Article

Extreme Value Index Estimation by Means of an Inequality Curve

Emanuele Taufer [1,*]**, Flavio Santi** [2]**, Pier Luigi Novi Inverardi** [1]**, Giuseppe Espa** [1] **and Maria Michela Dickson** [1]

[1] Department of Economics and Management, University of Trento, 38122 Trento, Italy; pierluigi.noviinverardi@unitn.it (P.L.N.I.); giuseppe.espa@unitn.it (G.E.); mariamichela.dickson@unitn.it (M.M.D.)
[2] Department of Economics, University of Verona, 37136 Verona, Italy; flavio.santi@univr.it
[*] Correspondence: emanuele.taufer@unitn.it

check for updates

Abstract: A characterizing property of Zenga (1984) inequality curve is exploited in order to develop an estimator for the extreme value index of a distribution with regularly varying tail. The approach proposed here has a nice graphical interpretation which provides a powerful method for the analysis of the tail of a distribution. The properties of the proposed estimation strategy are analysed theoretically and by means of simulations. The usefulness of the method will be tested also on real data sets.

Keywords: extreme value index; inequality curve; non-parametric estimation; bootstrap; regularly varying tail

1. Introduction

A distribution function F with survivor function $\bar{F} := 1 - F$ is regularly varying (RV) at infinity with index α, if there exists an $\alpha > 0$ such that $\forall x > 0$

$$\lim_{t \to \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}; \tag{1}$$

in this case we say that $\bar{F} \in RV_{-\alpha}$. In the extreme value (EV) literature it is typical to refer to the EV index $\gamma > 0$ with $\alpha = 1/\gamma$. Informally, we will say that the distribution has a Pareto tail or that the distribution is of the power-law type. Note that the case $1 < \alpha \le 2$ (or $1/2 \le \gamma < 1$) entails distributions with infinite variance and finite mean while the case $\alpha > 2$ (or $\gamma < 1/2$) entails distributions with finite mean and variance.

Precision in the analysis of the tail of a distribution allows to, for example, perform proper risk evaluation in finance, correcting empirical income distributions for various top-income measurement problems, or individuating a proper growth theory in economics or the biological sciences. For further examples of applications and deeper discussion see Clauset et al. [1], Jenkins [2] and Hlasny [3] with specific references to applications in income distributions and an overview of available models; see also Heyde and Kou [4] for a deep discussion of graphical methods for tail analysis.

The present paper will concentrate on estimation of the EV index γ. Probably the most well-known estimator of the EV index is the Hill [5] estimator, which exploits the k upper order statistics of a random sample through the formula

$$H(k) := H_{k,n} = k^{-1} \sum_{i=1}^{k} \log X_{(n-i+1)} - \log X_{(n-k)}, \tag{2}$$

where $X_{(i)}$ denotes the *i*-th order statistics from a sample of size *n* and $k = k(n) \leq n$ diverges to ∞ in an appropriate way. The Hill estimator has been thoroughly studied in the literature and several generalizations have been proposed. For a recent review of estimation procedures for the EV (or tail) index of a distribution see Gomes and Guillou [6].

Some recent approaches in tail or EV index estimation we would like to mention here are those of Brilhante et al. [7] which define a *moment of order p* estimator which reduces to the Hill estimator for $p = 0$ and Beran et al. [8] which define a *harmonic moment* tail index estimator. Recently Paulauskas and Vaičiulis [9] and Paulauskas and Vaičiulis [10] have connected in an interesting way some of the above approaches by defining parametric families of functions of the order statistics. Reduced bias (RB) versions of the above estimators have appeared in the literature, see for example Caeiro et al. [11], Gomes et al. [12] and Gomes et al. [13].

The main contribution of this paper consists in a new estimation procedure for EV index of a distribution satisfying (1) which relies on Zenga's inequality curve $\lambda(p)$, $p \in (0,1)$ (Zenga [14]).

The curve λ has the property of being constant for the Pareto Type I distribution, it has an intuitive graphical interpretation, it does not depend on location and it shows a nice regular behaviour when estimated. These properties will be discussed, analysed and extended in order to define our inferential strategies. Also it is important to point out that an inequality curve is defined for positive observations and hence we will implicitly assume that the right tail of a distribution is analysed. This is not really a restriction since if one wishes to consider the left tail it is enough to change sign to the data. Also, if the distribution is over the real line, tails can be considered separately and, under the symmetry assumption, absolute values of the data could be considered. The approach to estimation proposed here, directly connected to the inequality curve λ, has a nice and effective graphical interpretation which greatly helps in the analysis. Other graph-based methods are to be found in Kratz and Resnick [15], which exploit properties of the QQ-plot, and Grahovac et al. [16] which discuss an approach based on the asymptotic properties of the partition function, a moment statistic generally employed in the analysis of multi-fractality; see also Jia et al. [17] which analyse graphically and analytically the real part of the characteristic function at the origin.

We would like to point out here that the λ curve discussed by Zenga [14] does not coincide with the Zenga [18] curve originally indicated by the author with $I(p)$, $p \in (0,1)$ (more details in the next Section).

The paper is organized as follows: Section 2 introduces the curve λ and discusses its properties; Section 3 analyses the proposed estimation strategy and discusses some practical issues in applications. Finite sample performances are analysed in Sections 4 and 5 where applications with simulated and real data are considered. Proofs are postponed to the last Section.

2. The Proposed Estimator for the EV Index

Let *X* be a positive random variable with finite mean μ, distribution function *F*, and probability density *f*. The inequality curve $\lambda(p)$ has been first defined in Zenga [14]; with original notation:

$$\lambda(p) = 1 - \frac{\log(1 - Q(F^{-1}(p)))}{\log(1 - p)}, \quad 0 < p < 1, \tag{3}$$

where $F^{-1}(p) = \inf\{x \colon F(x) \geq p\}$ is the generalized inverse of *F* and $Q(x) = \int_0^x t f(t) dt / \mu$ is the first incomplete moment. *Q* can be defined as a function of *p* via the Lorenz curve

$$L(p) = Q(F^{-1}(p)) = \frac{1}{\mu} \int_0^p F^{-1}(t) dt. \tag{4}$$

See further Zenga [19] Arcagni and Porro [20] for a general introduction and analysis of $\lambda(p)$ for different distributions. It is worth to mention that the curve $\lambda(p)$ should not be confused with the inequality curve defined in Zenga [18], originally indicated as

$$I(p) = 1 - \frac{L(p)}{p} \frac{1-p}{1-L(p)}, \quad p \in (0,1). \tag{5}$$

The curve $I(p)$ has many nice properties and has been heavily studied in some recent literature; it is now commonly known as the Zenga curve $Z(p)$. For the sake of completeness in Zenga [14] the notation $Z(p)$ was originally used for another inequality curve based on quantiles, that is,

$$Z(p) = 1 - \frac{x_p}{x_p^*}, \quad p \in (0,1), \tag{6}$$

where $x_p = F^{-1}(p)$ and $x_p^* = Q^{-1}(p)$. As pointed out in Zenga [14] (without providing if and only if results) the curve λ is constant in p for type-I Pareto distributions, while the curve Z, as defined in Equation (6), is constant in p for Log-normal distributions. On the contrary, the curve I, as defined in (5), is not constant for any distribution, see Zenga [14] and Zenga [18] for further details. Turning back the attention to the curve λ, note that for a Pareto Type I distribution with

$$F(x) = 1 - (x/x_0)^{-\alpha}, \quad x \geq x_0, \quad x_0 > 0 \tag{7}$$

under the condition that $\alpha > 1$, the Lorenz curve has the form

$$L(p) = 1 - (1-p)^{1-\gamma}, \quad \gamma = 1/\alpha; \tag{8}$$

it follows that in this case $\lambda(p) = \gamma$, $p \in (0,1)$, that is $\lambda(p)$ is constant in p. This is actually an if-and-only-if result, which we formalize in the following lemma (see Section 7 for its proof).

