Electronic Journal of Statistics
Vol. 12 (2018) 224–248
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1394

Semi-parametric regression estimation of the tail index

Mofei Jia

International Business School Suzhou, Xi'an Jiaotong-Liverpool University, China
e-mail: mofei.jia@xjtlu.edu.cn

Emanuele Taufer and Maria Michela Dickson

Department of Economics and Management, University of Trento, Italy
e-mail: emanuele.taufer@unitn.it; mariamichela.dickson@unitn.it

Abstract: Consider a distribution F with regularly varying tails of index $-\alpha$. An estimation strategy for α, exploiting the relation between the behavior of the tail at infinity and of the characteristic function at the origin, is proposed. A semi-parametric regression model does the job: a nonparametric component controls the bias and a parametric one produces the actual estimate. Implementation of the estimation strategy is quite simple as it can rely on standard software packages for generalized additive models. A generalized cross validation procedure is suggested in order to handle the bias-variance trade-off. Theoretical properties of the proposed method are derived and simulations show the performance of this estimator in a wide range of cases. An application to data sets on city sizes, facing the debated issue of distinguishing Pareto-type tails from Log-normal tails, illustrates how the proposed method works in practice.

MSC 2010 subject classifications: Primary, 62G32; secondary 62J05.
Keywords and phrases: Tail index, heavy-tailed distributions, regular variation, empirical characteristic function, Zipf's law.

Received May 2017.

1. Introduction

Consider a random sample $X_1, X_2, \ldots, X_n$ with a distribution function F satisfying

$$\bar{F}(x) = x^{-\alpha} L(x), \quad x \to \infty, \tag{1}$$

where $\bar{F} = 1 - F$ and $L(x)$ is a slowly varying function, satisfying $L(tx)/L(x) \to 1$ as $x \to \infty$, for any $t > 0$. We will also say that $\bar{F} = 1 - F$ is regularly varying (RV) at infinity with index $-\alpha$, denoted as $\bar{F} \in RV_{-\alpha}$. The parameter $\alpha > 0$ is usually referred to as the tail index; alternatively, in the extreme value (EV) literature it is typical to refer to the EV index $\gamma > 0$ with $\alpha = 1/\gamma$.

The paper discusses an estimator of the tail index α which relies on the relationship between the tail of a distribution and the slope of the real part of the characteristic function (RCF) at the origin. This idea places itself out of the mainstream approaches for tail estimation, usually based on functions

of upper order statistics; moreover, the proposed method exploits all sample information in analyzing the slope of the empirical RCF (ERCF) at the origin, which, typically, has lower variability with respect to the largest order statistics. The estimation procedure proposed can be implemented exploiting standard regression analysis packages such as LM or GAM in R software and their tools in order to automatically select estimation parameters. Also, a full regression analysis of the ERCF can be exploited to better understand the phenomenon under scrutiny. All these issues will be discussed in turn; before doing so, a connection with previous relevant literature is given.

Probably the most well-known estimator of the tail index is the Hill [26] estimator, which exploits the k upper order statistics through the formula

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

where $X_{i:n}$ is the i-th order statistics from a sample of size n and $k = k(n) \to \infty$ in an appropriate way. The Hill estimator may suffer from high bias and is heavily dependent on the choice of k. It has been thoroughly studied [e.g., 32, 23, 22, 27, 40, 10].

Several generalizations of the Hill estimator and new approaches have appeared in the literature: for example Csorgo et al [8] present a general class of kernel estimators where the Hill estimator can be obtained by taking a uniform kernel; the family of j-moment ratio estimators [9], the smoothing Hill estimator [41], and the location invariant estimator [14]. Beirlant et al [2] propose the generalized Hill estimator; Brilhante et al [5] define a *moment of order p* estimator which reduces to the Hill estimator for $p = 0$; Beran et al [4] defines a *harmonic moment* tail index estimator. Recently Paulauskas and Vaičiulis [35] and Paulauskas and Vaičiulis [36] have connected in an interesting way some of the above approaches by defining parametric families of functions of the order statistics.

Other notable contributions are those of Pickands III [37], which proposed a simple estimator based on a linear combination of log-spacing order statistics, or the moment estimator of Dekkers et al [11]; estimators based on the growth rate of diverging statistics have been proposed by Meerschaert and Scheffler [34], Politis [39], McElroy and Politis [33]. Kratz and Resnick [29] exploit properties of the QQ-plot while a recent approach based on the asymptotic properties of the partition function, a moment statistic generally employed in the analysis of multifractality, has been introduced by Grahovac et al [21].

Reduced bias (RB) and optimized (for the choice of k) approaches are discussed in Caeiro et al [6], Gomes et al [18] and Gomes et al [20]; these approaches are particularly relevant for this paper where a RB and optimized estimator will be discussed. For further discussion and references one can consult Beirlant et al [3] and Gomes and Guillou [19].

Another reference relevant to our work is Welsh [43], which defines an estimator based on the ERCF evaluated at two points. This paper improves that approach in several ways: by defining a full regression analysis of the ERCF, by

providing a bias reduction component and an optimized choice of the points at which to evaluate the ERCF at the origin.

The paper is organized as follows: Section 2 presents the basic estimation strategy and analyzes its properties. Section 3 discusses an optimized and RB estimator and in Section 4 numerical comparisons with alternative RB and optimized estimators are presented. Section 5 gives some examples of applications to simulated and real data; an appendix contains the proofs of the results.

2. A simple regression estimator

2.1. Estimation strategy

The analysis of the tails of F will be based on the tail sum

$$H(x) = F(-x) + \bar{F}(x), \quad x > 0; \tag{3}$$

if $F(-x) \in RV_{-\alpha}$ and $\bar{F}(x) \in RV_{-\alpha}$, for example as in symmetric α-stable or t distributions, then $H(x) \in RV_{-\alpha}$. The same behavior of $H(x)$ will hold if F has a lighter left or right tail, e.g. $F(-x) \in RV_{-\beta}$ with $\beta > \alpha$, or exponentially decreasing or truncated. In all these cases, the estimator proposed will target α; the approach discussed will also allow to treat each tail separately.

Let ϕ denote the CF of X, i.e., for all real t, $\phi(t) = E[e^{itX}]$. Furthermore, let

$$\phi(t) = U(t) + iV(t), \tag{4}$$

where $U(t) = E[\cos(tX)]$ and $V(t) = E[\sin(tX)]$ are the real and imaginary parts of the CF, respectively. Integrating (4) by parts, one has

$$1 - U(t) = t \int_0^\infty \sin(tx) H(x)\, \mathrm{d}x, \tag{5}$$

that is, the RCF $U(t)$ depends only on the tail sum $H(x)$. Moreover if $H(x) \in RV_{-\alpha}$, $\alpha > 0$, from Pitman [38], it holds that, as $t \to 0$,

$$1 - U(t) \sim \begin{cases} s(\alpha) t^\alpha L(1/t), & \text{for} \quad 0 < \alpha < 2 \\ t^2 L_1(1/t), & \text{for} \quad \alpha = 2 \\ \frac{1}{2}\mu_2 t^2, & \text{for} \quad \alpha > 2 \end{cases} \tag{6}$$

where $\mu_2 = E(X^2)$ and $L_1(1/t) = \int_0^{1/t} xH(x)dx$ and $s(\alpha) = \pi/2[\Gamma(\alpha) \times \sin(\alpha\pi/2)]^{-1}$ if $\alpha > 0$ and $s(\alpha) = 1$ if $\alpha = 0$; $s(\alpha)$ is finite for any α which is not an even positive integer [38].

Remark 2.1. *We exploit relationship* (6) *to develop an estimator for* $0 < \alpha \leq 2$. *Note however that this restriction can be removed: if* X *has a distribution satisfying* (1) *with* $0 < \alpha \leq 4$, *then* $P(X^2 > x) = P(|X| > \sqrt{x}) = L(\sqrt{x})x^{-\alpha/2}$ *as* $x \to \infty$ *and relationship* (6) *can still be used to estimate the tail index for* $0 < \alpha \leq 4$. *In principle* (6) *can be used to estimate any* $\alpha > 0$ *although in practice the case* $0 < \alpha \leq 4$ *covers most situations of interest. Examples for these cases will be discussed below.*

Using an empirical version of $U(t)$, i.e. $U_n(t) = n^{-1}\sum_{j=1}^{n}\cos(tX_j)$ and evaluating $U_n(t)$ at points $t_1, t_2, \ldots, t_m$ (to be defined later), we use the regression equation

$$\log[(1 - U_n(t_j)] \sim \log[g(\alpha, t_j)] + \alpha \log t_j + \varepsilon_j \quad 0 < \alpha \leq 2, \tag{7}$$

where

$$\varepsilon_j = \log \frac{1 - U_n(t_j)}{1 - U(t_j)} \tag{8}$$

and, from (6), $g(\alpha, t) = s(\alpha)L(1/t)$ if $0 < \alpha < 2$ and $g(\alpha, t) = L_1(1/t)$ if $\alpha = 2$. By ordinary least squares (OLS), a simple estimator for α is

$$\hat{\alpha} = \frac{\sum_{j=1}^{m}(z_j - \bar{z})y_j}{S_{zz}}, \tag{9}$$

where $y_j = \log[(1 - U_n(t_j)]$, $z_j - \bar{z} = \log t_j - \frac{1}{m}\sum_{k=1}^{m}\log t_k$, and $S_{zz} = \sum_{k=1}^{m}(z_k - \bar{z})^2$. Note that, given (6), one should discard values of $\hat{\alpha}$ greater than 2. However we will retain the estimator unrestricted in order to discuss its properties. Application caveats will be discussed in Section 3 and Section 5. Also, $\hat{\alpha}$ implies that $g(\alpha, t)$ is treated as a constant even if it depends on t; this may actually be a source of non-negligible bias. A bias reduced version of $\hat{\alpha}$ will be developed in Section 3.

The estimator (9) for $m = 2$ corresponds to the ratio estimator discussed by Welsh (1986); as noted in the introduction, the approach discussed in this paper can be seen as an extension of that work. It will be seen that by adding more observations and taking care of the bias introduced by the fictitious intercept $g(\alpha, t)$ will lead to several advantages.

