	0.39	0.34	0.31	0.28	0.25	0.23	0.21	0.19	0.18	0.16	0.15	0.14	0.13	0.12	0.11
-2.00	1.00	1.07	0.72	0.58	0.49	0.43	0.38	0.34	0.30	0.27	0.25	0.22	0.21	0.19	0.17	0.16	0.15	0.14	0.13	0.12	0.11

$AREFF_a = 1$ $\quad$ $1 \le AREFF_a < 2$ $\quad$ $AREFF_a \ge 2$

Figure 3: Asymptotic relative efficiency (AREFF) indicator.

4.4. A GJ corrected-bias EI-estimator

Since the bias term of the aforementioned classical EI-estimator reveals two main components of different orders, as can be seen in (14), we need to use an affine combination of three EI-estimators, i.e. an order-2 GJ-statistic.

Let $\underline{X} = (X_1, \ldots, X_n)$ be a sample from F, and let $T_n = T_n(\underline{X}, F)$ be an estimator of a functional $\theta(F)$, or of an unknown parameter θ. If the bias of our estimator reveals two main terms that we would like to remove, the GJ methodology advises us to deal with three estimators with the same type of bias:

Definition 2 Given three estimators $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$ of θ, such that

$$E\left\{T_n^{(i)} - \theta\right\} = d_1(\theta)\ \varphi_1^{(i)}(n) + d_2(\theta)\ \varphi_2^{(i)}(n),\ i = 1,2,3,$$

the GJ-statistic (of order 2) is given by

$$T_n^{GJ} := \left\| \begin{array}{ccc} T_n^{(1)} & T_n^{(2)} & T_n^{(3)} \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{array} \right\| \Big/ \left\| \begin{array}{ccc} 1 & 1 & 1 \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{array} \right\|,$$

with $||A||$ denoting, as usual, the determinant of the matrix A.

Straightforwardly, one may state:

Proposition 1 T_n^{GJ} is unbiased for the estimation of θ.

Moreover, although the variance of T_n^{GJ} is always larger than the variance of the original estimators, the MSE of T_n^{GJ} is often smaller than that of any of the statistics $T_n^{(i)}$, $i = 1,2,3$.

Given the information on the bias of the EI-estimator $\hat{\theta}_n^{\mathrm{N}}(k)$, in (13), as stated in (14), let us consider, as in Gomes et al. (2008c), the levels k, $\lfloor \delta k \rfloor + 1$ and $\lfloor \delta^2 k \rfloor + 1$, depending on a *tuning parameter* δ, $0 < \delta < 1$, and the class of estimators,

$$\hat{\theta}_n^{\mathrm{GJ}(\delta)}(k) := \frac{(\delta^2+1)\,\hat{\theta}_n^{\mathrm{N}}(\lfloor \delta k \rfloor + 1) - \delta\left(\hat{\theta}_n^{\mathrm{N}}\left(\lfloor \delta^2 k \rfloor + 1\right) + \hat{\theta}_n^{\mathrm{N}}(k)\right)}{(1-\delta)^2}. \tag{19}$$

Among the members of this class, the aforementioned authors have been heuristically led to the choice $\delta = 1/4$. Distributional properties of

$$\hat{\theta}_n^{\mathrm{GJ}}(k) := \hat{\theta}_n^{\mathrm{GJ}(1/4)}(k) \tag{20}$$

have so far been obtained *only* through simulation techniques, and are next briefly presented.

In Figure 4, we picture the sample paths of $\hat{\theta}_n^{\mathrm{N}}(k)$, in (13), and $\hat{\theta}_n^{\mathrm{GJ}}(k) \equiv \hat{\theta}_n^{\mathrm{GJ}(1/4)}(k)$, in (20), with $\hat{\theta}_n^{\mathrm{GJ}(\delta)}(k)$ generally given in (19), for a stationary Fréchet(1) ARMAX sample of size $n = 5000$, with $\theta = 0.5$. Note the reasonably high stability of the sample path of the GJ EI-estimator for a wide range of k-values and around the target value $\theta = 0.5$, compared to that of Nandagopalan's estimator.

