

## Research Article

# Box-Cox Transformations and Bias Reduction in Extreme Value Theory

Lígia Henriques-Rodrigues <sup>1,2</sup> and M. Ivette Gomes <sup>3,4</sup>

<sup>1</sup>Department of Mathematics, School of Science and Technology, University of Évora, Évora, Portugal

<sup>2</sup>Research Center in Mathematics and Applications (CIMA), University of Évora, Évora, Portugal

<sup>3</sup>DEIO, Faculty of Sciences, University of Lisbon (FCUL), Lisbon, Portugal

<sup>4</sup>Statistics and Applications Center (CEAUL), University of Lisbon, Lisbon, Portugal

Correspondence should be addressed to Lígia Henriques-Rodrigues; [ligiahr@uevora.pt](mailto:ligiahr@uevora.pt)

Received 10 September 2021; Accepted 28 January 2022; Published 10 March 2022

Academic Editor: Christophe Chesneau

Copyright © 2022 Lígia Henriques-Rodrigues and M. Ivette Gomes. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Box-Cox transformations are used to make the data more suitable for statistical analysis. We know from the literature that this transformation of the data can increase the rate of convergence of the tail of the distribution to the generalized extreme value distribution, and as a byproduct, the bias of the estimation procedure is reduced. The reduction of bias of the Hill estimator has been widely addressed in the literature of extreme value theory. Several techniques have been used to achieve such reduction of bias, either by removing the main component of the bias of the Hill estimator of the extreme value index (EVI) or by constructing new estimators based on generalized means or norms that generalize the Hill estimator. We are going to study the Box-Cox Hill estimator introduced by Teugels and Vanroelen, in 2004, proving the consistency and asymptotic normality of the estimator and addressing the choice and estimation of the power and shift parameters of the Box-Cox transformation for the EVI estimation. The performance of the estimators under study will be illustrated for finite samples through small-scale Monte Carlo simulation studies.

## 1. Introduction and Preliminaries

In extreme value theory, one is often interested in the estimation of the extreme value index (EVI),  $\xi$ , a parameter controlling the heaviness of the right tail  $\bar{F} := 1 - F$ , with  $F$  the distribution function (d.f.) of the process under study. In a univariate independent and identically distributed (i.i.d.) setup, let us assume that we have access to a random sample of size  $n$  and use the notation  $X_{i:n}$  to denote the  $i$ -th ascending order statistic,  $1 \leq i \leq n$ . Similar to the central limit theorem, the extremal types theorem states that if there exists a nondegenerate limit distribution for a suitable normalization of the maximum,

$$P\left(\frac{X_{n:n} - b_n}{a_n} \leq x\right) \xrightarrow{n \rightarrow \infty} G_\xi(x), \quad (1)$$

being  $G_\xi(\cdot)$  an extreme value (EV) d.f. of the type:

$$G_\xi(x) \equiv \text{EV}_\xi(x) := \begin{cases} \exp\left(-(1 + \xi x)^{-1/\xi}\right), & 1 + \xi x > 0, \quad \text{if } \xi < 0, \\ \exp(-\exp(-x)), & x \in \mathbf{R}, \quad \text{if } \xi = 0. \end{cases} \quad (2)$$

From Equations (1) and (2), the d.f.  $F$  is said to pertain to the domain of attraction for maxima of  $\text{EV}_\xi$ , with  $\xi \in \mathbf{R}$ . Such a fact is denoted by  $F \in D_M(\text{EV}_\xi)$ . The distribution in Equation (2) encompasses three types of distributions: if  $\xi < 0$ , the right tail is light; i.e.,  $F$  has a finite right endpoint ( $x^F := \sup\{x : F(x) < 1\} < +\infty$ ); if  $\xi > 0$ , the right tail is heavy of a negative polynomial type and  $F$  has an infinite right endpoint; if  $\xi = 0$ , the right tail is of an exponential type and the right endpoint can be either finite or infinite.

In this paper, we are going to work with heavy-tailed models, i.e., models with  $\xi > 0$  that belong to the so-called Fréchet domain of attraction, denoted by  $D_M^+ := D_M(\text{EV}_\xi)_{\xi > 0}$ . These types of models are quite common in areas like biostatistics, computer science, finance, insurance, and statistical quality control, among others.

*1.1. Box-Cox Transformations in Extreme Value Theory.* Let  $X$  be a random variable (r.v.) with d.f.  $F \equiv F_X \in D_M(\text{EV}_\xi)$ , with probability density function  $f_X$ . The Box-Cox (BC) transformation [1] of  $X$  is a function of the parameter  $\sigma \in \mathbf{R}$  given by the following:

$$Y = \begin{cases} \frac{X^\sigma - 1}{\sigma}, & \sigma \neq 0, \\ \ln X, & \sigma = 0, \end{cases} \quad (3)$$

where  $\sigma = 1$  corresponds to a mere change in location,  $\sigma = 1/2$  to the square root transformation, and  $\sigma = -1$  to the reciprocal transformation.

The use of BC transformations in EV theory has been quite diverse. Bali [2] proposed the BC-GEV distribution to model the empirical distribution of returns as a generalization of the GEV distribution in Equation (2) and of the generalized Pareto distribution, with d.f.  $\text{GP}_\xi(x) = 1 + \ln G_\xi(x)$ . The d.f. of the BC-GEV distribution is as follows:

$$F_\xi(x; \sigma) = \frac{(G_\xi(x))^\sigma - 1}{\sigma} + 1. \quad (4)$$

Under a semiparametric framework, Teugels and Vanroelen [3] provided the theoretical framework through the use of regular variation (see Seneta [4] and Bingham et al. [5] for details on regular variation) to determine the optimal values of  $\sigma$  that maximize the rate of convergence of the second-order condition (to be detailed in Section 1.2) for values of  $\xi \geq 0$ . Another application of the BC transformation was presented in the work of Eastoe and Tawn [6], under a Bayesian framework. Wadsworth et al. [7] use the BC transformation in the inference procedure to account parametrically for the uncertainty surrounding the scale extrapolation in physical processes. Paulauskas and Vaiciulis [8, 9] used the BC transformation and considered estimators based on block scheme considering a family of functions of order statistics.

In the next theorem, the link between the domain of attraction of  $F_X$  and of the transformed variable  $Y$ , with d. f.  $F_Y$ , is established.

**Theorem 1** (Wadsworth et al. [7]). *Under the conditions such that convergence in Equation (1) holds, with norming sequences  $a_n$  and  $b_n$ , then with  $G_\xi(\cdot)$  the d.f. in Equation (2)*

$$P\left(\frac{Y_{n:n} - b_n^*}{a_n^*} \leq x\right) \xrightarrow{n \rightarrow \infty} G_{\xi^*}(y) \quad (5)$$

holds for some finite  $\xi^*$  where

$$\begin{aligned} b_n^* &= \frac{(b_n)^\sigma - 1}{\sigma}, \\ a_n^* &= a_n (b_n)^{\sigma-1}, \end{aligned} \quad (6)$$

and we can say that  $F_Y \in D_M(\text{EV}_{\xi^*})$ . Furthermore, if  $F_X$  is twice differentiable for sufficiently large  $x$ , then the limiting shape parameter  $\xi^*$  takes the form:

$$\xi^* = \xi + \lim_{x \rightarrow x^F} \frac{h(x)}{x} (\sigma - 1), \quad (7)$$

with  $h(x) = (1 - F(x))/f(x)$  the reciprocal hazard function of the parent distribution  $F_X$ . For any such distribution which has  $\xi \leq 0$ ,  $\xi^* = \xi$ , due to fact that  $h(x)/x$  tends then to zero, as  $x \rightarrow x^F$ .

*Remark 2.* Note that if  $\xi \leq 0$ , then  $\xi^* = \xi \leq 0$  and  $F_Y$  is in the same domain of attraction of  $F_X$ . If  $\xi > 0$ , i.e., if we are in the Fréchet domain of attraction,  $\xi^*$  can be different of  $\xi$  and  $F_Y$  can no longer belong to the Fréchet domain of attraction.

*1.2. First- and Second-Order Frameworks.* With  $F \in D_M^+$ , let  $F^{\leftarrow}$  denote the generalized inverse function of  $F$ , defined by  $F^{\leftarrow}(t) := \inf \{x : F(x) \geq t\}$ , let  $U$  be the associated (reciprocal) quantile function, defined as  $U(t) := F^{\leftarrow}(1 - 1/t)$ ,  $t \geq 1$ , and let  $\text{RV}_\alpha$  stand for the class of regularly varying functions at infinity with index  $\alpha$ , i.e., positive measurable functions  $g(\cdot)$  such that  $\lim_{t \rightarrow \infty} g(tx)/g(t) = x^\alpha$ , for all  $x > 0$ .

In this setup, we have the validity of the first-order condition (de Haan [10]):

$$F \in D_M^+ \iff 1 - F \in \text{RV}_{-1/\xi} \iff U \in \text{RV}_\xi. \quad (8)$$

To develop the asymptotic nondegenerate properties of the estimators under study, we need to assume the existence of an auxiliary function  $A(t)$ , with  $|A| \in \text{RV}_\rho$  [11], where  $\rho \leq 0$  is a shape second-order parameter, and for every  $x > 0$ , we often work with a second-order condition (SOC) of the type:

$$\lim_{t \rightarrow \infty} \frac{\ln U(tx) - \ln U(t) - \xi \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}. \quad (9)$$

Slightly more restrictively, we are going to work with models in the Fréchet domain of attraction ( $\xi > 0$ ) that belong to Hall-Welsh class of models [12], i.e., models with the second-order right tail expansion:

$$\begin{aligned} 1 - F(x) &:= \left(\frac{x}{C}\right)^{-1/\xi} \left\{ 1 + E_1 \left(\frac{x}{C}\right)^{-\rho/\xi} \right. \\ &\quad \left. + o\left(x^{-\rho/\xi + \varepsilon}\right) \right\}, \quad x \rightarrow \infty, \end{aligned} \quad (10)$$

with  $E_1 \neq 0$ ,  $\rho < 0$ , and  $C > 0$ , the first-order scale parameter. For these models, the second-order tail quantile function is as follows:

$$U(t) = Ct^\xi(1 + D_1 t^\rho + o(t^\rho)), \quad t \longrightarrow \infty, \quad (11)$$

with  $D_1 = \xi E_1$ . If the SOC in Equation (9) holds with  $A(t) = \xi \beta t^\rho$ , then  $D_1 = A(t)/(\rho t^\rho)$ . Most common heavy-tailed distributions, like the Fréchet, the generalized Pareto, the Burr and Student's  $t$ , belong to the class in Equation (11). The log-gamma and the log-Pareto models ( $\rho = 0$  in Equation (9)) do not belong to the class in Equation (11).

*Remark 3.* Under a parametric framework, Wadsworth et al. [7] showed that for models in Hall-Welsh class, a BC transformation of the data increases the rate of convergence of the tail of the distribution to the GEV distribution if  $\sigma = -\rho/\xi$ , and as a byproduct, the bias of the estimation procedure is reduced.

