

# TAIL PREPIVOTING FOR THE HILL ESTIMATOR

MARGARIDA BRITO, ANA CRISTINA MOREIRA FREITAS,  
AND JORGE MILHAZES FREITAS

ABSTRACT. It is well known that prepivoting reduces level error of confidence sets. We adapt this method to the context of the tail index estimation, introducing a procedure that we call tail prepivoting. We apply this procedure to the Hill estimator and establish its consistency.

## 1. INTRODUCTION

Let  $X_1, X_2, \dots$  be independent nonnegative random variables (r.v.) with continuous common cumulative distribution function (c.d.f.)  $F$ . Assume that  $1 - F$  is regularly varying in the upper tail, namely there exists  $0 < c < \infty$ , such that, for any  $t > 0$ ,

$$\limsup_{x \rightarrow +\infty} \frac{1 - F(tx)}{1 - F(x)} = t^{-c},$$

or equivalently

$$1 - F(x) = x^{-c}L(x), \text{ for } x > 0, \quad (1.1)$$

where  $L$  is a slowly varying function at infinity. Denoting by  $F^{-1}$  the left continuous inverse of  $F$ , i.e.,  $F^{-1}(s) := \inf\{x: F(x) \geq s\}$ , (1.1) is equivalent to

$$F^{-1}(1 - s) = s^{-1/c}\tilde{L}(s), \quad 0 < s < 1, \quad (1.2)$$

where  $\tilde{L}$  is a slowly varying function at zero (see e.g. de Haan (1970)).

The problem of estimating  $c$  or related tail indices has received considerable attention and common applications may be found in a large variety of domains, as for example in economics, applied finance, insurance, business, industry, telecommunications, traffic, geology, meteorology and hydrology. An important application to the field of risk theory was studied in Csörgő and Steinebach (1991), where the authors related the problem of estimating the adjustment coefficient with the estimation of a certain exponential tail index. We also mention an application to rainfalls developed by Carreau et al. (2009). Another

---

*Date:* February 22, 2016.

*2000 Mathematics Subject Classification.* 62G05, 62F12, 62G30.

*Key words and phrases.* Bootstrap; Parameter estimation; Prepivoting; Tail index.

important application in the domain of seismology was developed in Brito et al. (2015), where the authors estimated the tail parameters of the seismic moment distribution. In the past few years, Extreme Value Theory has been also applied to stochastic processes arising from chaotic dynamical systems. Initially, this flourished by giving a different approach and insight into the statistical properties of such erratic systems and specially about its extremal behaviour. It turned out that these statistical features are tied with the recurrence properties of the system (see Freitas et al. (2010)). Moreover, for systems carrying a strange attractor, Extreme Value Analysis proved to be useful to recover information about the geometry of the attractor itself. In this particular application of EVT, the estimation of tail index is rather important since it was shown in Lucarini et al. (2012) that it depends directly on the dimension of the attractor. This means that good estimates for the tail index are also important in this setting.

One of the most commonly used estimators for  $1/c$  is the Hill estimator (1975), given by

$$\hat{H}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \log X_{n-i+1,n} - \log X_{n-k_n,n},$$

where  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  denote the order statistics (o.s.) of the sample  $X_1, X_2, \dots, X_n$  and  $k_n$  is a sequence of positive integers satisfying

$$1 \leq k_n < n, \quad \lim_{n \rightarrow \infty} k_n = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} k_n/n = 0. \quad (1.3)$$

The asymptotic properties of  $\hat{H}(k_n)$  have been much studied. It is well known that, under some regularity conditions,  $\hat{H}(k_n)$  is strongly consistent (cf Deheuvels et al. (1988)) and asymptotically normal with convergence rate  $k_n^{-1/2}$  when properly normalized (cf Haeusler and Teugels (1985)).

