Statistics and Probability Letters 117 (2016) 128–135

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Competitive estimation of the extreme value index[☆]

M. Ivette Gomes[a,b,*], Lígia Henriques-Rodrigues[c,b]

[a] Universidade de Lisboa, Faculdade de Ciências (DEIO), Portugal
[b] CEAUL, Portugal
[c] Universidade de São Paulo, IME, Brazil

ARTICLE INFO

Article history:
Received 4 May 2016
Accepted 14 May 2016
Available online 25 May 2016

Keywords:
Heavy tails
PORT methodology
Statistical extreme value theory

ABSTRACT

The *mean-of-order-p* (MO_p) *extreme value index* (EVI) estimators are based on Hölder's mean of an adequate set of statistics, and generalize the classical Hill EVI-estimators, associated with $p = 0$. Such a class of estimators, dependent on the tuning parameter $p \in \mathbb{R}$, has revealed to be highly flexible, but it is not invariant for changes in location. To make the MO_p location-invariant, it is thus sensible to use the *peaks over a random threshold* (PORT) methodology, based upon the excesses over an adequate ascending order statistic. In this article, apart from an asymptotic comparison at optimal levels of the optimal MO_p class and some competitive EVI-estimators, like a Pareto probability weighted moments EVI-estimator, a few details on PORT classes of EVI-estimators are provided, enhancing their high efficiency both asymptotically and for finite samples.

1. Introduction

Given a sample of n *random variables* (RVs), $\underline{\mathbf{X}}_n := (X_1, \ldots, X_n)$, either independent, identically distributed or possible weakly dependent and stationary from a *cumulative distribution function* (CDF) F, let us denote by $X_{1:n} \le \cdots \le X_{n:n}$ the associated ascending order statistics. Let us further assume that there exist sequences of real constants, $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that the maximum, linearly normalized, i.e. $(X_{n:n} - b_n)/a_n$, converges in distribution to a non-degenerate RV. Then, the limiting CDF is compulsory of the type of an *extreme value* (EV) CDF, dependent upon a generalized shape parameter $\xi \in \mathbb{R}$, and with the functional form

$$\mathrm{EV}_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}),\ 1+\xi x > 0 & \text{if} \quad \xi \neq 0, \\ \exp(-\exp(-x)),\ x \in \mathbb{R} & \text{if} \quad \xi = 0. \end{cases} \tag{1.1}$$

It is then said that F belongs to the max-domain of attraction of EV_ξ, in (1.1), and the notation $F \in \mathcal{D}_{\mathcal{M}}\left(\mathrm{EV}_\xi\right)$ is commonly used in the field of *extreme value theory* (EVT). The parameter ξ, the so-called *extreme value index* (EVI), is the primary parameter of extreme events. The EVI measures the heaviness of the *right tail-function* (RTF), $\overline{F}(x) := 1 - F(x)$, and the heavier the right-tail the larger the EVI is. For Paretian-type RTFs ($\xi > 0$), i.e. the often called heavy RTFs, the most popular EVI-estimator is the Hill (H) EVI-estimator (Hill, 1975), the average of the k log-excesses,

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

☆ Research partially supported by FCT—Fundação para a Ciência e a Tecnologia, project UID/MAT/00006/2013 (CEA/UL).
* Correspondence to: Faculdade de Ciências de Lisboa, Bloco C6, Piso 4, 1749-016 Lisboa, Portugal.
E-mail address: ivette.gomes@fc.ul.pt (M.I. Gomes).

http://dx.doi.org/10.1016/j.spl.2016.05.012
0167-7152/

being thus the logarithm of the geometric mean (or mean-of-order-0) of

$$U_{ik} := X_{n-i+1:n}/X_{n-k:n}, \quad 1 \le i \le k < n, \tag{1.2}$$

i.e.

$$\hat{\xi}_k^{\mathrm{H}} = \hat{\xi}_k^{\mathrm{H}}(\underline{\mathbf{X}}_n) := \frac{1}{k}\sum_{i=1}^{k} V_{ik} = \ln\left(\prod_{i=1}^{k}\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{1/k}. \tag{1.3}$$

Brilhante et al. (2013) considered as basic statistics the *mean-of-order-p* (MO_p) of U_{ik}, in (1.2), for $p \ge 0$. More generally, Gomes and Caeiro (2014) considered the same statistics for $p \in \mathbb{R}$, i.e.