Let us consider a BC transformation such that  $X^* = X^\sigma + a$  and denote by  $(X_1^*, \dots, X_n^*)$  the transformed sample, with d.f.  $F^*$ , associated with the original sample  $(X_1, \dots, X_n)$ , where  $X_i^* = T(X_i)$ , with  $T(X) = X^\sigma + a$ . The first-order condition associated with the  $X^*$  is as follows:

$$\lim_{t \rightarrow \infty} \frac{U^*(tx)}{U^*(t)} = x^{\xi\sigma} = x^{\xi^*} \iff U^* \in \text{RV}_{\xi^* = \xi\sigma}. \quad (12)$$

Under the validity of Equation (12) and for some  $C^*$ ,  $C > 0$ , we have  $U^*(t) \sim C^* t^{\xi^*} = (Ct^\xi)^\sigma$ .

Teugels and Vanroelen [3] used the theory of extended regular variation (see de Haan and Ferreira [13]) to determine the optimal values of  $\sigma$  that maximize the rate of convergence of the second-order condition for values of  $\xi \geq 0$  and proved in Theorems 4.1, 4.2, and 4.3 that when  $\xi > 0$  and  $\xi + \rho > 0$ , a BC transformation has no effect on the SOC of  $U^*$  unless  $\sigma = -\rho/\xi$ ; if  $\xi > 0$  and  $\xi + \rho < 0$ , the BC transformation can have a negative or positive effect on the SOC of  $U^*$  and the positive effect happens when  $\sigma = -\rho/\xi$ . If  $\xi > 0$  and  $\xi + \rho = 0$ , for any value other than  $\sigma = -\rho/\xi = 1$ , there is no improvement in the speed of convergence of the SOC. Thus, the value  $\sigma$  that maximizes the rate of convergence of the SOC that will be considered in this work is as follows:

$$\sigma = -\frac{\rho}{\xi}. \quad (13)$$

This choice will enhance the bias reduction in the estimation procedure.

**1.3. Reduced Bias EVI Estimation.** For heavy right tails, i.e., whenever Equation (8) holds, the classical extreme value index (EVI) estimator is the Hill (H) estimator [14], with functional form

$$H_{n,k} := \frac{1}{k} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n}}{X_{n-k:n}} \right), \quad (14)$$

the average of the log-excesses, and consequently also the average of the  $k$  scaled log-spacings  $U_i := i(\ln X_{n-i+1:n} - \ln X_{n-i:n})$ ,  $1 \leq i \leq k$ .

But the Hill EVI estimator, in Equation (14), exhibits the usual trade-off between bias and variance, and therefore, the suitable accommodation of the asymptotic bias has been widely discussed among EVT practitioners. The first papers are due to Peng [15], Beirlant et al. [16], Feuerverger and Hall [17], and Gomes et al. [18], among others. Such a bias reduction usually provided more stable sample trajectories and a lower mean square error at the optimal level, but at the expense of an increase in variance, which did not allow the estimators proposed in the aforementioned papers to surpass the classical estimators for small  $k$  values. Further details can be found in Gomes and Guillou [19], among others.

For heavy-tailed models, Caeiro et al. [20, 21] and Gomes et al. [22, 23], among others, were able to overcome the bias-variance trade-off, keeping the variance of the proposed reduced bias EVI estimators equal to  $\xi^2/k$ , the variance of the Hill EVI estimator. For that reason, these estimators were coined minimum variance reduced bias (MVRB) EVI estimators. These classes of estimators depend upon the adequate consistent estimation of the vector of second-order parameters,  $(\beta, \rho) \in (\mathbf{R}, \mathbf{R}^-)$  in  $A(t) = \xi \beta t^\rho$  in a high level  $k_1$  of larger order than  $k$ , the number of top order statistics used in the EVI estimation, and surmount the classical EVI estimators for all  $k$  values.

Practical issues related with the estimation of  $(\beta, \rho)$  and algorithms are presented by Gomes and Pestana [24], among others. Among several classes of MVRB EVI estimators, we refer to the class introduced in Caeiro et al. [20], the corrected Hill (CH) estimator:

$$\text{CH}_{n,k} \equiv \text{CH}_{n,k}(\widehat{\beta}, \widehat{\rho}) := H_{n,k} \left( 1 - \frac{\widehat{\beta}}{1 - \widehat{\rho}} \left( \frac{n}{k} \right)^\rho \right), \quad (15)$$

with  $(\widehat{\beta}, \widehat{\rho})$  a set of adequate estimators of the second-order parameters  $(\beta, \rho)$  (see Section 3). The estimator in Equation (15) will be used in Section 4 for comparison with the new class of MVRB EVI estimators introduced in this paper. Other techniques have been used in the literature to reduce the bias of the Hill estimator, in Equation (14), and new estimators based on generalized means or norms that generalize the Hill estimator have been proposed (see Brillhante et al. [25], Caeiro et al. [26–29], Paulauskas and Vaciulis [8, 9], Beran et al. [30], and Beirlant et al. [31], among others).

In Section 2 of this paper, we present the BC Hill estimator and derive the consistency and asymptotic normality of the estimator under study assuming that the BC parameters are known. Next, in Section 3, we address the estimation of the BC parameters, and in Section 4, we present the finite sample behaviour of the BC Hill estimators and compare it with the corrected Hill estimator. Finally, in Section 5, some concluding remarks and future work are put forward.

## 2. The Box-Cox Hill Estimator

The Box-Cox Hill (BC Hill) estimator introduced by Teugels and Vanroelen [3] has the functional form:

$$H_{n,k}^{BC}(\sigma, a) := \frac{1}{k\sigma} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n}^\sigma + a}{X_{n-k:n}^\sigma + a} \right), \quad (16)$$

where  $a > 0$ , so that  $\{X_i^\sigma + a\}_{i=1}^n$  is strictly positive,  $\sigma = -\rho/\xi > 0$ , as in Equation (13) and,  $H_{n,k}^{BC}(\sigma, 0) \equiv H_{n,k}$  in Equation (14). The estimators  $H_{n,k}$  in Equation (14), and  $H_{n,k}^{BC}(\sigma, a)$ , in Equation (16), are both consistent for the estimation of  $\xi$ , for intermediate sequences of integers  $k$ , i.e., such that

$$k = k_n, \quad 1 \leq k < n, \quad k_n \rightarrow \infty, \quad k_n = o(n), \quad \text{as } n \rightarrow \infty. \quad (17)$$

Under the second-order framework in Equation (9), the asymptotical distributional representation

$$H_{n,k} \stackrel{d}{=} \xi + \frac{\xi}{\sqrt{k}} Z_k + \frac{1}{1-\rho} A\left(\frac{n}{k}\right) (1 + o_p(1)) \quad (18)$$

holds true [32], where  $Z_k = \sqrt{k}(\sum_{i=1}^k E_i/k - 1)$ , with  $\{E_i\}$  a sequence of i.i.d. standard exponential r.v.'s, is an asymptotically standard normal r.v.

*Remark 4.* The BC Hill estimator is also a generalization of the EVI estimator introduced by Gomes and Oliveira [33],

$$H_{n,k}^{(a)} \equiv H_{n,k}^{BC}(1, a) := \frac{1}{k} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n} + a}{X_{n-k:n} + a} \right), \quad (19)$$

where the shift  $a$  imposed to the data is a tuning parameter. The authors studied the adequate choice of  $a$  that enabled the reduction of the main component of the bias of the Hill's estimator, in Equation (14). The results obtained in the simulation studies for models with  $\sigma = 1$  are in line with the ones provided in Section 4.

*2.1. Asymptotic Behaviour of the Box-Cox Hill Estimator When  $\sigma$  and  $a$  Are Known.* In this section, we start by deriving the asymptotic behaviour of the BC Hill estimator assuming that the BC parameters,  $\sigma$  and  $a$ , are known. We then get an optimal value for  $a$  that allows us to obtain a new class of MVRB EVI estimators.

**Theorem 5.** *In Hall-Welsh class of models in Equation (11), for intermediate  $k$ , i.e., for  $k$  values such that Equation (17) holds, we have the validity of following distributional representation,*

$$H_{n,k}^{BC}(\sigma, a) \stackrel{d}{=} \xi + \frac{\xi}{\sqrt{k}} Z_k + \frac{A(n/k)}{1-\rho} \left( 1 - \frac{a}{\beta C^\sigma} \right) (1 + o_p(1)). \quad (20)$$

If  $\sqrt{k}A(n/k) \rightarrow \lambda$ , possibly nonnull, we can guarantee that

$$\sqrt{k} (H_{n,k}^{BC}(\sigma, a) - \xi) \xrightarrow[n \rightarrow \infty]{d} \xi N(0, 1) + \frac{\lambda}{1-\rho} \left( 1 - \frac{a}{\beta C^\sigma} \right), \quad (21)$$

with  $N(0, 1)$  denoting a standard normal r.v. If we further choose

$$a = \beta C^\sigma, \quad (22)$$

then

$$\sqrt{k} (H_{n,k}^{BC}(\sigma, a) - \xi) \xrightarrow[n \rightarrow \infty]{d} \xi N(0, 1). \quad (23)$$

*Proof.* The universal uniform transformation and from the definition of the reciprocal quantile function  $U(\cdot)$  in Equation (11), we get  $X_{i:n}^\sigma = {}^d U(Y_{i:n})$ , with  $Y$  a unit Pareto r.v. In what follows, we shall use the following relations: for  $j > i$ ,  $Y_{j:n}/Y_{i:n} = {}^d Y_{j-i:n-i}$ ,  $\ln Y_{i:n} = {}^d E_{i:n}$ , where  $E$  denotes a standard exponential r.v., and  $Y_{n-k:n} \sim {}^p n/k$ , whenever Equation (8) holds.

$$\begin{aligned} H_{n,k}^{BC}(\sigma, a) &= \frac{1}{k\sigma} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n}^\sigma + a}{X_{n-k:n}^\sigma + a} \right) \\ &\stackrel{d}{=} \frac{1}{k\sigma} \sum_{i=1}^k \ln \left( \frac{U^\sigma(Y_{n-i+1:n}) + a}{U^\sigma(Y_{n-k:n}) + a} \right) \\ &\stackrel{d}{=} \frac{1}{k\sigma} \sum_{i=1}^k \ln \left( \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} \right)^\sigma \\ &\quad + \frac{1}{k\sigma} \sum_{i=1}^k \ln \left( \frac{1 + a/U^\sigma(Y_{n-i+1:n})}{1 + a/U^\sigma(Y_{n-k:n})} \right) \\ &= \underbrace{\frac{1}{k} \sum_{i=1}^k \ln \left( \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} \right)}_{H_{n,k}} \\ &\quad + \frac{1}{k\sigma} \sum_{i=1}^k \left( \frac{a}{U^\sigma(Y_{n-i+1:n})} - \frac{a}{U^\sigma(Y_{n-k:n})} \right) \\ &= H_{n,k} + \frac{1}{k\sigma} \sum_{i=1}^k \frac{a}{U^\sigma(Y_{n-k:n})} \\ &\quad \cdot \left( -1 + \left\{ \frac{U(Y_{n-k:n})}{U(Y_{n-i+1:n})} \right\}^\sigma \right). \end{aligned} \quad (24)$$

For the second term in the right hand side, we know that under the validity of Equations (8) and (11), we can write  $U(Y_{n-k:n}) \sim U(n/k) = C(n/k)^\xi$  and  $U(Y_{n-i+1:n})/U(Y_{n-k:n})$

TABLE 1: First-order, second-order, and Box-Cox parameters for some heavy-tailed models.