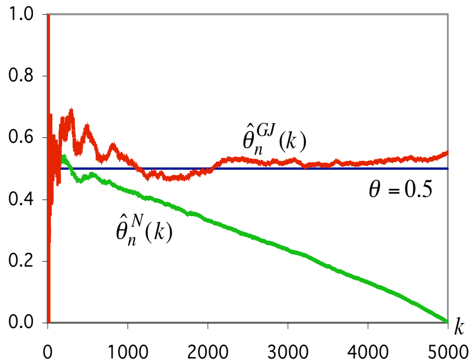

Figure 4: Sample paths of $\hat{\theta}_n^{\mathrm{N}}(k)$ and $\hat{\theta}_n^{\mathrm{GJ}}(k) \equiv \hat{\theta}_n^{\mathrm{GJ}(1/4)}(k)$, for a stationary Fréchet(1) ARMAX sample of size $n = 5000$, with $\theta = 0.5$.

In Figure 5, to exhibit the influence of the *tuning parameter* δ in the GJ EI-estimator, we present the expected values and MSEs of such an estimator, associated with $\delta = 0.1,\ 0.4$ and 0.5, for the

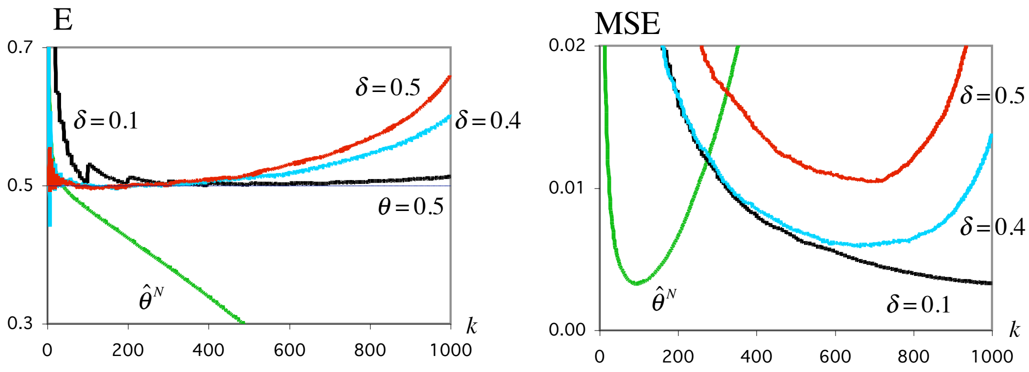

Figure 5: Expected value (E) and MSE of $\hat{\theta}_n^{\mathrm{GJ}(\delta)}(k)$, in (19), associated with $\delta = 0.1,\ 0.4$ and 0.5, for a stationary Fréchet(1) ARMAX sample of size $n = 5000$, with $\theta = 0.5$.

same structure as before.

Remark 2 The mean value stability around the target value θ, for a wide range of k-values, is true for all θ and for all models simulated in Gomes et al. (2008c). But for small θ, the GJ EI-estimator, $\hat{\theta}_n^{\mathrm{GJ}}$, in (20), may be not able to improve on the original EI-estimator, $\hat{\theta}_n^{\mathrm{N}}$, in (13), regarding MSE at optimal levels. Extra investment is thus needed in the 'optimal choice' of the three levels to be used in the building of a GJ EI-estimator or in the use of extra resampling or sub-sampling techniques, as initially performed in Gomes et al. (2008c). These authors have used simple subsampling techniques, briefly sketched in the following section, in order to attain a smaller MSE at optimal levels.

4.5. Effect of sampling frequency on the EI of an ARMAX process

From the articles of Robinson and Tawn (2000), Scotto, Turkman and Anderson (2003) and Martins and Ferreira (2004), among others, we get the following result for stationary sequences under **D** and **D''** conditions: If we consider a level $u = u_n$ such that (10) holds and the sub-sample $\mathbf{V} = \{X_{(n-1)T}\}_{n\geq 1}$ we have,

$$\theta_V = \theta_X + \lim_{n\to\infty} \frac{\sum_{i=0}^{T-2} P(X_0 \leq u < X_1, X_{T-i} > u)}{\tau/n}.$$

For ARMAX sequences,

$$\theta_{\mathbf{V}} = 1-(1-\theta_{\mathbf{X}})^T \iff \theta_{\mathbf{X}} = 1-(1-\theta_{\mathbf{V}})^{1/T}.$$

Sub-sampling may thus improve the performance of an EI-estimator. After the implementation of different subsampling algorithms, we here advance with the following simple algorithm.

