Journal of Computational and Applied Mathematics 436 (2024) 115347

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Reduced bias estimation of the shape parameter of the log-logistic distribution

Frederico Caeiro *, Ayana Mateus

Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa, Portugal
Centro de Matemática e Aplicações (CMA), Universidade Nova de Lisboa, Portugal

ARTICLE INFO

Article history:
Received 15 September 2022
Received in revised form 7 April 2023

Keywords:
Log-logistic distribution
Shape parameter
Bias-reduction
Hill estimator

ABSTRACT

In the literature, the log-logistic distribution is commonly presented with two parameters: one that governs the shape of the model, and the other that governs its scale. However, to make this model more suitable for data analysis, an additional location parameter can be added, resulting in the three-parameter or shifted log-logistic model. In this paper, we introduce a new estimator for the shape parameter of a three-parameter log-logistic distribution that reduces bias. We also derive various properties of the proposed estimator. Additionally, a simulation study and an application example to a real data set are conducted to examine the efficiency for finite sample sizes. The theoretical and simulated results confirm that our proposed estimation method performs significantly better than other estimation methods found in the literature.

1. Introduction

The log-logistic distribution has been successfully applied in different fields of research, including income distribution [1], medicine [2], environmental science [3], survival analysis [4], demographic data [5], forestry [6] or hydrology [7,8]. In the economic field, this probability model is known as the Fisk [1] distribution. The log-logistic model is related to the well-known logistic distribution. Specifically, let us consider two random variables X and T, which are related by the equation $T = \alpha \ln(X/\sigma)$, $\alpha, \sigma > 0$ and T has a logistic distribution with probability density function (p.d.f.)

$$g(t) = \frac{e^t}{(1+e^t)^2}, \quad t \in \mathbb{R}. \tag{1}$$

Then, X has a two-parameter log-logistic distribution with p.d.f. given by

$$f(x) = \frac{\frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1}}{\left(1+\left(\frac{x}{\sigma}\right)^{\alpha}\right)^2}, \quad x > 0, \tag{2}$$

where α is the shape parameter and σ is the scale parameter. Note that this model can also be derived from the quotient of two independent generalized gamma variables (Malik [9]). Additionally, the log-logistic distribution is a member of Burr's type XII [10] and Dagum's [11] family of distributions. For the special case where the scale parameter σ is equal to 1, Balakrishnan et al. [12] noticed the following relation

$$xf(x) = \alpha F(x)(1-F(x)), \quad \alpha > 0,$$

* Corresponding author.
E-mail addresses: fac@fct.unl.pt (F. Caeiro), amf@fct.unl.pt (A. Mateus).

https://doi.org/10.1016/j.cam.2023.115347

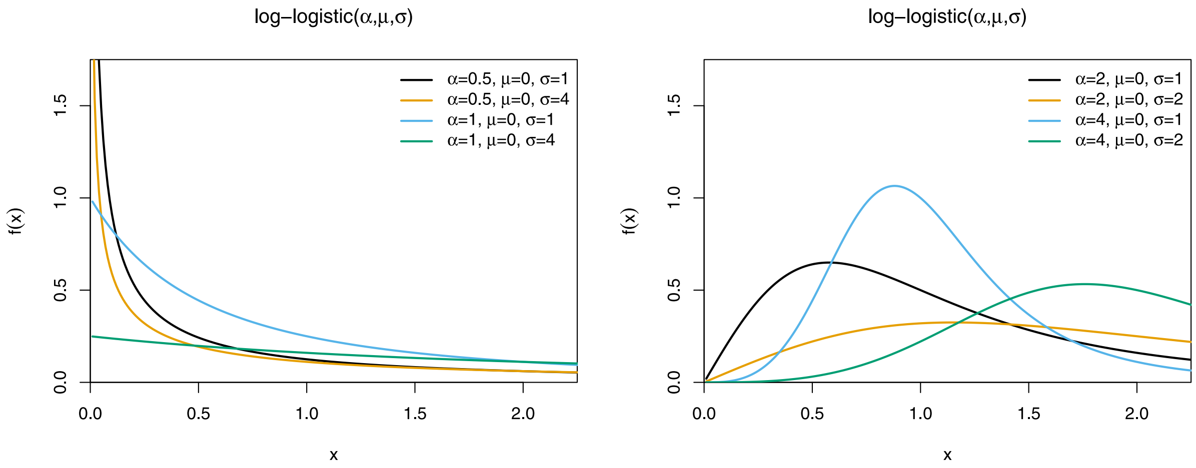

Fig. 1. Probability density function of the log-logistic distribution for $\alpha = 0.5,\ 1,\ \mu = 0,\ \sigma = 1,\ 4$ (left) and $\alpha = 2,\ 4,\ \mu = 0,\ \sigma = 1,\ 4$ (right).

where F is the distribution function (d.f.). The log-logistic p.d.f. in (2) is right skewed and can take different shapes. If $\alpha \le 1, f$ is decreasing, while if $\alpha > 1, f$ is unimodal with a mode at the value $\sigma(\frac{\alpha-1}{\alpha+1})^{1/\alpha}$. Moreover, the p.d.f. can have a similar shape to the log-normal d.f., however it has heavier tails and the advantage of being mathematically more tractable (Singh et al. [13]). The rth order moments about the origin only exists if $r < \alpha$ and are given by

$$E(X^r) = \sigma^r \mathrm{B}\left(1 - \frac{r}{\alpha}, 1 + \frac{r}{\alpha}\right),$$

where $\mathrm{B}(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ and $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$ represents, respectively, the Beta and the Gamma functions and a and b are positive real values. Thus, the mean of X only exist if $\alpha > 1$. Regarding the distribution of order statistics, the density function of the ith ascending order statistic from a random sample of size n, $X_{i:n}$, has a simple closed form. Clearly, the p.d.f. is

$$f_{X_{i:n}}(x) = \frac{\frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha i-1}}{\mathrm{B}(i, n-i+1)\left(1 + \left(\frac{x}{\sigma}\right)^{\alpha}\right)^{n+1}}, \quad x > 0,$$

which corresponds to the p.d.f. of a generalized beta distribution of the second kind (McDonald [14]).

Recent research has focused on new extensions of the two-parameter log-logistic model, in order to generate more suitable probability distributions for data modeling. One easy way to develop a new generalized distribution is by adding one or more parameters to an existing distribution. For a survey on generalizations of the log-logistic distribution see Muse et al. [15]. The primary focus of the current study will be directed towards one of these generalizations, namely, the three-parameter log-logistic distribution, which is alternatively referred to as the shifted log-logistic distribution or Pareto type III distribution (see Arnold [16]). A random variable X is said to have a three-parameter log-logistic distribution with the shape parameter $\alpha > 0$, the location parameter $\mu \in \mathbb{R}$, and with the scale parameter $\sigma > 0$, if its d.f. is given by

$$F(x|\alpha, \mu, \sigma) = 1 - \frac{1}{1 + \left(\frac{x-\mu}{\sigma}\right)^{\alpha}} = \frac{1}{1 + \left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}, \quad x > \mu. \tag{3}$$

and we denote this by $X \sim LL(\alpha, \mu, \sigma)$. The corresponding quantile function is

$$Q_X(p|\alpha, \mu, \sigma) = \mu + \sigma\left(\frac{p}{1-p}\right)^{1/\alpha}, \quad 0 \le p < 1. \tag{4}$$

The p.d.f. associated to the distribution function in (3) is then expressed as

$$f(x|\alpha, \mu, \sigma) = \frac{\alpha\left(\frac{x-\mu}{\sigma}\right)^{\alpha-1}}{\left(1 + \left(\frac{x-\mu}{\sigma}\right)^{\alpha}\right)^2}, \quad x > \mu, \tag{5}$$

If $\alpha = 1$, X has a location and scale beta prime distribution. Additionally, a three-parameter log-logistic model with $\alpha = 1$ and $\mu = \sigma$ is a Pareto type I model with shape parameter equal to 1 and scale parameter σ. Fig. 1 illustrate the p.d.f. in (5) for selected values of the shape and scale parameters. The location parameter is always equal to zero.

The estimation of the parameters of the log-logistic distribution has been discussed by many authors. For the two-parameter model we mention the papers [17–22], among others. Despite the fact that the two-parameter case has received greater attention, several estimation techniques for the three-parameter log-logistic model are presently available in the literature. Ahmad et al. [7] discussed several methods to estimate the three-parameter log-logistic distribution for flood frequency analysis and presented a new estimation method based on generalized least squares. Singh et al. [13] employed the principle of maximum entropy (POME) to derive a new method to estimate the three-parameter log-logistic distribution. El-Rahman and El Genidy [23] developed an algorithm of percentile roots which combine percentile equations with a measures of central tendency. Balakrishnan et al. [12] derived the best linear unbiased estimators (BLUE) for the location and scale parameters of log-logistic model, with a known shape parameter. In real data applications, it is unrealistic to assume that the shape parameter α is known, and it should be estimated. Recently, Ahsanullah and Alzaatreh [24] considered the BLUE for the location and scale parameters and estimated the shape parameter using the reciprocal of a location-invariant Hill-type estimator. This estimator achieves location-invariant property by shifting it by the sample minimum. Based on empirical studies, the same authors proposed a threshold of 10% of the sample size. More recently, Mateus and Caeiro [25] improved the estimation of the shape parameter of the log-logistic model with alternative reduced bias estimators. Such bias reduction results from the well-known trade-off between bias and variance. The main goal of this paper is to further improve the trade-off between bias and variance and introduce a more efficient estimator of the parameter α.

The organization of this paper is as follows. In the next section we review several competitive estimators and define a new reduced bias estimator for the shape parameter of a log-logistic distribution. In Section 3 we establish the asymptotic behavior of the shape estimators under consideration. In addition, we discuss the optimal sample fraction to be used by each estimator. Section 4 is devoted to a small Monte-Carlo simulation study. To illustrate the use of the new reduced bias estimator, an application to a real data set is provided in Section 5. Finally, some conclusions of the main results achieved with this research are presented in Section 6.

2. Estimation of the shape parameter

Let $(X_1, X_2, \ldots, X_n)$ be a sample of n independent and identically distributed random variables with a common three-parameter log-logistic d.f., given in (3). Let $(X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n})$ denote the associated sample of non-decreasing order statistics.