Model	First-order			Second-order		Box-Cox
	$\xi$	$C$	$\rho$	$\beta$	$\sigma$	$a$
Burr $_{\xi,\rho}$	$\xi$	1	$\rho$	1	$-\rho/\xi$	1
Frechet $_{\xi}$	$\xi$	1	-1	1/2	1/ $\xi$	1/2
Log-logistic $_{\theta,\xi}$	$\xi$	$\theta^{-\xi}$	-1	1	1/ $\xi$	1/ $\theta$
Generalized Pareto, GP $_{\xi}$	$\xi$	1/ $\xi$	$-\xi$	1	1	1/ $\xi$
Cauchy	1	1/ $\pi$	-2	2 $\pi^2/3$	2	2/3
Student $t_{\nu}$	1/ $\nu$	$\sqrt{\nu}/c_{\nu}$	-2/ $\nu$	$((\nu+1)c_{\nu}^2)/(\nu+2)$	2	$(\nu(\nu+1))/(\nu+2)$

$= {}^d Y_{k-i+1:k}^{-\xi} (1 + o_p(1))$ . Then,

$$H_{n,k}^{BC}(\sigma, a) = H_{n,k} + \frac{1}{\sigma} \frac{a}{C^{\sigma}} \left(\frac{n}{k}\right)^{-\xi\sigma} \frac{1}{k} \sum_{i=1}^k \left(-1 + Y_i^{-\xi\sigma}\right). \quad (25)$$

Since  $E[Y^a] = 1/(1-a)$ , for  $a < 1$ ,

$$\begin{aligned} H_{n,k}^{BC}(\sigma, a) &\stackrel{d}{=} \xi + \frac{\xi}{\sqrt{k}} Z_k + \frac{1}{1-\rho} A(n/k) (1 + o_p(1)) \\ &\quad + \frac{a}{C^{\sigma}} \left(\frac{n}{k}\right)^{-\xi\sigma} \frac{-\xi}{1+\xi\sigma} (1 + o_p(1)) \\ &\stackrel{d}{=} \xi + \frac{\xi}{\sqrt{k}} Z_k + \left(\frac{A(n/k)}{1-\rho} + \frac{a}{C^{\sigma}} \left(\frac{n}{k}\right)^{-\xi\sigma} \frac{-\xi}{1+\xi\sigma}\right) \\ &\quad \cdot (1 + o_p(1)). \end{aligned} \quad (26)$$

Using the parametrizations  $A(t) = \xi\beta t^{\rho}$  and  $\sigma = -\rho/\xi$ , the distributional representation in Equation (20) holds true. We then get an optimal  $a$  value, the value of  $a$  in Equation (22), such that that the dominant component of the bias of the BC Hill estimator is null, and Equation (23) follows.  $\square$

**Corollary 6.** *Under the first-order condition in Equation (12), and for the heavy-tailed models in Hall and Welsh class, in Equation (11), such that  $U^*(t) = c_1 t^{\xi^*} + c_2$ , with  $c_1$  and  $c_2$  real constants,  $c_1 \neq 0$ , the BC Hill estimator in Equation (16) is asymptotically unbiased for any  $k = k_n \rightarrow \infty$ , not necessarily intermediate.*

*Proof.* This result comes straightforwardly from the definition of the Box-Cox Hill estimator in Equation (16) and from the fact that we are working with heavy-tailed models in Hall and Welsh class, with quantile function  $U^*(t) = c_1 t^{\xi^*} + c_2$ ,  $c_1$  and  $c_2$  real constants,  $c_1 \neq 0$ .  $\square$

*Remark 7.* The heavy-tailed models in Hall and Welsh class that satisfy the property in Corollary 6 are the Burr $_{\xi,\rho}$ , the

log-logistic $_{\theta,\xi}$ , and the GP $_{\xi}$  parents. This behaviour will be illustrated in Section 4.

In Table 1, we show for some heavy-tailed models the optimal values of  $\sigma$  and  $a$ , in Equations (13) and (22), respectively, where  $c_{\nu} = (\nu\mathcal{B}(\nu/2, 1/2))^{1/\nu}$ , with  $\mathcal{B}$  the complete beta function.

*Remark 8.* Teugels and Vanroelen [3] considered two examples to show the bias reduction of the BC Hill estimator. In the first one, a Burr $_{\xi=1,\rho=-0.5}$  model was chosen. The power of the BC transformation is  $\sigma = 0.5$ , and the authors used  $a = 1$ , which is, according to Equation (22) and Table 1, the optimal value for bias reduction. In the second example, the authors considered a Cauchy model ( $\xi = 1$ ,  $\rho = -2$ ), with  $\sigma = 2$ . It was noticed that the choice of  $a$  is very important and a value around 0.5 was suggested as it was the value that had the strongest effect of minimizing the bias of the BC Hill estimator. We now know, from Equation (22) and Table 1, that the optimal value is  $a = 2/3$ .

### 3. Adaptive Estimation of the Box-Cox Parameters

The BC parameters depend on the estimation of the first-order scale parameter  $C$  and on the estimation of the second-order shape,  $\rho$ , and scale,  $\beta$ , parameters. In this work, we are going to consider a misspecification of the first-order scale parameter, considering that  $C = 1$ .

For the estimation of the second-order shape parameter,  $\rho$ , we consider the class of estimators introduced in [34] defined by the following:

$$\hat{\rho}(k; \tau) := - \frac{\left| 3 \left( T_n^{(\tau)}(k) - 1 \right) \right|}{\left| T_n^{(\tau)}(k) - 3 \right|}, \quad (27)$$

with  $\tau$  a real tuning parameter [35]:

$$T_n^{(\tau)}(k) := \begin{cases} \frac{\ln(M_n^{(1)}(k)) - 1/2 \ln(M_n^{(2)}(k)/2)}{1/2 \ln(M_n^{(2)}(k)/2) - 1/3 \ln(M_n^{(3)}(k)/6)}, & \text{if } \tau = 0, \\ \frac{(M_n^{(1)}(k))^\tau - (M_n^{(2)}(k)/2)^{\tau/2}}{(M_n^{(2)}(k)/2)^{\tau/2} - (M_n^{(3)}(k)/6)^{\tau/3}}, & \text{if } \tau = 1, \end{cases}$$

$$M_n^{(\alpha)}(k) := \frac{1}{k} \sum_{i=1}^k (\ln X_{n-i+1:n} - \ln X_{n-k:n})^\alpha. \quad (28)$$

Note that  $M_n^{(1)}(k) \equiv H_{n,k}$ , in Equation (14), and the results obtained in Fraga Alves et al. [34] and in Gomes and Martins [36] advise to consider  $\tau = 0$  for values of  $\rho \geq -1$  and  $\tau = 1$  for values of  $\rho > -1$ .

For the estimation of the second-order scale parameter,  $\beta$ , we are going to consider the estimator introduced in Gomes and Martins [36], given by the following:

$$\widehat{\beta}(k; \widehat{\rho}) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{d_k(-\widehat{\rho}) \times D_k(0) - D_k(-\widehat{\rho})}{d_k(\widehat{\rho}) \times D_k(-\widehat{\rho}) - D_k(-2\widehat{\rho})}, \quad (29)$$

where for any  $\alpha \geq 0$ ,

$$D_k(\alpha) := \frac{1}{k} \sum_{i=1}^k \left(\frac{i}{k}\right)^\alpha U_i, \quad (30)$$

$$d_k(\alpha) := \frac{1}{k} \sum_{i=1}^k \left(\frac{i}{k}\right)^\alpha,$$

with  $U_i = i\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$ ,  $1 \leq i \leq k$ , the scaled log-spacings, and  $\widehat{\rho}$  the  $\rho$  estimator in Equation (27).

Asymptotic considerations as well as simulated results led several authors to consider the estimation of the second-order estimators at a high level  $k_1$ , given by  $k_1 := \lceil n^{1-\varepsilon} \rceil$  where  $\varepsilon \sim 0^+$  and  $\lceil x \rceil$  denotes, as usual, the integer part of  $x$ . In the simulation studies in Section 4, the value  $\varepsilon = 0.001$  was used to compute the estimates of the second-order parameters. Under adequate general conditions, the classes of estimators in Equations (27) and (29) are asymptotically normally distributed and show highly stable sample paths as functions of  $k$ , the number of top order statistics used, for a wide range of large  $k$  values.

**Theorem 9** (Fraga Alves et al. [34] and Gomes and Martins [36]). *If the SOC in Equation (9) holds, for models in Hall and Welsh class with  $A(t) = \xi \beta t^\rho$ ,  $\rho < 0$ , if  $k$  is intermediate, i.e., if Equation (17) holds and if  $\sqrt{k}A(n/k) \xrightarrow{n \rightarrow \infty} \infty$ , then  $\widehat{\rho}(k)$  converges in probability towards  $\rho$ , as  $n \rightarrow \infty$ , and  $\widehat{\beta}(k; \rho)$  converges in probability towards  $\beta$ , as  $n \rightarrow \infty$ . Moreover, with  $W_k^* \sim^a \text{Normal}(0, 1)$ ,  $\bullet = R, B$ ,  $\sigma_\rho = \xi(1-\rho)^3 \sqrt{2\rho^2 - 2\rho + 1}/\rho$ , and  $\sigma_\beta = |\beta| (1-\rho) \sqrt{1-2\rho}/|\rho|$ ,*

$$\widehat{\rho}(k; \tau) \stackrel{d}{=} \rho + \frac{\sigma_\rho}{\sqrt{k}A(n/k)} W_k^R + o_p(1), \quad (31)$$

$$\widehat{\beta}(k; \rho) \stackrel{d}{=} \beta + \frac{\sigma_\beta}{\sqrt{k}A(n/k)} W_k^B + o_p(1).$$

As stated earlier, the estimates of  $\beta$  and  $\rho$  are computed at the high level  $k_1$  and the following results in Caeiro and Gomes [37] hold true.

**Theorem 10** (Caeiro and Gomes [37]). *Under the conditions of the previous theorem, with  $\widehat{\rho} = \widehat{\rho}(k_1; \tau)$  for any  $\tau$  and  $k_1$  such that*

$$\widehat{\rho} - \rho = o_p\left(\frac{1}{\ln n}\right), \quad \text{as } n \rightarrow \infty, \quad (32)$$

*$\widehat{\beta}(k; \widehat{\rho})$  is consistent for the estimation of  $\beta$ . Moreover, for models in Hall and Welsh class and with  $\widehat{\beta} := \widehat{\beta}(k_1, \widehat{\rho})$ , we may further say that  $\widehat{\beta} - \beta = O_p(\ln(n/k_1)(\widehat{\rho} - \rho))$ , at most of the order of  $(\widehat{\rho} - \rho) \ln n$ .*

In addition to the classes presented in Equations (27) and (29), we also refer the classes introduced in Goegebeur et al. [38], Ciuperca and Mercadier [39], Worms and Worms [40], de Wet et al. [41], and Henriques-Rodrigues et al. [42], among others, for the estimation of  $\rho$  and the classes introduced in Caeiro and Gomes [35] and Gomes et al. [43], for the estimation of  $\beta$ .