Lemma 1. *The curve $\lambda(p)$ defined in (3) is constant in p, $p \in (0,1)$, and equals $\gamma = 1/\alpha$ if, and only if, F satisfies (7) with $\alpha > 1$ or, equivalently, $\gamma < 1$.*

Lemma 1 could be exploited to derive a new approach to the estimation of the EV index $\gamma = 1/\alpha$ for the Pareto distribution. In order to define an estimator for the more general case where $\bar{F}$ satisfies (1) it is worth to analyse in more detail what is the behaviour of the Lorenz curve under the framework defined by (1). This will be done by considering the truncated random variable $Y = X|X > s$ with $X \sim F$, $F \in RV_{-1/\gamma}$. If G and g denote respectively the distribution function and the density of Y, note that $G(y) = \frac{F(y)-F(s)}{\bar{F}(s)}$ and $g(y) = f(y)/\bar{F}(s)$. Furthermore, setting $G(y) = p$ and inverting we have $G^{-1}(p) = F^{-1}(F(s) + p\bar{F}(s))$. A formal result on the Lorenz curve for Y is given in the next lemma.

Lemma 2. *Consider the random variable X with distribution function $F \in RV_{-1/\gamma}$ and absolutely continuous density f; define $Y = X|X > s$, $s > 0$, and let $L_Y(p)$ the Lorenz curve of Y. Then*

$$1 - L_Y(p) = (1-p)^{1-\gamma}, \quad p \in (0,1), \quad s \to \infty. \tag{9}$$

Remark 1. *Lemma 2 implies that the curve $\lambda(p)$, for the truncated random variable $Y = X|X > s$, with distribution satisfying (1), will be constant with value γ for all $p \in (0,1)$ if the truncation level s will be large enough. This fact can be exploited to derive a general estimator for the EV index for all distributions in the class (1) as long as $\gamma < 1$.*

Before arriving at a formal definition of the estimator, some preliminary quantities need to be defined. Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics of a random sample of size n from a distribution satisfying (1). Let $k = k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$. Define the estimator of the conditional Lorenz curve as

$$\hat{L}_k(p) = \frac{\sum_{j=1}^{i} X_{(n-k+j)}}{\sum_{j=1}^{k} X_{(n-k+j)}}, \quad \text{for} \quad \frac{i}{k} \leq p < \frac{i+1}{k}, \quad i = 1, \ldots, k-1. \tag{10}$$

After defining

$$\hat{\lambda}_{k,i} = \hat{\lambda}_k(p_i) = 1 - \frac{\log(1 - \hat{L}_k(p_i))}{\log(1 - p_i)} \tag{11}$$

the proposed estimator of γ is

$$\hat{\gamma}_k = \frac{1}{k-1} \sum_{i=1}^{k-1} \hat{\lambda}_{k,i}. \tag{12}$$

Remark 2. *The estimator defined in (12), based on a Lorenz curve computed on upper order statistics (defined by k), puts into practice the results of Lemma 1 and Lemma 2. Below we will discuss conditions under which (12) provides a consistent estimator of γ for the class of distributions satisfying (1). Guidance in the choice of k will be also discussed.*

Letting $\mathbb{I}_{(A)}$ denote the indicator function of the event A the above estimators are based on the non-parametric estimators

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(X_i \leq x)} \qquad Q_n(x) = \frac{\sum_{i=1}^{n} X_i \mathbb{I}_{(X_i \leq x)}}{\sum_{i=1}^{n} X_i}. \tag{13}$$

Under the Glivenko-Cantelli theorem it holds that $F_n(x) \to F(x)$ almost surely and uniformly in $0 < x < \infty$; under the assumption that $E(X) < \infty$, it holds that $Q_n(x) \to Q(x)$ almost surely and uniformly in $0 < x < \infty$ (Goldie [21]). F_n and Q_n are both step functions with jumps at $X_{(1)}, \ldots, X_{(n)}$. The jumps of F_n are of size $1/n$ while the jumps of Q_n are of size $X_{(i)}/T$ where $T = \sum_{i=1}^{n} X_{(i)}$.

Letting $F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\}$, we note that since $F_n^{-1}\left(\frac{n-k}{n}\right) = x_{(n-k)}$ and that $F_n^{-1}\left(F_n(X_{(n-k)}) + p\bar{F}_n(X_{(n-k)})\right) = x_{(n-k+i)}$ for $i/k \leq p < (i+1)/k$ we have the representation

$$\hat{L}_k(p) = \frac{\sum_{i=1}^{n} X_i \mathbb{I}_{(X_i > X_{(n-k)})} \mathbb{I}_{(X_i \leq X_{(n-k+i)})}}{\sum_{i=1}^{n} X_i \mathbb{I}_{(X_i > X_{(n-k)})}}, \quad \text{for} \quad \frac{i}{k} \leq p < \frac{i+1}{k}, \quad i = 1, \ldots, k-1. \tag{14}$$

Exploiting the above representation and the results of Goldie [21], uniform consistency of $\hat{L}_k(p)$ can be claimed. As far as uniform consistency of $\hat{\lambda}_k(p)$ we state the following lemma, which is proven in Section 7.

Lemma 3. *For $X_1, \ldots, X_n$ i.i.d. from a distribution F with $E(X) < \infty$; then*

$$\sup_{p \in (0,1)} |\hat{\lambda}_{k(n)}(p) - \lambda(p)| = o_p(1), \quad n \to \infty. \tag{15}$$

Following Lemma 2, graphical inspection of the tail of a distribution satisfying (1) can be carried out by analysing a graph with coordinates $(p_i, \hat{\lambda}_i)$, $i = 1, \ldots, n$ which will show a flat line with intercept around the value $\gamma = 1/\alpha$. Apart from the case of the Pareto distribution, for distributions satisfying (1), to observe a constant line with intercept close to $\gamma = 1/\alpha$ it is necessary to truncate the sample, that is, using only the upper order statistics $X_{(n-k+1)}, \ldots X_{(n)}$ when estimating λ.

As an example, Figure 1 reports the empirical curve $\hat{\lambda}_i$ as a function of p_i for some cases of interest at different truncation thresholds. There appear two distributions with tail satisfying (1), namely Pareto as defined by (7) and Fréchet (more formally defined below), both with tail index $\alpha = 2$. There appear also two distributions which do not satisfy (1), namely Log-normal with null location and standard deviation equal to 2 and Exponential with unit scale. Note that for Log-normal distribution the curve λ does not depend on location, while it does not depend on scale for the exponential distribution (Zenga [14]).

Inspection of the graphs reveals a remarkably regular behaviour of the curves; the Pareto case is constant (with some slight variations) for all level of truncation, while the Fréchet one becomes more and more constant with increasing levels of truncation. The Log-normal and Exponential cases show a slope in the curve at all levels of truncation.
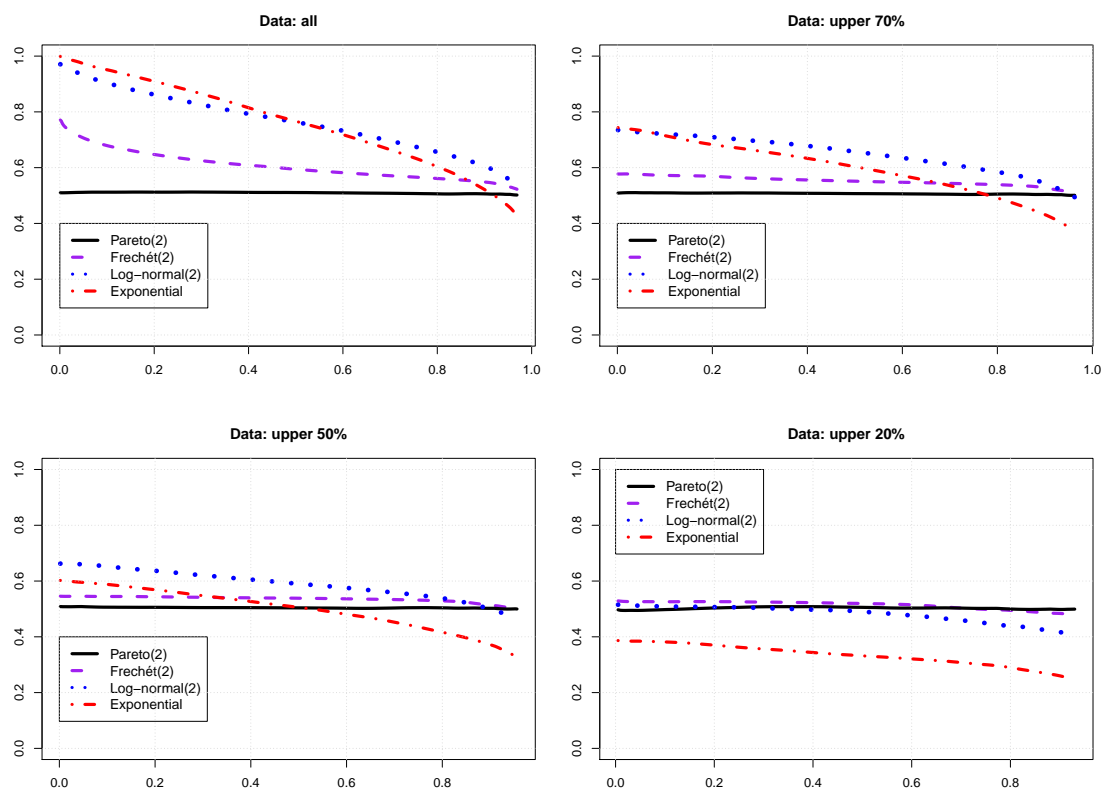