Algorithm 2.
Fix T *(possibly* $T = 2$*), and compute* $r = \lfloor n/T \rfloor$.

S1 *Consider, for $i = 1,\ldots,T$, $\mathbf{V}_i = \big(X_i, X_{T+i}, \cdots, X_{(r-1)T+i}\big)$, the T subsamples of size r, and compute the estimates $\hat{\theta}^{\mathrm{GJ}}_{\mathbf{V}_i}(j)$, $j = 1,2,\ldots,r-1$.*

S2 *Compute*

$$\hat{\theta}^{\mathrm{GJ}}_{sub|T}(k) = 1 - \frac{1}{T}\sum_{i=1}^{T}\big(1-\hat{\theta}^{\mathrm{GJ}}_{\mathbf{V}_i}(j)\big)^{1/T},$$

for thresholds $k = (j-1)T+1,\ldots,j\,T$, $j = 1,2,\ldots,r-1$.

The use of the previous algorithm in $\hat{\theta}^{\mathrm{GJ}(\delta)}$, in (19), with $\delta = 1/4$, enable us to achieve, at optimal levels, an MSE smaller than that of $\hat{\theta}^{\mathrm{N}}$, even for small values of θ, the most problematic ones, i.e., the ones for which the GJ EI-estimator had not been able to improve on the original estimator, regarding MSE at optimal levels. For small θ (here illustrated with $\theta = 0.2$), we are able to improve on the original estimator at optimal levels, when we consider the GJ statistic with $\delta = 1/4$, defined in (20), together with the use of subsampling techniques with $T = 2$ or 3.

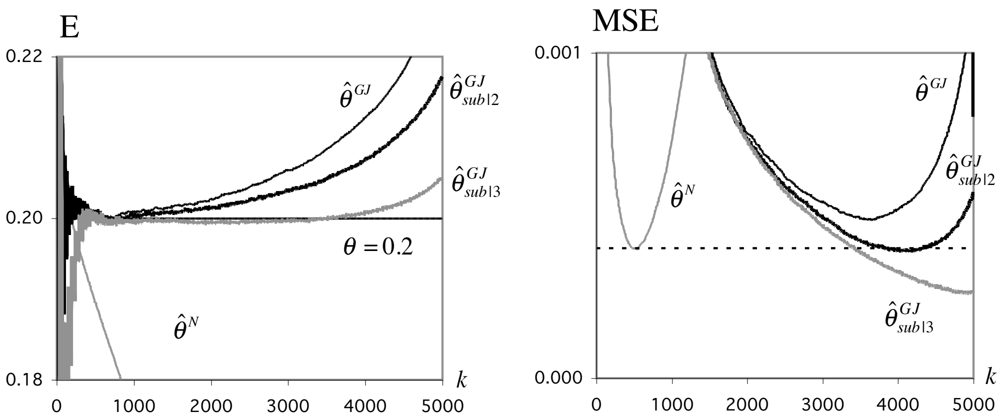

Figure 6: Behaviour of the GJ statistic with $\delta = 1/4$, together with the use of subsampling techniques with $T = 2$ and 3.

5. Case studies

5.1. The GJ EVI-estimation applied to insurance data

We next consider an illustration of the performance of the estimators under study, through the analysis of automobile claim amounts exceeding 1,200,000 Euro over the period 1988-2001, gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re), with a size $n = 371$. This data set was already studied in Beirlant, Goegebeur, Segers and Teugels (2004), Beirlant, Figueiredo, Gomes and Vandewalle (2008) and Vandewalle and Beirlant

(2006), as an example of an application to excess-of-loss reinsurance rating and heavy-tailed distributions in car insurance. It is clear from the box-and-whiskers plot in Figure 7 that data have been left-censored and that the right tail of the underlying model is quite heavy.

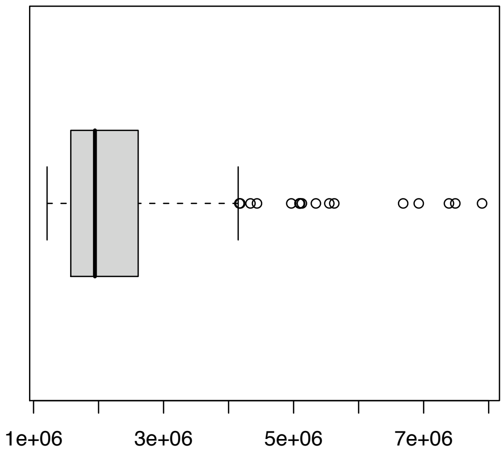

Figure 7: Box-and-whiskers plot associated with Secura data.