In this paper, we are going to work with adaptive estimates of the BC parameters in Equations (13) and (22). So, for the estimation of  $\sigma$ , we consider the following:

$$\widehat{\sigma} = -\frac{\widehat{\rho}(k_1)}{H_{n, \widehat{k}_0^H}}, \quad (33)$$

with  $\widehat{\rho}$  the  $\rho$  estimator in Equation (27),  $k_1$  the high level used in the estimation of the second-order parameters, and  $H_{n, \widehat{k}_0^H}$  the Hill estimator computed at the level  $\widehat{k}_0^H := k_0^H(\widehat{\beta}, \widehat{\rho})$ , with  $k_0^H(\beta, \rho) := \arg \min_k \text{AMSE}(H_{n,k})$  [44], given by the following:

$$k_0^H(\beta, \rho) := \left( \frac{(1-\rho)^2 n^{-2\rho}}{-2\rho\beta^2} \right)^{1/(1-2\rho)}. \quad (34)$$

For the estimation of  $a$  and with the misspecification  $C = 1$ , we are going to consider:

$$\widehat{a} = \widehat{\beta}(k_1), \quad (35)$$

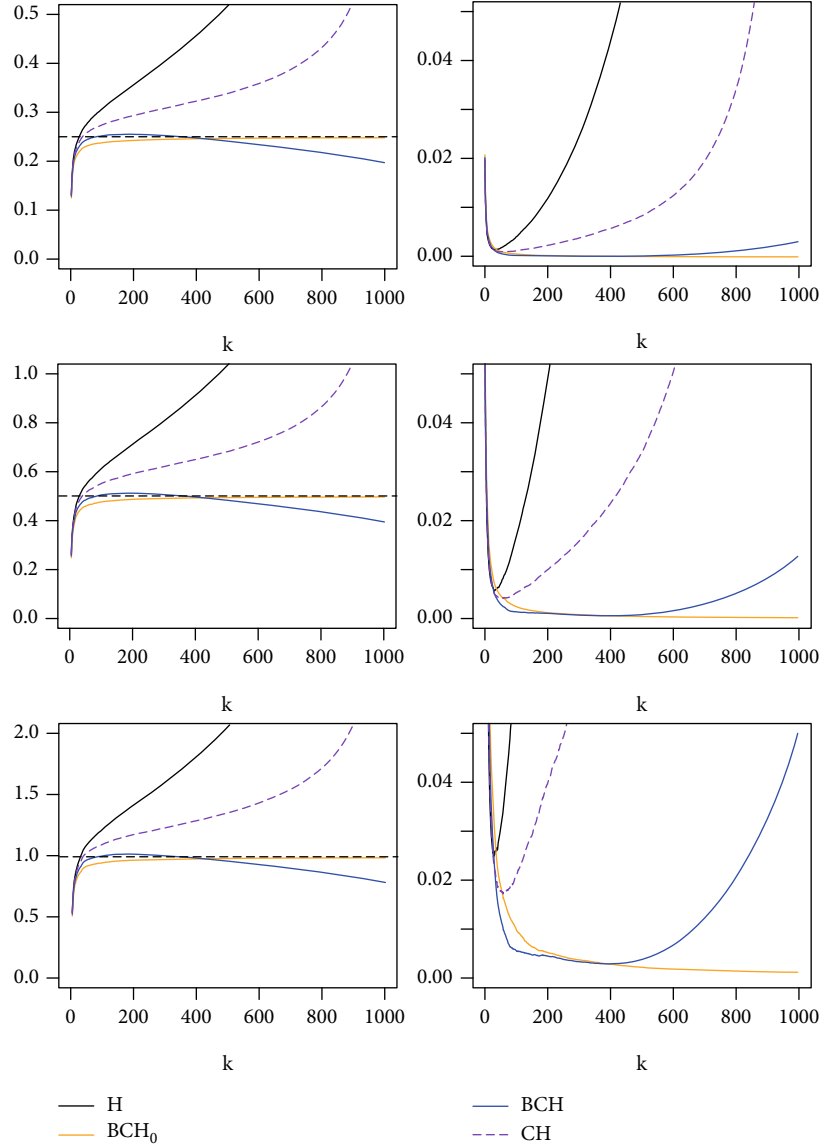


FIGURE 1: Mean values (left) and MSEs (right) of the estimators under study for samples of size  $n = 1000$  from a Burr $_{\xi, \rho=0.5}$  parent with  $\xi = 0.25$  (top),  $\xi = 0.5$  (middle), and  $\xi = 1$  (bottom).

with  $\hat{\beta}$  the  $\beta$  estimator in Equation (29). The associated estimators of  $\sigma$ ,  $\hat{\sigma}(k) = \hat{\rho}(k)/\hat{\xi}(k)$  and  $a$ , considering  $C = 1$ , i.e.  $\hat{a}(k) = \hat{\beta}(k)$ , which are consistent for the estimation of  $\sigma$  and  $a$ , respectively, for sequences of  $k$  values such that Equation (17) holds. The BC Hill estimator, in Equation (16), assuming that the BC parameters are adaptively estimated has the functional form:

$$H_{n,k}^{\text{BCH}} \equiv H_{n,k}^{\text{BC}}(\hat{\sigma}, \hat{a}) := \frac{1}{k\hat{\sigma}} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n}^{\hat{\sigma}}}{X_{n-k:n}^{\hat{\sigma}} + \hat{a}} \right), \quad (36)$$

with  $\hat{\sigma}$  and  $\hat{a}$  presented in Equations (33) and (35), respectively. The consistency of the BCH estimator in Equation (36) for the estimation of the EVI,  $\xi$ , is presented in the next theorem.

**Theorem 11.** *In Hall-Welsh class of models in Equation (11), for intermediate  $k$ , i.e., for  $k$  values such that Equation (17) holds, with  $\hat{\sigma}$  and  $\hat{a}$  presented in Equations (33) and (35), consistent estimators of  $\sigma$  and  $a$ , respectively, then  $H_{n,k}^{\text{BC}}(\hat{\sigma}, \hat{a}) \xrightarrow{\mathcal{P}} \xi$ .*

*Proof.* The delta method is not going to be directly applied since the partial derivatives of the estimator  $H_{n,k}^{\text{BC}}(\hat{\sigma}, \hat{a})$ , in Equation (36), in order to  $(\sigma, a)$  are complex. We first note that  $X_{n-i+1:n} > X_{n-k:n}$ ,  $1 \leq i \leq k < n$ , with  $\hat{\sigma}$  a consistent estimator of  $\sigma > 0$  and  $\hat{a}$  a consistent estimator of  $a > 0$ , and defining  $\hat{a}^* := r\hat{a}$ , with  $r > 0$  (a consistent estimator of  $a^* = ra$ ), then

$$\left( \frac{X_{n-i+1:n} + \hat{a}^*}{X_{n-k:n} + \hat{a}^*} \right)^{\hat{\sigma}} \leq \frac{X_{n-i+1:n} + \hat{a}}{X_{n-k:n} + \hat{a}} < \left( \frac{X_{n-i+1:n}}{X_{n-k:n}} \right)^{\hat{\sigma}}. \quad (37)$$

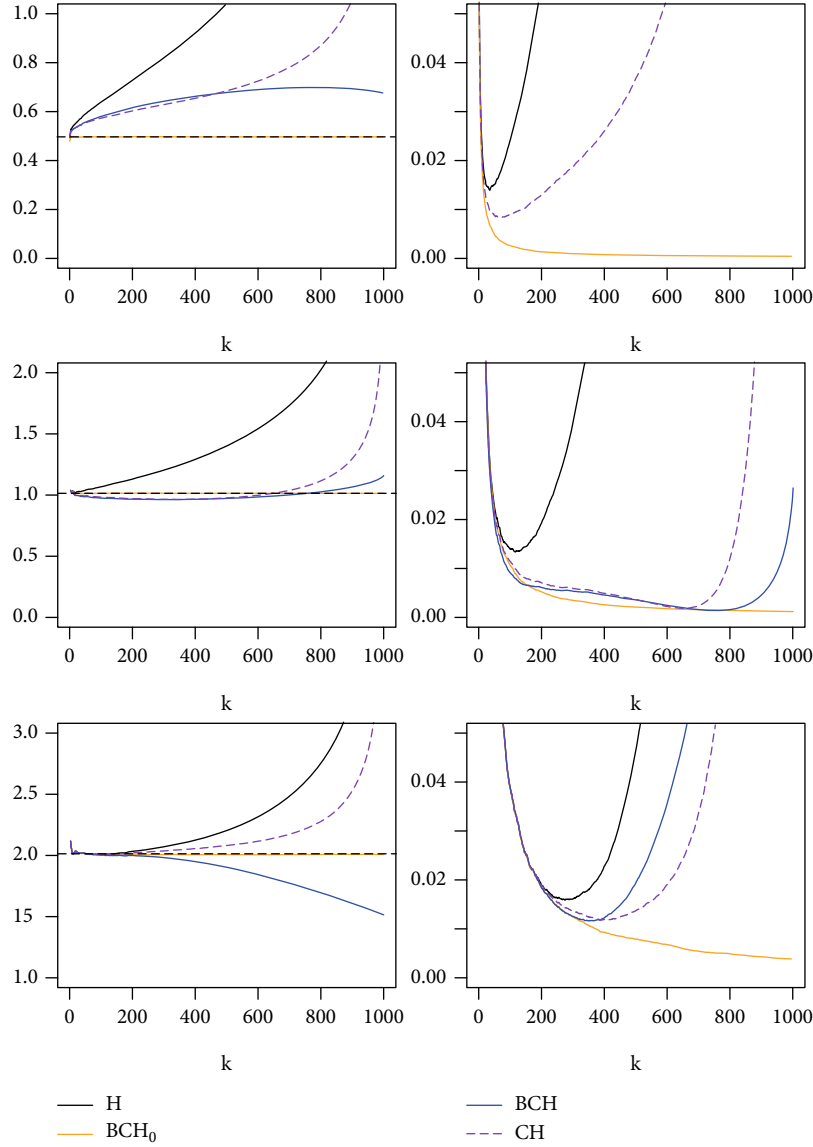


FIGURE 2: Mean values (left) and MSEs (right) of the estimators under study for samples of size  $n = 1000$  from  $GP_\xi$  parent with  $\xi = 0.5$  (top),  $\xi = 1$  (middle), and  $\xi = 2$  (bottom).