Figure 1. Plot of $\hat{\lambda}_i$ (y-axis) as a function of p_i (x-axis), $i = 1, \dots, n$, for Pareto, Fréchet, Log-normal and Exponential distributions at various levels of truncation. Sample size $n = 1000$.

3. Asymptotic Properties of $\hat{\gamma}_k$

3.1. Consistency

Exploiting some theoretical results given Section 7 (see the proof Lemma 2 for details), one can write

$$\frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{\lambda}_k(p_i) - \lambda(p_i)) = (\hat{\gamma}_n - \gamma) - \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{\log\left[H_U([(1-p_i)\overline{F}(s)]^{-1})/H_U([\overline{F}(s)]^{-1})\right]}{\log(1-p_i)} \quad (16)$$

where $H_U(s)$ is a slowly varying function, that is, it holds that $\lim_{s \to \infty} \frac{H_U(sx)}{H_U(s)} = 1$ for any $x > 0$.

To analyze in more detail the second term on the *r.h.s.* of the above equation we assume a second order condition which is quite common in the EV theory (Caeiro et al. [11] and Gomes et al. [13]), that is,

$$\lim_{s \to \infty} \log H_U([(1-p)\overline{F}(s)]^{-1}) - \log H_U([\overline{F}(s)]^{-1}) = \frac{(1-p_i)^\rho - 1}{\rho} A([\overline{F}(s)]^{-1}), \qquad (17)$$

where $A(t) = \gamma \beta t^\rho$, $C > 0$, $\gamma > 0$, $\rho < 0$, $\beta \neq 0$. To evaluate the *r.h.s.* of (16), as $n \to \infty$, set $s = F^{-1}\left(\frac{(n-k)}{n}\right)$ from which $\overline{F}(s) = k/n$. Then, an asymptotic evaluation requires to evaluate the expression

$$\gamma \beta \left(\frac{k}{n}\right)^{-\rho} \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{(1-p_i)^\rho - 1}{\rho \log(1-p_i)}. \qquad (18)$$

Note that the asymptotic behaviour of the sum in (18) is governed by $p_i \to 0$. Expanding in Taylor series the numerator and using $\log(1-x) \sim -x$ as $x \to 0$,

$$\left(\frac{k}{n}\right)^{-\rho} \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{(1-p_i)^\rho - 1}{\rho \log(1-p_i)} \sim \left(\frac{k}{n}\right)^{-\rho} \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{1 - \rho p_i + \frac{1}{2}\rho(\rho-1)p_i^2 - 1}{-\rho p_i}$$
$$\sim \left(\frac{k}{n}\right)^{-\rho} \left(1 + \frac{1}{4}(1-\rho)(k-1)\right), \qquad (19)$$

which is $o(1)$ as $k \to \infty$ if $k = n^\delta$, $0 < \delta < 1$ with $\delta < -\frac{\rho}{1-\rho}$. For example if $\rho = -1/2$ then one can choose $0 < \delta < 1/3$ in order to have an asymptotically bias free estimator. There exists valid estimation methods for ρ and β implemented also in R (see Section 4 for details).

Lemma 4. *Under the conditions of Lemma 3 and condition (17), with $k = n^\delta$, $0 < \delta < -\frac{\rho}{1-\rho}$*

$$|\hat{\gamma}_k - \gamma| = o_P(1) \quad as \quad n \to \infty. \qquad (20)$$

3.2. Asymptotic Distribution

To provide an operational distributional result, we exploit a result of Csorgo and Mason [22]. Given $X_1, \ldots, X_n$ i.i.d. with $X \sim F$, $F \in RV_{-1/\gamma}$, let $S_j = E_1 + E_2 + \ldots E_j$, where E_j's are i.i.d. Exponential with unit scale parameter (Exp(1)) random variables; then, for fixed k,

$$\frac{1}{n^\gamma H(1/n)} \sum_{j=1}^{k} X_{(n-j+1)} \to_D \sum_{j=1}^{k} (S_j)^{-\gamma}, \quad n \to \infty. \qquad (21)$$

Note that $S_j \sim \Gamma(j, 1)$ and that if $Y = S_j^{-\gamma}$ then Y has a Generalized Inverse Gamma (GIG) distribution with density

$$f(y) = \gamma \frac{1}{\Gamma(j)} x^{-j\gamma - 1} e^{-\left(\frac{1}{y}\right)^\gamma}, \qquad j, \gamma > 0. \qquad (22)$$

Compare with Mead [23] setting $\lambda = 0$, $\alpha = j$, $\beta = \gamma$, $\theta = 1$. Using this general result, a simple and fast parametric bootstrap procedure can be implemented in order to obtain the full asymptotic distribution of $\hat{\gamma}$ and from it estimates of the standard error and confidence intervals.

Remark 3. *Since $\hat{\gamma}_k$ is consistent, the above procedure is consistent for the asymptotic distribution of the estimator (12).*

Once the bootstrap distribution is available it can be used for variance and confidence intervals estimation. Simulations show that the approximation works quite well already for small sample sizes and for different k values. Clearly the precision depends on a good preliminary estimator of γ which

is the only parameter needed in determining the distribution; this is however a typical feature of asymptotic results for estimators.

Figure 2, considering different sample sizes ($n = 100, 500, 1000, 2000$) from a Fréchet(2) distribution, shows the histograms of true distribution of $\hat{\gamma}_k$ (obtained by the Montecarlo method with 2000 iterations) and the bootstrap distribution obtained by Algorithm 1. The value of γ used in Algorithm 1 has been randomly selected from the 1000 central values estimated in the connected Montecarlo experiment. Sampling of $\hat{\gamma}_k$ has been done independently in each of the 4 experiments in the graph.

Algorithm 1 Bootstrap for the asymptotic distribution of $\hat{\gamma}_k$.

1: Given the data, get the estimated value $\hat{\gamma}_k$ using formulae (10) to (12).

2: Generate k i.i.d. Exp(1) random variables $E_1, \ldots, E_k$ and form the partial sums $S_j = \sum_{i=1}^{j} E_i$.

3: Obtain the bootstrap estimate, say $\hat{\gamma}^*$, using estimator (12) applied to the data $S_1^{-\hat{\gamma}}, \ldots, S_k^{-\hat{\gamma}}$.

4: Repeat the previous steps a large number of times to get the asymptotic distribution of $\hat{\gamma}_k$ for given k and given sample size.