2.1. Location invariant Hill-type estimator

Ahsanullah and Alzaatreh [24] noted that the reduced log-logistic distribution, $F(x|\alpha, 0, 1)$, has a Pareto-type tail (see also Fisk [1]). More precisely,

$$1 - F(x|\alpha, 0, 1) = x^{-\alpha} l(x), \tag{6}$$

with α the tail or Pareto index and $l(x) = (1 + x^{-\alpha})^{-1}$, $x > 0$ a slowly varying function at infinity that measures the departure of F to the Pareto Type I distribution. Moreover, since $l(x)$ admits the Taylor expansion $l(x) = 1 - x^{-\alpha} + o(x^{-\alpha})$, as $x \to \infty$, the reduced log-logistic model belongs to Hall's class (see [26], Eq. (1)) of Pareto-type models with survival function,

$$1 - F(x) = \left(\frac{x}{c}\right)^{-1/\xi} \left(1 + d\left(\frac{x}{c}\right)^{\rho/\xi} + o\left(x^{\rho/\xi}\right)\right), \quad x \to \infty \tag{7}$$

with $\xi = 1/\alpha > 0$, $c = 1$, $d = -1$ and $\rho = -1$. The class of models verifying (6) or (7) plays an important role in tail inference. Such models are in the domain of attraction for maxima of the Fréchet extreme value distribution. The shape parameter ξ in (7) is known as the extreme value index and is a crucial parameter in tail inference. In the statistics literature, numerous estimators have been proposed for ξ, or equivalently, for the shape parameter α. For a general overview of the available estimators we refer to [27–30]. When dealing with Pareto-type models, one of the most popular estimator for the extreme value index is the one introduced by Hill [31]. This estimator is defined as the average of the log-excesses over the threshold $u = X_{n-k:n} > 0$,

$$H(k) = \frac{1}{k} \sum_{i=1}^{k} \ln X_{n-i+1:n} - \ln X_{n-k:n}, \qquad k = 1, 2, \ldots, n-1. \tag{8}$$

where $k + 1$ represents the number of upper order statistics used in the estimation. Due to theoretical motives, the threshold k is often assumed to be intermediate, i.e., $k = k_n \in [1, n-1]$ is assumed to be a sequence of positive integers satisfying

$$k \to \infty \quad \text{and} \quad k/n \to 0, \quad \text{as} \qquad n \to \infty. \tag{9}$$

Many of the estimators that have been suggested to estimate the parameter ξ, including the one in (8), are only scale-invariant. A change in the location can modify the estimator behavior (for more information see papers [32–34]). The properties of the Hill estimator and the fact that the d.f. in (3) has a location parameter, lead Ahsanullah and Alzaatreh [24] to propose the estimation of the shape parameter α with the following location-invariant Hill-type estimator,

$$\hat{\alpha}^H(k) = \frac{1}{\tilde{H}(k)}, \tag{10}$$

with

$$\tilde{H}(k)=\frac{1}{k}\sum_{i=1}^{k}\ln\frac{X_{n-i+1:n}-X_{1:n}}{X_{n-k:n}-X_{1:n}},\quad 1\leq k\leq n-2. \tag{11}$$

Note that the estimator $\tilde{H}(k)$ in (10) belongs to the class of estimators discussed in [32,35,36]. Furthermore, the estimator in (10) share some properties with $\hat{\alpha}^{H}(k)$ in (8). In particular, the variance decreases and the absolute bias increases, as the value k increases. Therefore, the choice of k leads to a trade-off between the bias and the variance of the estimator. Regarding the choice of the parameter k, Ahsanullah and Alzaatreh [24] proposed $k=\lfloor n/10\rfloor$, if $n>100$, where $\lfloor x\rfloor$ denotes the integer part of x.

2.2. A generalized Jackknife estimator

Despite its wide use, the Hill estimator is difficult to apply in real-world data situations due to its significant bias. This issue motivated several researchers to construct alternative reduced bias estimators, less sensitive to the choice of the threshold k. We refer to the first reduced bias estimators of the extreme value index in papers [37,38] (see also the papers [39,40] for a general overview on bias reduction). Bias reduction of tail parameter estimators typically requires the estimation of tail second order parameters such as ρ in (7). Several authors identify challenges in the estimation of ρ and assume $\rho=-1$ (see [41], section 4.5 for related discussions), which also corresponds to the value of ρ for the log-logistic model. In the context of Pareto Type tails, Gomes et al. [42] (see also [43]) considered the following Generalized Jackknife (GJ) estimator of the extreme value index

$$GJ(k)=2MR(k)-H(k)=\frac{M^{(2)}(k)}{M^{(1)}(k)}-M^{(1)}(k),\quad k=1,2,\dots,n-1, \tag{12}$$

and

$$M^{(j)}(k)=\frac{1}{k}\sum_{i=1}^{k}\left(\ln X_{n-i+1:n}-\ln X_{n-k:n}\right)^{j},\quad j>0,\quad (M^{(1)}(k)\equiv H(k)), \tag{13}$$

are the moments of order j of the log-excesses. The estimator in (12) belongs to the class of GJ statistic introduced in Peng [38] and is based on the Hill and the moments ratio (MR) estimators (Danielsson et al. [44]). Moreover, both $H(k)$ and MR(k) are members of Lehmer's mean-of-order-p class of extreme value index estimators studied in papers [30,45,46]. Mateus and Caeiro [25] modified the generalized Jackknife statistic in (12) and proposed the following location invariant GJ estimator for the shape parameter of a log-logistic model

$$\hat{\alpha}^{GJ}(k)=\frac{\tilde{M}^{(1)}(k)}{\tilde{M}^{(2)}(k)-(\tilde{M}^{(1)}(k))^{2}},\quad k=1,2,\dots,n-2, \tag{14}$$

with

$$\tilde{M}^{(j)}(k)=\frac{1}{k}\sum_{i=1}^{k}\left(\ln\frac{X_{n-i+1:n}-X_{1:n}}{X_{n-k:n}-X_{1:n}}\right)^{j},\quad j>0, \tag{15}$$

the moments of order j of the shifted log-excesses. The estimator $\hat{\alpha}^{GJ}(k)$ has a null dominant component of asymptotic bias and a higher asymptotic variance than $\hat{\alpha}^{H}(k)$ in (10) (Mateus and Caeiro [25]).

2.3. A new reduced bias estimator

Although the GJ estimator in (14) is asymptotically unbiased, it has the side effect of having a much higher variance than the estimator $\hat{\alpha}^{H}(k)$ in Eq. (10). Therefore, it seems relevant to study more efficient estimators for the shape parameter α. Mateus and Caeiro ([25], Proposition 1) derived the non-degenerated asymptotic behavior of $\tilde{H}(k)$ and computed the first-order term of the asymptotic bias. Hence, a simple reduced-bias estimator of the shape parameter can then be constructed as

$$\hat{\alpha}^{RBH}(k)=\frac{1}{\tilde{H}^{RB}(k)}, \tag{16}$$

where

$$\tilde{H}^{RB}(k)=\frac{1}{k}\sum_{i=1}^{k}\ln\frac{X_{n-i+1:n}-X_{1:n}}{X_{n-k:n}-X_{1:n}}\times\left(1-\frac{k}{2n}\right),\quad 1\leq k\leq n-2 \tag{17}$$

is a reduced-bias estimator of $1/\alpha$.

3. Asymptotic properties of the estimators

In this section, we present the asymptotic results for all estimators of the shape parameter under consideration. Higher-order bias terms will be computed to evaluate the asymptotic bias of the reduced-bias estimators. We achieve this goal by assuming that k is an intermediate sequence of integer values, satisfying (9).

3.1. Asymptotic normality of the estimators

The following two theorems extends the results in Proposition 1 and Proposition 2 from Mateus and Caeiro [25].

Theorem 3.1. *Assume that k is an intermediate sequence of integers satisfying (9). Then, the following distributional representation*

$$\hat{\alpha}^H(k) \overset{d}{=} \alpha\left(1-\frac{Z_k}{\sqrt{k}}+O_p\left(\frac{\sqrt{k}}{n}\right)-\frac{k}{2n}-\frac{1}{12}\left(\frac{k}{n}\right)^2(1+o_p(1))\right)+\left(\frac{k}{n}\right)^{\frac{1}{\alpha}}o_p(1) \tag{18}$$

is valid, where Z_k is an asymptotically standard normal random variable. Consequently, the asymptotic variance of $\hat{\alpha}^H(k)$ is α^2/k and the dominant term of asymptotic bias is always negative and equal to $-k/(2n)$. Moreover, if $\sqrt{k}(k/n)\to c$, then

$$\sqrt{k}(\hat{\alpha}^H(k)-\alpha) \overset{d}{\longrightarrow} N\left(-\alpha c/2, \alpha^2\right). \tag{19}$$

Theorem 3.2. *Assume the conditions of Theorem 3.1. Then, the following distributional representation*

$$\hat{\alpha}^{GJ}(k) \overset{d}{=} \alpha\left(1-\frac{\sqrt{5}Z_k^{GJ}}{\sqrt{k}}+O_p\left(\frac{\sqrt{k}}{n}\right)+\frac{1}{36}\left(\frac{k}{n}\right)^2(1+o_p(1))\right)+\left(\frac{k}{n}\right)^{\frac{1}{\alpha}}o_p(1). \tag{20}$$

holds, where Z_k^{GJ} is an asymptotically standard normal random variable. Consequently, the asymptotic variance of $\hat{\alpha}^{GJ}(k)$ is $5\alpha^2/k$ and the dominant term of asymptotic bias is positive and equal to $\frac{1}{36}\left(\frac{k}{n}\right)^2$. Additionally, if $\sqrt{k}(k/2n)\to c$, then

$$\sqrt{k}(\hat{\alpha}^{GJ}(k)-\alpha) \overset{d}{\longrightarrow} N\left(\frac{\alpha c}{36}, 5\alpha^2\right). \tag{21}$$

Clearly, the best rate of convergence is achieved when $c \neq 0$.

Finally, we establish the asymptotic normality of the proposed new estimator for the shape parameter in (16).