Consider then the normalized estimator

$$T_n := ck_n^{1/2} \left( \hat{H}(k_n) - \frac{1}{c} \right). \quad (1.4)$$

Let  $\ell_n$  be a nondecreasing sequence of positive integers and denote by  $W_1, W_2, \dots, W_{\ell_n}$  the  $\ell_n$  exceedances of the random level  $\log X_{n-\ell_n,n}$ , that is

$$W_i := \log X_{n-\ell_n+i,n} - \log X_{n-\ell_n,n}, \quad 1 \leq i \leq \ell_n. \quad (1.5)$$

The tail bootstrap is based on resampling, with replacement, the sample  $W_1, W_2, \dots, W_{\ell_n}$  instead of the initial sample  $X_1, X_2, \dots, X_n$ . If we

take  $\ell_n = k_n$ ,  $n \in \mathbb{N}$ , then we may write

$$\widehat{H}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} W_i.$$

By the above representation for  $\widehat{H}(k_n)$ , we can derive the corresponding bootstrap version in an obvious way. Consider the sample  $\mathbf{W} = (W_1, \dots, W_{k_n})$  and let  $\mathbf{W}^* = (W_1^*, \dots, W_{k_n}^*)$  be a sample drawn with replacement from  $W_1, \dots, W_{k_n}$ . The tail bootstrap version of the estimator is given by

$$\widehat{H}^*(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} W_i^*. \quad (1.6)$$

The standardised sequence  $\frac{1}{\widehat{H}(k_n)} k_n^{1/2} (\widehat{H}^*(k_n) - \widehat{H}(k_n))$  converges weakly to the same limit as that of  $ck_n^{1/2} (\widehat{H}(k_n) - \frac{1}{c})$ , as stated in Bacro and Brito (1998).

Motivated by least squares considerations, Brito and Freitas (2006, 2008) introduced some consistent estimators for  $c$ . We consider here:

$$\widetilde{R}(k_n) = \left( \frac{1}{k_n} \sum_{i=1}^{k_n} W_i^2 - \left( \frac{1}{k_n} \sum_{i=1}^{k_n} W_i \right)^2 \right)^{-1/2}.$$

Observe that  $\widetilde{R}^{-2}(k_n)$  is the sample variance of  $\mathbf{W}$  which makes its use appropriate to standardise the bootstrap version of the root- $n$   $T_n$ :

$$T_n^* := \widetilde{R}(k_n) k_n^{1/2} (\widehat{H}^*(k_n) - \widehat{H}(k_n)). \quad (1.7)$$

Using the result above of Bacro and Brito (1998) we immediately obtain that the tail bootstrap root- $n$   $T_n^*$  defined in (1.7) converges weakly to the same limit as that of  $T_n$ .

Beran (1987) introduced the notion of prepivoting, which corresponds to the transformation of a root by its estimated bootstrap cumulative c.d.f. He has also shown that this procedure produces smaller error than the original bootstrap introduced by Efron (1979). The idea is that by prepivoting the root we get a “smoother” new root, in the sense that the dependence on the original distribution of the population is reduced. Then one naturally expects that the bootstrap approximation of the sampling distribution of this new prepivoted root is sharper than the corresponding one for the original root.

In the context of the tail estimation is rather important to improve the accuracy of the approximations for the estimators’ distribution. In order to obtain such better approximations, we adapt the original prepivoting method of Beran to this context of tail estimation.

To illustrate the validity of this adjusted version of prepivotting, that we call tail prepivotting, we show its consistency when applied to the common Hill's estimator. Although there are much more recent and better performing estimators, our purpose here is just to show that the method works and produces indeed better approximations for the Hill's estimator, which is still the base of comparison for almost every new estimator or new method introduced. We note that our approach may be applied to estimators that may be represented in terms of ratios of uniform o.s. as considered in (2.1), like the geometric-type estimator introduced in Brito and Freitas (2003). Moreover, we believe that the method should work and provide better approximations for more general estimators depending only on the higher o.s. of the sample, even if they cannot be written in terms of such ratios, in which case, the argument would need to be adapted. One of the advantages of this tail prepivotting procedure that we introduce here is that it reduces error, as it can be seen in the simulation results for the Hill estimator presented below, in spite of the small number of observations considered, which are actually used for resampling.