(**a**) n = 100

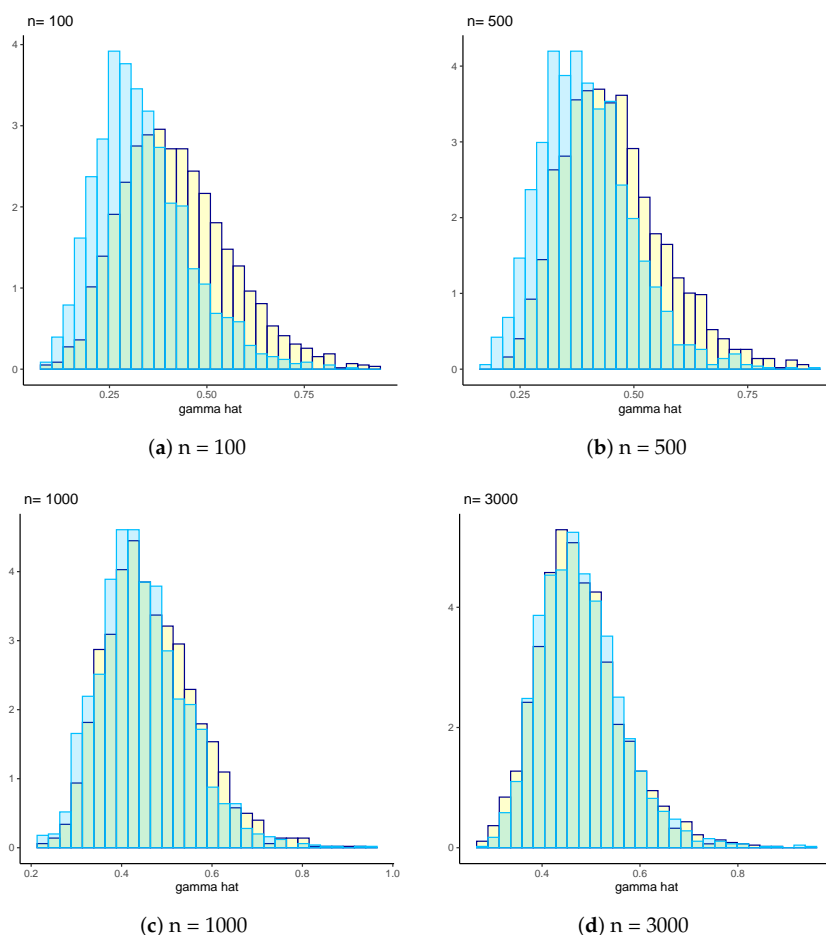(**b**) n = 500

(**c**) n = 1000

(**d**) n = 3000

Figure 2. Histograms of the empirical distribution of $\hat{\gamma}_k$ for selected sample sizes; $k = n^{0.5}$; data samples are generated from a Fréchet(2) distribution. Yellow: values obtained by Montecarlo simulations (2000 iterations); blue: values obtained by Algorithm 1 (2000 iterations). The value of γ used has been selected randomly from a pool of estimated values.

3.3. Selecting k

The estimator $\hat{\gamma}_k$, like many tail estimators, requires the choice of k, the number of upper order statistics to be used in estimation. Lemma 4 provides some indication on how to do so; estimation of the required parameters governing the second order conditions can be carried on quite straightforwardly (see discussion in the next section).

In order to arrive at a data-driven procedure to define the fraction of upper order statistics for estimation of the EV index, consider the linear equation

$$\lambda(p) = \beta_0 + \beta_1 p. \tag{23}$$

From Lemmas 1 and 2, it follows that for a type-I Pareto distribution and for for a truncated random variable satisfying (1), as $s \to \infty$ is large enough, distribution F satisfying (7) with $\gamma = 1/\alpha$, in the above equation one has $\beta_0 = \gamma$ and $\beta_1 = 0$. Considering a sample version of (23): given a random sample $X_1, \ldots, X_n$, using the notation established in the previous Section, write

$$\hat{\lambda}_{k,i} = \beta_0 + \beta_1 p_i + \varepsilon_i \quad i = 1, \ldots, k-1, \tag{24}$$

where $\varepsilon_i = \hat{\lambda}_i - \lambda_i$. Note that the proposed estimator (12) can be interpreted as the intercept estimate in model (24) exploiting the information that $\beta_1 = 0$. More formally, using ordinary least squares, define the estimators

$$\hat{\gamma}_k = \hat{\beta}_0 = \frac{1}{k-1} \sum_{i=1}^{k-1} \hat{\lambda}_{k,i}, \qquad \hat{\beta}_1 = \sum_{i=1}^{k-1} \frac{\hat{\lambda}_{k,i}(p_i - \bar{p})}{S_p^2} = \sum_{i=1}^{k-1} \hat{\lambda}_{k,i} c(p_i), \tag{25}$$

where $\bar{p}$ is the mean of the p_i's and $S_p^2 = \sum_i^{k-1}(p_i - \bar{p})^2$, $c(p_i) = (p_i - \bar{p})/S_p^2$. Lemma 3 implies

Proposition 1. *Under the conditions of Lemma 4 it holds that* $\hat{\beta}_1 \to_P 0$.

Following this reasoning, one can define a procedure based on the graph of $(p_i, \hat{\lambda}_{k,i})$ for different levels of truncation: we observe the fraction of upper order statistics which gets the smallest, in absolute value, regression coefficient $\hat{\beta}_1$ from the regression (25).

Algorithm 2 Data-driven estimator $\hat{\gamma}_{Opt}$.

1: Given a random sample of size n, order the data and consider sub-samples defined by the $(1-p)$-th fraction of upper order statistics. In our simulations the values $p = 0.1 \cdot i$, $i = 0, 1, \ldots, 9$ were considered. However, in order to avoid using sub-samples with too few observations when n is small an upper bound of the form $0.5 + 0.4 \max(0, (n-100)/n)$ is imposed to the sequence $p = 0.1 \cdot i$, $i = 0, 1, \ldots, 9$. For example, when $n \leq 100$, at least 50% of the largest order statistics is used.

2: For each sub-sample estimate $\hat{\gamma}_k$ and $\hat{\beta}_1$.

3: Define $\hat{\gamma}_{Opt}$ as the estimate $\hat{\gamma}_k$ obtained for the sub-sample which has the lowest value of $|\hat{\beta}_1|$.

In the next two Sections, the performance of the proposed estimation strategy is analysed on simulated and real data.

4. Numerical Comparisons

In this section we will evaluate the performance of $\hat{\gamma}_k$ with respect to some alternative estimators of the EV (or tail) index. As far as the estimator for γ is concerned, beyond considering the estimator $\hat{\gamma}_{Opt}$, the estimator $\hat{\gamma}_k$ with different levels of truncation of the data is considered. In the tables, $\hat{\gamma}_{1-p}$ indicates the estimator $\hat{\gamma}_k$, with $1-p$ indicating the fraction of upper order statistics used in estimation; the notation $\hat{\gamma}_{All}$ indicates the case where all the sample data are used in estimation.

Numerical comparisons will be carried out with respect to some reduced bias (RB) competitors (Caeiro et al. [11], Gomes et al. [12]) based on Hill (Hill [5]), generalized Hill (Beirlant et al. [24]), moment (Dekkers et al. [25]) and moment of order p (Gomes et al. [13]) estimators; optimized with respect to the choice of k as discussed in Gomes et al. [13].

RB estimation of γ for the above mentioned alternative estimators is based on external estimation of additional parameters (ρ, β) (refer to Gomes et al. [26] and Gomes et al. [13] for further details). In our comparisons the following RB-versions are used:

(1)   RB-Hill estimator, outperforming $H(k)$ (defined in (2)) for all k

$$\bar{H}(k) = H(k) \left(1 - \hat{\beta}(n/k)^{\hat{\rho}} / (1 - \hat{\rho}) \right). \tag{26}$$

(2)   RB-Moment estimator, denoted by MM in the tables,

$$\bar{M}(k) = M(k) \left(1 - \hat{\beta}(n/k)^{\hat{\rho}} / (1 - \hat{\rho}) \right) - \hat{\beta}\hat{\rho}(n/k)^{\hat{\rho}} / (1 - \hat{\rho})^2, \tag{27}$$

   with

$$M(k) = M_k^{(1)} + \frac{1}{2} \left[1 - (M_k^{(2)} / (M_k^{(1)})^2 - 1)^{-1} \right] \tag{28}$$

   and $M_k^{(j)} = \sum_{i=1}^{k} (\ln X_{(n-i+1)} - \ln X_{(n-k)})^j, j \geq 1$.