Regarding the EVI-estimation, note that whereas the Hill EVI-estimator is unbiased for the estimation of ξ when the underlying model is a strict Pareto model, it always exhibits a relevant bias when we have only Pareto-like tails, as happens here and can be seen in Figure 8.

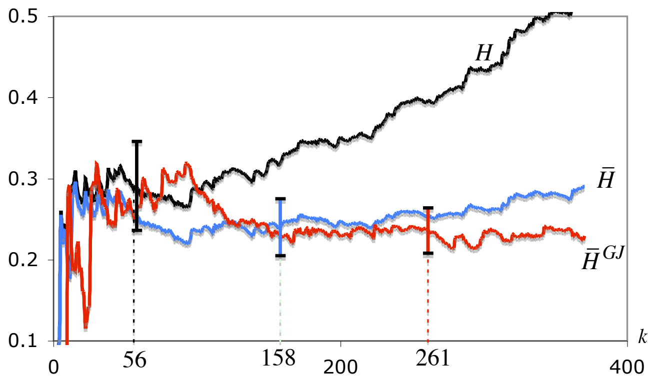

Figure 8: Estimates and 95% bootstrap confidence intervals of the extreme value index ξ for the Secura Belgian Re data.

The corrected-bias estimators, which are 'asymptotically unbiased', have a smaller bias, exhibit more stable sample paths as functions of k, and enable us to take a decision upon the estimate of ξ to be used, even with the help of any heuristic stability criterion, like the 'largest run' method suggested in Gomes and Figueiredo (2006). A bootstrap algorithm, not detailed here, but fully sketched in Gomes et al. (2012), helps us to provide an adaptive choice of the optimal sample fraction for corrected-bias EVI-estimators. Using that algorithm, we got $\hat{k}_{0|\mathrm{H}} = 56$, $\hat{k}_{0|\overline{H}} = 158$, $\hat{k}_{0|\overline{\mathrm{H}}^{\mathrm{GJ}}} = 261$, and the EVI-estimates $\mathrm{H}^* = 0.286$, $\overline{\mathrm{H}}^* = 0.240$ and $\overline{\mathrm{H}}^{\mathrm{GJ}^*} = 0.236$, the values pictured in Figure 8. The associated bootstrap 95% confidence intervals were $(0.236, 0.346)$, $(0.205, 0, 275)$ and $(0.208, 0.264)$, with sizes 0.110, 0.070 and 0.056, respectively for the Hill, the corrected-Hill and the generalised jackknife. Indeed, both bootstrap confidence intervals and asymptotic confidence intervals are easily associated with the estimates presented, the smallest size (with a high coverage probability) being related to $\overline{\mathrm{H}}^{\mathrm{GJ}^*}$.

5.2. The GJ EI-estimation applied to financial data

We now consider the performance of the above mentioned estimators in the analysis of Euro-UK Pound daily exchange rates from January 4, 1999 until December 14, 2004, already considered in Gomes et al. (2008c). Working with the $n_0 = 725$ positive log-returns, we picture as an illustration, the sample paths of $\hat{\theta}^{\mathrm{N}}(k)$ and $\hat{\theta}^{\mathrm{GJ}}(k)$, as functions of k. A bootstrap algorithm of the type of the one devised for EVI-estimation led us to $\hat{k}_{0|\mathrm{N}} = 45$, $\hat{k}_{0|GJ} = 230$, pictured in Figure 9, and the EI-estimates $\hat{\theta}^{\mathrm{N}*} = 0.96$ and $\hat{\theta}^{\mathrm{GJ}*} = 1$. The associated bootstrap 95% confidence intervals (truncated at one), also pictured at Figure 9, were $(0.885, 1)$ and $(0.934, 1)$, with sizes 0.115 and 0.066, the smallest size being the one associated with $\hat{\theta}^{\mathrm{GJ}}$.