Remark 2.2. *Our method, in the context of i.i.d. data, is analogous to that used by Geweke and Porter-Hudak [16] in the estimation of long-memory time series models. In that paper, the relation between the spectral density function at the origin and the covariance function at infinity is exploited to estimate the long-memory parameter of the series. See also Robinson [42] and Hassler et al [25].*

2.2. Properties of $\hat{\alpha}$

In order to understand the behavior of $\hat{\alpha}$ we need to analyze in more detail the error $\varepsilon_j = \log[1 + \frac{U(t_j) - U_n(t_j)}{1 - U(t_j)}]$. Consider first the term

$$\varepsilon'_j = \frac{U(t_j) - U_n(t_j)}{1 - U(t_j)}; \tag{10}$$

noting that $Cov(U_n(t_i), U_n(t_j)) = (2n)^{-1}\{U(t_i + t_j) + U(t_i - t_j) - 2U(t_i)U(t_j)\}$ and using (6), as $t_j \downarrow 0$, we have (see also Welsh [43]),

$$Var(\varepsilon'_j) = \begin{cases} \frac{1}{nt_j{}^{\alpha}L(1/t_j)}\frac{(2-2^{\alpha-1})}{s(\alpha)}(1 + o(1)) + O(\frac{1}{n}), & \text{for} \quad 0 < \alpha < 2 \\ O(\frac{1}{n}) & \text{for} \quad \alpha \geq 2. \end{cases} \tag{11}$$

To make sure $Var(\varepsilon'_j) < \infty \ \forall j$ and $0 < \alpha \leq 2$, nt_j^α needs to diverge as $n \to \infty$ for all t_j's. A way to evaluate the ERCF at the origin is to choose $t_j = j/n^\eta$ with $j = 1, 2, ..., m$, $m = \lfloor n^\delta \rfloor$ for some $0 < \delta < \eta < 1$. To make $nt_j^\alpha \to \infty$ hold as $n \to \infty$ for all $0 < \alpha < 2$, the condition $0 < \delta < \eta \leq 1/2$ is necessary. Moreover, to have as many points as possible in the regression, $t_j = j/\sqrt{n}$ with $j = 1, 2, ..., m$, $m = \lfloor n^\delta \rfloor$ and $0 < \delta < 1/2$ seems a sensible choice. Thus, a Taylor expansion for $\varepsilon_j = \log[1 + \varepsilon'_j] = \varepsilon'_j - \frac{(\varepsilon'_j)^2}{2} + \ldots$, for $0 < \alpha < 2$ and $nt_j^\alpha \to \infty$, leads to the expression (see also Welsh [43])

$$\varepsilon_j = \varepsilon'_j + o_p((nt_j^\alpha)^{-\frac{1}{2}}). \tag{12}$$

Similarly, let $\lambda_{ij} = t_j/t_i$ $t_i > t_j$, then for $t_i \to 0$,

$$Cov(\varepsilon'_i, \varepsilon'_j) = \begin{cases} \frac{1}{nt_j{}^\alpha L(1/t_j)} \frac{(2+2\lambda_{ij}^\alpha - (1-\lambda_{ij})^\alpha - (1+\lambda_{ij})^\alpha)}{2s(\alpha)}(1 + o(1)) + O(\frac{1}{n}), & \text{for} \quad 0 < \alpha < 2 \\ O(\frac{1}{n}) & \text{for} \quad \alpha \geq 2. \end{cases} \tag{13}$$

The above results can be exploited in order to calculate bias and variance of $\hat{\alpha}$, which are summarized in the following theorem.

Theorem 2.3. *If $H(x) \in RV_{-\alpha}$ then, for $t_j = j/\sqrt{n}$, $j = 1, \ldots, m$, $m = \lfloor n^\delta \rfloor$, $0 < \delta < 1/2$, as $n \to \infty$,*

$$E(\hat{\alpha} - \alpha) = o(n^{\frac{\alpha - 2 - 2\alpha\delta}{4}}) + O(g_1(\alpha, n)) \quad \textit{for} \quad 0 < \alpha \leq 2, \tag{14}$$

$$Var(\hat{\alpha}) = \begin{cases} \frac{2}{s(1)} \frac{n^{-1/2}}{mL(\sqrt{n}/m)}\{1 + o(1)\}, & \textit{if } \alpha = 1, \\ \frac{2(A_\alpha - A_\alpha C_\alpha)}{s(\alpha)} \frac{n^{\alpha/2-1}}{m^\alpha L(\sqrt{n}/m)}\{1 + o(1)\}, & \textit{if } 0 < \alpha < 2 \ \& \ \alpha \neq 1, \\ O(1/n) + O((nm^2)^{-1} log^2 m), & \textit{for} \quad \alpha \geq 2, \end{cases} \tag{15}$$

where A_α and C_α are finite constants (details in the appendix) and

$$g_1(\alpha, n) = \begin{cases} s(\alpha)\frac{1}{S_{zz}} \sum_{j=1}^m (z_j - \bar{z}) L(1/t_j), & \textit{for} \quad 0 < \alpha < 2, \\ s(\alpha)\frac{1}{S_{zz}} \sum_{j=1}^m (z_j - \bar{z}) L_1(1/t_j), & \textit{for} \quad \alpha = 2, \\ 0, & \textit{for} \quad \alpha > 2. \end{cases} \tag{16}$$

Note that from Lemma 7.3 in the Appendix, $S_{zz} = m + O(log^2 m)$; since $L(x) \in RV_0$, then for some $0 < c < \delta$, $\lim_{n\to\infty} L(\sqrt{n}/j)/n^c = 0 \ \forall j \in \{1, 2, \ldots, \lfloor n^\delta \rfloor\}$. Hence, a rough evaluation of bias and variance from Theorem 2.3 allows to see that $\hat{\alpha}$ provides a consistent estimator of $\alpha \in (0, 2]$ for any choice of $m = \lfloor n^\delta \rfloor$ and any slowly varying function L.

A more precise statement can be made by placing further conditions on $L(x)$. A standard assumption in the context of tail index estimation is that (see Hall [23], Hall and Welsh [24])

$$L(x) = C[1 + Dx^{-\beta} + o(x^{-\beta})], \quad as \quad x \to \infty, \tag{A1}$$

where $C > 0, \beta > 0$ and D is a non-zero real number. With this, and restriction to the case $0 < \alpha < 2$, it is possible to obtain exact rates of convergence of $\hat{\alpha}$ and its asymptotic normality. Define $\theta = \min(\alpha + \beta, 2)$, Theorems 2.4 to 2.6 give results on $\hat{\alpha}$ under (*A*1).

Theorem 2.4. *Suppose $H(x) \in RV_{-\alpha}$, $0 < \alpha < 2$, (A1) holds, and $\hat{\alpha}$ is defined in (9) with $t_j = j/\sqrt{n}$ for $j = 1, 2, ..., m$, $m = \lfloor n^{\delta} \rfloor$ with $0 < \delta < 1/2$. Then as $n \to \infty$,*

$$E(\hat{\alpha}) - \alpha = O\left(n^{(\delta - 1/2)(\theta - \alpha)}\right) + o\left(n^{\frac{\alpha - 2 - 2\alpha\delta}{4}}\right).$$

Since $\theta > \alpha$ and $0 < \alpha < 2$, the bias goes to zero as $n \to \infty$ regardless the value of δ. The small o term may dominate the bias for small δ values.

Theorem 2.5. *Suppose $H(x) \in RV_{-\alpha}$, $0 < \alpha < 2$, (A1) holds, and $\hat{\alpha}$ is defined in (9) with $t_j = j/\sqrt{n}$ for $j = 1, 2, ..., m$, $m = \lfloor n^{\delta} \rfloor$ with $0 < \delta < 1/2$. Then as $n \to \infty$,*

$$Var(\hat{\alpha}) = O\left(n^{\frac{\alpha}{2} - 1 - \delta\alpha}\right).$$

The exponent of n, $\frac{\alpha}{2} - 1 - \delta\alpha$, is always less than 0, therefore the variance of $\hat{\alpha}$ converges regardless of the value of δ in $(0, 1/2)$. Combining Theorems 2.4 and 2.5, the mean squared error (MSE) is

$$MSE(\hat{\alpha}) = O\left(n^{2(\delta - 1/2)(\theta - \alpha)}\right) + O\left(n^{\frac{\alpha}{2} - 1 - \delta\alpha}\right), \tag{17}$$

as $n \to \infty$, under (*A*1) with $0 < \alpha < 2$. Similarly, the MSE goes to zero as $n \to \infty$ for all $0 < \delta < 1/2$. Note that in this case the small o term of the bias is always dominated by the variance term. With regard to the explicit size of the MSE, it is determined by the term that dominates on the right-hand side of (17). The difference between these two orders is $2\delta\theta - \theta - \delta\alpha + \frac{\alpha}{2} + 1$. If this value is less than zero, i.e., $\delta < \frac{1}{2} - \frac{1}{2\theta - \alpha}$, the second term dominates in the MSE. More precisely, if $\theta = 2$, $2\theta - \alpha$ is always between 2 and 4 for all $0 < \alpha < 2$ and the solution of $\delta < \frac{1}{2} - \frac{1}{2\theta - \alpha}$ exists under the condition $0 < \delta < 1/2$. If $\theta = \alpha + \beta$, a combination of small α and β makes $\frac{1}{2} - \frac{1}{2\theta - \alpha}$ negative, which conflicts with $\delta > 0$, then the first term dominates.

A general conclusion about the asymptotic normality of $n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}}(\hat{\alpha} - E(\hat{\alpha}))$ is stated in the following theorem.

Theorem 2.6. *Suppose $H(x) \in RV_{-\alpha}$, $0 < \alpha < 2$, (A1) holds, and $\hat{\alpha}$ is defined in (9) with $t_j = j/\sqrt{n}$ for $j = 1, 2, ..., m = \lfloor n^{\delta} \rfloor$ with $0 < \delta < 1/2$. Then as $n \to \infty$,*

$$n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}}(\hat{\alpha} - E(\hat{\alpha})) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{18}$$

where $\sigma^2 < \infty$ is a real constant (given in the proof).