Theorem 3.3. *Assume the conditions of Theorem 3.1. Then, we have the following distributional representation*

$$\hat{\alpha}^{RBH}(k) \overset{d}{=} \alpha\left(1-\frac{Z_k}{\sqrt{k}}+O_p\left(\frac{\sqrt{k}}{n}\right)-\frac{1}{12}\left(\frac{k}{n}\right)^2(1+o_p(1))\right)+\left(\frac{k}{n}\right)^{\frac{1}{\alpha}}o_p(1), \tag{22}$$

where Z_k is an asymptotically standard normal random variable. If $\sqrt{k}(k/n)\to c$, then

$$\sqrt{k}(\hat{\alpha}^H(k)-\alpha) \overset{d}{\longrightarrow} N\left(0, \alpha^2\right). \tag{23}$$

This implies that $\hat{\alpha}^{RBH}(k)$ has lower order bias terms and the same asymptotic variance as $\hat{\alpha}^H(k)$ in (10). Thus, we expect $\hat{\alpha}^{RBH}(k)$ to be superior to $\hat{\alpha}^H(k)$, for every k, with respect to mean squared error.

3.2. Selection of the threshold

In practical applications, the threshold is fundamental to yield accurate estimates and it must be chosen before applying any of the aforementioned shape parameter estimators. Let us denote the optimal threshold by

$$k_0(n) = \arg\min_k \mathrm{MSE}(\hat{\alpha}(k)),$$

with MSE standing for mean squared error. When the MSE is unknown, a typical approach is to choose the threshold through the minimization of the asymptotic mean squared error (AMSE),

$$\hat{k}_0 = \hat{k}_0(n) = \arg\min_k \mathrm{AMSE}(\hat{\alpha}(k)) = k_0(n)(1+o(1)).$$

Since such choice depends on asymptotic arguments, it may only be reliable when the sample size becomes large. Alternative methods for selecting the threshold can be found in Refs. [47–49]. From (18) it follows that the AMSE of $\hat{\alpha}^H(k)$ is given by

$$AMSE(\hat{\alpha}^H(k)) = \alpha^2\left(\frac{1}{k}+\frac{k^2}{4n^2}\right). \tag{24}$$

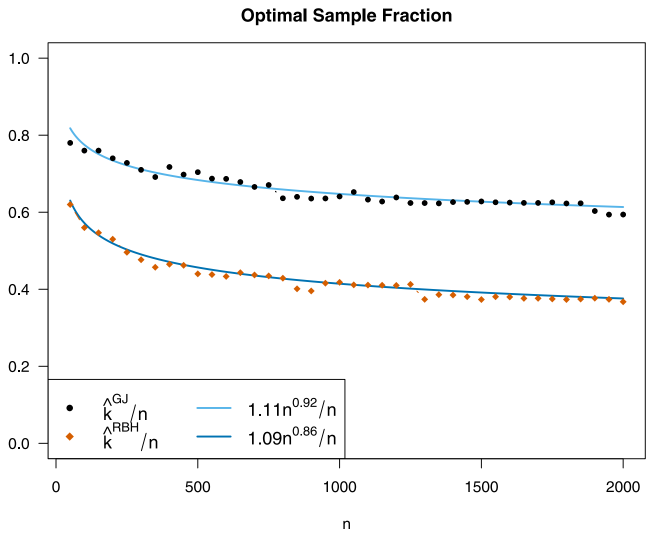

Fig. 2. Empirical optimal sample fraction for $\hat{\alpha}^{GJ}(k)$ and $\hat{\alpha}^{RBH}(k)$ and the corresponding power regression curves.

Then, the level k_A that minimizes the AMSE in Eq. (24) is

$$\hat{k}_A^H = \arg\min[AMSE(\hat{\alpha}^H(k))] = (2n^2)^{1/3}. \tag{25}$$

Regarding the reduced-bias estimators in (14) and (16), research shows that the optimal level that minimizes the corresponding AMSE is often unreliable for small sample sizes. This poor approximation can be justified by the difference between exact and asymptotic distributional behaviour of the estimators. We propose a more reliable approximation, based on empirical results. We follow the approach of Mateus and Caeiro [25] and propose the following algorithm.

Algorithm 3.1 (*Empirical Threshold Selection*)**.**

1. Generate 5000 samples of size n from a log-logistic distribution, with n taking values between 50 and 2000, with step 50.
2. Let $\hat{\alpha}(k, i, n)$ denote the estimates based on the ith sample of size n. For each sample size,
   - compute $\hat{\alpha}(k, i, n)$, $k = 1, 2, \ldots, n-2$, $i = 1, \ldots, 5000$.
   - compute the empirical Mean Squared Error as a function of k.
   - Obtain the level $\hat{k}(n)$ that minimizes the empirical Mean Squared Error.
3. Finally, perform a Power Regression with $\hat{k}_0(n)$ the response variable and the sample size as the predictor variable. The regression coefficients are the vector values (a_1, a_2). The empirical threshold is then given by

$$\hat{k}_E = \hat{k}_E(n) = \lfloor a_1 n^{a_2} \rfloor, \tag{26}$$

   where $\lfloor x \rfloor$ denotes the integer part of x.

Remark 3.1. The agreement between the empirical threshold and the fitted curve, given in Algorithm 3.1, can be changed with the modification of the values in step 1.

We applied Algorithm 3.1 to samples of the log-logistic model with parameters $(\alpha, \mu, \sigma) = (1,1,1)$ and the aforementioned reduced bias estimators $\hat{\alpha}^{GJ}(k)$ and $\hat{\alpha}^{RBH}(k)$. The algorithm provided the empirical thresholds

$$k_E^{GJ} = \lfloor 1.11n^{0.92} \rfloor \quad \text{for} \quad \hat{\alpha}^{GJ}(k) \qquad \text{and} \qquad \hat{k}_E^{RBH} = \lfloor 1.09n^{0.86} \rfloor \quad \text{for} \quad \hat{\alpha}^{RBH}(k). \tag{27}$$

Fig. 2 presents the empirical optimal sample fraction and the corresponding regression curve for both estimators. The overall agreement between empirical and fitted curves is quite good.

3.3. Proofs

We first provide several Lemmas that are useful in the derivation of the main results. Let $E_{1:n} \le E_{2:n} \le \cdots \le E_{n:n}$ be the order statistics from n mutually independent and identically distributed exponential random variables $E_1, E_2, \ldots, E_n$, with a common d.f. $F_E(x) = 1 - e^{-x}$, $x > 0$.

Lemma 3.1 (*Balakrishnan and Basu [50]*). *Considering the convention that $E_{0:n} \equiv 0$, we have*

$$E_{j:n} - E_{i:n} \stackrel{d}{=} E_{j-i:n-i}, \quad 1 \le i \le j \le n.$$

Lemma 3.2 (*Girard [51]*). *Suppose k is an intermediate sequence, i.e., (9) holds. Then,*

$$\frac{E_{n-i+1:n}}{\ln(n/i)} \xrightarrow{p} 1, \quad i = 1, \ldots, k.$$

Lemma 3.3 (*Araújo et al. [36]*). *Assume that the quantile function Q satisfies the following second order regular variation condition*

$$\lim_{t\to\infty} \frac{\frac{Q(1-1/tx)}{Q(1-1/t)} - x^{\xi}}{A(t)} = x^{\xi}\frac{x^{\rho}-1}{\rho}, \tag{28}$$

for all $x > 1$, where $\xi > 0$ and ρ are, respectively, positive and negative real numbers and the function $A(t)$ satisfies for any $x > 0$,

$$\lim_{t\to\infty} \frac{A(tx)}{A(t)} = x^{\rho}.$$

Then, for any intermediate sequence k, such that (9) holds,

$$\frac{1}{k}\sum_{i=1}^{k}\left(\ln\frac{X_{n-i+1:n} - X_{[np]+1:n}}{X_{n-k:n} - X_{[np]+1:n}}\right)^{r} - \frac{1}{k}\sum_{i=1}^{k}\left(\ln\frac{X_{n-i+1:n} - \chi_p}{X_{n-k:n} - \chi_p}\right)^{r} = o_p\left(\frac{1}{Q(1-k/n)}\right), \quad r = 1, 2,$$

where χ_p denotes the quantile of order p $(0 \le p < 1)$ for the random variable X.

Lemma 3.4 (*Gomes et al. [42]*). *Suppose (9) holds and define*

$$P_k = \sqrt{k}\frac{1}{k}\sum_{i=1}^{k}(E_i - 1) \tag{29}$$

and

$$Q_k = \sqrt{k}\frac{1}{k}\sum_{i=1}^{k}(E_i^2 - 2). \tag{30}$$

Then, (P_k, Q_k) is asymptotically bivariate normal distributed with null mean, with variances 1 and 20, respectively, and covariance equal to 4.

3.4. Proof of Theorem 3.1

Consider first the random variable

$$T_1(k) = \frac{1}{k}\sum_{i=1}^{k}\ln\frac{X_{n-i+1:n} - \mu}{X_{n-k:n} - \mu}. \tag{31}$$

Next, define the random variable $\tilde{X} = (X - \mu)/\sigma \sim LL(\alpha, 0, 1)$ with quantile function denoted by $\tilde{Q}(p) = Q_X(p|\alpha, 0, 1)$, $0 \le p < 1$. From the inverse probability integral transform we have that $\tilde{X} \stackrel{d}{=} \tilde{Q}(1 - e^{-E}) = (e^{E}(1 - e^{-E}))^{1/\alpha}$. Also, since $\tilde{Q}$ is monotonically increasing, $\tilde{X}_{i:n} \stackrel{d}{=} \tilde{Q}(1 - e^{-E_{i:n}})$, $1 \le i \le n$. Consequently

$$\begin{aligned} T_1(k) &\stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\ln\frac{\tilde{Q}(1 - e^{-E_{n-i+1:n}})}{\tilde{Q}(1 - e^{-E_{n-k:n}})} \\ &= \frac{1}{k\,\alpha}\sum_{i=1}^{k}\ln\frac{e^{E_{n-i+1:n}}(1 - e^{-E_{n-i+1:n}})}{e^{E_{n-k:n}}(1 - e^{-E_{n-k:n}})} \\ &= \frac{1}{\alpha}\left(\frac{1}{k}\sum_{i=1}^{k}(E_{n-i+1:n} - E_{n-k:n}) + \frac{1}{k}\sum_{i=1}^{k}\ln\frac{1 - e^{-E_{n-i+1:n}}}{1 - e^{-E_{n-k:n}}}\right). \end{aligned}$$