If  $\widehat{\sigma} \geq 1$ , we can choose  $r = 1$ , and if  $0 < \widehat{\sigma} < 1$ , the value of  $r$  depends upon the underlying model in study. So, we can write the following:

$$\begin{aligned} H_{n,k}^{\widehat{a}^*} &:= \frac{1}{k} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n} + \widehat{a}^*}{X_{n-k:n} + \widehat{a}^*} \right) \leq \frac{1}{k\widehat{\sigma}} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n}^{\widehat{\sigma}}}{X_{n-k:n}^{\widehat{\sigma}} + \widehat{a}} \right) \\ &< \frac{1}{k} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n}}{X_{n-k:n}} \right) = H_{n,k}. \end{aligned} \quad (38)$$

The estimator in the upper bound is the Hill estimator,  $H_{n,k}$ , in Equation (14), consistent for the estimation of  $\xi$ . For

the estimator in the lower bound, the delta method leads to the following:

$$H_{n,k}^{\widehat{a}^*} \stackrel{d}{=} H_{n,k}^{a^*} + \frac{\partial H_{n,k}^{a^*}}{\partial a^*} (\widehat{a}^* - a^*) (1 + o_p(1)), \quad (39)$$

with

$$H_{n,k}^{a^*} = \frac{1}{k} \sum_{i=1}^k \ln \left( \frac{X_{n-i+1:n} + a^*}{X_{n-k:n} + a^*} \right). \quad (40)$$

The estimator  $H_{n,k}^{a^*}$  is of the same type of the estimator  $H_{n,k}^{(a)}$ , in Equation (19), and is therefore consistent for the



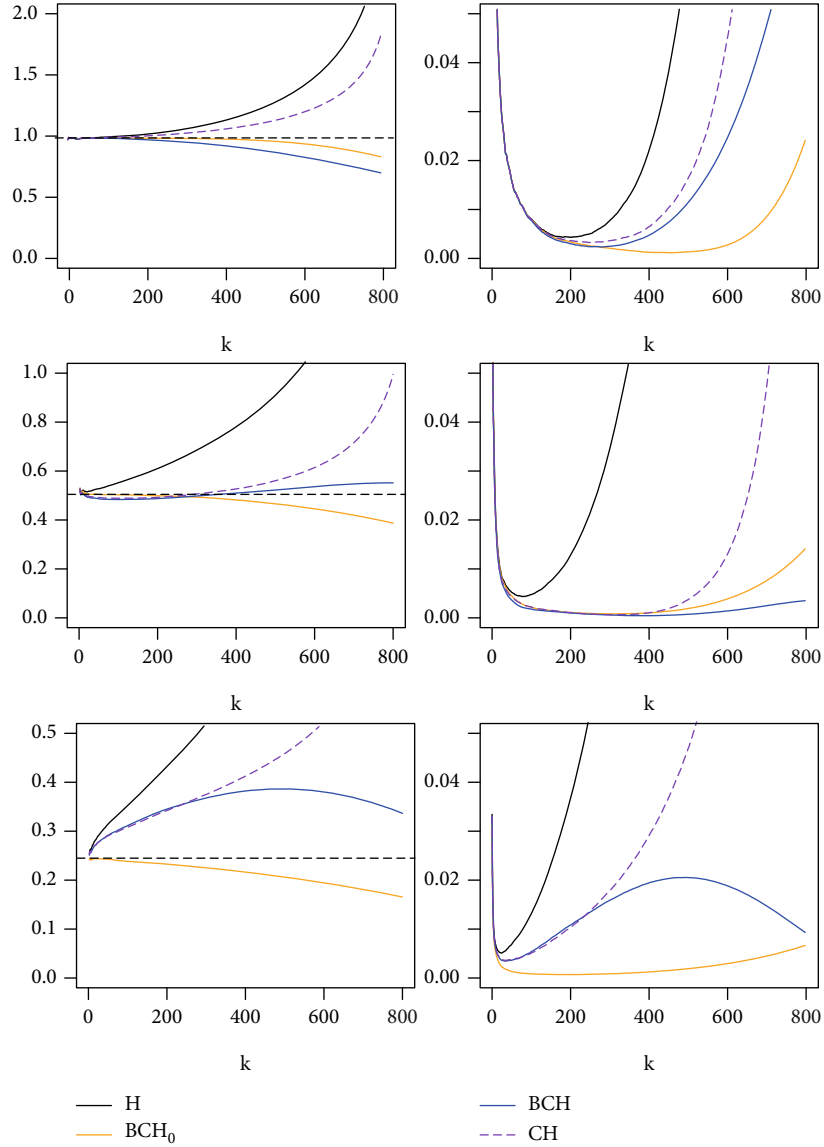


FIGURE 3: Mean values (left) and MSEs (right) of the estimators under study for samples of size  $n = 2000$  from a Cauchy parent (top), a Student  $t_2$  parent (middle), and a Student  $t_4$  parent (bottom).

estimation of  $\xi$  as proven in Section 2 of Gomes and Oliveira [33]. The term  $\partial H_{n,k}^{a^*} / \partial a^*$  can be written as follows:

$$\begin{aligned} \frac{\partial H_{n,k}^{a^*}}{\partial a^*} &= \frac{1}{k} \sum_{i=1}^k \frac{X_{n-k:n} - X_{n-i+1:n}}{X_{n-i+1:n} X_{n-k:n}} \\ &= \frac{1}{k} \sum_{i=1}^k \frac{X_{n-k:n} / X_{n-i+1:n} - 1}{X_{n-k:n}}. \end{aligned} \quad (41)$$

Let us next use the same kind of arguments referred to in the proof of Theorem 5:  $X_{i:n} = {}^d U(Y_{i:n})$ , with  $Y$  a unit Pareto r.v. and  $U(\cdot)$  the quantile function in Equation (11), for  $j > i$ ,  $Y_{j:n} / Y_{i:n} = {}^d Y_{j-i:n-i}$ ,  $\ln Y_{i:n} = {}^d E_{i:n}$ , where  $E$  denotes a stan-

dard exponential r.v., and  $Y_{n-k:n} \sim {}^p n/k$ , whenever Equation (8) holds. Then, we have the following:

$$\frac{\partial H_{n,k}^{a^*}}{\partial a^*} \underset{p}{\approx} \frac{1}{kU(n/k)} \frac{-\xi}{1+\xi}, \quad (42)$$

and we can write the following:

$$H_{n,k}^{\hat{a}^*} \stackrel{d}{=} H_{n,k}^{a^*} - \frac{(\hat{a}^* - a^*)}{kU(n/k)} \frac{\xi}{1+\xi} (1 + o_p(1)). \quad (43)$$

Working with  $k$  values such that Equation (17) holds and given that  $\hat{a}^*$  is a consistent estimator of  $a^*$  and  $H_{n,k}^{a^*}$  is consistent for the estimation of  $\xi$ , the estimator  $H_{n,k}^{\hat{a}^*}$  is also

TABLE 2: Box-Cox estimates and associated 95% CIs based on 10 replicates with 1000 runs each.

	200	500	1000	2000	5000
Burr parent with $\xi = 0.25$ and $\rho = -0.5$					
$\sigma = 2$	$2.2759 \pm 0.0060$	$2.4373 \pm 0.0031$	$2.5529 \pm 0.0044$	$2.6605 \pm 0.0034$	$2.8112 \pm 0.0027$
$a = 1$	$1.0114 \pm 0.0001$	$1.0116 \pm 0.0001$	$1.0114 \pm 0.0000$	$1.0120 \pm 0.0001$	$1.0125 \pm 0.0000$
Burr parent with $\xi = 0.5$ and $\rho = -0.5$					
$\sigma = 1$	$1.1404 \pm 0.0030$	$1.2206 \pm 0.0035$	$1.2758 \pm 0.0026$	$1.3321 \pm 0.0025$	$1.4035 \pm 0.0018$
$a = 1$	$1.0113 \pm 0.0001$	$1.0115 \pm 0.0001$	$1.0115 \pm 0.0000$	$1.0121 \pm 0.0000$	$1.0124 \pm 0.0000$
Burr parent with $\xi = 1$ and $\rho = -0.5$					
$\sigma = 0.5$	$0.5696 \pm 0.0009$	$0.6113 \pm 0.0010$	$0.6375 \pm 0.0011$	$0.6658 \pm 0.0012$	$0.7034 \pm 0.0006$
$a = 1$	$1.0114 \pm 0.0001$	$1.0114 \pm 0.0001$	$1.0114 \pm 0.0000$	$1.0121 \pm 0.0000$	$1.0125 \pm 0.0000$
GP parent with $\xi = 0.5$					
$\sigma = 1$	$1.1426 \pm 0.0047$	$1.2199 \pm 0.0027$	$1.2760 \pm 0.0030$	$1.3318 \pm 0.0024$	$1.4045 \pm 0.0012$
$a = 2$	$1.0114 \pm 0.0001$	$1.0115 \pm 0.0001$	$1.0115 \pm 0.0001$	$1.0121 \pm 0.0001$	$1.0125 \pm 0.0000$
GP parent with $\xi = 1$					
$\sigma = 1$	$0.7774 \pm 0.0027$	$0.7998 \pm 0.0017$	$0.8166 \pm 0.0013$	$0.8347 \pm 0.0013$	$0.8611 \pm 0.0008$
$a = 1$	$0.9913 \pm 0.0001$	$0.9925 \pm 0.0001$	$0.9930 \pm 0.0001$	$0.9928 \pm 0.0001$	$0.9927 \pm 0.0000$
GP parent with $\xi = 2$					
$\sigma = 1$	$1.0880 \pm 0.0018$	$1.1165 \pm 0.0010$	$1.1331 \pm 0.0015$	$1.1409 \pm 0.0013$	$1.1519 \pm 0.0006$
$a = 0.5$	$1.0133 \pm 0.0005$	$1.0152 \pm 0.0003$	$1.0157 \pm 0.0001$	$1.0166 \pm 0.0002$	$1.0172 \pm 0.0001$
Cauchy parent with $\xi = 1$					
$\sigma = 2$	$1.9266 \pm 0.0043$	$2.0056 \pm 0.0042$	$2.0507 \pm 0.0039$	$2.0859 \pm 0.0020$	$2.1143 \pm 0.0019$
$a = 2/3$	$1.0524 \pm 0.0005$	$1.0536 \pm 0.0002$	$1.0551 \pm 0.0001$	$1.0554 \pm 0.0001$	$1.0577 \pm 0.0001$
Student $t_2$ parent with $\xi = 0.5$					
$\sigma = 2$	$1.3044 \pm 0.0046$	$1.3636 \pm 0.0048$	$1.3944 \pm 0.0028$	$1.4171 \pm 0.0018$	$1.4397 \pm 0.0022$
$a = 1.5$	$1.0133 \pm 0.0002$	$1.0143 \pm 0.0001$	$1.0146 \pm 0.0001$	$1.0148 \pm 0.0000$	$1.0155 \pm 0.0000$
Student $t_4$ parent with $\xi = 0.25$					
$\sigma = 2$	$1.7947 \pm 0.0082$	$1.9771 \pm 0.0078$	$2.0919 \pm 0.0056$	$2.1934 \pm 0.0031$	$2.3023 \pm 0.0040$
$a = 10/3$	$1.0236 \pm 0.0002$	$1.0242 \pm 0.0001$	$1.0245 \pm 0.0001$	$1.0247 \pm 0.0001$	$1.0257 \pm 0.0000$

consistent for the estimation of  $\xi$  and the consistency of  $H_{n,k}^{\text{BC}}(\hat{\sigma}, \hat{a})$  follows.  $\square$

*Remark 12.* In the paper by Gomes and Oliveira [33], the tuning parameter  $a$  of the estimator  $H_{n,k}^{(a)}$  presented in Remark 4 was estimated by the order statistic  $1 - X_{1:n}$ . According to the authors, the estimator  $H_{n,k}^{(1-X_{1:n})}$  is consistent for the estimation of  $\xi$ , with asymptotic distributional properties similar to those of the Hill estimator, for models in  $D_M^+$  with finite left endpoint.

*Remark 13.* The estimator  $H_{n,k}^{\hat{a}^*}$  has the same functional form of the shifted Hill estimator introduced in Aban and Meerschaert [45]. The shifted Hill estimator was obtained for shifted Pareto parents of the type  $F(x) = 1 - c(x+a)^{-1/\xi}$  and is a conditional maximum likelihood estimator of  $\xi$ . The estimator is consistent for the estimation of  $\xi$  even though  $\hat{a}$  is not consistent for the estimation of  $a$  under a

semiparametric framework (see Remark 2.1 of Gomes et al. [46]).