From this theorem and Theorem 2.4 we deduce that, as $n \to \infty$,

$$n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2) + \frac{D_1(\theta - \alpha)}{Cs(\alpha)(\theta - \alpha + 1)^2} n^{\delta\theta - \frac{\theta}{2} - \frac{\delta\alpha}{2} + \frac{\alpha}{4} + \frac{1}{2}}\{1 + o(1)\},$$

where, using notation of (*A1*), $D_1 = CD(\alpha+\beta)$ if $\alpha+\beta<2$, otherwise D_1 is not specified unless the form of the remainder in (*A1*) is known.

If the second term on the right-hand side of the above equation is finite, i.e., $\delta\theta - \frac{\theta}{2} - \frac{\delta\alpha}{2} + \frac{\alpha}{4} + \frac{1}{2} \le 0$, the distribution of $\hat{\alpha}-\alpha$ is normal. That is, if the condition $0<\delta<\frac{1}{2}-\frac{1}{2\theta-\alpha}$ holds, as $n\to\infty$,

$$n^{\frac{1}{2}-\frac{\alpha}{4}+\frac{\delta\alpha}{2}}(\hat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

At point $\delta = \frac{1}{2}-\frac{1}{2\theta-\alpha}$ if it exists, as $n\to\infty$,

$$n^{\frac{1}{2}-\frac{\alpha}{4}+\frac{\delta\alpha}{2}}(\hat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}(0,\sigma^2) + \frac{D_1(\theta-\alpha)}{Cs(\alpha)(\theta-\alpha+1)^2}.$$

As we can see, the bias term can be quite relevant in the case of $\hat{\alpha}$. Typically bias reduction tools are based on external estimation of parameter of $L(x)$ of assumption *A1*, see, e.g. Gomes et al [18]. Given the regression context of $\hat{\alpha}$, we will instead resort to a bias reduction approach based on a semi-parametric component in the regression equation. This will allow consideration of a general $L(x)$, going beyond the case considered by *A1*.

3. Bias reduction and choice of m

Trying to contemplate a bias reducing approach, consider again equation (7): a non-parametric component will be introduced in order to approximate the unknown slowly varying function in $g(\alpha,t)$ (recall from section 2.1 that $g(\alpha,t) = s(\alpha)L(1/t)$ if $0<\alpha<2$ and $g(\alpha,t)=L_1(1/t)$ if $\alpha=2$). The new regression equation is

$$y_j \sim \alpha_0 + f(t_j) + \alpha\log t_j + \varepsilon_j \qquad 0<\alpha\le 2, \tag{19}$$

where $f(t)=\log L(1/t)$ (or $\log L_1(1/t)$). The semi-parametric regression equation (19) can be solved by approximating $f(t)$ using a smooth spline $h(t)$, which minimizes, for a given $\lambda\ge 0$,

$$\sum_{j=1}^{m}(y_j-\alpha_0-\alpha\log(t_j)-h(t_j))^2 + \lambda\int[h''(u)]^2du \tag{20}$$

see, e.g. Wood [45]. Practical estimation of (19) can be carried out using standard software for generalized additive models. Below, an estimation procedure which uses R is described. The GAM() function of the MGCV package, which can handle smooth terms and parametric ones as required in (19), is used. Estimation of (20) will be handled using the default approach defined for GAM(), which is based on a thin plate regression spline as discussed in Wood [44].

Practical selection of λ is typically carried out using generalized cross-validation (GCV) or restricted maximum likelihood (REML), see, e.g. Marra and Wood [31]. Although errors of model (19) are not i.i.d. we found both criteria to work very well, in particular REML.

To complete the estimation procedure we would require to set the value of m, i.e. set the number of points at which the ERCF is evaluated at the origin. With this, we define a procedure to obtain a RB estimator of α optimized, according to GCV or REML, with respect to the choice of m (i.e. δ).

ALGORITHM 1

1. Define $t_j = j/\sqrt{n}$, $j = 1, 2, \ldots, m = \lfloor n^{\delta} \rfloor$ with a large value of $\delta \in (0, 1/2)$. In our simulations we have chosen $\delta = 0.45$.
2. For $m = 6, 7, \ldots \lfloor n^{0.45} \rfloor$.
   a. Estimate model (19) with $h(t)$ defined as in (20). In our simulations GAM() was used.
   b. Select λ that minimizes a GCV or REML criterion.
3. Among the models selected in Step 2, select that with lower GCV or REML.

Set the estimate obtained from ALGORITHM 1 as $\hat{\alpha}^*$, then the proposed estimate of α for $0 < \alpha \leq 2$ is

$$\hat{\alpha} = \min(\hat{\alpha}^*, 2). \tag{21}$$

As noted, the range of α for which we can apply our procedure can be extended by transformation of the data. If $\hat{\alpha}$ provides a value equal to 2 it may indicate that the true value of α is above that level. It is possible to run again the estimation procedure with the squared data in order to see if a higher value of α is appropriate. Examples of this situation will be presented in Section 4 and Section 5.

4. Numerical comparisons

The estimation strategy proposed here can be cast in the class of optimized RB estimators; in our case optimization refers to the choice of m, i.e an appropriate value for $0 < \delta < 1/2$.

Numerical comparisons will then be carried out with respect to some RB competitors (Caeiro et al [6], Gomes et al [17]) based on Hill [26], generalized Hill [2], moment [11] and moment of order p [20] estimators; optimized with respect to the choice of k as discussed in Gomes et al [20].

For the above mentioned estimators, second order conditions are typically imposed in order to have a non-degenerate behavior under a semi-parametric model. Furthermore, when dealing with bias reduction, a more restrictive class (the so-called third order condition) is imposed; more specifically, the tail quantile function $U(x) := F^{-1}(1 - 1/x)$, $x \geq 1$, is $RV_{1/\alpha}$ with a slowly varying function $L(x)$ as defined by (*A1*) re-parametrized as $L(x) = C(1 + \beta x^{\rho}/(\alpha\rho) + o(x^{\rho}))$ as $x \to \infty$. (Note that in this re-parametrization β of *A1* is set to $-\rho$.)

This implies that, for example, Log-gamma, Log-Pareto and Pareto distributions are excluded while EV and t-distributions are included. RB estimation of

α is based on external estimation of (ρ, β); for further details see Gomes et al [18] and Gomes et al [20]. In our comparisons the following RB-versions are used:

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

$$\bar{H}(k) = H(k)\left(1 - \hat{\beta}(n/k)^{\hat{\rho}}/(1-\hat{\rho})\right). \tag{22}$$

where $H(k)$ is the Hill estimator given in (2).

2) RB-Moment estimator [11], denoted by MM in the tables,

$$\bar{M}(K) = M(k)\left(1 - \hat{\beta}(n/k)^{\hat{\rho}}/(1-\hat{\rho})\right) - \hat{\beta}\hat{\rho}(n/k)^{\hat{\rho}}/(1-\hat{\rho})^2, \tag{23}$$

with

$$M(k) = M_k^{(1)} + \frac{1}{2}\left[1 - (M_k^{(2)}/(M_k^{(1)})^2 - 1)^{-1}\right] \tag{24}$$

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 (23) applied to

$$GH(k) = \sum_{i=1}^{k}(\ln UH_{j,n} - \ln UH_{k,n}) \tag{25}$$

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

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

$$\bar{H}_p(k) = H_p(k)\left(1 - \frac{\hat{\beta}(1 - pH_p(k))}{1 - \hat{\rho} - pH_p(k)}\left(\frac{n}{k}\right)^{\hat{\rho}}\right) \tag{26}$$

with $H_p(k) = (1 - A_p^{-p}(k))/p$, $A_p(k) = \left(\sum_{i=1}^{k} U_{ik}^p/k\right)^{1/p}$, $U_{ik} = X_{n-i+1:n}/X_{n-k:n}$, $1 \leq i \leq k < n$. Denoted by MOP(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 [20].

Computations of the above estimators have been performed using the package EVT0 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 [20]

$$\hat{k} = \min\left(n-1, \lfloor((1 - \varphi(\hat{\rho}) - \hat{\rho})^2 n^{-2\hat{\rho}}/(-2\hat{\rho}\hat{\beta}^2(1 - 2\varphi(\hat{\rho})))\rfloor^{1/(1-2\hat{\rho})} + 1\right), \tag{27}$$

where $\lfloor x \rfloor$ is the integer part of x and $\varphi(\rho) = 1 - (\rho + \sqrt{\rho^2 - 4\rho + 2})/2$.

n	100	200	500	1000	2000
Hill	0.58 (2.41)	0.60 (2.81)	0.17 (1.41)	0.25 (1.95)	0.03 (0.78)
MOP(.5)	0.54 (1.02)	0.60 (1.02)	0.19 (0.99)	0.26 (1.00)	0.03 (0.99)
MOP(1)	0.88 (1.07)	0.75 (1.12)	0.34 (0.81)	0.60 (0.56)	0.10 (0.45)
GH	0.71 (1.00)	0.66 (1.00)	0.19 (1.01)	0.27 (1.00)	0.04 (1.00)
MM	0.59 (0.72)	0.67 (1.00)	0.21 (0.99)	0.27 (1.00)	0.04 (0.99)
GCV	-0.07 (3.72)	-0.05 (4.96)	-0.06 (3.06)	-0.06 (4.71)	-0.04 (2.17)
REML	-0.06 (4.22)	-0.01 (5.76)	0.00 (3.44)	-0.01 (5.36)	-0.02 (2.48)