(3)   RB-Generalized Hill estimator, $\bar{GH}(k)$, denoted GH in the tables, with the same bias correction as in (27) applied to

$$GH(k) = \sum_{i=1}^{k} (\ln UH(j) - \ln UH(k)) \tag{29}$$

   with $UH(j) = X_{(n-j)} H(k)\, 1 \leq j \leq k$.

(4)   RB-MOP (moment of order p) estimator, for $0 < p < \alpha$ (the case $p = 0$ reduces to the Hill estimator) defined by

$$\bar{H}_p(k) = H_p(k) \left(1 - \frac{\hat{\beta}(1 - pH_p(k))}{1 - \hat{\rho} - pH_p(k)} \left(\frac{n}{k} \right)^{\hat{\rho}} \right), \tag{30}$$

   with $H_p(k) = (1 - A_p^{-p}(k))/p$, $A_p(k) = \left(\sum_{i=1}^{k} U_{ik}^p / k \right)^{1/p}$, $U_{ik} = X_{(n-i+1)} / X_{(n-k)}, 1 \leq i \leq k < n$. Denoted by MP_p in the tables. In this case p is a tuning parameter which will be set, in our simulations, equal to 0.5 and 1. For an estimated optimal value of p based on a preliminary estimator of α see Gomes et al. [13].

Computations of the above estimators have been performed using the package `evt0` (Manjunat and Caeiro [27]) in R. More precisely, $GH(k)$ and $M(k)$ are obtained using the function `other.EVI()` respectively with the options `GH` and `MO`. Estimation of the parameters (ρ, β) for the bias correction terms can be obtained from the function `mop()`. RB-Hill and RB-MOP estimates are directly obtained by the function `mop()` by appropriately specifying a value of p and the option `RB-MOP`. In order to optimize the choice of k we used the formula [13]

$$\hat{k} = \min \left(n - 1, \lfloor ((1 - \varphi(\hat{\rho}) - \hat{\rho})^2 n^{-2\hat{\rho}} / (-2\hat{\rho}\hat{\beta}^2 (1 - 2\varphi(\hat{\rho})))) ^{1/(1-2\hat{\rho})} \rfloor + 1 \right), \tag{31}$$

where $\lfloor x \rfloor$ is the integer part of x and $\varphi(\rho) = 1 - (\rho + \sqrt{\rho^2 - 4\rho + 2})/2$. For the comparisons, the following distributions are used:

(1) Pareto distribution, as defined in (7). Random numbers from this distribution are simply generated in R using the function `runif()` and inversion of F.

(2) Fréchet distribution with $F(x) = \exp(-x^{-\alpha})$, $x \geq 0$, denote by Fréchet(α). This distribution is simulated in R using the function `rfrechet()` from the package evd (Stephenson [28]) with shape parameter set equal to α.

(3) Burr distribution with $F(x) = 1 - (1 + x^{\alpha})^{-1}$, indicated with Burr($\alpha$). This distribution is simulated in R using the function `rburr()` from the package actuar (Dutang et al. [29]) with the parameter `shape1` set to 1 and `shape2` set equal to α.

(4) Symmetric stable distribution with index of stability α, $0 < \alpha < 2$, indicated with Stable(α):= Stable($\alpha, \beta = 0, \mu = 0, \sigma = 1$); where β, μ and σ indicate, respectively, asymmetry, location and scale. This distribution is simulated in R using the function `rstable()` from the package `stabledist` (Wuertz et al. [30]). For this distribution only the positive observed data are used in estimation.

Tables 1–8 contain the empirical RMSE (Root-MSE) and the relative RMSE, with respect to $\hat{\gamma}_{Opt}$, of the estimators, that is, for *any* of the evaluated estimators, say $\hat{\gamma}$, then

$$\text{RMSE}(\hat{\gamma}) = \sqrt{\hat{\text{E}}(\hat{\gamma} - \gamma)^2}, \quad \text{Rel-RMSE}(\hat{\gamma}) = \frac{\text{RMSE}(\hat{\gamma})}{\text{RMSE}(\hat{\gamma}_{Opt})}.$$

Note that a Rel-RMSE greater than one implies a worse performance of the estimator with respect to $\hat{\gamma}_{Opt}$. $\hat{\text{E}}$ denotes the empirical expected value, that is, the mean over the Montecarlo experiments. For each sample size $n = 50, 100, 200, 300, 500$, and 1000; 1000 Montecarlo replicates were generated. Computations have been carried out with R version 3.5.1 and each experiment, that is, given a chosen distribution and a chosen n, has been initialized using `set.seed(3)`. Numerical results representative for each distribution are reported in the tables. More tables with other choices of parameters can be found in the on-line Supplementary Materials accompanying this paper.

Table 1. RMSE of the estimators for the Pareto(4) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	$MP_{0.5}$	MP_1	*GH*	*MM*
50	0.051	0.040	0.047	0.054	0.066	0.294	0.258	0.228	2.660	0.893
100	0.038	0.031	0.036	0.042	0.052	0.245	0.198	0.171	0.925	4.882
300	0.026	0.018	0.021	0.025	0.032	0.284	0.253	0.227	5.657	0.734
500	0.023	0.014	0.017	0.020	0.025	0.250	0.226	0.205	2.852	0.631
1000	0.016	0.010	0.012	0.014	0.018	0.159	0.142	0.128	0.755	0.530

Table 2. Relative RMSE of the estimators for the Pareto(4) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	$MP_{0.5}$	MP_1	*GH*	*MM*
50	1.000	0.788	0.914	1.054	1.282	5.716	5.012	4.426	51.761	17.372
100	1.000	0.812	0.947	1.106	1.389	6.481	5.241	4.519	24.468	129.156
300	1.000	0.668	0.792	0.936	1.215	10.713	9.532	8.574	213.453	27.687
500	1.000	0.627	0.737	0.864	1.105	10.956	9.912	8.982	125.105	27.689
1000	1.000	0.624	0.745	0.879	1.140	10.121	9.057	8.178	48.096	33.771

Table 3. RMSE of the estimators for the Fréchet(1.5) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	*MP*$_{0.5}$	*MP*$_1$	*GH*	*MM*
50	0.114	0.098	0.095	0.116	0.153	0.186	0.185	0.174	0.183	16.691
100	0.091	0.099	0.081	0.092	0.119	0.141	0.142	0.136	0.138	518.904
300	0.079	0.098	0.067	0.068	0.084	0.095	0.097	0.100	0.094	15.324
500	0.073	0.098	0.060	0.057	0.068	0.075	0.077	0.082	0.075	18.848
1000	0.065	0.100	0.060	0.054	0.060	0.059	0.063	0.073	0.059	22.018

Table 4. Relative RMSE of the estimators for the Fréchet(1.5) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	*MP*$_{0.5}$	*MP*$_1$	*GH*	*MM*
50	1.000	0.855	0.832	1.011	1.335	1.625	1.620	1.522	1.603	145.897
100	1.000	1.086	0.889	1.012	1.312	1.547	1.566	1.495	1.514	5702.241
300	1.000	1.238	0.841	0.859	1.059	1.195	1.228	1.262	1.187	193.238
500	1.000	1.344	0.828	0.784	0.930	1.033	1.058	1.136	1.030	259.609
1000	1.000	1.540	0.931	0.832	0.927	0.915	0.971	1.131	0.918	339.784

Table 5. RMSE of the estimators for the Burr(2) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	*MP*$_{0.5}$	*MP*$_1$	*GH*	*MM*
50	0.111	0.228	0.130	0.106	0.112	0.125	0.124	0.119	0.105	7.446
100	0.100	0.229	0.126	0.097	0.096	0.114	0.113	0.110	0.104	6.218
300	0.073	0.228	0.120	0.084	0.071	0.084	0.084	0.087	0.080	1.778
500	0.066	0.227	0.118	0.080	0.063	0.071	0.071	0.076	0.068	1.266
1000	0.054	0.226	0.117	0.078	0.055	0.055	0.056	0.063	0.053	0.326