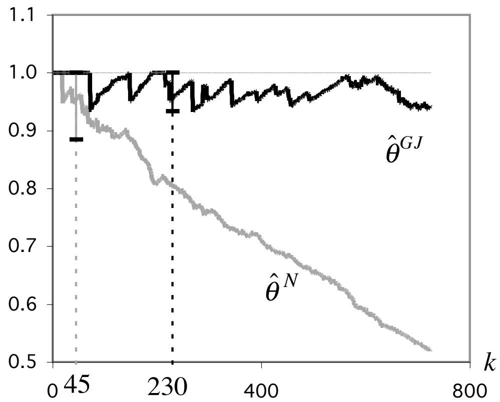

Figure 9: Sample paths of $\hat{\theta}^{\mathrm{N}}(k)$ and $\hat{\theta}^{\mathrm{GJ}}(k)$, as functions of k, and 95% bootstrap confidence intervals of the extremal index θ for the Euro-UK Pound log-returns under study.

The high stability of the GJ EI-estimates, around the value $\theta = 1$ (the estimate chosen by the bootstrap algorithm), is quite clear. Such a stability appears both for small and large values of k, whereas for the classical estimates the sample path exhibits a very small stability region.

6. Some overall conclusions

1. The most attractive features of the GJ estimators are their stable sample paths (for a wide region of k values), close to the target value, and the 'bath-tub' MSE patterns.

2. Regarding the EI-estimators, the choice $\delta = 1/4$ (heuristically based) in $\hat{\theta}_n^{\mathrm{GJ}(\delta)}(k)$, defined in (19), provides sample paths with a high stability. However reduction of MSE at optimal levels, relative to the original $\hat{\theta}_n^{\mathrm{N}}$ is not always achieved. Such an objective can be attained only with the extra use of a subsampling algorithm. Further investment is thus welcome.

3. Again: the insensitivity of the mean value and sample path to changes in k is indeed the nicest feature of these GJ-estimators.

4. And the tail bootstrap has revealed to be of high importance in the choice of the optimal threshold, in the sense of minimal MSE.

Acknowledgements. This research was partially supported by National Funds through **FCT** – Fundação para a Ciência e a Tecnologia, through projects PEst-OE/MAT/UI0006/2014 and PEst-OE/AGR/UI0239/2014. We also would like to thank the reviewer for helpful corrections and suggestions.

References

ALPUIM, M. T. (1989). An extremal markovian sequence. *Journal of Appl. Probab.*, **26**, 219–232.

BEIRLANT, J., CAEIRO, F., AND GOMES, M. I. (2012). An overview and open research topics in the field of statistics of univariate extremes. *Revstat*, **10** (1), 1–31.

BEIRLANT, J., DIERCKX, G., GOEGEBEUR, Y., AND MATTHYS, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, **2**, 177–200.

BEIRLANT, J., FIGUEIREDO, F., GOMES, M. I., AND VANDEWALLE, B. (2008). Improved reduced-bias tail index and quantile estimators. *Journal of Statistical Planning and Inference*, **138** (6), 1851–1870.

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., AND TEUGELS, J. L. (2004). *Statistics of Extremes. Theory and Applications*. Wiley.

BINGHAM, N. H., GOLDIE, C. M., AND TEUGELS, J. L. (1987). *Regular Variation*. Cambridge Univ. Press.

CAEIRO, F. AND GOMES, M. I. (2014). On the bootstrap methodology for the estimation of the tail sample fraction. *In* GILLI, M., GONZALEZ-RODRIGUEZ, G., AND NIETO-REYES, A. (Editors) *Proceedings of COMPSTAT 2014*. pp. 545–552.

CAEIRO, F., GOMES, M. I., AND HENRIQUES-RODRIGUES, L. (2009). Reduced-bias tail index estimators under a third order framework. *Communications in Statistics—Theory & Methods*, **38** (7), 1019–1040.

CAEIRO, F., GOMES, M. I., AND PESTANA, D. (2005). Direct reduction of bias of the classical hill estimator. *Revstat*, **3** (2), 113–136.

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. *Journal of Multivariate Analysis*, **76**, 226–248.

DE HAAN, L. (1984). Slow variation and characterization of domains of attraction. *In* DE OLIVEIRA, J. T. (Editor) *Statistical Extremes and Applications*. D. Reidel, Dordrecht, Holland, pp. 31–48.

DE HAAN, L. AND PENG, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, **52**, 60–70.

DRAISMA, G., DE HAAN, L., PENG, L., AND PEREIRA, T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme value index. *Extremes*, **2** (4), 367–404.