TABLE 1

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. Pareto$(3,1)$ *distribution. Results based on 1000 replications.*

n	100	200	500	1000	2000
Hill	0.58 (2.95)	0.60 (3.44)	0.17 (1.73)	0.25 (2.39)	0.03 (0.95)
MOP(.5)	0.58 (1.02)	0.60 (1.01)	0.17 (1.00)	0.26 (1.00)	0.03 (1.00)
MOP(1)	0.53 (0.98)	0.54 (1.02)	0.19 (0.99)	0.27 (1.01)	0.04 (0.96)
GH	-1.66 (0.29)	0.71 (1.00)	0.21 (0.99)	0.28 (1.00)	0.04 (0.99)
MM	0.75 (0.98)	0.80 (0.99)	1.20 (0.18)	0.30 (1.00)	0.04 (0.99)
GCV	0.04 (3.98)	0.03 (4.60)	-0.01 (2.53)	-0.03 (3.76)	-0.02 (1.64)
REML	0.02 (4.37)	0.04 (5.11)	0.02 (2.81)	-0.01 (4.23)	0.00 (1.83)

TABLE 2

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. Pareto$(1,1.5)$ *distribution. Results based on 1000 replications.*

For the comparisons, the following distributions are used:

1) Pareto distributions, with $F(x) = 1 - \left(\frac{x_m}{x}\right)^{\alpha}$, $x \geq x_m$, denoted by Pareto(x_m, α). Random numbers from this distribution are simply generated in R using the function RUNIF() and inversion of F. The cases Pareto(3,1) and Pareto(1.5,1) are analyzed.
2) Symmetric stable distributions 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 library STABLEDIST. The cases Stable (1.3), Stable(1.9) are considered.
3) Extreme value distributions with $F(x) = exp(-(1+x/\alpha)^{-\alpha})$, $1+x/\alpha > 0$ denoted by $EV(\alpha)$. This distribution is simulated in R using the function RGEV() from the library FEXTREMES with shape parameter XI set equal to α. The cases EV(3) and EV(4) are analyzed; i.e. $\alpha > 2$ is considered. Following remark 2.1, the squared data have been used in estimation.
4) Fréchet 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 library EVD with shape parameter set equal to α. Examples with Fréchet(1) and Fréchet(1.5) are presented.

Tables 1 to 8 contain results of simulations. For each sample size $n = 100, 200$, 500, 1000, 2000, $M = 1000$ Monte-Carlo replicates were generated. Computations have been carried out with R version 3.3.1 and each experiment, i.e. given

n	100	200	500	1000	2000
Hill	0.37 (0.86)	0.31 (0.76)	0.26 (0.66)	0.21 (0.59)	0.17 (0.53)
MOP(.5)	0.37 (1.00)	0.30 (1.00)	0.24 (1.01)	0.19 (1.01)	0.15 (1.02)
MOP(1)	0.43 (0.98)	0.35 (0.98)	0.28 (0.97)	0.22 (0.96)	0.18 (0.95)
GH	-1.96 (0.08)	0.44 (0.58)	0.28 (0.78)	0.19 (0.88)	0.14 (0.91)
MM	-0.37 (0.19)	-0.06 (0.11)	9.44 (0.03)	-1.05 (0.13)	-1.54 (0.11)
GCV	0.19 (0.92)	0.15 (0.84)	0.12 (0.77)	0.04 (0.69)	0.04 (0.75)
REML	0.20 (1.03)	0.17 (0.96)	0.16 (0.87)	0.10 (0.82)	0.08 (0.80)

TABLE 3

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. Stable(1.3) *distribution. Results based on 1000 replications.*

n	100	200	500	1000	2000
Hill	0.77 (1.32)	0.88 (1.37)	1.07 (1.46)	1.13 (1.49)	1.14 (1.49)
MOP(.5)	0.80 (0.99)	0.90 (0.99)	1.07 (1.00)	1.10 (1.01)	1.09 (1.02)
MOP(1)	0.84 (0.97)	0.93 (0.98)	1.06 (0.99)	1.06 (1.02)	1.03 (1.04)
GH	23.64 (0.04)	-3.30 (0.10)	1.30 (0.19)	0.19 (0.17)	0.59 (0.34)
MM	-0.26 (0.96)	-0.44 (0.36)	-0.19 (2.40)	-0.20 (2.42)	-0.20 (2.39)
GCV	0.06 (2.42)	0.03 (1.80)	0.03 (2.35)	0.04 (3.11)	0.04 (2.64)
REML	0.05 (2.90)	0.04 (2.65)	0.03 (2.76)	0.03 (3.06)	0.02 (3.07)

TABLE 4

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. Stable(1.9) *distribution. Results based on 1000 replications.*

a chosen distribution and a chosen n, has been initialized using SET.SEED(1). The acronyms for the alternative estimators have been defined at the beginning of the section. GCV and REML indicate the estimators defined in ALGORITHM 1. The tables report:

a) values not in parentheses, all rows: the empirical relative bias, calculated as (with $\hat{\mathrm{E}}(\hat{\alpha})$ indicating the mean of the 1000 Monte-Carlo Estimates):

$$\frac{\hat{\mathrm{E}}(\hat{\alpha}) - \alpha}{\alpha}. \tag{28}$$

b) Values in parentheses, first row: the empirical RMSE (Root-MSE) for the RB-Hill estimator, i.e. $\sqrt{\hat{\mathrm{E}}(\hat{\alpha} - \alpha)^2}$.

c) Values in parentheses, all other rows: the relative RMSE of all other estimators with respect to that of the RB-Hill estimator calculated as

$$\frac{RMSE_H}{RMSE_E} \tag{29}$$

where the suffix E stands for the estimator considered in turn and H stands for RB-Hill. A value of the ratio greater than one indicates smaller efficiency of the RB-Hill with respect to estimator E.

In the simulations, for a better comparison of the results with the other estimators, we have used the value of $\hat{\alpha}$ provided by the estimation procedure

n	100	200	500	1000	2000
Hill	-0.31 (1.04)	-0.28 (0.97)	-0.26 (0.92)	-0.24 (0.87)	-0.22 (0.84)
MOP(.5)	-0.29 (1.03)	-0.27 (1.03)	-0.24 (1.03)	-0.23 (1.03)	-0.21 (1.02)
MOP(1)	-0.26 (1.08)	-0.24 (1.08)	-0.22 (1.07)	-0.20 (1.07)	-0.19 (1.06)
GH	0.19 (0.25)	0.42 (0.13)	0.05 (0.39)	0.05 (0.49)	0.01 (0.74)
MM	-0.36 (0.18)	-0.64 (0.56)	-0.60 (0.63)	-0.56 (0.67)	-0.55 (0.65)
GCV	0.28 (0.91)	0.30 (0.91)	0.30 (0.89)	0.30 (0.87)	0.29 (0.84)
REML	0.27 (0.95)	0.30 (0.91)	0.31 (0.89)	0.26 (0.92)	0.26 (0.92)

TABLE 5

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. EV(3) *distribution. Results based on 1000 replications. Squared data are used in this example for REML and GCV estimators.*

n	100	200	500	1000	2000
Hill	-0.43 (1.35)	-0.40 (1.29)	-0.38 (1.24)	-0.35 (1.20)	-0.33 (1.17)
MOP(.5)	-0.42 (1.02)	-0.39 (1.02)	-0.36 (1.02)	-0.34 (1.02)	-0.32 (1.02)
MOP(1)	-0.39 (1.05)	-0.37 (1.05)	-0.34 (1.05)	-0.32 (1.04)	-0.31 (1.04)
GH	0.47 (0.11)	0.22 (0.27)	-0.05 (0.25)	-0.02 (0.40)	-0.08 (0.23)
MM	-0.06 (0.16)	-0.70 (0.69)	-0.66 (0.77)	-0.65 (0.75)	-0.65 (0.73)
GCV	0.01 (1.66)	0.00 (1.63)	0.00 (2.15)	0.00 (2.46)	0.00 (2.75)
REML	0.01 (1.68)	0.01 (1.91)	-0.01 (2.15)	-0.04 (2.27)	-0.03 (2.56)

TABLE 6

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. EV(4) *distribution. Results based on 1000 replications. Squared data are used in this example for REML and GCV estimators.*

without truncating it at 2. Looking at the results in the Tables, we can try to summarize as follows:

a) REML has generally better efficiency with respect to GCV.
b) REML and GCV are much better with respect to all other estimators for Pareto, Stable (1.9) and EV(4) cases; is comparable for Stable(1.3) and EV(3) while is worse with respect to Hill and MOP estimators in the case of Fréchet distributions.
c) In the case of EV distributions, with $\alpha > 2$, REML and GCV show a good performance.
d) The relative bias of REML and GCV is nearly always quite low with respect to the other estimators.

Overall, considering a larger pool of simulations results we have that GCV and REML estimators compare well in general with other estimators. They are clearly well suited for cases of α quite close to 2 where it can be much more efficient with respect to other estimation strategies. The extremely low bias of REML and GCV shows that probably an overall gain could be achieved by acting on the bias-variance trade-off. This is difficult to tackle in automatic simulations, however, as it will be shown in the next Section, a careful regression analysis of the ERCF at the origin can greatly help in determining the optimal value of m to use.

n	100	200	500	1000	2000
Hill	0.08 (0.51)	0.04 (0.44)	0.01 (0.35)	0.00 (0.30)	0.00 (0.26)
MOP(.5)	0.11 (0.97)	0.07 (0.96)	0.04 (0.94)	0.02 (0.93)	0.01 (0.91)
MOP(1)	0.28 (0.86)	0.24 (0.82)	0.20 (0.74)	0.17 (0.70)	0.15 (0.64)
GH	0.16 (0.70)	0.10 (0.79)	0.05 (0.82)	0.03 (0.87)	0.02 (0.88)
MM	-0.90 (0.14)	0.74 (0.09)	-0.43 (0.11)	0.02 (0.19)	-0.12 (0.23)
GCV	-0.02 (0.71)	-0.02 (0.65)	-0.04 (0.59)	-0.03 (0.59)	-0.04 (0.61)
REML	-0.04 (0.81)	-0.04 (0.81)	-0.04 (0.72)	-0.02 (0.70)	-0.02 (0.70)