Using Lemma 3.1 and the Taylor expansion $\ln(1+x) \sim x - \frac{x^2}{2}$, as $x \to 0$,

$$T_1(k) \stackrel{d}{=} \frac{1}{\alpha}\left(\frac{1}{k}\sum_{i=1}^{k}E_{k-i+1:k} + \frac{1}{k}\sum_{i=1}^{k}(e^{-E_{n-k:n}} + \frac{1}{2}e^{-2E_{n-k:n}} - e^{-E_{n-i+1:n}} - \frac{1}{2}e^{-2E_{n-i+1:n}})\right)(1 + o_p(1))$$

$$\overset{d}{=} \frac{1}{\alpha}\left(\frac{1}{k}\sum_{i=1}^{k} E_i + e^{-E_{n-k:n}}\frac{1}{k}\sum_{i=1}^{k}(1-e^{-E_{k-i+1:k}}) + \frac{1}{2}e^{-2E_{n-k:n}}\frac{1}{k}\sum_{i=1}^{k}(1-e^{-2E_{k-i+1:k}})(1+o_p(1))\right)$$
$$= \frac{1}{\alpha}\left(\frac{1}{k}\sum_{i=1}^{k} E_i + e^{-E_{n-k:n}}(1-\frac{1}{k}\sum_{i=1}^{k}e^{-E_i}) + \frac{1}{2}e^{-2E_{n-k:n}}(1-\frac{1}{k}\sum_{i=1}^{k}e^{-2E_i})(1+o_p(1))\right)$$

From the weak law of large numbers for independent and identically distributed random variables, $\frac{1}{k}\sum_{i=1}^{k} e^{-tE_i} \overset{p}{\longrightarrow} \frac{1}{1+t}$, $t = 1, 2$. Also by Lemmas 3.1, 3.2 and 3.4 we have

$$T_1(k) \overset{d}{=} \frac{1}{\alpha}\left(1 + \frac{Z_k}{\sqrt{k}} + \frac{1}{2}e^{-\ln(n/k)} + \frac{1}{3}e^{-2\ln(n/k)}(1+o_p(1))\right)$$

where $Z_k = P_k$ is an asymptotic standard normal random variable by the central limit theorem.

Note that condition (28) holds for the quantile function in (4) and $X_{1:n} \overset{p}{\longrightarrow} \chi_0 = \mu$. Then, by Lemma 3.3,

$$\tilde{H}(k) = \tilde{M}^{(1)}(k) = \frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{n-i+1:n} - X_{1:n}}{X_{n-k:n} - X_{1:n}} \overset{d}{=} T_1(k) + \left(\frac{k}{n}\right)^{\frac{1}{\alpha}} o_p(1), \tag{32}$$

and using a second order Taylor expansion for $(1-x)^{-1}$, $x \to 0$, (18) and (19) follows straightforwardly.

3.5. Proof of Theorem 3.2

Let us introduce the random variable

$$T_2(k) = \frac{1}{k}\sum_{i=1}^{k}\left(\ln \frac{X_{n-i+1:n} - \mu}{X_{n-k:n} - \mu}\right)^2. \tag{33}$$

Repeating the same arguments, presented in the proof of Theorem 3.1 we have

$$T_2(k) \overset{d}{=} \frac{2}{\alpha^2}\left(1 + \frac{\sqrt{5}Z_k^{(2)}}{\sqrt{k}} + O_p\left(\frac{\sqrt{k}}{n}\right) + \frac{3}{4}\left(\frac{k}{n}\right) + \frac{11}{18}\left(\frac{k}{n}\right)^2 (1+o_p(1))\right)$$

where $Z_k^{(2)} = Q_k/\sqrt{20}$, with Q_k defined in Lemma 3.4, is an asymptotic standard normal random variable. Then, by Lemma 3.3,

$$\tilde{M}^{(2)}(k) \overset{d}{=} T_2(k) + \left(\frac{k}{n}\right)^{\frac{1}{\alpha}} o_p(1).$$

Next, it follows

$$\begin{aligned}\frac{1}{\tilde{M}^{(2)}(k) - (\tilde{M}^{(1)}(k))^2} \overset{d}{=} \alpha^2 &\left(1 - \frac{2\sqrt{5}Z_k^{(2)} - 2Z_k}{\sqrt{k}} + O_p\left(\frac{\sqrt{k}}{n}\right) - \frac{1}{2}\left(\frac{k}{n}\right) - \frac{1}{18}\left(\frac{k}{n}\right)^2\right)(1+o_p(1)) \\ &+ \left(\frac{k}{n}\right)^{\frac{1}{\alpha}} o_p(1).\end{aligned} \tag{34}$$

Theorem 3.2 now follows by combining the results in Eqs. (32) and (34) and denoting

$$Z_k^{GJ} = \frac{2\sqrt{5}Z_k^{(2)} - 3Z_k}{\sqrt{5k}}.$$

3.6. Proof of Theorem 3.3

First, from Theorem 3.1 we have

$$\begin{aligned}\tilde{H}^{RB}(k) &\overset{d}{=} T_1(k)\left(1 - \frac{k}{2n}\right) + \left(\frac{k}{n}\right)^{\frac{1}{\alpha}} o_p(1) \\ &\overset{d}{=} \frac{1}{\alpha}\left(1 + \frac{Z_k}{\sqrt{k}} + \frac{1}{12}\left(\frac{k}{n}\right)^{-2}(1+o_p(1))\right) + \left(\frac{k}{n}\right)^{\frac{1}{\alpha}} o_p(1)\end{aligned}$$

with $T_1(k)$ given in (31). Now, an application of Taylor expansion to $1/(1-x)$ gives the result in (18) and the (23) follows straightforwardly.

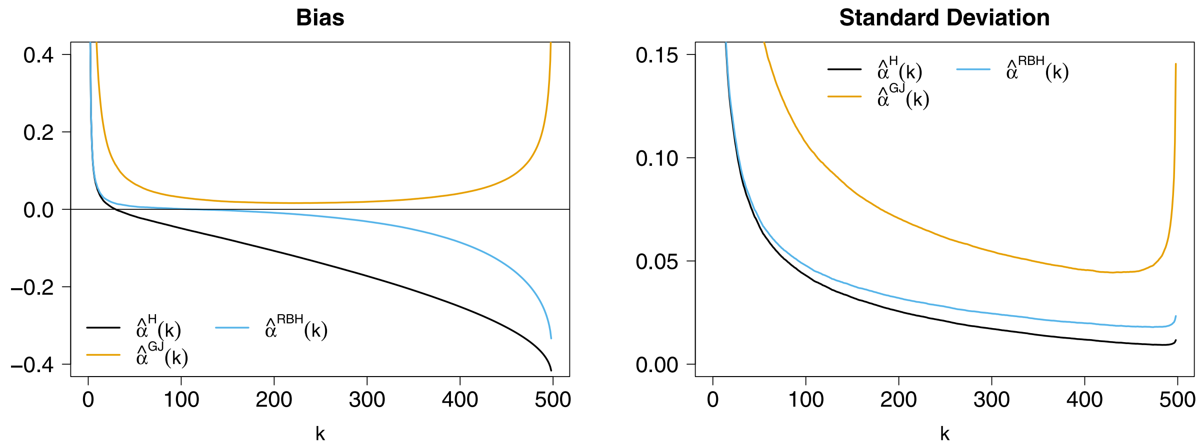

Fig. 3. Simulated bias (left) and standard deviation (right) of the estimators of α, versus k, for samples of size $n = 500$ from the log-logistic distribution with $(\alpha, \mu, \sigma) = (0.5, 1, 1)$.

4. Simulation study

In this section, we use Monte Carlo simulation to evaluate and compare the finite sample performance of the proposed new estimator of the shape parameter α in (16) with the estimators in (10) and (14). All the computations were done using R [52] software. We generated 10 000 samples, with sample sizes $n = 20, 50, 75, 100, 200, 500, 750, 1000, 2000, 5000$, from the three-parameter log-logistic model, through the inverse-transform method. The following parameter combinations were considered

- Case 1: $(\alpha, \mu, \sigma) = (0.5, 1, 1)$;
- Case 2: $(\alpha, \mu, \sigma) = (1.5, 1, 1)$;
- Case 3: $(\alpha, \mu, \sigma) = (2, 1, 1)$;
- Case 4: $(\alpha, \mu, \sigma) = (3, 1, 1)$;

Let $\hat{\alpha}^{\bullet}(k)$, with $\bullet \in \{H, GH, RBH\}$, denote any of the aforementioned estimators under study. Since the three estimators are all location and scale invariant, the choice of those two parameters is irrelevant. For each sample, the estimates of α are first computed for every k. We then computed the simulated bias and the root mean squared error (RMSE) for each k, where

$$Bias(\hat{\alpha}^{\bullet}(k)) = \bar{\alpha}^{\bullet}(k) - \alpha, \qquad SD(\hat{\alpha}^{\bullet}(k)) = \frac{1}{10\,000}\sum_{i=1}^{10\,000}(\hat{\alpha}^{\bullet}(k) - \bar{\alpha}^{\bullet}(k))^2 \tag{35}$$

with $\bar{\alpha}^{\bullet}(k) = \frac{1}{10\,000}\sum_{i=1}^{10\,000}\hat{\alpha}^{\bullet}(k)$ and the root mean squared error (RMSE)

$$RMSE(\hat{\alpha}^{\bullet}(k)) = \frac{1}{10\,000}\sum_{i=1}^{10\,000}(\hat{\alpha}^{\bullet}(k) - \alpha)^2.$$

We have further computed the simulated optimum level

$$\hat{k}_0^{(\bullet)} = \arg\min_k \mathrm{RMSE}[\hat{\alpha}^{\bullet}(k)], \tag{36}$$

and the simulated characteristics in (35) at the optimal level in (36). Furthermore, since in practical applications the optimal value of k is unknown, we also obtained the simulated bias, SD and RMSE of the estimators at the adaptive level proposed in Eqs. (25) and (27). For simplicity of notation, hereafter we shall denote

$$\hat{\alpha}_0^{H} = \hat{\alpha}^{H}(\hat{k}_0^{H}), \qquad \hat{\alpha}_0^{GJ} = \hat{\alpha}^{GJ}(\hat{k}_0^{GJ}), \qquad \text{and} \qquad \hat{\alpha}_0^{RBH} = \hat{\alpha}^{GJ}(\hat{k}_0^{RBH}), \tag{37}$$

$$\hat{\alpha}_A^{H} = \hat{\alpha}^{H}(\hat{k}_A^{H}), \qquad \hat{\alpha}_E^{GJ} = \hat{\alpha}^{RBH}(\hat{k}_E^{GJ}), \qquad \text{and} \qquad \hat{\alpha}_E^{RBH} = \hat{\alpha}^{RBH}(\hat{k}_E^{RBH}). \tag{38}$$

In Figs. 3 to 6, we present the Monte Carlo estimates of the bias (left) and SD (right), with respect to k, for the three different estimators and samples of size $n = 500$. A good performance is assessed by the flatness of the curve of the mean value in a large continuous region of values of k, close to the true value of α, as well as by a small SD.