Table 6. Relative RMSE of the estimators for the Burr(2) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	*MP*$_{0.5}$	*MP*$_1$	*GH*	*MM*
50	1.000	2.053	1.171	0.951	1.011	1.128	1.112	1.069	0.946	66.962
100	1.000	2.294	1.258	0.971	0.966	1.141	1.134	1.099	1.043	62.240
300	1.000	3.099	1.635	1.149	0.971	1.144	1.150	1.180	1.089	24.226
500	1.000	3.440	1.794	1.222	0.953	1.076	1.083	1.152	1.036	19.215
1000	1.000	4.150	2.154	1.431	1.015	1.009	1.028	1.158	0.980	5.982

Table 7. RMSE of the estimators for the Stable(1.1) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	*MP*$_{0.5}$	*MP*$_1$	*GH*	*MM*
50	0.372	0.292	0.355	0.375	0.394	0.420	0.413	0.417	0.408	4.398
100	0.335	0.267	0.324	0.337	0.345	0.374	0.365	0.366	0.365	3.163
300	0.297	0.239	0.289	0.295	0.292	0.343	0.323	0.312	0.338	4.996
500	0.275	0.226	0.271	0.275	0.268	0.332	0.305	0.288	0.328	9.098
1000	0.251	0.211	0.252	0.252	0.241	0.325	0.292	0.263	0.321	3.132

Table 8. Relative RMSE of the estimators for the Stable(1.1) distribution; 1000 Montecarlo replications.

n	$\hat{\gamma}_{Opt}$	$\hat{\gamma}_{All}$	$\hat{\gamma}_{0.7}$	$\hat{\gamma}_{0.5}$	$\hat{\gamma}_{0.3}$	*Hill*	*MP*$_{0.5}$	*MP*$_1$	*GH*	*MM*
50	1.000	0.785	0.954	1.007	1.058	1.128	1.110	1.120	1.095	11.814
100	1.000	0.795	0.966	1.005	1.030	1.117	1.089	1.092	1.089	9.431
300	1.000	0.807	0.972	0.993	0.984	1.157	1.088	1.052	1.139	16.827
500	1.000	0.821	0.985	0.999	0.976	1.207	1.111	1.048	1.191	33.097
1000	1.000	0.842	1.003	1.006	0.963	1.294	1.167	1.048	1.282	12.492

Trying to summarize the results we note the general good performance of the estimators based on the curve λ defined in this paper for which the gain in efficiency can be substantial. We note also the

actual usefulness of $\hat{\gamma}_{Opt}$ for practical applications since it is able to individuate appropriate levels of truncation for different distributions although an actual knowledge of the optimal level of truncation would obtain higher efficiency.

Turning to the single cases, one can note that the $\hat{\gamma}_{Opt}$ outperforms all the other estimators for the Pareto distribution where relative efficiency (see Table 2), is always greater than 4. For the case of the Pareto distribution, $\hat{\gamma}_{All}$ would be the most efficient choice, as expected.

In the case of the Fréchet distribution $\hat{\gamma}_{Opt}$ is always more efficient than all competitors test for smaller sample sizes (see Table 4); as sample size increases the gain in efficiency decreases and maybe slightly lower in some cases.

The performance of $\hat{\gamma}_{Opt}$ in the case of the Burr distribution is comparable to that of the competitors, with relative RMSE (see Table 6) slightly smaller or greater than one depending on the case considered.

In the case of the Symmetric stable distribution, the performance of $\hat{\gamma}_{Opt}$ is slightly better than all alternative estimators in all cases (see Table 8). The *MM* estimator turns out to be quite efficient for the stable distribution with α closer to 2 (see the on-line Supplementary Materials).

We note that the *MM* and *GH* estimators, computed with the package evt0, has shown some illogical results in some instances with extremely high values of the RMSE, typically for some specific sample sizes, after several checks, we could not figure out the reason of such results.

5. Examples

Here we concentrate on six real data examples that have been used in the literature to discuss methods to detect a power-law in the tail of the underlying distribution. These data have all been thoroughly analysed, for example, in Clauset et al. [1]. The following data sets are analysed here:

1. The frequency of occurrence of unique words in the novel Moby Dick by Herman Melville (Newman [31]).
2. The severity of terrorist attacks worldwide from February 1968 to June 2006, measured as the number of deaths directly resulting (Clauset et al. [32]).
3. The sizes in acres of wildfires occurring on U.S. federal land between 1986 and 1996 (Newman [31]).
4. The intensities of earthquakes occurring in California between 1910 and 1992, measured as the maximum amplitude of motion during the quake (Newman [31]).
5. The frequencies of occurrence of U.S. family names in the 1990 U.S. Census (Clauset et al. [1]).
6. Peak gamma-ray intensity of solar flares between 1980 and 1989 (Newman [31]).

Figure 3 provides the estimated λ curves for the six examples, either considering the whole data and selected percentages of the upper order statistics. The range of λ may vary in the graphs in order to have a better detail of the path of the curves.

On each of the data-set we apply Algorithm 2 in order to select the optimal number of k in computing $\hat{\gamma}_{Opt}$; with the given estimate we apply Algorithm 1 in order to compute a 95% confidence interval for the estimate.

Next we apply a testing procedure to evaluate if the graphs in Figure 3, for the k chosen by Algorithm 1, can be considered "enough flat" in order to support the hypothesis that the data come from a distribution within the class (1). A bootstrap test setting $H_0 : \beta_1 = 0$ in model (24) has been developed in Taufer et al. [33].

For comparison we apply also the testing procedure for the power-law hypothesis developed by Clauset et al. [1].

Table 9 reports analytical results on estimated values, 95% confidence intervals, the fraction of upper order statistics used and the p-values of the testing procedures.

Trying to summarize briefly the results we would say that the conclusions about the presence of a Pareto-type tail in the distributions coincide fully with the conclusions of Clauset et al. [1], that is: clear evidence of a power law distribution fitting the data is for the Moby Dick and Terrorism data-sets.

For the others there is no convincing evidence. We point out that for the contrasting p-values for the Solar Flares data, Clauset et al. [1] suggest a power tail with an exponential cut-off at a certain point. Given the characteristics of the graphs based on the λ curve this feature cannot be noticed in our analysis.

As far as the estimated values of γ, the values of the estimators obtained here are substantially lower with respect to those obtained by Clauset et al. [1] (which uses the Hill estimator). Given the good performance in the simulations of $\hat{\gamma}_{Opt}$ in comparison to the Hill estimator, the values in Table 9, at least for the Moby Dick and Terrorism data-set can be considered reliable.
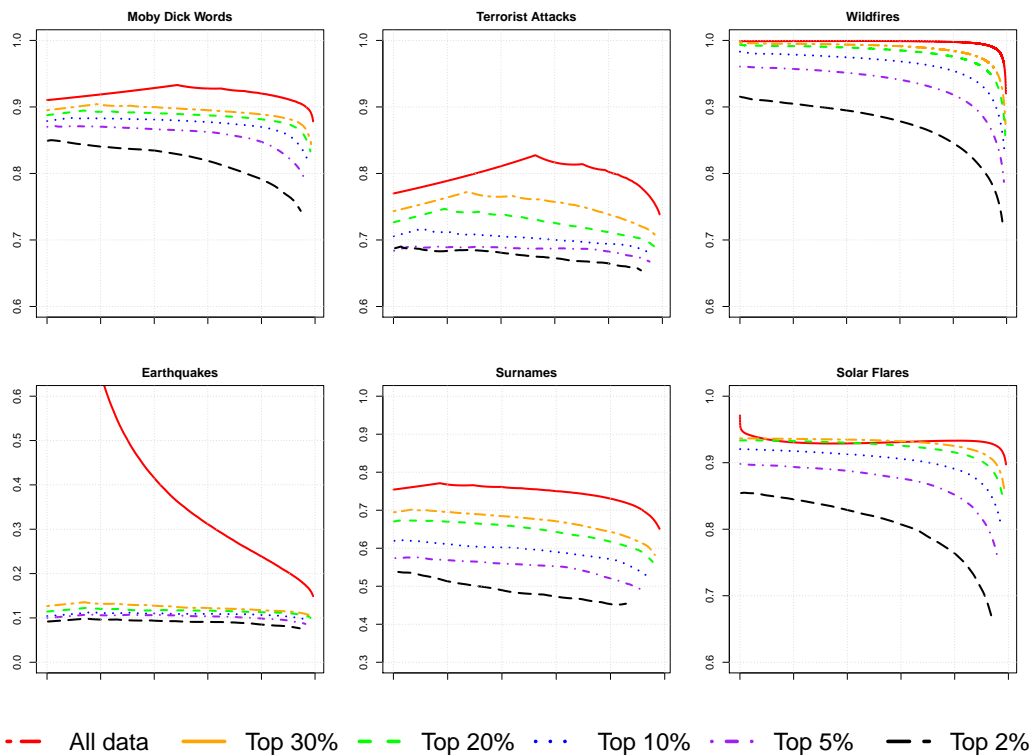
Figure 3. Plot of the estimated λ curves for the six dataset: all data and selected percentages of upper order statistics.