TABLE 7

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. Frechet(1) *distribution. Results based on 1000 replications.*

n	100	200	500	1000	2000
Hill	0.08 (0.63)	0.04 (0.54)	0.01 (0.43)	0.00 (0.37)	0.00 (0.32)
MOP(0.5)	0.09 (0.99)	0.06 (0.98)	0.02 (0.97)	0.01 (0.97)	0.00 (0.97)
MOP(1)	0.15 (0.96)	0.11 (0.94)	0.07 (0.9)	0.04 (0.88)	0.03 (0.85)
GH	0.08 (0.29)	0.16 (0.52)	0.08 (0.70)	0.04 (0.77)	0.03 (0.79)
MM	-3.77 (0.05)	0.34 (0.08)	-0.49 (0.10)	-1.04 (0.14)	-0.84 (0.22)
GCV	0.01 (0.74)	0.00 (0.70)	-0.02 (0.60)	-0.03 (0.63)	-0.03 (0.54)
REML	-0.01 (0.87)	0.00 (0.79)	0.02 (0.68)	-0.01 (0.65)	-0.01 (0.61)

TABLE 8

Hill estimator (first row): relative bias and (RMSE); other estimators: relative bias and (relative RMSE) wrt to the Hill estimator. Frechet(1.5) *distribution. Results based on 1000 replications.*

5. Examples of application

5.1. Simulated Pareto, Fréchet and Normal distributions

This example considers simulated data from the Pareto(1,0.1), the Fréchet(1) and $N(0,1)$ distributions. In all cases 1000 pseudo-random values were generated. To estimate α we choose $\delta = 0.45$ which gives $m = 22$ and apply the estimation procedure described in ALGORITHM 1 using either REML or GCV criterion in order to select the optimal m and smoothing.

Considering the Pareto and Fréchet cases first, a plot of $\log(t_j)$ and $\log(1 - U_n(t_j))$ for selected values of t_j is in Figure 1. The circle points (red in color) are those used in estimation, i.e. $t_j = j/\sqrt{n}$, $j = 1, \ldots, m$ while the cross ones (blue in color) are additional points used to depict a fuller range of the behavior of $\log(1 - U_n(t))$.

For the the Pareto(1,0.1) distribution, when REML is used as selection criterion, one has an estimate $\hat{\alpha} = 0.104$, indicating $m = 22$ as an optimal choice. The estimates for different choices of m range from a minimum of 0.02, when $m = 6$, to a maximum of 0.106, when $m = 20$. If a GCV selection criterion is considered the estimate given is 0.072 for $m = 22$. The estimates for different choices of m range from a minimum of -0.09, when $m = 9$, to a maximum of 0.106, when $m = 20$ (same as in REML).

For the Fréchet(1) case, estimation with REML, gives an estimate of $\hat{\alpha} = 1.006$, indicating $m = 22$ as an optimal choice. The estimates for different choices

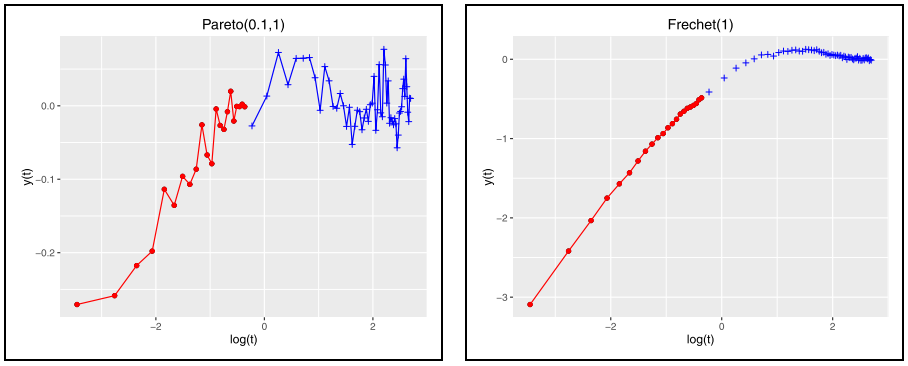

FIG 1. *Plot of* $\log(t)$ *and* $\log(1-U_n(t))$ *for the simulated data from the Pareto(1,0.1) and the Fréchet(1) distributions. Red points: points used in estimation. Blue crosses: points not used in the estimation procedure*

of m range from a minimum of 0.947, when $m = 6$, to a maximum of 1.012, when $m = 13$. If a GCV selection criterion is considered, the estimate given is 0.970 for $m = 12$. The estimates for different choices of m range from a minimum of 0.940, when $m = 6$, to a maximum of 1.022, when $m = 10$.

Consider now the case of the Normal distribution. Figure 2 reports the values of $\log(t)$ and $\log(1-U_n(t))$ where the latter has been calculated: firstly, using the generated data (left graph) and, secondly, using the squared normal data (right graph). In both cases one expects the estimates be close to 2 as the distributions have finite variance.

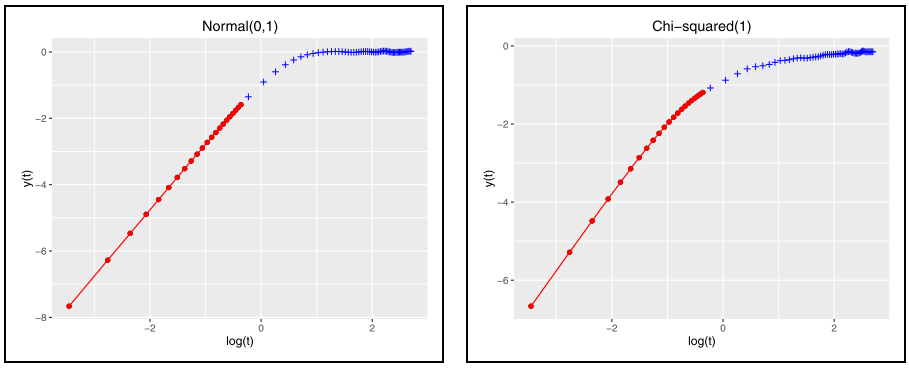

FIG 2. *Plot of* $\log(t)$ *and* $\log(1-U_n(t))$ *for the simulated data from the Normal(0,1) distribution (left graph) and for the squared data from the Normal(0,1). Red points: points used in estimation. Blue crosses: points not used in the estimation procedure.*

In more detail, considering the generated (not squared) data, estimation with REML, gives an estimate of $\hat{\alpha} = 1.987$, indicating $m = 22$ as an optimal choice.

The estimates for different choices of m range from a minimum of 1.987, when $m = 22$, to a maximum of 2.001, when $m = 6$. If a GCV selection criterion is considered, the estimate given is $\hat{\alpha} = 2.001$ for $m = 10$. The estimates for different choices of m range from a minimum of 1.990, when $m = 17$, to a maximum of 2.002, when $m = 6$.

Considering the squared data, estimation with REML, gives an estimate of $\hat{\alpha} = 2.006$, indicating $m = 22$ as an optimal choice. The estimates for different choices of m range from a minimum of 1.943, when $m = 14$, to a maximum of 2.008, when $m = 16$. If a GCV selection criterion is considered, the estimate given is $\hat{\alpha} = 2.002$ for $m = 9$. The estimates for different choices of m range from a minimum of 1.939, when $m = 22$, to a maximum of 2.013, when $m = 6$.

5.2. *City sizes*

There is an ongoing debate on the specific regularity on the distribution of city sizes: Zipf's law (a heavy-tailed distribution with $\alpha = 1$) or a log-normal distribution. Early studies show that Zipf's law generally holds well for data on the largest cities in a country (see Gabaix and Ioannides [15] and references therein). However, recent developments by Eeckhout [12] extend the analysis to the entire range of city sizes and demonstrate that the size of all populated places in the US follows a log-normal distribution. Levy [30] argues that significant deviations from a log-normal distribution exist in the top range of the largest cities by using graphical analysis. Several potential problems with Levy's studies are pointed out by Eeckhout [13].

The data set used by Eeckhout [12] and Levy [30] containing the entire distribution with population ranging from 1 to several millions is analyzed here. The "US Census 2000 places" dataset is based on 25,358 units of "city, town, borough, or village". To estimate α we choose $\delta = 0.4$ which gives $m = 57$.

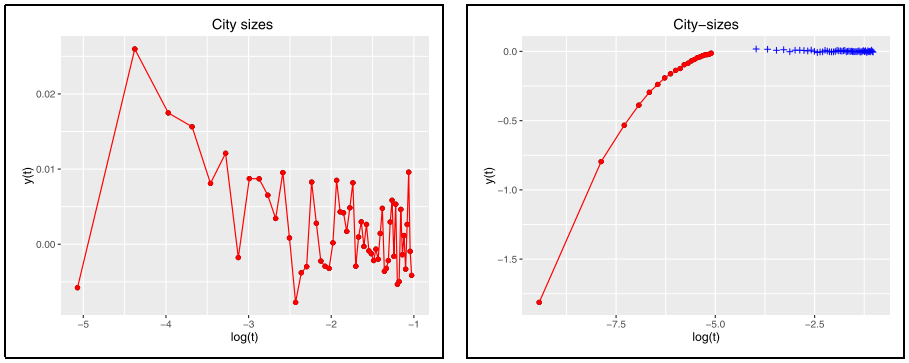

FIG 3. *Plot of* $\log(t)$ *and* $\log(1-U_n(t))$ *for the City sizes data. Left graph: plot with* $t_j = j/\sqrt{n}$, $j = 1, \ldots, 57$. *Right graph: plot with* $t_j = 7.97e-05 + j \cdot 0.0003$, $j = 0, \ldots, 20$. *Red points: points used in estimation. Blue crosses: points not used in the estimation procedure*

Looking at Figure 3, left graph, it appears that $\log(1-U_n(t))$, when plotted against $\log(t_j)$ for $t_j = j/\sqrt{n}$, $j = 1,\ldots,57$, straightens only in very close proximity to 0. For this situation it seems sensible to analyze further the problem by getting closer to 0. We select manually some points by setting $t_j = 7.97e-05+j\cdot 0.0003$, $j = 0,\ldots,20$; the plot of $\log(1-U_n(t))$ for these points plus some additional ones (the blue crosses) are in the right graph of Figure 3, a more regular behavior of the function can be observed now.