For the four parameters combination here considered, $\hat{\alpha}^{GJ}(k)$ and $\hat{\alpha}^{RBH}(k)$ always yields a smaller (absolute) bias than $\hat{\alpha}^{H}(k)$. In addition, $\hat{\alpha}^{GJ}(k)$ provides a much wider region of estimates with near zero bias. Regarding the SD, we have

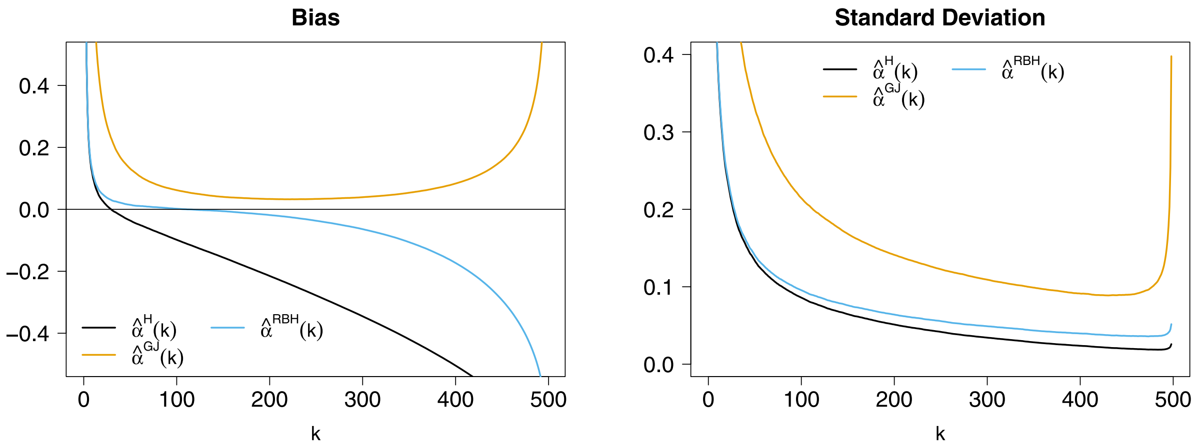

Fig. 4. Simulated bias (left) and standard deviation (right) of the estimators of α, versus k, for samples of size $n = 500$ from the log-logistic distribution with $(\alpha, \mu, \sigma) = (1, 1, 1)$.

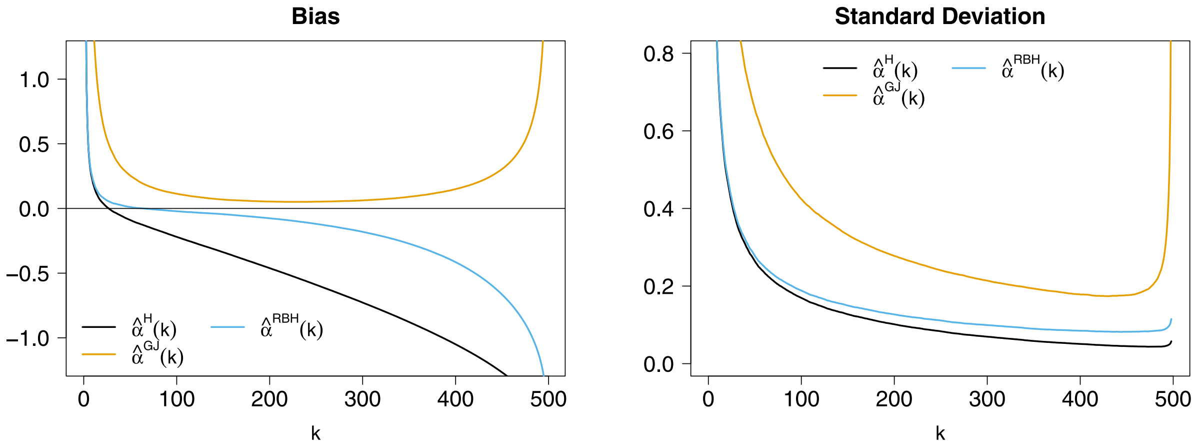

Fig. 5. Simulated bias (left) and standard deviation (right) of the estimators of α, versus k, for samples of size $n = 500$ from the log-logistic distribution with $(\alpha, \mu, \sigma) = (2, 1, 1)$.

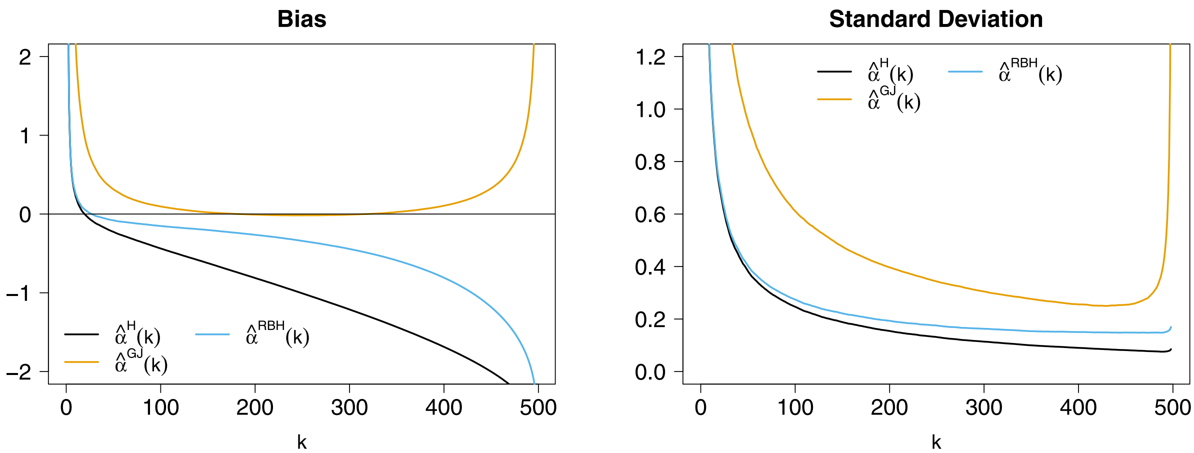

Fig. 6. Simulated bias (left) and standard deviation (right) of the estimators of α, versus k, for samples of size $n = 500$ from the log-logistic distribution with $(\alpha, \mu, \sigma) = (3, 1, 1)$.

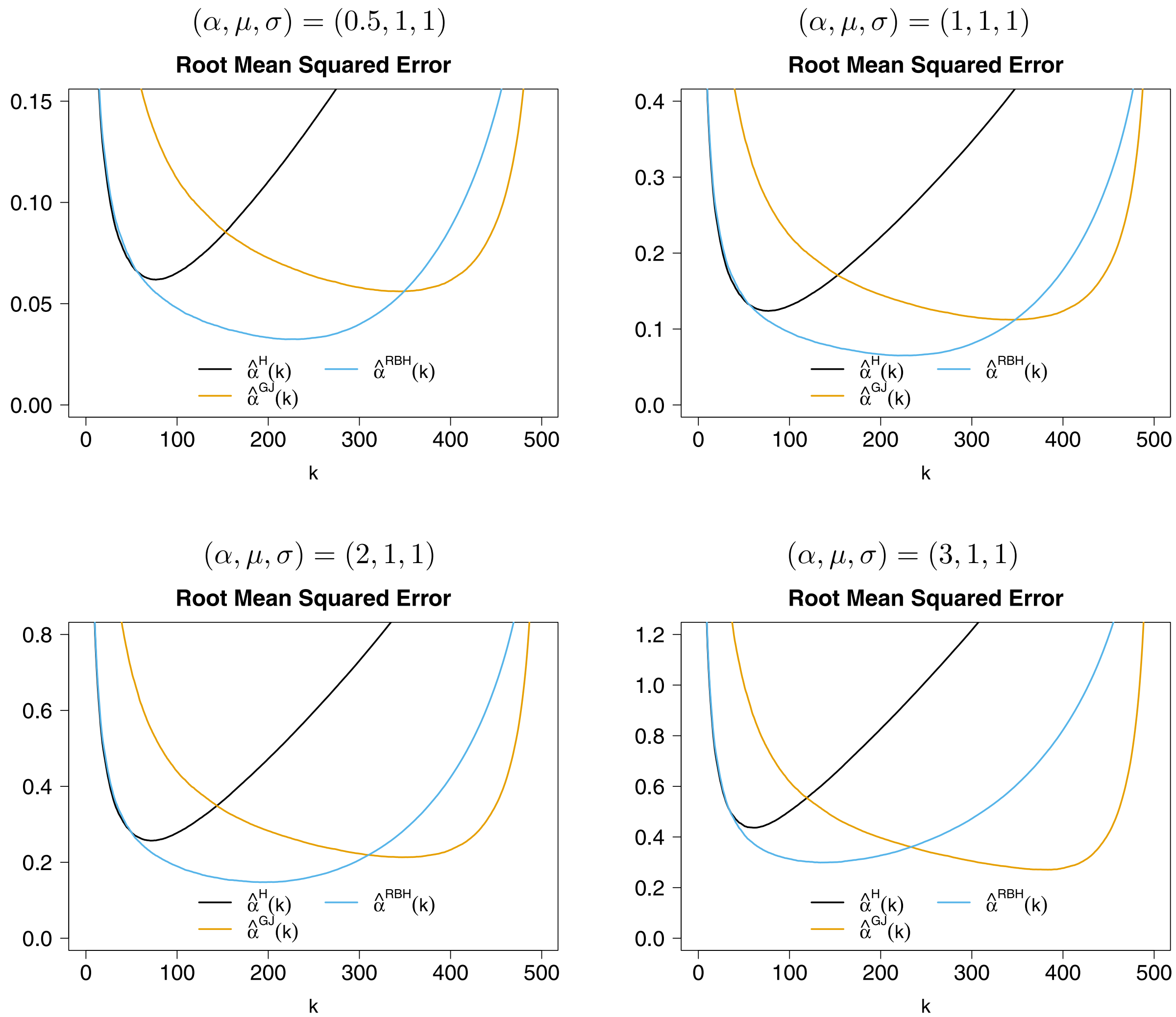

Fig. 7. Simulated RMSE of the estimators of α, versus k, for samples of size $n = 500$ from the log-logistic distribution for several parameter combinations.