#### 4. Finite Sample Behaviour of the Estimators

To assess the finite sample behaviour of the BC Hill estimators, in Equations (16) and (36), we have implemented a multisample Monte Carlo simulation experiment of size  $1000 \times 10$  (see Gomes and Oliveira [47], for further details on multisample simulation) for sample sizes  $n = 200, 500, 1000, 2000,$  and  $5000$ , from the following models:

- (1) The Burr  $\xi, \rho$  model with d.f.  $F(x) = 1 - (1 + x^{-\rho/\xi})^{1/\rho}$ ,  $x > 0$ ,  $\xi = \{0.25, 0.5, 1\}$ , and  $\rho = -0.5$ ,
- (2) The generalized Pareto model, GP  $\xi$ , with d.f.  $F(x) = 1 - (1 + \xi x)^{-1/\xi}$ ,  $x \geq -1/\xi$ , and  $\xi = \{0.25, 0.5, 1\}$ ,
- (3) The Cauchy model, with d.f.  $F(x) = 1/2 + 1/\pi \arctan x$ ,  $x \in \mathbf{R}$ , and  $\xi = 1$ ,

TABLE 3: Optimal sample fraction and associated 95% CIs based on 10 replicates with 1000 runs each.

	200	500	1000	2000	5000
Burr parent with $\xi = 0.25$ and $\rho = -0.5$					
H	0.0645 ± 0.0887	0.0446 ± 0.0555	0.0357 ± 0.0991	0.0319 ± 0.0342	0.0290 ± 0.0342
CH	0.1225 ± 0.1311	0.0740 ± 0.1267	0.0635 ± 0.1261	0.0501 ± 0.0834	0.0427 ± 0.0847
BCH	0.4730 ± 0.1929	0.4242 ± 0.3135	0.4001 ± 0.1553	0.3971 ± 0.1352	0.3917 ± 0.2153
Burr parent with $\xi = 0.5$ and $\rho = -0.5$					
H	0.0670 ± 0.0751	0.0448 ± 0.0643	0.0360 ± 0.0261	0.0308 ± 0.0748	0.0291 ± 0.0449
CH	0.1335 ± 0.1371	0.0748 ± 0.1199	0.0594 ± 0.0796	0.0497 ± 0.0782	0.0409 ± 0.0951
BCH	0.5010 ± 0.1320	0.4238 ± 0.2908	0.4030 ± 0.1653	0.3903 ± 0.1700	0.3922 ± 0.1698
Burr parent with $\xi = 1$ and $\rho = -0.5$					
H	0.0645 ± 0.0729	0.0452 ± 0.0456	0.0352 ± 0.0553	0.0325 ± 0.0645	0.0294 ± 0.0514
CH	0.1265 ± 0.1547	0.0840 ± 0.1294	0.0614 ± 0.1679	0.0490 ± 0.1167	0.0426 ± 0.0862
BCH	0.4830 ± 0.2557	0.4194 ± 0.3387	0.3977 ± 0.2099	0.3946 ± 0.2314	0.3939 ± 0.1347
GP parent with $\xi = 0.5$					
H	0.0620 ± 0.0592	0.0414 ± 0.0523	0.0382 ± 0.0708	0.0304 ± 0.0609	0.0288 ± 0.0353
CH	0.1295 ± 0.1783	0.0844 ± 0.1299	0.0615 ± 0.1447	0.0473 ± 0.0936	0.0407 ± 0.0871
BCH	0.1470 ± 0.0973	0.0844 ± 0.1299	0.0579 ± 0.0859	0.0445 ± 0.1082	0.0366 ± 0.0935
GP parent with $\xi = 1$					
H	0.1745 ± 0.1247	0.1352 ± 0.0839	0.1182 ± 0.1179	0.1076 ± 0.1555	0.0993 ± 0.1192
CH	0.6810 ± 0.1458	0.6810 ± 0.1148	0.6839 ± 0.1014	0.6799 ± 0.0662	0.6777 ± 0.1206
BCH	0.7815 ± 0.1170	0.7802 ± 0.1236	0.7839 ± 0.1599	0.7895 ± 0.1218	0.7932 ± 0.1745
GP parent with $\xi = 2$					
H	0.3825 ± 0.1465	0.3384 ± 0.2355	0.3048 ± 0.3279	0.2863 ± 0.4116	0.2681 ± 0.3300
CH	0.5695 ± 0.1330	0.5244 ± 0.1528	0.4578 ± 0.3913	0.4362 ± 0.4030	0.3929 ± 0.4868
BCH	0.4260 ± 0.1367	0.3756 ± 0.1937	0.3342 ± 0.2768	0.3050 ± 0.2726	0.2606 ± 0.4894
Cauchy parent with $\xi = 1$					
H	0.4540 ± 0.011	0.3418 ± 0.0151	0.2971 ± 0.0073	0.2510 ± 0.0094	0.2032 ± 0.0060
CH	0.5637 ± 0.0180	0.4399 ± 0.0121	0.3909 ± 0.0094	0.3260 ± 0.0100	0.2693 ± 0.0110
BCH	0.5820 ± 0.0197	0.4473 ± 0.0117	0.3947 ± 0.0083	0.3416 ± 0.0059	0.2897 ± 0.0053
Student $t_2$ parent with $\xi = 0.5$					
H	0.2345 ± 0.0125	0.1554 ± 0.0064	0.1139 ± 0.0072	0.0956 ± 0.0048	0.0709 ± 0.0036
CH	0.5449 ± 0.0171	0.4729 ± 0.0133	0.4211 ± 0.0113	0.4009 ± 0.0061	0.3779 ± 0.0060
BCH	0.6799 ± 0.0349	0.6077 ± 0.0123	0.5371 ± 0.0125	0.4912 ± 0.0081	0.4479 ± 0.0065
Student $t_4$ parent with $\xi = 0.25$					
H	0.0943 ± 0.0089	0.0597 ± 0.0058	0.0457 ± 0.0030	0.0312 ± 0.0018	0.0208 ± 0.0007
CH	0.1633 ± 0.0153	0.1020 ± 0.0109	0.0670 ± 0.0058	0.0461 ± 0.0032	0.0301 ± 0.0010
BCH	0.2126 ± 0.0150	0.1176 ± 0.0090	0.0713 ± 0.0048	0.0460 ± 0.0033	0.0266 ± 0.0024

(4) The Student  $t_\nu$  model, with d.f.  $F(x) = \Gamma((\nu + 1)/2) / \Gamma(\nu/2) \sqrt{\pi\nu} \int_{-\infty}^x (1 + (y^2/\nu))^{-(\nu+1)/2} dy$ ,  $x \in \mathbf{R}$ , and  $\nu > 0$ , where  $\Gamma$  denotes the (complete) gamma function,  $\xi = \{0, 5, 0.25\}$  ( $\xi = 1/\nu$ ).

In Figures 1–3, we present the mean values,  $E(\cdot)$ , and mean squared errors,  $MSE(\cdot)$ , for the models under study provided by the Hill estimator,  $H$ , in Equation (14), the BC

Hill r.v. in Equation (16), assuming that the BC parameters are known and denoted by  $BCH_0$  in the figures, the BC Hill estimator, BCH, in Equation (36), assuming that the BC parameters are adaptively estimated by Equations (33) and (35), and the MVRB EVI estimator, CH, in Equation (15) for comparison. From Figures 1–3, we can state that:

- (i) The BCH estimator beats the Hill estimator for all simulated models and for all  $k$  values, as it presents

TABLE 4: Simulated mean values at optimal levels and associated 95% CIs based on 10 replicates with 1000 runs each.

	200	500	1000	2000	5000
Burr parent with $\xi = 0.25$ and $\rho = -0.5$					
H	0.2872 $\pm$ 0.0042	0.2702 $\pm$ 0.0023	0.2597 $\pm$ 0.0032	0.2563 $\pm$ 0.0010	0.2519 $\pm$ 0.0006
CH	0.2827 $\pm$ 0.0020	0.2681 $\pm$ 0.0022	0.2640 $\pm$ 0.0017	0.2579 $\pm$ 0.0011	0.2534 $\pm$ 0.0007
BCH	<b>0.2438</b> $\pm$ 0.0005	<b>0.2459</b> $\pm$ 0.0008	<b>0.2478</b> $\pm$ 0.0003	<b>0.2487</b> $\pm$ 0.0002	<b>0.2495</b> $\pm$ 0.0002
Burr parent with $\xi = 0.5$ and $\rho = -0.5$					
H	0.5775 $\pm$ 0.0068	0.5404 $\pm$ 0.0047	0.5235 $\pm$ 0.0021	0.5096 $\pm$ 0.0035	0.5054 $\pm$ 0.0018
CH	0.5687 $\pm$ 0.0053	0.5363 $\pm$ 0.0038	0.5254 $\pm$ 0.0031	0.5154 $\pm$ 0.0020	0.5051 $\pm$ 0.0014
BCH	<b>0.4870</b> $\pm$ 0.0011	<b>0.4921</b> $\pm$ 0.0011	<b>0.4953</b> $\pm$ 0.0007	<b>0.4979</b> $\pm$ 0.0004	<b>0.4991</b> $\pm$ 0.0003
Burr parent with $\xi = 1$ and $\rho = -0.5$					
H	1.1476 $\pm$ 0.0135	1.0805 $\pm$ 0.0076	1.0446 $\pm$ 0.0085	1.0285 $\pm$ 0.0047	1.0125 $\pm$ 0.0029
CH	1.1324 $\pm$ 0.0097	1.0830 $\pm$ 0.0067	1.0544 $\pm$ 0.0087	1.0273 $\pm$ 0.0064	1.0099 $\pm$ 0.0034
BCH	<b>0.9751</b> $\pm$ 0.0027	<b>0.9854</b> $\pm$ 0.0020	<b>0.9918</b> $\pm$ 0.0014	<b>0.9952</b> $\pm$ 0.0013	<b>0.9980</b> $\pm$ 0.0008
GP parent with $\xi = 0.5$					
H	0.5705 $\pm$ 0.0062	0.5431 $\pm$ 0.0037	0.5265 $\pm$ 0.0047	0.5126 $\pm$ 0.0023	0.5045 $\pm$ 0.0014
CH	<b>0.5662</b> $\pm$ 0.0039	<b>0.5344</b> $\pm$ 0.0042	<b>0.5265</b> $\pm$ 0.0039	<b>0.5081</b> $\pm$ 0.0024	<b>0.5044</b> $\pm$ 0.0015
BCH	0.5688 $\pm$ 0.0034	0.5458 $\pm$ 0.0040	0.5286 $\pm$ 0.0031	0.5161 $\pm$ 0.0028	0.5073 $\pm$ 0.0020
GP parent with $\xi = 1$					
H	1.0686 $\pm$ 0.0083	1.0384 $\pm$ 0.0033	1.0224 $\pm$ 0.0038	1.0151 $\pm$ 0.0031	1.0070 $\pm$ 0.0015
CH	1.0036 $\pm$ 0.0046	<b>1.0009</b> $\pm$ 0.0020	<b>1.0003</b> $\pm$ 0.0009	<b>1.0001</b> $\pm$ 0.0010	<b>0.9999</b> $\pm$ 0.0007
BCH	<b>0.9974</b> $\pm$ 0.0027	0.9981 $\pm$ 0.0013	0.9989 $\pm$ 0.0008	0.9993 $\pm$ 0.0006	0.9998 $\pm$ 0.0004
GP parent with $\xi = 2$					
H	2.0798 $\pm$ 0.0057	2.0474 $\pm$ 0.0072	2.0294 $\pm$ 0.0046	2.0197 $\pm$ 0.0053	2.0089 $\pm$ 0.0022
CH	<b>2.0626</b> $\pm$ 0.0045	<b>2.0449</b> $\pm$ 0.0029	<b>2.0268</b> $\pm$ 0.0034	<b>2.0177</b> $\pm$ 0.0030	<b>2.0084</b> $\pm$ 0.0014
BCH	1.8812 $\pm$ 0.0034	1.9084 $\pm$ 0.0037	1.9254 $\pm$ 0.0028	1.9327 $\pm$ 0.0019	1.9406 $\pm$ 0.0008
Cauchy parent with $\xi = 1$					
H	1.0794 $\pm$ 0.0050	1.0547 $\pm$ 0.0045	1.0411 $\pm$ 0.0027	1.0307 $\pm$ 0.0024	1.0212 $\pm$ 0.0014
CH	<b>1.0656</b> $\pm$ 0.0053	1.0487 $\pm$ 0.0028	1.0386 $\pm$ 0.0031	1.0282 $\pm$ 0.0023	1.0204 $\pm$ 0.0018
BCH	0.9236 $\pm$ 0.0031	<b>0.9524</b> $\pm$ 0.0025	<b>0.9627</b> $\pm$ 0.0018	<b>0.9731</b> $\pm$ 0.0012	<b>0.9823</b> $\pm$ 0.0008
Student $t_2$ parent with $\xi = 0.5$					
H	0.5858 $\pm$ 0.0035	0.5556 $\pm$ 0.0035	0.5414 $\pm$ 0.0023	0.5359 $\pm$ 0.0021	0.5267 $\pm$ 0.0018
CH	0.5176 $\pm$ 0.0017	0.5101 $\pm$ 0.0019	0.5059 $\pm$ 0.0014	0.5049 $\pm$ 0.0007	0.5020 $\pm$ 0.0006
BCH	<b>0.5060</b> $\pm$ 0.0035	<b>0.5067</b> $\pm$ 0.0016	<b>0.5052</b> $\pm$ 0.0008	<b>0.5040</b> $\pm$ 0.0006	<b>0.5019</b> $\pm$ 0.0006
Student $t_4$ parent with $\xi = 0.25$					
H	0.3374 $\pm$ 0.0040	0.3163 $\pm$ 0.0036	0.3058 $\pm$ 0.0026	0.2957 $\pm$ 0.0012	0.2861 $\pm$ 0.0008
CH	0.3230 $\pm$ 0.0037	0.3069 $\pm$ 0.0035	0.2961 $\pm$ 0.0016	<b>0.2889</b> $\pm$ 0.0014	0.2818 $\pm$ 0.0011
BCH	<b>0.3161</b> $\pm$ 0.0029	<b>0.3054</b> $\pm$ 0.0019	<b>0.2958</b> $\pm$ 0.0017	0.2894 $\pm$ 0.0016	<b>0.2814</b> $\pm$ 0.0016