We adapt the prepivotting procedure to this context of tail index estimation, by considering the new tail prepivotted root- $n$

$$\mathcal{T}_n := P[T_n^* \leq T_n | (X_{n-k_n, n}, \mathbf{W})], \quad (1.8)$$

whose asymptotic c.d.f. is the uniform distribution,  $U(0, 1)$ , and its tail bootstrap version:

$$\mathcal{T}_n^* := P[T_n^{**} \leq T_n^* | (X_{n-k_n, n}, \mathbf{W}^*)],$$

where

$$T_n^{**} = \tilde{R}^*(k_n) k_n^{1/2} (\hat{H}^{**}(k_n) - \hat{H}^*(k_n)),$$

$$\hat{H}^{**}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} W_i^{**},$$

$$\tilde{R}^*(k_n) = \left( \frac{1}{k_n} \sum_{i=1}^{k_n} (W_i^*)^2 - \left( \frac{1}{k_n} \sum_{i=1}^{k_n} W_i^* \right)^2 \right)^{-1/2}$$

and  $\mathbf{W}^{**} = (W_1^{**}, \dots, W_{k_n}^{**})$  is a sample drawn with replacement from  $W_1^*, \dots, W_{k_n}^*$ .

We will establish the consistency of the tail prepivotting procedure. Namely, we show that the root- $n$   $\mathcal{T}_n^*$  converges weakly to the same limit as that of  $\mathcal{T}_n$ .

The paper is organized as follows. In Section 2 we state the main results. In Section 3 we present a small simulation study in order to analyse and compare the asymptotic, empirical and tail bootstrap c.d.f.s for  $T_n$  and  $\mathcal{T}_n$ . In Section 4 we prove the results of the paper.

## 2. STATEMENT OF RESULTS

In the sequel,  $\xrightarrow{D}$  and  $\stackrel{D}{=}$  stand, respectively, for convergence and equality in distribution. Moreover, we denote the standard normal c.d.f. by  $\Phi$ , its density by  $\phi$  and the uniform  $U(0, 1)$  c.d.f. by  $\mathcal{U}$ .

Before presenting the results it is convenient to introduce some notation. We assume that  $U_1, U_2, \dots$  is a sequence of independent uniform  $U(0, 1)$  r.v.. We denote the o.s. of the sample  $(U_1, U_2, \dots, U_n)$  by  $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$ . Since  $X_i \stackrel{D}{=} F^{-1}(U_i)$ ,  $i \geq 1$ , the  $k_n$  exceedances of the random level  $\log X_{n-k_n,n}$ , defined in (1.5), may be written in the following form, without loss of generality,

$$W_i = \log F^{-1}(U_{n-k_n+i,n}) - \log F^{-1}(U_{n-k_n,n}), \quad 1 \leq i \leq k_n.$$

As in Bacro and Brito (1998) (cf Theorem 1) we shall make use of the following equivalent representation for  $W_i$ ,  $i = 1, \dots, k_n$ :

$$W_i = -\frac{1}{c} \log Y_i + \log \frac{\tilde{L}(Y_i(1 - U_{n-k_n,n}))}{\tilde{L}(1 - U_{n-k_n,n})}, \quad (2.1)$$

where

$$Y_i = \frac{1 - U_{n-k_n+i,n}}{1 - U_{n-k_n,n}}, \quad \text{for } i = 1, \dots, k_n. \quad (2.2)$$

We recall also that  $(Y_i)_{1 \leq i \leq k_n}$  is distributed as the vector of the o.s. of an i.i.d.  $k_n$ -sample from an uniform  $U(0, 1)$  distribution. Now, we define

$$J(k_n) := k_n^{1/2} \sup_{Y_{k_n} \leq y \leq 1} \left| \log \frac{\tilde{L}(y(1 - U_{n-k_n,n}))}{\tilde{L}\left(\frac{k_n}{n}\right)} \right|. \quad (2.3)$$

In the following result we show the validity of the tail prepivoting, that is, we show that  $\mathcal{T}_n^*$  converges weakly to the uniform c.d.f., in the sense of (2.5). Recall that the weak limit of the root- $n$   $\mathcal{T}_n$  defined in (1.8) is the uniform c.d.f. because  $T_n^*$  converges to the same weak limit of  $T_n$ .