For the other data-sets, since the null hypotheses of a power law has significant p-values, the estimated γ should be discarded and it becomes of interest to select an alternative model by using, for example a likelihood ratio test as discussed in Clauset et al. [1] to which the interested reader is referred.

Table 9. Sample size, estimated γ and 95% confidence intervals for the six data-sets. Fraction of upper order statistics used $(1 - p)$ and *p*-values of the testing procedures defined in Taufer et al. [33] (Sig[1]) and Clauset et al. [1] (Sig[2]). Asterisk indicates significant p-values.

Dataset	n	$\hat{\gamma}$	0.95-CI	$1 - p$	Sig[1]	Sig[2]
Moby Dick	18,855	0.90	(0.80–0.95)	0.4	0.224	0.49
Terrorism	9101	0.80	(0.73–0.87)	1.0	0.184	0.68
Wildfires	203,785	0.99	(0.90–0.99)	1.0	0.012 *	0.05 *
Earthquakes	19,302	0.22	(0.21–0.23)	0.5	0.000 *	0.00 *
Surnames	2753	0.74	(0.67–0.84)	1.0	0.000 *	0.00 *
Solar flares	12,773	0.96	(0.81–0.97)	0.2	0.038 *	1.00

6. Conclusions

An estimation strategy for the tail index of a distribution in class (1) has been defined starting from a characterizing property of Zenga's inequality curve λ. On the basis of the theoretical properties of the estimator $\hat{\gamma}_k$ two simple bootstrap procedures have been obtained: the first provides a general result for the asymptotic distribution of $\hat{\gamma}_k$ and the second gives a data-driven procedure to determine the optimal value of k. Simulations show the good performance of $\hat{\gamma}_k$ and the implementation algorithm.

The data-driven optimized estimator often outperforms optimized (with respect to bias) competing estimation strategies. The gain in efficiency is substantial in the case of Pareto distributions.

The graph of the λ curve associated with the estimator provides a valid support in the analysis of real data.

7. Proofs

Proof of Lemma 1. It is trivially verified that if F satisfies (7) then $\lambda(p) = 1/\alpha$. Suppose now that $\lambda(p) = k$, $p \in (0, 1)$, where k is some constant. Then it must hold that $1 - L(p) = (1 - p)^k$ or equivalently, after some algebraic manipulation,

$$\int_0^p F^{-1}(u)du = \mu[1 - (1 - p)^k] \tag{32}$$

Taking derivatives on both sides we have that

$$\frac{d}{dp} \int_0^p F^{-1}(u)du = \frac{d}{dp}\mu[1 - (1 - p)^k], \tag{33}$$

which gets

$$F^{-1}(p) = \mu k(1 - p)^{k-1}$$

from which, setting $x_p = F^{-1}(p)$, which implies $p = F(x_p)$, it follows that, after some further elementary manipulations,

$$\left(\frac{x_p}{\mu k}\right)^{1/(k-1)} = 1 - F(x_p).$$

Setting $1/(k - 1) = -\alpha$, properly normalized, the above F follows (7). $\square$

Proof of Lemma 2. Let G and g denote respectively the distribution function and the density of $Y = X|X > s$ and note that $G(y) = P(Y \leq y) = \frac{F(y) - F(s)}{\overline{F}(s)}$ and $g(y) = f(y)/\overline{F}(s)$. Setting $G(y) = p$ and inverting we have $G^{-1}(p) = F^{-1}(F(s) + p\overline{F}(s))$. Note that

$$E(Y) = E(X|X > s) = \frac{E(X\mathbb{I}_{(X>s)})}{\overline{F}(s)}$$

also, setting $s = F^{-1}(p)$,

$$E(X\mathbb{I}_{(X>s)}) = \int_s^\infty xf(x)dx = \int_p^1 F^{-1}(u)du$$

Consider the Lorenz curve for Y, $L(p)$ and $\overline{L}(p) = 1 - L(p)$, let $t > 0$ and $s = G^{-1}(p)$. Then

$$\overline{L}(p) = \frac{E(Y\mathbb{I}_{(Y>G^{-1}(p))})}{E(Y)} = \frac{E(X\mathbb{I}_{(X>F^{-1}(F(s)+p\overline{F}(s)))})}{E(X|X > s)}.$$

Consider first the numerator of the above expression

$$E(Y\mathbb{I}_{(Y>G^{-1}(p))}) = \int_p^1 G^{-1}(u)du$$

$$= \int_p^1 F^{-1}(F(s) + u\bar{F}(s))du$$

$$= \frac{1}{\bar{F}(s)} \int_{F(s)+p\bar{F}(s)}^1 F^{-1}(t)dt$$

after setting $t = F(s) + u\bar{F}(s)$. Next, to link to the function $U(w) = F^{-1}(1 - 1/w)$, set $t = 1 - 1/w$; the above term, as $s \to \infty$, by Karamata's theorem (see De Haan and Ferreira [34], p. 363), becomes

$$\frac{1}{\bar{F}(s)} \int_{[(1-p)\bar{F}(s)]^{-1}}^{\infty} U(w)\frac{1}{w^2}dw$$

$$= (1 - \gamma)^{-1}(1 - p)^{1-\gamma}\bar{F}(s)^{-\gamma}H_U([(1 - p)\bar{F}(s)]^{-1}), \quad (34)$$

since as $s \to \infty$, $\bar{F}(s) \to 0$. H_U is a slowly varying function. Next consider the denominator; similar computations bring to

$$E(X|X > s) = \int_0^1 G^{-1}(u)du$$

$$= \int_0^1 F^{-1}(F(s) + u\bar{F}(s))du \qquad (35)$$

$$= \frac{1}{\bar{F}(s)} \int_{[\bar{F}(s)]^{-1}}^{\infty} U(w)\frac{1}{w^2}dw$$

which, as $s \to \infty$, converges to

$$(1 - \gamma)^{-1}\bar{F}(s)^{-\gamma}H_U([\bar{F}(s)]^{-1}).$$

Finally, putting together the results one has

$$\bar{L}(p) = (1 - p)^{1-\gamma}, \quad s \to \infty \qquad (36)$$

since $\frac{L_U([(1-p)\bar{F}(s)]^{-1})}{L_U([\bar{F}(s)]^{-1})} \to 1$ as $s \to \infty$, by the properties of slowly varying functions. □

Proof of Lemma 3. Note that $\lim_{n\to\infty} \sup_{p\in(0,1)} |L_n(p) - L(p)| = o_P(1)$ and that p lies in a compact interval. $\lambda_n(p)$ is continuous transformation of $L_n(p)$; it follows that for fixed $p \in (0,1)$, $\lim_{n\to\infty} |\lambda_n(p) - \lambda(p)| = o_P(1)$.