smaller absolute bias and the minimum MSE is smaller than the minimum MSE of the Hill estimator;

- (ii) For the simulated Burr  $\xi, \rho$  models, the BCH estimator underestimates the true value of  $\xi$  for  $k$  values larger than 400. The MSE exhibit the usual shape being the minimum values smaller than the ones of the CH estimator. Note that if  $X$  has a Burr  $\xi, \rho$  distribution, then  $X^{-\rho/\xi} + 1$  has a strict Pareto distri-

bution and this could justify the excellent results of the  $BCH_0$  statistic;

- (iii) For the GP  $\xi$  models with  $\xi = \{0.5, 1\}$ , the mean values of the CH and BCH estimators almost overlap for a wide region of  $k$  values ( $k < 500$ , when  $\xi = 0.5$  and  $k < 600$ , when  $\xi = 1$ ), and from that  $k$  value forward, the bias of the BCH estimator is smaller than the one of the CH estimator. When  $\xi$

TABLE 5: Simulated MSE at optimal levels and associated 95% CIs based on 10 replicates with 1000 runs each.

	200	500	1000	2000	5000
Burr parent with $\xi = 0.25$ and $\rho = -0.5$					
H	0.0062 ± 0.0003	0.0028 ± 0.0001	0.0015 ± 0.0000	0.0008 ± 0.0000	0.0003 ± 0.0000
CH	0.0037 ± 0.0001	0.0018 ± 0.0001	0.0010 ± 0.0000	0.0006 ± 0.0000	0.0002 ± 0.0000
BCH	<b>0.0007</b> ± 0.0000	<b>0.0003</b> ± 0.0000	<b>0.0001</b> ± 0.0000	<b>0.0001</b> ± 0.0000	<b>0.0000</b> ± 0.0000
Burr parent with $\xi = 0.5$ and $\rho = -0.5$					
H	0.0251 ± 0.0009	0.0115 ± 0.0003	0.0063 ± 0.0002	0.0032 ± 0.0001	0.0013 ± 0.0000
CH	0.0145 ± 0.0005	0.0074 ± 0.0002	0.0043 ± 0.0001	0.0023 ± 0.0000	0.0010 ± 0.0000
BCH	<b>0.0028</b> ± 0.0001	<b>0.0012</b> ± 0.0000	<b>0.0006</b> ± 0.0000	<b>0.0003</b> ± 0.0000	<b>0.0001</b> ± 0.0000
Burr parent with $\xi = 1$ and $\rho = -0.5$					
H	0.1000 ± 0.0031	0.0462 ± 0.0014	0.0250 ± 0.0007	0.0124 ± 0.0004	0.0051 ± 0.0001
CH	0.0580 ± 0.0017	0.0289 ± 0.0007	0.0171 ± 0.0005	0.0088 ± 0.0003	0.0038 ± 0.0001
BCH	<b>0.0116</b> ± 0.0002	<b>0.0050</b> ± 0.0001	<b>0.0025</b> ± 0.0001	<b>0.0013</b> ± 0.0000	<b>0.0005</b> ± 0.0000
GP parent with $\xi = 0.5$					
H	0.0251 ± 0.0007	0.0114 ± 0.0004	0.0061 ± 0.0001	0.0032 ± 0.0001	0.0013 ± 0.0000
CH	0.014 ± 0.0006	0.0073 ± 0.0002	0.0041 ± 0.0001	<b>0.0022</b> ± 0.0001	<b>0.0009</b> ± 0.0000
BCH	<b>0.0108</b> ± 0.0004	<b>0.0064</b> ± 0.0002	<b>0.0039</b> ± 0.0001	<b>0.0022</b> ± 0.0001	<b>0.0009</b> ± 0.0000
GP parent with $\xi = 1$					
H	0.0314 ± 0.0009	0.0144 ± 0.0003	0.0077 ± 0.0001	0.0041 ± 0.0001	0.0018 ± 0.0001
CH	0.0076 ± 0.0002	0.0030 ± 0.0001	0.0015 ± 0.0001	0.0008 ± 0.0000	0.0003 ± 0.0000
BCH	<b>0.0058</b> ± 0.0002	<b>0.0024</b> ± 0.0000	<b>0.0012</b> ± 0.0000	<b>0.0006</b> ± 0.0000	<b>0.0002</b> ± 0.0000
GP parent with $\xi = 2$					
H	0.0551 ± 0.0016	0.0243 ± 0.0005	0.0128 ± 0.0003	0.0067 ± 0.0003	0.0027 ± 0.0001
CH	<b>0.0373</b> ± 0.0009	<b>0.0165</b> ± 0.0002	<b>0.0089</b> ± 0.0002	<b>0.0047</b> ± 0.0002	<b>0.0020</b> ± 0.0000
BCH	0.0469 ± 0.0006	0.0237 ± 0.0004	0.0144 ± 0.0005	0.0095 ± 0.0003	0.0059 ± 0.0001
Cauchy parent with $\xi = 1$					
H	0.0365 ± 0.0011	0.0168 ± 0.0005	0.0094 ± 0.0003	0.0053 ± 0.0001	0.0025 ± 0.0001
CH	0.0283 ± 0.0008	0.0130 ± 0.0004	0.0073 ± 0.0002	0.0041 ± 0.0001	0.0020 ± 0.0001
BCH	<b>0.0227</b> ± 0.0005	<b>0.0099</b> ± 0.0003	<b>0.0057</b> ± 0.0002	<b>0.0031</b> ± 0.0001	<b>0.0015</b> ± 0.0000
Student $t_2$ parent with $\xi = 0.5$					
H	0.0236 ± 0.0005	0.0117 ± 0.0003	0.0069 ± 0.0002	0.0043 ± 0.0001	0.0022 ± 0.0000
CH	0.0064 ± 0.0002	0.0025 ± 0.0001	0.0013 ± 0.0000	0.0007 ± 0.0000	0.0003 ± 0.0000
BCH	<b>0.0041</b> ± 0.0001	<b>0.0017</b> ± 0.0000	<b>0.0009</b> ± 0.0000	<b>0.0005</b> ± 0.0000	<b>0.0002</b> ± 0.0000
Student $t_4$ parent with $\xi = 0.25$					
H	0.0210 ± 0.0006	0.0114 ± 0.0003	0.0071 ± 0.0002	0.0048 ± 0.0001	0.0028 ± 0.0001
CH	0.0124 ± 0.0004	0.0070 ± 0.0001	0.0046 ± 0.0001	0.0032 ± 0.0001	<b>0.0020</b> ± 0.0001
BCH	<b>0.0080</b> ± 0.0002	<b>0.0055</b> ± 0.0001	<b>0.0041</b> ± 0.0001	<b>0.0031</b> ± 0.0001	<b>0.0020</b> ± 0.0001

= 2, the BCH estimator underestimates  $\xi$  for almost all  $k$  values, whereas the CH estimator overestimates  $\xi$  for the same  $k$  values. Regarding the minimum MSE, both estimators (CH and BCH) present similar values;

(iv) For the Cauchy model, the behaviour of CH and the BCH estimator is similar to the behaviour observed for the GP<sub>2</sub> parent;

(v) For the Student  $t_\nu$  models ( $\nu = 2, 4$ ), the sample paths of the mean values of the CH and BCH estimators overlap from small up to moderate values of  $k$ , and from that value forward, the bias of the BCH estimator is smaller. Regarding the minimum MSE, both estimators (CH and BCH) present similar values.

In Table 2, we present the estimates of the BC parameters  $\sigma$  and  $a$ , as described in Section 3, for the aforementioned

models. Information on the 95% confidence intervals (CIs), computed on the basis of 10 replicates with 1000 runs each, is also provided. The simulated data used to support the findings of this study are available from the corresponding author upon request.

Next, and based on the 1000 runs, we computed the simulated optimum level, i.e., the level  $\hat{k}_0 := \arg \min_k \text{MSE}(\hat{\xi}(k))$ , with  $\hat{\xi} = \text{H, CH, BCH}$ , and present in Tables 3–5 the simulated values of the optimal sample fraction (the optimal level divided by the number of positive elements in the sample), the mean value of each estimator computed at  $\hat{k}_0$ , and the MSE of each estimator also computed at  $\hat{k}_0$ . For each model, the smallest absolute bias and the smallest MSE are written in bold. Information on the 95% CIs, computed on the basis of 10 replicates with 1000 runs each, is also provided.