$SD(\hat{\alpha}^{RBH}(k)) \leq SD(\hat{\alpha}^{H}(k)) < SD(\hat{\alpha}^{GJ}(k))$, for every k. Moreover, Fig. 7 shows that the new estimator $\hat{\alpha}^{RBH}(k)$, in (16), perform better than $\hat{\alpha}^{H}(k)$ for all values of α here considered and than $\hat{\alpha}^{GJ}(k)$ if $\alpha < 3$.

We now compare the performance of the estimators at their simulated optimal level. The simulated optimal value of k provides a benchmark of the best possible performance obtainable with each estimator of α. However, in practical applications such optimal level may not be achieved. The simulated values of the mean value, SD and RMSE, computed at the simulated optimal threshold are presented in Table 1. It appears that the proposed estimator $\hat{\alpha}^{RBH}(k)$ has usually the smallest (absolute) bias, SD and RMSE, except when $\alpha = 3$. In Table 2 we provide the simulated values of the mean value, SD and RMSE, computed at the level provided by the adaptive selection procedure. Notice that such simulated values are not very far from the corresponding values in Table 1, specially if $\alpha \leq 1$. This confirms the efficiency of our approach to obtain the value of the parameter k.

5. Data analysis

We now analyze one data set to illustrate the use of the proposed estimation method $\hat{\alpha}^{RBH}(k)$, for the shape parameter α of the log-logistic model. Based on the estimated value of α, we computed the estimated values of the location and scale parameters using the BLUE estimators discussed in detail in Balakrishnan et al. [12]. The product moments needed to for the covariance between $X_{i:n}$ and $X_{j:n}$, $i < j$ were computed using the integrate() function in R. The data set, taken from Hall et al. [53], represents the peak concentration of a toxic gas cloud released into the ambient flow, measured at a fixed location and are as follows:

Table 1
Simulated mean values, SD and RMSE of the estimators $\hat{\alpha}^{H}$, $\hat{\alpha}^{GJ}$ and $\hat{\alpha}^{RBH}$, at their simulated optimal threshold, for a log-logistic distribution.

n	20	50	75	100	200	500	750	1000	2000	5000
	$(\alpha, \mu, \sigma) = (0.5, 1, 1)$									
$Bias(\hat{\alpha}_0^{H})$	−0.1023	−0.0746	−0.0634	−0.0588	−0.0463	−0.0353	−0.0328	−0.0277	−0.0236	−0.0179
$Bias(\hat{\alpha}_0^{GJ})$	0.1993	0.0927	0.0720	0.0617	0.0411	0.0252	0.0208	0.0187	0.0131	0.0093
$Bias(\hat{\alpha}_0^{RBH})$	−0.0372	−0.0359	−0.0271	−0.0231	−0.0198	−0.0118	−0.0121	−0.0105	−0.0079	−0.0055
$SD(\hat{\alpha}_0^{H})$	0.1334	0.1024	0.0899	0.0815	0.0676	0.0509	0.0436	0.0419	0.0329	0.0246
$SD(\hat{\alpha}_0^{GJ})$	0.3010	0.1564	0.1232	0.1052	0.0759	0.0501	0.0416	0.0362	0.0274	0.0182
$SD(\hat{\alpha}_0^{RBH})$	0.1199	0.0748	0.0643	0.0570	0.0424	0.0302	0.0247	0.0222	0.0166	0.0116
$RMSE(\hat{\alpha}_0^{H})$	0.1681	0.1267	0.1100	0.1005	0.0819	0.0619	0.0546	0.0502	0.0405	0.0304
$RMSE(\hat{\alpha}_0^{GJ})$	0.3610	0.1818	0.1427	0.1220	0.0863	0.0561	0.0465	0.0408	0.0304	0.0204
$RMSE(\hat{\alpha}_0^{RBH})$	0.1255	0.0830	0.0697	0.0615	0.0468	0.0324	0.0275	0.0246	0.0184	0.0128
	$(\alpha, \mu, \sigma) = (1, 1, 1)$									
$Bias(\hat{\alpha}_0^{H})$	−0.2214	−0.1537	−0.1293	−0.1192	−0.0933	−0.0707	−0.0657	−0.0555	−0.0473	−0.0359
$Bias(\hat{\alpha}_0^{GJ})$	0.4148	0.1904	0.1477	0.1230	0.0809	0.0507	0.0418	0.0375	0.0262	0.0186
$Bias(\hat{\alpha}_0^{RBH})$	−0.0826	−0.0668	−0.0568	−0.0517	−0.0383	−0.0242	−0.0240	−0.0213	−0.0160	−0.0110
$SD(\hat{\alpha}_0^{H})$	0.2603	0.2034	0.1791	0.1626	0.1350	0.1017	0.0872	0.0838	0.0657	0.0491
$SD(\hat{\alpha}_0^{GJ})$	0.6106	0.3127	0.2465	0.2122	0.1530	0.1001	0.0832	0.0725	0.0548	0.0364
$SD(\hat{\alpha}_0^{RBH})$	0.2454	0.1570	0.1299	0.1134	0.0863	0.0604	0.0498	0.0444	0.0331	0.0231
$RMSE(\hat{\alpha}_0^{H})$	0.3417	0.2549	0.2209	0.2017	0.1641	0.1238	0.1092	0.1005	0.0810	0.0608
$RMSE(\hat{\alpha}_0^{GJ})$	0.7382	0.3661	0.2874	0.2452	0.1731	0.1122	0.0931	0.0816	0.0608	0.0409
$RMSE(\hat{\alpha}_0^{RBH})$	0.2590	0.1706	0.1418	0.1246	0.0944	0.0651	0.0552	0.0492	0.0368	0.0256
	$(\alpha, \mu, \sigma) = (2, 1, 1)$									
$Bias(\hat{\alpha}_0^{H})$	−0.5092	−0.3348	−0.2908	−0.2558	−0.2100	−0.1462	−0.1219	−0.1116	−0.0906	−0.0762
$Bias(\hat{\alpha}_0^{GJ})$	0.6637	0.3010	0.2368	0.2051	0.1366	0.0874	0.0729	0.0681	0.0523	0.0329
$Bias(\hat{\alpha}_0^{RBH})$	−0.2928	−0.1892	−0.1704	−0.1497	−0.1115	−0.0697	−0.0626	−0.0564	−0.0363	−0.0261
$SD(\hat{\alpha}_0^{H})$	0.5463	0.4264	0.3695	0.3461	0.2740	0.2113	0.1901	0.1741	0.1394	0.0980
$SD(\hat{\alpha}_0^{GJ})$	1.1492	0.5862	0.4671	0.4016	0.2934	0.1945	0.1624	0.1408	0.1051	0.0723
$SD(\hat{\alpha}_0^{RBH})$	0.5097	0.3418	0.2819	0.2489	0.1873	0.1300	0.1076	0.0950	0.0732	0.0497
$RMSE(\hat{\alpha}_0^{H})$	0.7468	0.5421	0.4702	0.4303	0.3452	0.2570	0.2258	0.2068	0.1662	0.1241
$RMSE(\hat{\alpha}_0^{GJ})$	1.3271	0.6589	0.5236	0.4509	0.3236	0.2133	0.1780	0.1563	0.1174	0.0794
$RMSE(\hat{\alpha}_0^{RBH})$	0.5878	0.3907	0.3294	0.2905	0.2179	0.1475	0.1245	0.1105	0.0817	0.0561
	$(\alpha, \mu, \sigma) = (3, 1, 1)$									
$Bias(\alpha_0^{H})$	−0.8949	−0.6212	−0.4997	−0.4607	−0.3522	−0.2819	−0.2381	−0.2166	−0.1613	−0.1253
$Bias(\alpha_0^{GJ})$	0.5721	0.1805	0.1562	0.1033	0.0959	0.0773	0.0525	0.0486	0.0386	0.0247
$Bias(\alpha_0^{RBH})$	−0.6246	−0.4483	−0.3998	−0.3489	−0.2718	−0.1863	−0.1671	−0.1554	−0.1144	−0.0806
$SD(\alpha_0^{H})$	0.8547	0.6656	0.6216	0.5595	0.4653	0.3335	0.2992	0.2752	0.2296	0.1668
$SD(\alpha_0^{GJ})$	1.5579	0.7956	0.6300	0.5551	0.3967	0.2596	0.2202	0.1933	0.1451	0.0988
$SD(\alpha_0^{RBH})$	0.8182	0.5623	0.4697	0.4289	0.3276	0.2338	0.1944	0.1692	0.1347	0.0934
$RMSE(\alpha_0^{H})$	1.2374	0.9105	0.7975	0.7247	0.5836	0.4366	0.3823	0.3502	0.2806	0.2086
$RMSE(\alpha_0^{GJ})$	1.6596	0.8158	0.6491	0.5646	0.4082	0.2708	0.2264	0.1993	0.1501	0.1019
$RMSE(\alpha_0^{RBH})$	1.0293	0.7191	0.6168	0.5528	0.4256	0.2990	0.2563	0.2297	0.1767	0.1234

12.100, 8.757, 6.678, 5.702, 5.796 11.280 16.160 10.550, 5.072, 9.010, 1.701, 5.670, 3.026, 7.262, 3.497, 6.804 11.500, 7.655, 6.045, 3.215, 9.074 12.890, 6.806, 6.086, 7.087, 5.292, 5.072, 5.229, 6.458, 5.859, 7.056, 7.119 11.750, 7.568 15.800, 6.273, 9.041 14.900, 4.993 10.020, 7.025, 2.523, 5.742, 7.941, 4.316, 10.840, 8.927, 7.087, 7.403, 6.962, 4.777, 9.055, 4.007 14.030, 7.591, 6.587, 7.560, 2.646 13.480 11.440, 8.870, 7.341, 7.340, 7.844 13.990, 8.757, 4.694, 3.704 11.530, 5.765, 7.656, 3.938, 2.849, 3.150, 9.185, 9.344, 6.832, 9.293, 9.926, 6.928 10.920 10.460, 6.418, 7.818, 6.286, 5.513 15.380, 6.117, 3.451, 5.171, 6.806 11.050, 8.456, 8.554 11.040 11.040 10.250 13.650 16.910, 7.825.