**Theorem 1.** *Assume that  $F$  satisfies (1.1) and  $k_n$  is a sequence of positive integers such that (1.3) holds. If we suppose that, as  $n \rightarrow \infty$ ,*

$$k_n^{1/2} \sup_{1/k_n \leq y \leq 1} \left| \log \left( \frac{\tilde{L}(yt^{k_n/n})}{\tilde{L}\left(\frac{k_n}{n}\right)} \right) \right| \rightarrow 0 \quad (2.4)$$

*uniformly in  $t$  on compact sets of  $(0, \infty)$ , then, for all  $x \in (0, 1)$ ,*

$$P[\mathcal{T}_n^* \leq x | (X_{n-k_n,n}, \dots, X_{n,n})] \xrightarrow{P} x \quad \text{as } n \rightarrow \infty. \quad (2.5)$$

In order to provide an example of application of the result above, we consider the important model introduced by Hall (1982):

$$1 - F(x) = Dx^{-c}\{1 + O(x^{-\beta})\}, \quad x > 0, \quad (2.6)$$

where  $D$  and  $\beta$  are positive constants. This model can also be written in the equivalent form

$$F^{-1}(1-s) = s^{-1/c} \tilde{L}(s), \quad 0 < s < 1, \quad (2.7)$$

where

$$\tilde{L}(s) = D^{1/c} \{1 + O(s^{\beta/c})\}.$$

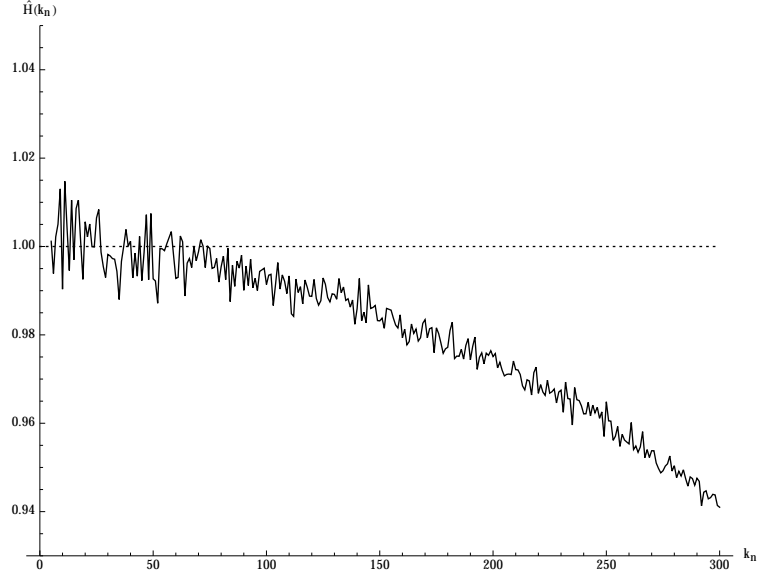
**Corollary 1.** *Assume that (2.6) holds and  $k_n$  is a sequence of positive integers satisfying (1.3). For sequences  $k_n \rightarrow \infty$  such that  $k_n = o(n^{2\beta/(2\beta+c)})$ , we have that, for all  $x \in (0, 1)$ ,*

$$P[\mathcal{T}_n^* \leq x | (X_{n-k_n, n}, \dots, X_{n, n})] \xrightarrow{P} x \quad \text{as } n \rightarrow \infty.$$

### 3. SIMULATION STUDY

We perform here a small simulation study where we considered the Hall's family (2.7) with  $c = 1$  and  $\tilde{L}(s) = 1 + s^2$ .