Table 2 shows that the estimation of  $a$  works well for models with first-order scale parameter,  $C = 1$ , such as the Burr and  $\text{GP}_1$ , as expected and that the estimation of  $\sigma$  could be improved in some parents. However, from Tables 4 and 5, we noticed that the precise estimation of  $\sigma$  has more impact on the behaviour of the BCH estimator in some models, like the  $\text{GP}_{0.25}$  and  $\text{GP}_{0.5}$  than on other models.

At optimal levels, the BCH estimator is always better than the CH estimator for all simulated Burr parents, and in this case, the misspecification  $C = 1$  has no impact on the BCH estimator since  $C = 1$  (see Table 1). The simulated mean values, at optimal levels, of the CH estimator are better than the ones of the BCH estimator for all GP parents, but the BCH estimator has competitive MSEs for the  $\text{GP}_{0.5}$  and  $\text{GP}_1$  parents. For the Cauchy and Student  $t_2$  and  $t_4$  parents, the BCH estimator is better than the CH estimator for almost all sample sizes, even with the misspecification  $C = 1$ .

## 5. Concluding Remarks and Future Work

In this paper, we have studied the BC Hill estimator introduced by Teugels and Vanroelen, in 2004. We proved the consistency and asymptotic normality of the estimators and with the appropriate choice of the BC parameters we were able to obtain a new class of MVRB EVI estimators. We considered the adaptive estimation of the BC parameters, and the finite sample behaviour of the estimators was illustrated by a Monte Carlo simulation study. From the simulation study, we conclude that the BC Hill estimator performs better than the corrected Hill estimator for several heavy-tailed models in terms of simulated mean values and simulated MSEs. Furthermore, the estimation procedure presents some robustness to the estimation of the BC parameters, as can be depicted by the misspecification of the first-order scale parameter,  $C = 1$ . Note that several heavy-tailed models satisfy this criterion. The estimation of the BC parameters could be improved by deriving the maximum likelihood estimators, a topic under development and out of the scope of this paper. Other topics of interest for future research, out of the scope of this paper, are the determination of the optimal level  $k$  used to compute the BC EVI esti-

mates and the use of the BC Hill estimator to derive new classes of estimators for other parameters of extreme events such as high quantiles and the second-order shape parameter,  $\rho$ .

## Data Availability

The simulated data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no potential conflict of interests.

## Acknowledgments

This research was partially supported by the national funds through FCT, Fundação para a Ciência e a Tecnologia, projects UIDB/04674/2020 (CIMA) and UIDB/0006/2020 (CEA/UL).

## References

- [1] G. E. P. Box and D. R. Cox, “An analysis of transformations,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, no. 2, pp. 211–243, 1964.
- [2] T. G. Bali, “The generalized extreme value distribution,” *Economics Letters*, vol. 79, no. 3, pp. 423–427, 2003.
- [3] J. L. Teugels and G. Vanroelen, “Box-Cox transformations and heavy-tailed distributions,” *Journal of Applied Probability*, vol. 41, no. A, pp. 213–227, 2004.
- [4] E. Seneta, *Regularly Varying Functions*, Springer-Verlag, Berlin, 1976.
- [5] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*, Cambridge University Press, 1989.
- [6] E. F. Eastoe and J. A. Tawn, “Modelling non-stationary extremes with application to surface level ozone,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 58, no. 1, pp. 25–45, 2009.
- [7] J. L. Wadsworth, J. A. Tawn, and P. Jonathan, “Accounting for choice of measurement scale in extreme value modeling,” *The Annals of Applied Statistics*, vol. 4, no. 3, pp. 1558–1578, 2010.
- [8] V. Paulauskas and M. Vaiciulis, “On an improvement of Hill and some other estimators,” *Lithuanian Mathematical Journal*, vol. 53, no. 3, pp. 336–355, 2013.
- [9] V. Paulauskas and M. Vaiciulis, “A class of new tail index estimators,” *Annals of the Institute of Statistical Mathematics*, vol. 69, pp. 461–487, 2017.
- [10] L. de Haan, *Slow Variation and Characterization of Domains of Attraction*, Springer Netherlands, Dordrecht, 1984.
- [11] J. Geluk and L. de Haan, *Regular Variation, Extensions and Tauberian Theorems*, Centre for Mathematics and Computer Science, CWI Tract 40, Amsterdam The Netherlands, 1987.
- [12] P. Hall and A. Welsh, “Adaptive estimates of parameters of regular variation,” *The Annals of Statistics*, vol. 13, no. 1, pp. 331–341, 1985.
- [13] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*, Springer, New York, 2006.
- [14] B. M. Hill, “A simple general approach to inference about the tail of a distribution,” *The Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

- [15] L. Peng, "Asymptotically unbiased estimators for the extreme-value index," *Statistics & Probability Letters*, vol. 38, no. 2, pp. 107–115, 1998.
- [16] J. Beirlant, G. Dierckx, Y. Goegebeur, and G. Matthys, "Tail index estimation and an exponential regression model," *Extremes*, vol. 2, no. 2, pp. 177–200, 1999.
- [17] A. Feuerverger and P. Hall, "Estimating a tail exponent by modelling departure from a Pareto distribution," *The Annals of Statistics*, vol. 27, no. 2, pp. 760–781, 1999.
- [18] M. I. Gomes, M. J. Martins, and M. Neves, "Alternatives to a semi-parametric estimator of parameters of rare events—the jackknife methodology," *Extremes*, vol. 3, no. 3, pp. 207–229, 2000.
- [19] M. I. Gomes and A. Guillo, "Extreme value theory and statistics of univariate extremes: a review," *International Statistical Review*, vol. 83, no. 2, pp. 263–292, 2015.
- [20] F. Caeiro, M. I. Gomes, and D. Pestana, "Direct reduction of bias of the classical Hill estimator," *Revstat - Statistical Journal*, vol. 3, no. 2, pp. 113–136, 2005.
- [21] F. Caeiro, M. I. Gomes, and L. Henriques-Rodrigues, "Reduced-bias tail index estimators under a third order framework," *Communications in Statistics—Theory and Methods*, vol. 38, no. 7, pp. 1019–1040, 2009.
- [22] M. I. Gomes, M. J. Martins, and M. Neves, "Improving second order reduced bias extreme value index estimation," *Revstat - Statistical Journal*, vol. 5, no. 2, pp. 177–207, 2007.
- [23] M. I. Gomes, L. de Haan, and L. Henriques-Rodrigues, "Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 31–52, 2008.
- [24] M. I. Gomes and D. Pestana, "A sturdy reduced-bias extreme quantile (VaR) estimator," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 280–292, 2007.
- [25] M. F. Brillhante, M. I. Gomes, and D. Pestana, "A simple generalisation of the Hill estimator," *Computational Statistics & Data Analysis*, vol. 57, no. 1, pp. 518–535, 2013.
- [26] F. Caeiro, M. I. Gomes, J. Beirlant, and T. de Wet, "Mean-of-order  $p$  reduced-bias extreme value index estimation under a third-order framework," *Extremes*, vol. 19, no. 4, pp. 561–589, 2016.
- [27] F. Caeiro, L. Henriques-Rodrigues, and D. Prata Gomes, "A simple class of reduced bias kernel estimators of extreme value parameters," *Computational and Mathematical Methods*, vol. 1, no. 3, article e1025, 2019.
- [28] F. Caeiro, L. Henriques-Rodrigues, M. I. Gomes, and I. Cabral, "Minimum-variance reduced-bias estimation of the extreme value index: a theoretical and empirical study," *Computational and Mathematical Methods*, vol. 2, no. 4, article e1101, 2020.
- [29] F. Caeiro, A. Mateus, and L. Soltane, "A class of weighted Hill estimators," *Computational and Mathematical Methods*, vol. 3, no. 6, article e1167, 2021.
- [30] J. Beran, D. Schell, and M. Stehlík, "The harmonic moment tail index estimator: asymptotic distribution and robustness," *Annals of the Institute of Statistical Mathematics*, vol. 66, no. 1, pp. 193–220, 2013.
- [31] J. Beirlant, G. Maribe, and A. Verster, "Using shrinkage estimators to reduce bias and MSE in estimation of heavy tails," *Revstat - Statistical Journal*, vol. 17, no. 1, pp. 91–108, 2019.
- [32] L. de Haan and L. Peng, "Comparison of tail index estimators," *Statistica Neerlandica*, vol. 52, no. 1, pp. 60–70, 1998.
- [33] M. I. Gomes and O. Oliveira, "How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events," *Communications in Statistics - Simulation and Computation*, vol. 32, no. 4, pp. 1005–1028, 2003.
- [34] M. I. Fraga Alves, M. I. Gomes, and L. de Haan, "A new class of semi-parametric estimators of the second order parameter," *Portugaliae Mathematica*, vol. 60, no. 2, pp. 193–213, 2003.
- [35] F. Caeiro and M. I. Gomes, "A new class of estimators of a "scale" second order parameter," *Extremes*, vol. 9, pp. 193–211, 2006.
- [36] M. I. Gomes and M. J. Martins, "Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter," *Extremes*, vol. 5, no. 1, pp. 5–31, 2002.
- [37] F. Caeiro and M. I. Gomes, "Semi-parametric second-order reduced-bias high quantile estimation," *TEST*, vol. 18, no. 2, pp. 392–413, 2009.
- [38] Y. Goegebeur, J. Beirlant, and T. de Wet, "Kernel estimators for the second order parameter in extreme value statistics," *Journal of Statistical Planning and Inference*, vol. 140, no. 9, pp. 2632–2652, 2010.
- [39] G. Ciuperca and C. Mercadier, "Semi-parametric estimation for heavy tailed distributions," *Extremes*, vol. 13, pp. 55–87, 2010.
- [40] J. Worms and R. Worms, "Estimation of second order parameters using probability weighted moments," *ESAIM: Probability and Statistics*, vol. 16, pp. 97–113, 2012.
- [41] T. de Wet, Y. Goegebeur, and M. Munch, "Asymptotically unbiased estimation of a second order tail parameter," *Statistics & Probability Letters*, vol. 82, no. 3, pp. 565–573, 2012.
- [42] L. Henriques-Rodrigues, M. I. Gomes, M. I. Fraga Alves, and C. Neves, "PORT-estimation of a shape second-order parameter," *Revstat - Statistical Journal*, vol. 12, no. 3, pp. 299–328, 2014.
- [43] M. I. Gomes, L. Henriques-Rodrigues, H. Pereira, and D. Pestana, "Tail index and second-order parameters' semi-parametric estimation based on the log-excesses," *Journal of Statistical Computation and Simulation*, vol. 80, no. 6, pp. 653–666, 2010.
- [44] P. Hall, "On some simple estimates of an exponent of regular variation," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 44, no. 1, pp. 37–42, 1982.
- [45] I. B. Aban and M. M. Meerschaert, "Shifted Hill's estimator for heavy tails," *Communications in Statistics - Simulation and Computation*, vol. 30, no. 4, pp. 949–962, 2001.
- [46] M. I. Gomes, F. Figueiredo, and L. Henriques-Rodrigues, "The link between the excesses over a high threshold and the shifted Hill estimator," *Bulletin of the International Statistical Institute (ISI)*, vol. LXII, pp. 3590–3593, 2007.
- [47] M. I. Gomes and O. Oliveira, "The bootstrap methodology in statistics of extremes—choice of the optimal sample fraction," *Extremes*, vol. 4, no. 4, pp. 331–358, 2001.