Table 2
Simulated mean values, SD and RMSE of the estimators $\hat{\alpha}_A^H$, $\hat{\alpha}_E^{GJ}$ and $\hat{\alpha}_E^{RBH}$, for a log-logistic distribution.

n	20	50	75	100	200	500	750	1000	2000	5000
	$(\alpha,\mu,\sigma)=(0.5,1,1)$									
$Bias(\hat{\alpha}_A^H)$	−0.1023	−0.0746	−0.0634	−0.0588	−0.0479	−0.0364	−0.0324	−0.0295	−0.0235	−0.0177
$Bias(\hat{\alpha}_E^{GJ})$	0.2202	0.0961	0.0703	0.0590	0.0395	0.0242	0.0203	0.0177	0.0140	0.0101
$Bias(\hat{\alpha}_E^{RBH})$	−0.0372	−0.0267	−0.0244	−0.0231	−0.0180	−0.0134	−0.0117	−0.0107	−0.0084	−0.0065
$SD(\hat{\alpha}_A^H)$	0.1334	0.1024	0.0899	0.0815	0.0665	0.0502	0.0439	0.0407	0.0330	0.0248
$SD(\hat{\alpha}_E^{GJ})$	0.3053	0.1547	0.1245	0.1070	0.0770	0.0508	0.0420	0.0369	0.0270	0.0179
$SD(\hat{\alpha}_E^{RBH})$	0.1199	0.0795	0.0655	0.0570	0.0432	0.0296	0.0249	0.0221	0.0163	0.0112
$RMSE(\hat{\alpha}_A^H)$	0.1681	0.1267	0.1100	0.1005	0.0820	0.0620	0.0546	0.0503	0.0405	0.0304
$RMSE(\hat{\alpha}_E^{GJ})$	0.3764	0.1821	0.1430	0.1222	0.0865	0.0562	0.0466	0.0409	0.0304	0.0205
$RMSE(\hat{\alpha}_E^{RBH})$	0.1255	0.0839	0.0699	0.0615	0.0468	0.0325	0.0275	0.0246	0.0184	0.0129
	$(\alpha,\mu,\sigma)=(1,1,1)$									
$Bias(\hat{\alpha}_A^H)$	−0.2214	−0.1537	−0.1293	−0.1192	−0.0964	−0.0730	−0.0649	−0.0590	−0.0470	−0.0354
$Bias(\hat{\alpha}_E^{GJ})$	0.4773	0.1984	0.1438	0.1202	0.0798	0.0485	0.0407	0.0354	0.0281	0.0201
$Bias(\hat{\alpha}_E^{RBH})$	−0.1180	−0.0668	−0.0568	−0.0517	−0.0383	−0.0275	−0.0238	−0.0218	−0.0170	−0.0130
$SD(\hat{\alpha}_A^H)$	0.2603	0.2034	0.1791	0.1626	0.1329	0.1003	0.0878	0.0814	0.0660	0.0495
$SD(\hat{\alpha}_E^{GJ})$	0.6393	0.3096	0.2489	0.2138	0.1539	0.1015	0.0839	0.0738	0.0540	0.0358
$SD(\hat{\alpha}_E^{RBH})$	0.2311	0.1570	0.1299	0.1134	0.0863	0.0592	0.0499	0.0442	0.0327	0.0224
$RMSE(\hat{\alpha}_A^H)$	0.3417	0.2549	0.2209	0.2017	0.1642	0.1241	0.1092	0.1006	0.0810	0.0609
$RMSE(\hat{\alpha}_E^{GJ})$	0.7978	0.3677	0.2875	0.2453	0.1733	0.1125	0.0932	0.0818	0.0608	0.0410
$RMSE(\hat{\alpha}_E^{RBH})$	0.2595	0.1706	0.1418	0.1246	0.0944	0.0653	0.0553	0.0493	0.0368	0.0259
	$(\alpha,\mu,\sigma)=(2,1,1)$									
$Bias(\hat{\alpha}_A^H)$	−0.6085	−0.4028	−0.3317	−0.2989	−0.2311	−0.1668	−0.1457	−0.1310	−0.1022	−0.0753
$Bias(\hat{\alpha}_E^{GJ})$	0.8065	0.3175	0.2289	0.1933	0.1297	0.0802	0.0681	0.0596	0.0485	0.0357
$Bias(\hat{\alpha}_E^{RBH})$	−0.5043	−0.2980	−0.2440	−0.2140	−0.1499	−0.0979	−0.0813	−0.0720	−0.0530	−0.0370
$SD(\hat{\alpha}_A^H)$	0.4682	0.3828	0.3423	0.3139	0.2596	0.1980	0.1739	0.1616	0.1314	0.0988
$SD(\hat{\alpha}_E^{GJ})$	1.2474	0.5818	0.4715	0.4081	0.2975	0.1988	0.1649	0.1455	0.1069	0.0712
$SD(\hat{\alpha}_E^{RBH})$	0.4238	0.2986	0.2505	0.2212	0.1703	0.1177	0.0994	0.0883	0.0654	0.0448
$RMSE(\hat{\alpha}_A^H)$	0.7677	0.5557	0.4767	0.4335	0.3476	0.2589	0.2268	0.2080	0.1664	0.1242
$RMSE(\hat{\alpha}_E^{GJ})$	1.4854	0.6628	0.5242	0.4515	0.3246	0.2143	0.1784	0.1572	0.1174	0.0796
$RMSE(\hat{\alpha}_E^{RBH})$	0.6587	0.4218	0.3497	0.3077	0.2268	0.1531	0.1283	0.1139	0.0842	0.0581
	$(\alpha,\mu,\sigma)=(3,1,1)$									
$Bias(\hat{\alpha}_A^H)$	−1.1872	−0.8273	−0.6956	−0.6290	−0.4895	−0.3527	−0.3064	−0.2751	−0.2136	−0.1545
$Bias(\hat{\alpha}_E^{GJ})$	0.7705	0.2018	0.1154	0.0885	0.0420	0.0126	0.0094	0.0060	0.0072	0.0063
$Bias(\hat{\alpha}_E^{RBH})$	−1.1141	−0.7495	−0.6396	−0.5749	−0.4325	−0.3039	−0.2594	−0.2331	−0.1791	−0.1278
$SD(\hat{\alpha}_A^H)$	0.6244	0.5280	0.4785	0.4432	0.3714	0.2873	0.2538	0.2368	0.1942	0.1468
$SD(\hat{\alpha}_E^{GJ})$	1.7222	0.7918	0.6461	0.5639	0.4174	0.2839	0.2374	0.2107	0.1567	0.1055
$SD(\hat{\alpha}_E^{RBH})$	0.5792	0.4269	0.3648	0.3269	0.2576	0.1820	0.1550	0.1386	0.1046	0.0726
$RMSE(\hat{\alpha}_A^H)$	1.3414	0.9814	0.8443	0.7694	0.6144	0.4549	0.3978	0.3630	0.2887	0.2131
$RMSE(\hat{\alpha}_E^{GJ})$	1.8866	0.8170	0.6563	0.5708	0.4195	0.2842	0.2376	0.2108	0.1569	0.1057
$RMSE(\hat{\alpha}_E^{RBH})$	1.2557	0.8625	0.7363	0.6613	0.5034	0.3542	0.3022	0.2712	0.2074	0.1470

Randomness of the sample values was not rejected with the runs test, available in randtests software package [54]. Hankin and Lee [55] considered for this data the Davies distribution, defined by the quantile function

$$Q_D(p|c,\lambda_1,\lambda_2)=\frac{c\,p^{\lambda_1}}{(1-p)^{\lambda_2}},\quad 0\le p<1,\quad c,\lambda_1,\lambda_2>0.$$

For this data set, Hankin and Lee computed the maximum likelihood estimates $\hat{c}=8.326$, $\hat{\lambda}_1=0.319$ and $\hat{\lambda}_2=0.181$. The Kolmogorov–Smirnov (K–S) distance between the fitted and empirical d.f. is 0.0637. For the three-parameter log-logistic distribution, we considered the shape parameter estimator $\hat{\alpha}_E^{RBH}$ in (38) and the BLUE estimators of the location and scale

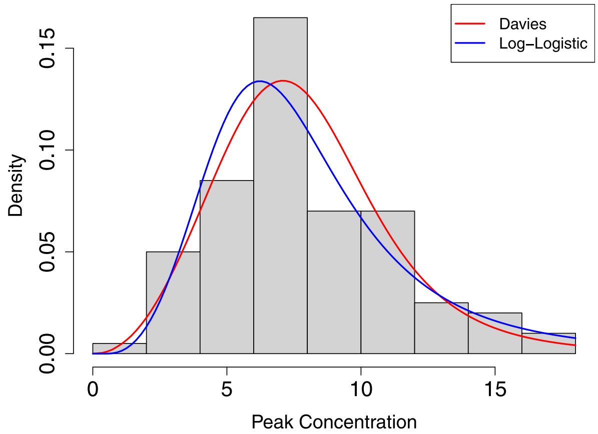

Fig. 8. Histogram and fitted p.d.f. of the Davies and log-logistic distributions.

parameters. The estimated parameters values are $\hat{\alpha} = 3.363$, $\hat{\mu} = 0.501$ and $\hat{\sigma} = 6.872$. The corresponding K–S distance is 0.0578. Fig. 8 displays the histogram and the Davies and log-logistic estimated p.d.f.'s. Therefore, based on the K–S values and Fig. 8, the log-logistic distribution provided a good fit to this data set and a closer fit to the empirical d.f. than the Davies distribution.