$$\mathrm{M}_p(k) = \begin{cases} \left(\frac{1}{k}\sum_{i=1}^{k} U_{ik}^p\right)^{1/p} & \text{if} \quad p \ne 0, \\ \left(\prod_{i=1}^{k} U_{ik}\right)^{1/k} & \text{if} \quad p = 0, \end{cases}$$

and, with $\hat{\xi}_k^{\mathrm{H}_0} \equiv \hat{\xi}_k^{\mathrm{H}}$, defined in (1.3), the class of MO_p EVI-estimators,

$$\hat{\xi}_k^{\mathrm{H}_p} = \hat{\xi}_k^{\mathrm{H}_p}(\underline{\mathbf{X}}_n) := \begin{cases} \left(1 - \mathrm{M}_p^{-p}(k)\right)/p & \text{if} \quad p < 1/\xi, \\ \ln \mathrm{M}_0(k) = \hat{\xi}_k^{\mathrm{H}} & \text{if} \quad p = 0. \end{cases} \tag{1.4}$$

The class of MO_p EVI-estimators in (1.4) depends on the tuning parameter $p \in \mathbb{R}$, it is highly flexible, it is scale-invariant but it is not invariant for changes in location, a property enjoyed by the EVI itself. To make the MO_p EVI-estimators in (1.4) location-invariant, it is sensible to use the *peaks over a random threshold* (PORT) methodology, introduced in Araújo-Santos et al. (2006), and further computationally studied in Gomes et al. (2008a). The PORT methodology is based on the sample of excesses over the random threshold $X_{n_q:n}$, $0 \le q < 1$, $n_q := \lfloor nq \rfloor + 1$, where $\lfloor x \rfloor$ denotes the integer part of x, i.e. it is based on the sample of size $n^{(q)} = n - n_q$,

$$\underline{\mathbf{X}}_n^{(q)} := \left(X_{n:n} - X_{n_q:n}, \ldots, X_{n_q+1:n} - X_{n_q:n}\right). \tag{1.5}$$

The PORT-MO_p EVI-estimators are thus estimators with the same functional form of the EVI-estimators in (1.4), but with the original sample $\underline{\mathbf{X}}_n$ replaced by the sample of excesses $\underline{\mathbf{X}}_n^{(q)}$, in (1.5). Consequently, the PORT-MO_p EVI-estimators are given by $\hat{\xi}_k^{\mathrm{H}_p^{(q)}} := \hat{\xi}_k^{\mathrm{H}_p}(\underline{\mathbf{X}}_n^{(q)})$.

Remark 1.1. We can have $q = 0$ whenever $F(\cdot)$ has a finite left endpoint (the random level can then be the minimum). Note that the choice $q = 0$ is appealing in practice, but should be used with care. Such a random threshold can indeed lead to under-estimation and even inconsistency (see, for instance Gomes et al., 2008a). Generally, we can have $0 < q < 1$ (the random level is then an empirical quantile).

As a measure of comparison, the recent and promising *Pareto probability weighted moments* (PPWM) (see Caeiro and Gomes, 2011; Caeiro et al., 2014) will be considered. The PPWM EVI-estimators are consistent for $\xi < 1$, depend on the statistics,

$$\hat{a}_0(k) := \frac{1}{k}\sum_{i=1}^{k} X_{n-i+1:n}, \qquad \hat{a}_1(k) := \frac{1}{k}\sum_{i=1}^{k}\frac{i-1}{k-1}X_{n-i+1:n},$$

and are defined by

$$\hat{\xi}_k^{\mathrm{PPWM}} = 1 - \frac{\hat{a}_1(k)}{\hat{a}_0(k) - \hat{a}_1(k)}, \quad 1 \le k < n. \tag{1.6}$$

PORT-PPWM EVI-estimators (see Caeiro et al., 2016) will also be included in the comparative studies to be developed in this article. In Section 2 a few details on second-order frameworks in EVT, reduced-bias estimation and asymptotic behavior of the estimators will be provided. Section 3 is dedicated to the finite sample properties of the EVI-estimators under study as well as their PORT-versions, done through a large-scale simulation study. Section 4 is devoted to a few final comments on the advantages of PORT EVI-estimators and on possible choices of the vector of tuning parameters.

2. Second-order frameworks for heavy RTFs, reduced-bias and PORT EVI-estimation

Let us consider for the reciprocal *right tail quantile function* (RTQF) associated with F, the notation $U(t) := F^{\leftarrow}(1 - 1/t)$, with $F^{\leftarrow}(y) := \inf\{x : F(x) \ge y\}$. For heavy right-tails, it is usual to work under the validity of a first-order condition