For this distribution, we generated 1000 samples of size  $n = 1000$  which gave rise to an empirical approximation of the c.d.f. of the r.v.  $T_{1000}$ . We considered the value  $k_n = 70$  for the number of high o.s. taken into account. This choice corresponds to a value in the range where the Hill's estimator presents small bias, as it can be seen in Figure 1, and which are not too small to produce satisfactorily accurate approximations for the c.d.f.s of  $T_{1000}$  and of the prepivoted root  $\mathcal{T}_{1000}$  (we may note that, in real cases, where the true value of  $1/c$  is unknown, we may choose a reasonable number of o.s.  $k_n$  in the range where the estimator presents a stabilised behaviour). For each sample we resampled the  $k_n$  largest o.s. 1000 times in order to obtain both an approximation for the tail bootstrap c.d.f. of  $T_n$  and an empirical approximation of the c.d.f. of the r.v.  $\mathcal{T}_n$ . Finally, each tail bootstrap sample, obtained by resampling the  $k_n$  largest o.s. of the original sample, was resampled 1000 times to produce an approximation of the tail bootstrap c.d.f. of  $\mathcal{T}_n$ .

FIGURE 1.  $\hat{H}(k_n)$  for different values of  $k_n$ 

| $x$     | $ \hat{G}_{T_n} - \Phi $ | $ \hat{G}_{T_n} - \hat{G}_{T_n}^* $ | $x$    | $ \hat{G}_{\mathcal{T}_n} - \mathcal{U} $ | $ \hat{G}_{\mathcal{T}_n} - \hat{G}_{\mathcal{T}_n}^* $ |
|---------|--------------------------|-------------------------------------|--------|---|---|
| -2.3263 | 0.0080                   | 0.0039                              | 0.0100 | 0.0030                                    | 0.0007  |
| -1.9600 | 0.0070                   | 0.0010                              | 0.0250 | 0.0030                                    | 0.0006  |
| -1.6449 | 0.0010                   | 0.0066                              | 0.0500 | 0.0040                                    | 0.0072  |
| -1.4395 | 0.0090                   | 0.0025                              | 0.0750 | 0.0030                                    | 0.0009  |
| -1.2816 | 0.0010                   | 0.0051                              | 0.1000 | 0.0010                                    | 0.0030  |
| -0.8416 | 0.0210                   | 0.0184                              | 0.2000 | 0.0130                                    | 0.0102  |
| -0.5244 | 0.0170                   | 0.0078                              | 0.3000 | 0.0110                                    | 0.0036  |
| -0.2533 | 0.0260                   | 0.0134                              | 0.4000 | 0.0220                                    | 0.0118  |
| 0.0000  | 0.0180                   | 0.0044                              | 0.5000 | 0.0110                                    | 0.0001  |
| 0.2533  | 0.0220                   | 0.0101                              | 0.6000 | 0.0050                                    | 0.0048  |
| 0.5244  | 0.0050                   | 0.0037                              | 0.7000 | 0.0070                                    | 0.0040  |
| 0.8416  | 0.0100                   | 0.0132                              | 0.8000 | 0.0090                                    | 0.0118  |
| 1.2816  | 0.0080                   | 0.0048                              | 0.9000 | 0.0040                                    | 0.0017  |
| 1.4395  | 0.0070                   | 0.0025                              | 0.9250 | 0.0030                                    | 0.0002  |
| 1.6449  | 0.0100                   | 0.0046                              | 0.9500 | 0.0020                                    | 0.0019  |
| 1.9600  | 0.0090                   | 0.0038                              | 0.9750 | 0.0020                                    | 0.0014  |
| 2.3263  | 0.0070                   | 0.0031                              | 0.9900 | 0.0020                                    | 0.0009  |

TABLE 1. Differences between the asymptotic, empirical and tail bootstrap c.d.f.s for  $T_n$  and  $\mathcal{T}_n$  with  $k_n = 70$ .

In Table 1 we present the differences between the asymptotic c.d.f. and the empirical c.d.f. as well as between the later and the tail bootstrap c.d.f. of both  $T_n$  and  $\mathcal{T}_n$ , for the case  $k_n = 70$ . We used the notation

$\hat{G}_{T_n}$ ,  $\hat{G}_{\mathcal{T}_n}$  for the empirical c.d.f. of  $T_n$  and  $\mathcal{T}_n$ , respectively, and  $\hat{G}_{T_n}^*$ ,  $\hat{G}_{\mathcal{T}_n}^*$  for the respective tail bootstrap c.d.f. approximations.