DREES, H. (1998). A general class of estimators of the extreme value index. *Journal Statistical Planning and Inference*, **98**, 95–112.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7** (1), 1–26.

FEUERVERGER, A. AND HALL, P. (1999). Estimating a tail exponent by modelling departure from a pareto distribution. *Annals of Statistics*, **27**, 760–781.

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), 193–214.

GELUK, J. AND DE HAAN, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.

GNEDENKO, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, **44**, 423–453.

GOMES, M. I., CAEIRO, F., HENRIQUES-RODRIGUES, L., AND MANJUNATH, B. G. (2014). Bootstrap methods in statistics of extremes. *In* LONGIN, F. (Editor) *Handbook of Extreme Value Theory and Its Applications to Finance and Insurance*. John Wiley & Sons, in press.

GOMES, M. I., CANTO-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.

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. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70** (1), 31–52.

GOMES, M. I. AND FIGUEIREDO, F. (2006). Bias reduction in risk modelling: semi-parametric quantile estimation. *Test*, **15** (2), 375–396.

GOMES, M. I., FIGUEIREDO, F., AND NEVES, M. M. (2012). Adaptive estimation of heavy right tails: the bootstrap methodology in action. *Extremes*, **15**, 463–489.

GOMES, M. I. AND GUILLOU, A. (2014). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review*, **doi:10.1111/insr.12058**.

GOMES, M. I., HALL, A., AND MIRANDA, C. (2008c). Subsampling techniques and the jackknife methodology in the estimation of the extremal index. *Journal of Computational Statistics and Data Analysis*, **52** (4), 2022–2041.

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.

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.

GOMES, M. I., MARTINS, M. J., AND NEVES, M. M. (2002). Generalized jackknife semi-parametric estimators of the tail index. *Portugaliae Mathematica*, **59** (4), 393–408.

GOMES, M. I., MARTINS, M. J., AND NEVES, M. M. (2007). Improving second order reduced bias extreme value index estimation. *Revstat*, **5** (2), 177–207.

GOMES, M. I., MARTINS, M. J., AND NEVES, M. M. (2013). Generalised jackknife-based estimators for univariate extreme-value modelling. *Communications in Statistics: Theory and Methods*, **42** (7), 1227–1245.

GOMES, M. I., MENDONÇA, S., AND PESTANA, D. (2011). Adaptive reduced-bias tail index and var estimation via the bootstrap methodology. *Communications in Statistics—Theory and Methods*, **40** (16), 2946–2968.

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.

GOMES, M. I. AND PESTANA, D. (2007). A sturdy reduced-bias extreme quantile (var) estimator. *Journal of the American Statistical Association*, **102** (477), 280–292.

GRAY, H. L. AND SCHUCANY, W. R. (1972). *The Generalized Jackknife Statistic*. Marcel Dekker.

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, **3**, 1163–1174.

LEADBETTER, M. R. (1973). On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **28**, 289–303.

LEADBETTER, M. R., LINDGREN, G., AND ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Series*. Springer-Verlag, New York.

LEADBETTER, M. R. AND NANDAGOPALAN, S. (1989). On exceedance point processes for stationary sequences under mild oscillation restrictions. *In* HÜSLER, J. AND REISS, R.-D. (Editors) *Statistical Extremes and Applications*. Springer-Verlag, New York, pp. 69–80.

MARTINS, A. P. AND FERREIRA, A. (2004). The extremal index of sub-sampled processes. *Journal of Statistical Planning and Inference*, **124**, 145–152.

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

PENG, L. (1998). Asymptotically unbiased estimator for the extreme-value index. *Statistics and Probability Letters*, **38** (2), 107–115.

REISS, R.-D. AND THOMAS, M. (2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser Verlag, Basel.

ROBINSON, M. E. AND TAWN, J. A. (2000). Extremal analysis of processes sampled at different frequencies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**, 117–135.

SCOTTO, M., TURKMAN, K. F., AND ANDERSON, C. W. (2003). Extremes of some sub-sampled time series. *Journal of Time Series Analysis*, **24** (5), 505–512.

VANDEWALLE, B. AND BEIRLANT, J. (2006). On univariate extreme value statistics and the estimation of reinsurance premiums. *Insurance: Mathematics and Economics*, **38**, 441–459.

Manuscript received, 2014-10-28, revised, 2014-12-12, accepted, 2015-01-02.