Applying the estimation procedure using the points $t_j = 7.97e-05+j\cdot 0.0003$, $j = 0,\ldots,20$ we obtain $\alpha = 0.818$ for $m = 6$ (note that here $m = 21$) if GCV is used and $\alpha = 0.752$ for $m = 21$ if REML is used. Estimates, using the GCV criterion range from a minimum of 0.686 when $m = 21$ to a maximum of 0.839 when $m = 7$, while in the case of REML, estimates range from a minimum of 0.752 when $m = 21$ to a 0.836 when $m = 7$.

As we can see from this example, in practical applications, a full regression analysis of the ERCF close to the origin may reveal more details and give more insights on the phenomenon. In this case it appears more sensible to select a value of m not too large given the red part of the graph on the right of Figure 3; GCV, which suggests $m = 6$, seems to be the method to use here. Also, the result supports the presence of a Pareto tail, in the case of a log-normal distribution one would expect to have an estimate of α close to the value of 2.

6. Conclusions

For any distribution the rate of decay in the tails is related to the RCF at the origin. This fact is exploited to introduce a novel methodology for estimating the tail index of a distribution based on semi-parametric regression on the ERCF at the origin. The approach is conceptually similar to regression of the periodogram on frequency ordinates, which is used in time series to estimate the long memory parameter.

Consistency of a simple regression estimator is discussed under minimal assumptions, while rates of convergence and asymptotic normality are studied under the assumption that the slowly varying function $L(x)$ in (1) has a finite limit as $x \to \infty$. It turns out that a bias term may be severe for certain parameter values.

A semi-parametric regression equation with a smooth spline component approximating L in order to account for bias is implemented; this allows to keep at minimum assumptions on the unknown distribution underlying the data. Alternatively, under assumption *A*1, one could consider external estimation of parameters in L (e.g. Gomes et al [18]); this approach is not pursued here.

Simulated results indicate that $\hat{\alpha}$ has lower MSE than other estimators in a large variety of cases and bias is generally quite small. Practical implementation is supported by GCV or REML and in the examples the estimated values do not show high variability with respect to the choice of m. In real applications, a full regression analysis of the ERCF at the origin can help in determining the optimal estimation strategy. An application to the debated city-sizes distribution gives a clear answer in favor of a Pareto-type tail behavior.

The restriction $0 < \alpha \leq 2$ might not be too severe since, in practical applications, one is often interested to detect low values of the tail index. Moreover, by transformations, extension of the range of values that can be estimated is always possible.

Since the method proposed works extremely well for values of α close to 2 it could be interesting to use it as a starting point to try developing a testing procedure for the hypothesis of finite variance in the data.

7. Proofs

For the proofs, some preliminary lemmas are needed. In the following, for lighter notation, set $z_j - \bar{z} = a_j$.

Lemma 7.1. *Assume that* $0 < \beta < 1$, *then as* $m \to \infty$ *[7],*

$$\frac{1}{m}\sum_{j=1}^{m} j^{-\beta}\log j - \Big(\frac{1}{m}\sum_{j=1}^{m} j^{-\beta}\Big)\Big(\frac{1}{m}\sum_{j=1}^{m}\log j\Big) = \frac{-\beta}{(1-\beta)^2}m^{-\beta} + O(m^{-1}\log m).$$

$$\frac{1}{m}\sum_{j=1}^{m}(\log j)^2 - \Big(\frac{1}{m}\sum_{j=1}^{m}\log j\Big)^2 = 1 + O\{m^{-1}(\log m)^2\}.$$

Lemma 7.2. *For* $0 < \beta < 2$, *then as* $m \to \infty$,

$$\frac{1}{m}\sum_{j=1}^{m}\log j = \log m - 1 + \frac{\log m}{2m} + O(1/m).$$

$$\frac{1}{m}\sum_{j=1}^{m} j^{\beta}\log j - \Big(\frac{1}{m}\sum_{j=1}^{m}\log j\Big)\Big(\frac{1}{m}\sum_{j=1}^{m} j^{\beta}\Big) = \frac{\beta}{(\beta+1)^2}m^{\beta} + O(m^{\beta-1}\log m).$$

Lemma 7.3. *As* $m \to \infty$,

$$\sum_{j=1}^{m}(\log j - \frac{1}{m}\sum_{k=1}^{m}\log k)^2 j^{-\beta} = \begin{cases} (\frac{2}{(1-\beta)^3} + \frac{1}{1-\beta} - \frac{2}{(1-\beta)^2})m^{1-\beta}\{1+o(1)\}, & \text{if } 0<\beta<1, \\ \frac{1}{3}\log^3 m\{1+o(1)\}, & \text{if } \beta = 1, \\ O\{\log^2 m\}, & \text{if } 1<\beta<2. \end{cases}$$

$$\sum_{j=1}^{m}(\log j - \frac{1}{m}\sum_{k=1}^{m}\log k)^2 j^{\beta} = \begin{cases} m + O((\log m)^2), & \text{if } \beta = 0, \\ (\frac{2}{(1+\beta)^3} + \frac{1}{1+\beta} - \frac{2}{(1+\beta)^2})m^{1+\beta}\{1+o(1)\}, & \text{if } \beta > 0. \end{cases}$$

Also, for $\beta > 0$,

$$\sum_{j=1}^{m} j^{\beta} \log j = \frac{1}{\beta+1} m^{\beta+1} \log m\{1 + o(1)\}.$$

$$\sum_{j=1}^{m} j^{\beta} = \frac{1}{\beta+1} m^{\beta+1}\{1 + o(1)\}.$$

Lemmas 7.2 and 7.3 follow from Euler-Maclaurin-type integral approximations to series [1]; see Jia [28] for further details.

For the proof of Theorem 2.3 recall that $\lambda_{ij} = t_j/t_i = j/i,\ t_i > t_j$. Since $|\lambda_{ij}| < 1$, one has absolute convergence of the expansion for $0 < \alpha < 2\ \&\ \alpha \neq 1$

$$\begin{aligned}(1+\lambda_{ij})^{\alpha} + (1-\lambda_{ij})^{\alpha} &= \sum_{r=0}^{\infty} \binom{\alpha}{r} (\lambda_{ij})^r + \sum_{r=0}^{\infty} (-1)^r \binom{\alpha}{r} (\lambda_{ij})^r \\ &= 2 + 2\sum_{r=1}^{\infty} \binom{\alpha}{2r} (\lambda_{ij})^{2r}. \end{aligned} \tag{30}$$

For $\alpha = 1$,

$$(1+\lambda_{ij})^{\alpha} + (1-\lambda_{ij})^{\alpha} = 2. \tag{31}$$

Proof of Theorem 2.3. Consider first the bias term, which is given by

$$E(\hat{\alpha} - \alpha) = \frac{1}{S_{zz}} \sum_{j=1}^{m} a_j g(\alpha, t_j) + \frac{1}{S_{zz}} \sum_{j=1}^{m} a_j E(\varepsilon_j) \tag{32}$$

As $n \to \infty$, by Lemma 7.1,

$$S_{zz} = m\{1 + O\{m^{-1}(\log m)^2\}\}. \tag{33}$$

and, for the second term on the right-hand side of (32), by Lemma 7.1, as $m \to \infty$ and $n \to \infty$,

$$\begin{aligned}\sum_{j=1}^{m} a_j (n t_j^{\alpha})^{-\frac{1}{2}} &= m n^{\frac{\alpha-2}{4}} \Big\{ \Big(\frac{1}{m} \sum_{j=1}^{m} j^{-\alpha/2} \log j \Big) - \Big(\frac{1}{m} \sum_{j=1}^{m} \log j \Big) \Big(\frac{1}{m} \sum_{j=1}^{m} j^{-\alpha/2} \Big) \Big\} \\ &= m n^{\frac{\alpha-2}{4}} \Big[-\frac{\alpha}{2} (1 - \frac{\alpha}{2})^{-2} (m^{-\frac{\alpha}{2}}) + \{O(m^{-1} \log m)\} \Big]. \end{aligned}$$

With (33) and setting $m = n^{\delta}$ we obtain that

$$\frac{\sum_{j=1}^{m} a_j o_p((n t_j^{\alpha})^{-\frac{1}{2}})}{S_{zz}} = o\{n^{\frac{\alpha-2}{4}} m^{-\frac{\alpha}{2}}\} = o\{n^{\frac{\alpha-2-2\alpha\delta}{4}}\}. \tag{34}$$

Consider now the variance of $\hat{\alpha}$ given by

$$Var(\hat{\alpha}) = \frac{1}{S_{zz}^2} \sum_{j=1}^{m} a_j^2 Var(\varepsilon_j) + \frac{2}{S_{zz}^2} \sum_{i=2}^{m} \sum_{j=1}^{i-1} a_i a_j Cov(\varepsilon_i, \varepsilon_j). \tag{35}$$

Ignoring higher order terms we have

$$\frac{1}{S_{zz}^2}\sum_{j=1}^{m} a_j^2 Var(\varepsilon_j') \sim \frac{n^{\alpha/2-1}}{m^2 L(\sqrt{n}/m)} \frac{2-2^{\alpha-1}}{s(\alpha)} \sum_{j=1}^{m} (\log j - \frac{1}{m}\sum_{k=1}^{m}\log k)^2 j^{-\alpha}. \tag{36}$$