To prove uniform consistency of $\lambda_n(p)$ we need to show it is equicontinuous. For this note that $\lambda_n(p)$ depends on p stochastically only through $L_n(p)$, which is uniformly continuous. Hence for any $\delta > 0$ such that $|p_1 - p_2| < \delta$, it is possible to find an n_0, not depending on p_1, p_2 such that for $n > n_0$, $\varepsilon, \eta > 0$, one has

$$P(|\lambda_n(p_1) - \lambda_n(p_2))| > \varepsilon) < \eta.$$

□

Proof of Lemma 4. We have

$$\lim_{n\to\infty} |\gamma_n - \gamma| = \lim_{n\to\infty} \left| \frac{1}{k} \sum_{i=1}^{k} \hat{\lambda}(p_i) - \gamma \right|$$

$$\lim_{n\to\infty} \le \left| \frac{1}{k} \sum_{i=1}^{k} (\lambda_n(p_i) - \lambda(p_i)) \right| + \left| \frac{1}{k} \sum_{i=1}^{k} \lambda(p_i) - \gamma \right|$$

$$\le \lim_{n\to\infty} \sup_{p\in(0,1)} |\hat{\lambda}(p) - \lambda(p)| + \left| \frac{1}{k} \sum_{i=1}^{k} \frac{\log\left[L_U([(1-p)\overline{F}(s)]^{-1}) / L_U([\overline{F}(s)]^{-1}) \right]}{\log(1-p_i)} \right|$$

$$= o_P(1).$$

by using Lemma 3 and condition (17) with $k = n^\delta$, $0 < \delta < -\frac{\rho}{1-\rho}$. □

References

1. Clauset, A.; Shalizi, C.R.; Newman, M.E. Power-law distributions in empirical data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]
2. Jenkins, S.P. Pareto Distributions, Top Incomes, and Recent Trends in UK Income Inequality. *Economica* **2016**. Available online: https://www.econstor.eu/bitstream/10419/145258/1/dp10124.pdf (accessed on 30 September 2020).
3. Hlasny, V. *Parametric Representation of the Top of Income Distributions: Options, Historical Evidence and Model Selection (No. 90)*; Department of Economics, Tulane University: New Orleans, LA, USA, 2020.
4. Heyde, C.C.; Kou, S.G. On the controversy over tailweight of distributions. *Oper. Res. Lett.* **2004**, *32*, 399–408. [CrossRef]
5. Hill, B.M. A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **1975**, *3*, 1163–1174. [CrossRef]
6. Gomes, M.I.; Guillou, A. Extreme value theory and statistics of univariate extremes: A review. *Int. Stat. Rev.* **2015**, *83*, 263–292. [CrossRef]
7. Brilhante, M.F.; Gomes, M.I.; Pestana, D. A simple generalisation of the Hill estimator. *Comput. Stat. Data Anal.* **2013**, *57*, 518–535. [CrossRef]
8. Beran, J.; Schell, D.; Stehlík, M. The harmonic moment tail index estimator: Asymptotic distribution and robustness. *Ann. Inst. Stat. Math.* **2014**, *66*, 193–220. [CrossRef]
9. Paulauskas, V.; Vaičiulis, M. A class of new tail index estimators. *Ann. Inst. Stat. Math.* **2017**, *69*, 461–487. [CrossRef]
10. Paulauskas, V.; Vaičiulis, M. On an improvement of Hill and some other estimators. *Lith. Math. J.* **2013**, *53*, 336–355. [CrossRef]
11. Caeiro, F.; Gomes, M.I.; Pestana, D. Direct reduction of bias of the classical Hill estimator. *Revstat* **2005**, *3*, 111–136.
12. Gomes, M.I.; de Haan, L.; Henriques-Rodrigues, L. Tail index estimation for heavy tailed models: Accomodation of bias in weighted log-excesses. *J. R. Stat. Soc. B* **2008**, *70*, 31–52.
13. Gomes, M.I.; Brilhante, M.F.; Pestana, D. New reduced-bias estimators of a positive extreme value index. *Commun. Stat.-Simul. Comput.* **2016**, *45*, 833–862. [CrossRef]

14. Zenga, M. Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito. *Giornale degli Economisti e Annali di Economia* **1984**, *5*, 301–326.

15. Kratz, M.F.; Resnick, S.I. The QQ-estimator and heavy tails. *Commun. Stat. Stoch. Model.* **1996**, *12*, 699–724.

16. Grahovac, D.; Jia, M.; Leonenko, N.N.; Taufer, E. Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data. *Stat. J. Theor. Appl. Stat.* **2015**, *49*, 1221–1242. [CrossRef]

17. Jia, M.; Taufer, E.; Dickson, M. Semi-parametric regression estimation of the tail index. *Electron. J. Stat.* **2018**, *12*, 224–248. [CrossRef]

18. Zenga, M. Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Stat. Appl.* **2007**, *1*, 3–27. [CrossRef]

19. Zenga, M. Concentration curves and Concentration indexes derived from them. In *Income and Wealth Distribution, Inequality and Poverty*; Dagum, C., Zenga, M., Eds.; Springer: Berlin/Heidelberg, Germany, 1990; pp. 94–110.

20. Arcagni A.; Porro F. A comparison of income distributions models through inequality curves. *Stat. Appl.* **2016**, *XIV*, 123–144.

21. Goldie, C.M. Convergence theorems for empirical Lorenz curves and their inverses. *Adv. Appl. Probab.* **1977**, *9*, 765–791.

22. Csorgo, S.; Mason, D.M. The asymptotic distribution of sums of extreme values from a regularly varying distribution. *Ann. Probab.* **1986**, *14*, 974–983. [CrossRef]

23. Mead, M.E. Generalized Inverse Gamma Distribution and its Application in Reliability. *Commun. Stat. Theory Methods* **2015**, *44*, 1426–1435. [CrossRef]

24. Beirlant, J.; Vynckier, P.; Teugels, J.L. Excess functions and estimation of the extreme-value index. *Bernoulli* **1996**, *2*, 293–318. [CrossRef]

25. Dekkers, A.L.M.; Einmahl, J.H.J.; de Haan, L. A moment estimator for the index of an extreme-value distribution. *Ann. Stat.* **1989**, *17*, 1833–1855. [CrossRef]

26. Gomes, M.I.; Figueiredo, F.; Neves, M.M. Adaptive estimation of heavy right tails: Resampling-based methods in action. *Extremes* **2012**, *15*, 463–489. [CrossRef]

27. Manjunath, B.G.; Caeiro, F. evt0:Mean of Order p, Peaks over Random Threshold Hill and High Quantile Estimates R Package Version 1.1-3. Available online: https://cran.r-project.org/web/packages/evt0 (accessed on 31 August 2020) [CrossRef]

28. Stephenson, A.G. evd: Extreme Value Distributions. *R News* **2002**, *2*, 31–32. Available online: https://CRAN.R-project.org/doc/Rnews/ (accessed on 31 August 2020).

29. Dutang, C.; Goulet, V.; Pigeon, M. actuar: An R Package for Actuarial Science. *J. Stat. Softw.* **2008**, *25*, 1–37. Available online: http://www.jstatsoft.org/v25/i07 (accessed on 31 August 2020).

30. Wuertz, D.; Maechler, M.; Rmetrics Core Team Members. Stabledist: Stable Distribution Functions. R Package Version 0.7-1. 2016. Available online: https://CRAN.R-project.org/package=stabledist (accessed on 31 August 2020).

31. Newman, M.E. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351.

32. Clauset, A.; Young, M.; Gleditsch, K.S. On the frequency of severe terrorist events. *J. Confl. Resolut.* **2007**, *51*, 58–87. [CrossRef]

33. Taufer, E.; Santi, F.; Espa, G.; Dickson, M.M. Goodness-of-fit test for Pareto and Log-normal distributions based on inequality curves. 2020, Unpublished. [CrossRef]

34. De Haan, L.; Ferreira, A. *Extreme Value Theory: An Introduction*; Springer Science & Business Media: New York, NY, USA, 2007.