6. Conclusion

In this research, we present a new reduced-bias estimator of the shape parameter of the three-parameter log-logistic model. Its asymptotic normality has been demonstrated, and an adaptive selection procedure for the parameter k has been provided. The efficiency of our approach was illustrated in a simulation study. The results given here show that the new reduced-bias estimator usually outperforms the other estimators studied, in terms of bias and RMSE. The usefulness of this novel estimator was illustrated by applying it to a real-world data set.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank the referees for their comments and suggestions that led to an improvement of the manuscript.

Funding

Work partially funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., Portugal, under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications).

References

[1] P.R. Fisk, The graduation of income distributions, Econometrica 29 (2) (1961) 171–185.
[2] Z. Shayan, S. Ayatollahi, N. Zare, A parametric method for cumulative incidence modeling with a new four-parameter log-logistic distribution, Theor. Biol. Med. Model. 8 (1) (2011) 1–11.
[3] M.E. Girmay, D. Chikobvu, Quantifying South Africa's sulphur dioxide emission efficiency in coal-powered electricity generation by fitting the three-parameter log-logistic distribution, J. Energy South. Afr. 28 (1) (2017) 91–103.
[4] R.C. Gupta, O. Akman, S. Lvin, A study of log-logistic model in survival analysis, Biom. J. 41 (4) (1999) 431–443.
[5] J. Brüederl, A. Diekmann, The log-logistic rate model: Two generalizations with an application to demographic data, Sociol. Methods Res. 24 (2) (1995) 158–186.
[6] F. Ogana, Comparison of a modified log-logistic distribution with established models for tree height prediction, J. Res. For. Wildl. Environ. 10 (2) (2018) 49–55.
[7] M. Ahmad, C. Sinclair, A. Werritty, Log-logistic flood frequency analysis, J. Hydrol. 98 (3–4) (1988) 205–224.
[8] P.M. Rowinski, W.G. Strupczewski, V.P. Singh, A note on the applicability of log-gumbel and log-logistic probability distributions in hydrological analyses: I. Known pdf, Hydrol. Sci. J. 47 (1) (2002) 107–122.
[9] H.J. Malik, Exact distribution of the quotient of independent generalized Gamma variables, Canad. Math. Bull. 10 (3) (1967) 463–465.
[10] I.W. Burr, Cumulative frequency functions, Ann. Math. Stat. 13 (2) (1942) 215–232.
[11] C. Dagum, A new model for personal income distribution: Specification and estimation, Econ. Appl. 30 (1977) 413–437.

[12] N. Balakrishnan, H. Malik, S. Puthenpura, Best linear unbiased estimation of location and scale parameters of the log-logistic distribution, Comm. Statist. Theory Methods 16 (12) (1987) 3477–3495.
[13] V. Singh, H. Guo, F. Yu, Parameter estimation for 3-parameter log-logistic distribution (LLD3) by pome, Stoch. Hydrol. Hydraul. 7 (3) (1993) 163–177.
[14] J.B. McDonald, Some generalized functions for the size distribution of income, Econometrica 52 (3) (1984) 647.
[15] A.H. Muse, S.M. Mwalili, O. Ngesa, On the log-logistic distribution and its generalizations: A survey, Int. J. Stat. Probab. 10 (3) (2021) 93–125.
[16] B.C. Arnold, Pareto Distributions, in: Monographs on Statistics and Applied Probability, Chapman and Hall/CRC, 2015.
[17] F. Ashkar, S. Mahdi, Fitting the log-logistic distribution by generalized moments, J. Hydrol. 328 (3) (2006) 694–703, The ICWRER - Symposium in Dresden, Germany.
[18] R.R.L. Kantam, A.V. Roa, G.S. Roa, Optimum group limits for estimation in scaled log-logistic distribution from a grouped data, Statist. Papers 46 (3) (2005) 359–377.
[19] Y. Mert Kantar, Generalized least squares and weighted least squares estimation methods for distributional parameters, Revstat–Stat. J. 13 (8) (2015) 263–282.
[20] K. Abbas, Y. Tang, Objective Bayesian analysis for log-logistic distribution, Comm. Statist. Simulation Comput. 45 (8) (2016) 2782–2791.
[21] J. Reath, J. Dong, M. Wang, Improved parameter estimation of the log-logistic distribution with applications, Comput. Statist. 33 (1) (2017) 339–356.
[22] X. He, W. Chen, W. Qian, Maximum likelihood estimators of the parameters of the log-logistic distribution, Statist. Papers 61 (5) (2020) 1875–1892.
[23] D. Abd El-Rahman, M. El Genidy, Three parameters estimation of log-logistic distribution using algorithm of percentile roots, in: The 54th Annual Conference on Statistics, Computer Science and Operation Research, 2019.
[24] M. Ahsanullah, A. Alzaatreh, Parameter estimation for the log-logistic distribution based on order statistics, Revstat – Stat. J. 16 (4) (2018) 429–443.
[25] A. Mateus, F. Caeiro, Improved shape parameter estimation for the three-parameter log-logistic distribution, Comput. Math. Methods 2022 (2022) 1–13.
[26] P. Hall, On some simple estimates of an exponent of regular variation, J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (1) (1982) 37–42.
[27] J. Panaretos, Z. Tsourti, Extreme value index estimators and smoothing alternatives: A critical review, in: J. Panaretos (Ed.), Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century, Laurence Erlbaum, New Jersey, USA, 2003, pp. 141–160.
[28] J. Beirlant, K. Herrmann, J. Teugels, Estimation of the extreme value index, in: Extreme Events in Finance, John Wiley & Sons, Inc., 2016, pp. 97–115.
[29] I. Fedotenkov, A review of more than one hundred Pareto-tail index estimators, Statistica 80 (3) (2020) 245–299.
[30] I. Cabral, F. Caeiro, M.I. Gomes, On the comparison of several classical estimators of the extreme value index, Comm. Statist. Theory Methods 51 (1) (2022) 179–196.
[31] B.M. Hill, A simple general approach to inference about the tail of a distribution, Ann. Statist. 3 (5) (1975) 1163–1174.
[32] M.I. Gomes, O. Oliveira, How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events, Comm. Statist. Simulation Comput. 32 (4) (2003) 1005–1028.
[33] F. Caeiro, M.I. Gomes, Minimum-variance reduced-bias tail index and high quantile estimation, Revstat – Stat. J. 6 (1) (2008) 1–20.
[34] C. Liu, C. Ling, A location-invariant non-positive moment-type estimator of the extreme value index, Comm. Statist. Theory Methods 48 (5) (2018) 1166–1176.
[35] M. Fraga Alves, A location invariant hill-type estimator, Extremes 4 (3) (2001) 199–217.
[36] P. Araújo Santos, M. Fraga Alves, M.I. Gomes, Peaks over random threshold methodology for tail index and high quantile estimation, Revstat – Stat. J. 4 (3) (2006) 227–247.
[37] H. Drees, Refined pickands estimators with bias correction, Comm. Statist. Theory Methods 25 (4) (1996) 837–851.
[38] L. Peng, Asymptotically unbiased estimators for the extreme-value index, Statist. Probab. Lett. 38 (2) (1998) 107–115.
[39] J. Beirlant, F. Caeiro, M.I. Gomes, An overview and open research topics in statistics of univariate extremes, Revstat – Stat. J. 10 (1) (2012) 1–31.
[40] M.I. Gomes, A. Guillou, Extreme value theory and statistics of univariate extremes: a review, Internat. Statist. Rev. 83 (2) (2015) 263–292.
[41] J. Beirlant, Y. Goegebeur, J. Segers, J.L. Teugels, Statistics of Extremes: Theory and Applications, John Wiley & Sons, 2004.
[42] M.I. Gomes, M.J. Martins, M. Neves, Alternatives to a semi-parametric estimator of parameters of rare events–the jackknife methodology, Extremes 3 (3) (2000) 207–229.
[43] P. Jordanova, Z. Fabián, P. Hermann, L. Střelec, A. Rivera, S. Girard, S. Torres, M. Stehlík, Weak properties and robustness of t-hill estimators, Extremes 19 (4) (2016) 591–626.
[44] J. Danielsson, D.W. Jansen, C.G. de Vries, The method of moments ratio estimator for the tail shape parameter, Comm. Statist. Theory Methods 25 (4) (1996) 711–720.
[45] H. Penalva, M.I. Gomes, F. Caeiro, M.M. Neves, A couple of non reduced bias generalized means in extreme value theory: An asymptotic comparison, Revstat–Stat. J. 18 (3) (2020) 281–298.
[46] H. Penalva, M.I. Gomes, F. Caeiro, M.M. Neves, Lehmer's mean-of-order-p extreme value index estimation: a simulation study and applications, J. Appl. Stat. 47 (13–15) (2020) 2825–2845.
[47] C. Caballero-Megido, J. Hillier, D. Wyncoll, L. Bosher, B. Gouldby, Technical note: comparison of methods for threshold selection for extreme sea levels, J. Flood Risk Manage. 11 (2) (2017) 127–140.
[48] F. Caeiro, M. Gomes, Threshold selection in extreme value analysis, in: D.K. Dey, J. Yan (Eds.), Extreme Value Modeling and Risk Analysis, Chapman and Hall/CRC, 2015, pp. 69–86.
[49] L.F. Schneider, A. Krajina, T. Krivobokova, Threshold selection in univariate extreme value analysis, Extremes 24 (4) (2021) 881–913.
[50] N. Balakrishnan, A.P. Basu, Exponential Distribution: Theory, Methods and Applications, CRC Press, 1995.
[51] S. Girard, A hill type estimator of the Weibull tail-coefficient, Comm. Statist. Theory Methods 33 (2) (2004) 205–234.
[52] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022, URL https://www.R-project.org/.
[53] D. Hall, R. Waters, G. Marsland, S. Upton, Repeat variability in instantaneously released heavy gas clouds-some wind tunnel model experiments, Technical Report LR 804 (PA), Warren Spring Laboratory, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2BX, 1991.
[54] A. Mateus, F. Caeiro, An r implementation of several randomness tests, AIP Conf. Proc. 1618 (1) (2014) 531–534.
[55] R.K.S. Hankin, A. Lee, A new family of non-negative distributions, Aust. N. Z. J. Stat. 48 (1) (2006) 67–78.