Combining Equations (13), (33) and (31) or (30), for $\alpha = 1$,

$$\begin{aligned}
&\frac{2}{S_{zz}^2}\sum_{i=2}^{m}\sum_{j=1}^{i-1} a_i a_j Cov(\varepsilon_i', \varepsilon_j') \\
&\quad\sim \frac{2n^{-1/2}}{m^2 L(\sqrt{n}/m)s(1)} \sum_{i=2}^{m}\sum_{j=1}^{i-1} (\log i - \frac{1}{m}\sum_{k=1}^{m}\log k)(\log j - \frac{1}{m}\sum_{k=1}^{m}\log k) i^{-1} \\
&\quad\sim \frac{2n^{-1/2}}{m^2 L(\sqrt{n}/m)s(1)} \sum_{i=2}^{m} (\log i - \frac{1}{m}\sum_{k=1}^{m}\log k) i^{-1} \sum_{j=1}^{i-1} (\log j - \frac{1}{m}\sum_{k=1}^{m}\log k) \\
&\quad\sim \frac{2n^{-1/2}}{m^2 L(\sqrt{n}/m)s(1)} \sum_{i=2}^{m} (\log i - \frac{1}{m}\sum_{k=1}^{m}\log k)(i-1) i^{-1} (\log (i-1) \\
&\qquad - \frac{1}{m}\sum_{k=1}^{m}\log k) \\
&\quad\sim \frac{2n^{-1/2}}{m^2 L(\sqrt{n}/m)s(1)} \sum_{i=1}^{m} (\log i - \frac{1}{m}\sum_{k=1}^{m}\log k)^2 \\
&\quad\sim \frac{2n^{-1/2}}{mL(\sqrt{n}/m)s(1)}
\end{aligned} \tag{37}$$

for $0 < \alpha < 2$ & $\alpha \neq 1$,

$$\begin{aligned}
&\frac{2}{S_{zz}^2}\sum_{i=2}^{m}\sum_{j=1}^{i-1} a_i a_j Cov(\varepsilon_i', \varepsilon_j') \\
&\quad\sim \frac{2n^{\alpha/2-1}}{m^2 L(\sqrt{n}/m)s(\alpha)} \sum_{i=2}^{m}\sum_{j=1}^{i-1} (\log i - \frac{1}{m}\sum_{k=1}^{m}\log k)(\log j - \frac{1}{m}\sum_{k=1}^{m}\log k) \\
&\qquad \left[i^{-\alpha} - j^{-\alpha} \sum_{r=1}^{\infty} \binom{\alpha}{2r} (\lambda_{ij})^{2r} \right] \\
&\quad= \frac{2n^{\alpha/2-1}}{m^2 L(\sqrt{n}/m)s(\alpha)} \Big[\sum_{i=2}^{m}\sum_{j=1}^{i-1} (\log i - \frac{1}{m}\sum_{k=1}^{m}\log k)(\log j - \frac{1}{m}\sum_{k=1}^{m}\log k)(i^{-\alpha}) \\
&\qquad - \sum_{i=2}^{m}\sum_{j=1}^{i-1} (\log i - \frac{1}{m}\sum_{k=1}^{m}\log k)(\log j - \frac{1}{m}\sum_{k=1}^{m}\log k) \sum_{r=1}^{\infty} \binom{\alpha}{2r} j^{2r-\alpha} i^{-2r} \Big] \\
&\quad= \frac{2n^{\alpha/2-1}}{m^2 L(\sqrt{n}/m)s(\alpha)} \Big[T_1 - T_2 \Big].
\end{aligned}$$

Similar to (37),

$$T_1 \sim \sum_{i=1}^{m} (\log i - \frac{1}{m}\sum_{k=1}^{m} \log k)^2 i^{-\alpha+1}, \tag{38}$$

and

$$\begin{aligned} T_2 &= \sum_{r=1}^{\infty} \binom{\alpha}{2r} \sum_{i=2}^{m} (\log i - \frac{1}{m}\sum_{k=1}^{m} \log k) i^{-2r} \sum_{j=1}^{i-1} (\log j - \frac{1}{m}\sum_{k=1}^{m} \log k) j^{2r-\alpha} \\ &\sim \sum_{r=1}^{\infty} \binom{\alpha}{2r} \sum_{i=2}^{m} (\log i - \frac{1}{m}\sum_{k=1}^{m} \log k) \frac{i^{-2r}(i-1)^{2r-\alpha+1}}{2r-\alpha+1} (\log(i-1) \\ &\quad - \frac{1}{m}\sum_{k=1}^{m} \log k) \\ &\sim \sum_{r=1}^{\infty} \binom{\alpha}{2r} \frac{1}{2r-\alpha+1} \sum_{i=1}^{m} (\log i - \frac{1}{m}\sum_{k=1}^{m} \log k)^2 i^{-\alpha+1} \\ &\sim \left[\frac{2}{(2-\alpha)^3} + \frac{1}{2-\alpha} - \frac{2}{(2-\alpha)^2}\right] m^{-\alpha+2} \sum_{r=1}^{\infty} \binom{\alpha}{2r} \frac{1}{2r-\alpha+1}. \end{aligned} \tag{39}$$

For α is not a non-negative integer, as $r \to \infty$,

$$\binom{\alpha}{r} = \frac{(-1)^r}{\Gamma(-\alpha) r^{(1+\alpha)}} \{1 + o(1)\}.$$

It can be proven that $\sum_{r=1}^{\infty} \binom{\alpha}{2r} \frac{1}{2r-\alpha+1}$ converges. Therefore, combing T_1, T_2, and Equation (36), the variance of $\hat{\alpha}$ is

$$Var(\hat{\alpha}) = \begin{cases} \frac{2}{s(1)} \frac{n^{-1/2}}{mL(\sqrt{n}/m)} \{1 + o(1)\}, & \text{if } \alpha = 1, \\ \frac{2(A_\alpha - A_\alpha C_\alpha)}{s(\alpha)} \frac{n^{\alpha/2-1}}{m^\alpha L(\sqrt{n}/m)} \{1 + o(1)\}, & \text{if } 0 < \alpha < 2 \ \& \ \alpha \neq 1. \end{cases} \tag{40}$$

where $A_\alpha = \frac{2}{(2-\alpha)^3} + \frac{1}{2-\alpha} - \frac{2}{(2-\alpha)^2}$ and $\sum_{r=1}^{\infty} \binom{\alpha}{2r} \frac{1}{2r-\alpha+1} \to C_\alpha$. Therefore, the order of the variance of $\hat{\alpha}$ is $\frac{n^{\alpha/2-1}}{m^\alpha L(\sqrt{n}/m)}$ for all $0 < \alpha < 2$.

In the case that $\alpha \geq 2$ evaluation of (35) yields

$$Var(\hat{\alpha}) = O\left(\frac{1}{n}\right) + O\left(\frac{1}{nm^2} \log^2 m\right) \qquad \square \tag{41}$$

The proofs of theorems 2.4 and 2.5 are similar to that of Theorem 2.3, using Assumption *A1* and can be found in Jia [28].

Proof of Theorem 2.6. From (32), $\hat{\alpha} - E(\hat{\alpha})$ is equal to $\frac{\sum_{j=1}^{m} a_j \varepsilon_j}{S_{zz}}$. Combined with (12), it can be rewritten as

$$\hat{\alpha} - E(\hat{\alpha}) = \sum_{j=1}^{m} \frac{a_j \left(\frac{U_n(t_j) - U(t_j)}{1 - U(t_j)} + o_p((nt_j^{\alpha})^{-\frac{1}{2}}) \right)}{S_{zz}}$$
$$= \frac{1}{n} \frac{\sum_{j=1}^{m} a_j \sum_{i=1}^{n} \frac{[\cos(X_i t_j) - U(t_j)]}{[1 - U(t_j)]}}{S_{zz}} + \frac{\sum_{j=1}^{m} a_j o_p((nt_j^{\alpha})^{-\frac{1}{2}})}{S_{zz}}. \quad (42)$$

For the second term on the right-hand side of (42) see (34). We change the order of the summations in the first term on the right-hand side of (42). The first term is then rewritten as $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{a_j(\cos(X_i t_j) - U(t_j))}{S_{zz}[1 - U(t_j)]}$. Define the random variable $W_{i,m} = \sum_{j=1}^{m} \frac{a_j(\cos(X_i t_j) - U(t_j))}{S_{zz}[1 - U(t_j)]}$. Note that $W_{i,m}$'s with $i = 1, 2, \ldots, n$ for a given m are i.i.d. random variables with null mean.

As far as $Var(W_{i,m})$ is concerned, it can be derived by noting that, for $t_k > t_j$ and $m = n^{\delta}$, as $n \to \infty$,

$$Cov\Big(\frac{\cos(X_i t_k) - U(t_k)}{1 - U(t_k)}, \frac{\cos(X_i t_j) - U(t_j)}{1 - U(t_j)}\Big)$$
$$= E\Big(\frac{[\cos(X_i t_k) - U(t_k)][\cos(X_i t_j) - U(t_j)]}{[1 - U(t_k)][1 - U(t_j)]}\Big)$$
$$= \frac{U(t_k + t_j) + U(t_k - t_j) - 2U(t_k)U(t_j)}{2[1 - U(t_k)][1 - U(t_k)]}, \quad (43)$$

which is equal to $\lim_{n\to\infty} nCov(\varepsilon'_k, \varepsilon'_j)$.

Similarly, $Var(\frac{\cos(X_i t_j) - U(t_j)}{1 - U(t_j)})$ is equal to $\lim_{n\to\infty} nVar(\varepsilon'_j)$.

Hence, with calculation similar to those used to derive $Var(\hat{\alpha})$, we have

$$Var(W_{i,m}) = \begin{cases} \frac{2}{s(1)} \frac{n^{1/2}}{mL(\sqrt{n}/m)}\{1 + o(1)\}, & \text{if } \alpha = 1, \\ \frac{2(A_\alpha - A_\alpha C_\alpha)}{s(\alpha)} \frac{n^{\alpha/2}}{m^{\alpha} L(\sqrt{n}/m)}\{1 + o(1)\}, & \text{if } 0 < \alpha < 2 \ \& \ \alpha \neq 1. \end{cases} \quad (44)$$

Note that the normalization given in Theorem 2.6 is appropriate since, under A1, $L(x) \to C$ as $x \to \infty$.

Next, for $m = n^{\delta}$, define the random variable $W^*_{i,n^{\delta}} = \frac{1}{n^{\frac{\alpha}{4} - \frac{\delta\alpha}{2}}} W_{i,n^{\delta}}$ and let $\sigma^2 = \lim_{n\to\infty} Var(W^*_{i,n^{\delta}})$. Then denote $S_n = \sum_{i=1}^{n} W^*_{i,n^{\delta}}$ and $\sigma_n^2 = Var(\sum_{i=1}^{n} W^*_{i,n^{\delta}}) = \sum_{i=1}^{n} Var(W^*_{i,n^{\delta}}) = nVar(W^*_{1,n^{\delta}}) = n\sigma^2_{W^*}$.

The next step is to show that the Lindeberg condition holds. The sequence of $W^*_{i,n^{\delta}}$ satisfies the Lindeberg condition if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} E[W^{*2}_{i,n^{\delta}}; \{|W^*_{i,n^{\delta}}| > \epsilon\sigma_n\}] = 0. \quad (45)$$

Since

$$E[W^{*2}_{i,n^{\delta}}; \{|W^*_{i,n^{\delta}}| > \epsilon\sigma_n\}] = E[W^{*2}_{1,n^{\delta}}; \{|W^*_{1,n^{\delta}}| > \epsilon\sigma_n\}],$$

then the term inside the limit on the left-hand side of (45) can be written as

$$\frac{1}{\sigma_n^2}\sum_{i=1}^{n} E[W_{i,n^\delta}^{*2};\{|W_{i,n^\delta}^*|>\epsilon\sigma_n\}] = \frac{1}{n\sigma_{W^*}^2} nE[W_{1,n^\delta}^{*2};\{|W_{1,n^\delta}^*|>\epsilon\sigma_{W^*}\sqrt{n}\}]$$
$$= \frac{1}{\sigma_{W^*}^2} E[W_{1,n^\delta}^{*2};\{|W_{1,n^\delta}^*|>\epsilon\sigma_{W^*}\sqrt{n}\}].$$

Note that, by Chebyshev's inequality, as $n\to\infty$, for every $\epsilon>0$,

$$P(|W_{1,n^\delta}^*|\geq\epsilon\sigma_{W^*}\sqrt{n}) \leq \frac{Var(W_{1,n^\delta}^*)}{\epsilon^2\sigma_{W^*}^2 n} = \frac{\sigma_{W^*}^2}{\epsilon^2\sigma_{W^*}^2 n}\to 0.$$

Since $W_{i,n^\delta}^{*2}\mathbf{I}\{|W_{i,n^\delta}^*|>\epsilon\sigma_n\}\leq W_{i,n^\delta}^{*2}$ and $E(W_{i,n^\delta}^{*2})<\infty$, by the dominated convergence theorem, the Lindeberg condition (45) holds. Therefore, the sequence of W_{i,n^δ}^* satisfies the central limit theorem, i.e., as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{i,n^\delta}^* \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

By Slutsky's theorem, as $n\to\infty$,

$$\frac{\sqrt{n}}{n^{\frac{\alpha}{4}-\frac{\delta\alpha}{2}}}(\hat{\alpha}-E(\hat{\alpha})) \quad \xrightarrow{d} \quad \mathcal{N}(0,\sigma^2). \tag{46}$$

This completes the proof of Theorem 2.6. □

Acknowledgement

The authors would like to thank two anonymous referees and the associate editor for their careful reading and constructive comments which led to an improved version of the paper.

References

[1] Abramowitz, M., Stegun, I. A. (1965) *Handbook of Mathematical Functions*, Dover, New York.
[2] Beirlant, J., Vynckier, P., & Teugels, J. L. (1996). Excess functions and estimation of the extreme-value index. *Bernoulli*, 2(4), 293–318. MR0177136
[3] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2006) *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester. MR1440271
[4] Beran, J., Schell, D., Stehlík, M. (2014) The harmonic moment tail index estimator: asymptotic distribution and robustness. *Annals of the Institute of Statistical Mathematics* 66(1), 193–220. MR2108013

[5] Brilhante, M. F., Gomes, M. I., & Pestana, D. (2013). A simple generalisation of the Hill estimator. *Computational Statistics & Data Analysis*, 57(1), 518–535. MR3147551
[6] Caeiro, F., Gomes, M. I., Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* 3(2), 111–136. MR2981106
[7] Chan, G., Hall, P., Poskitt, D. S. (1995) Periodogram-based estimators of fractal properties. *The Annals of Statistics* 23(5), 1684–1711. MR2259357
[8] Csorgo, S., Deheuvels, P., & Mason, D. (1985). Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, 13(3), 1050–1077. MR1370303
[9] Danielsson, J., Jansen, D. W., De vries, C. G. (1996) The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics-Theory and Methods* 25(4), 711–720. MR0803758
[10] de Haan, L., Resnick, S. (1998) On asymptotic normality of the Hill estimator. *Communications in Statistics Stochastic Models* 14(4), 849–866. MR1380614
[11] Dekkers, A. L. M., Einmahl, J. H. J., de Haan, L. (1989) A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics* 17(4), 1833–1855. MR1631534
[12] Eeckhout, J. (2004) Gibrat's law for (all) cities. *The American Economic Review* 94(5), 1429–1451. MR1026315
[13] Eeckhout, J. (2009) Gibrat's law for (all) cities: reply. *The American Economic Review* 99(4), 1676–1683.
[14] Fraga Alves, M. (2001) A location invariant hill-type estimator. *Extremes* 4(3), 199–217.
[15] Gabaix, X., Ioannides YM (2004) The evolution of city size distributions. *Handbook of Regional and Urban Economics* 4, 2341–2378. MR1907061
[16] Geweke, J., Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4(4), 221–238.
[17] Gomes, M. I., de Haan, L., Henriques-Rodrigues, L. (2008) Tail index estimation for heavy tailed models: accomodation of bias in weighted log-excesses. *J. Royal Statistical Society B*, 70(1), 31–52. MR0738585
[18] Gomes, M. I., Figueiredo, F., & Neves, M. M. (2012). Adaptive estimation of heavy right tails: resampling-based methods in action. *Extremes*, 15(4), 463–489. MR2412630
[19] Gomes, M. I., & Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review*, 83(2), 263–292. MR2992550
[20] Gomes, M. I., Brilhante, M. F., & Pestana, D. (2016). New reduced-bias estimators of a positive extreme value index. *Communications in Statistics-Simulation and Computation*, 45(3), 833–862. MR3377081
[21] Grahovac, D., Jia, M., Leonenko, N. N., Taufer, E. (2015) Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data. *Statistics: A Journal of Theoretical and Applied Statistics* 49, 1221–1242. MR3415217
[22] Haeusler, E., Teugels, J. L. (1985) On asymptotic normality of Hill's esti-

mator for the exponent of regular variation. *The Annals of Statistics* 13(2), 743–756. MR3415217

[23] Hall, P. (1982) On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society Series B* (Methodological) 44(1), 37–42. MR0790569

[24] Hall, P., Welsh, A. H. (1985) Adaptive estimates of parameters of regular variation. *The Annals of Statistics* 13(1), 331–341. MR0655370

[25] Hassler, U., Marmol, F., & Velasco, C. (2006). Residual log-periodogram inference for long-run relationships. *Journal of Econometrics*, 130(1), 165–207. MR0773171

[26] Hill, B. M. (1975) A simple general approach to inference about the tail of a distribution. *The Annuals of Statistics* 3(5), 1163–1174. MR2208905

[27] Hsing, T. (1991) On tail index estimation using dependent data. *The Annals of Statistics* 19(3), 1547–1569. MR0378204

[28] Jia, M. (2014). *Heavy-tailed Phenomena and Tail Index Inference*. Ph.D. Thesis, University of Trento. MR1126337

[29] Kratz, M. F., Resnick, S. I. (1996) The QQ-estimator and heavy tails. *Comm. Statist. Stochastic Models* 12(4), 699–724.

[30] Levy, M. (2009) Gibrat's law for (all) cities: comment. *The American Economic Review* 99(4), 1672–1675. MR1410853

[31] Marra, G., & Wood, S.N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55(7), 2372–2387.

[32] Mason, D. M. (1982) Laws of large numbers for sums of extreme values. *The Annals of Probability* 10(3), 754–764. MR2786996

[33] McElroy, T., Politis, D.N. (2007) Moment-based tail index estimation. *Journal of Statistical Planning and Inference* 137(4), 1389–1406. MR0659544

[34] Meerschaert, M. M., Scheffler, H. P. (1998) A simple robust estimator for the thickness of heavy tails. *Journal of Statistical Planning and Inference* 71(1), 19–34. MR2301486

[35] Paulauskas, V., & Vaičiulis, M. (2017). A class of new tail index estimators. *Annals of the Institute of Statistical Mathematics*, 69(2), 461–487. MR1651847

[36] Paulauskas, V., & Vaičiulis, M. (2013). On an improvement of Hill and some other estimators. *Lithuanian Mathematical Journal*, 53(3), 336–355. MR3611528

[37] Pickands III, J. (1975) Statistical inference using extreme order statistics. *The Annals of Statistics* 3(1), 119–131. MR3097309

[38] Pitman, E. J. G. (1968) On the behaviour of the characteristic function of a probability distribution in the neighbourhood of the origin. *Journal of the Australian Mathematical Society* 8(3), 423–443. MR0423667

[39] Politis, D. N. (2002) A new approach on estimation of the tail index. *Comptes Rendus Mathematique* 335(3), 279–282. MR0231423

[40] Resnick, S., Stărică, C. (1995) Consistency of Hill's estimator for dependent data. *Journal of Applied Probability* 32(1), 139–167. MR1933674

[41] Resnick, S., Stărică, C. (1997) Smoothing the Hill estimator. *Advances in Applied Probability* 29(1), 271–293.
[42] Robinson, P. M. (1995) Log-periodogram regression of time series with long range dependence. *Ann. Statist.* 23(3), 1048–1072. MR1432940
[43] Welsh, A. H. (1986) On the use of the empirical distribution and characteristic function to estimate parameters of regular variation. *Australian Journal of Statistics* 28(2), 173–181. MR1345214
[44] Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B* (Statistical Methodology), 65(1), 95–114. MR0860462
[45] Wood, S. N. (2006). *Generalized additive models: an introduction with R.* CRC press. MR1959095