In Figures 2 and 3, the dashed line has been used for the asymptotic approximations of the c.d.f. of  $T_n$  (standard normal c.d.f.) and the prepivoted  $\mathcal{T}_n$  (uniform c.d.f.); the solid line was kept for the approximations given by the respective empirical c.d.f.s; the dotted line was left for the tail bootstrap approximations.

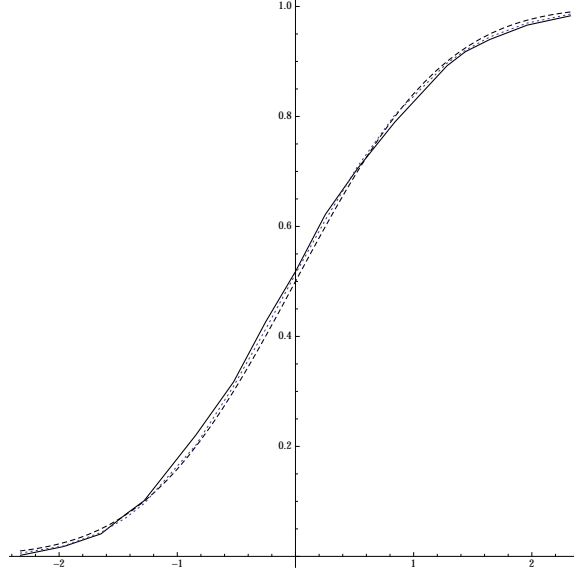


FIGURE 2. Approximations for the c.d.f. of  $T_n$ ,  $k_n = 70$ .

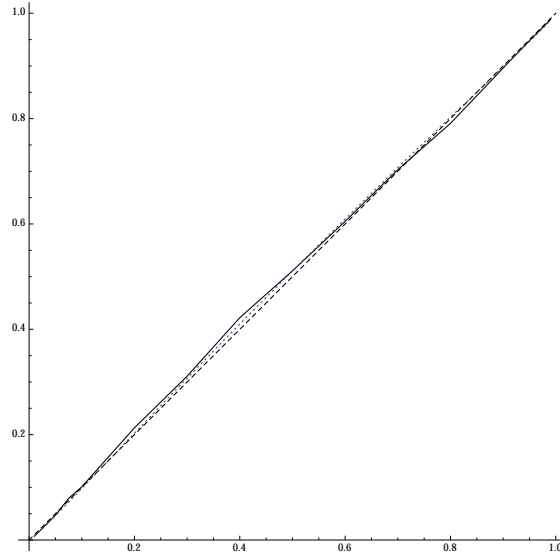


FIGURE 3. Approximations for the c.d.f. of  $\mathcal{T}_n$ ,  $k_n = 70$ .



As it can be seen, there is a good adjustment between the asymptotic normal c.d.f. and the empirical c.d.f. of  $T_n$  as well as between the later and tail bootstrap c.d.f. approximation. Moreover, the tail bootstrap approximation is in general better than the asymptotic one, which is particularly clear in the tails of the distributions.

We may observe that there is a very good adjustment between the tail bootstrap c.d.f. of  $\mathcal{T}_n$  and the uniform c.d.f., and the tail bootstrap c.d.f. seems to provide a better approximation of the empirical c.d.f. of  $\mathcal{T}_n$ , when compared to the asymptotic one.

#### 4. PROOFS

We will start by introducing some general notation. Let  $\mathcal{Z}_1, \mathcal{Z}_2, \dots$  be an arbitrary sequence of r.v.. We denote by  $\overline{\mathcal{Z}}$  the sample mean of  $\mathcal{Z} = (\mathcal{Z}_1, \dots, \mathcal{Z}_n)$ , that is,

$$\overline{\mathcal{Z}} = \frac{1}{n} \sum_{i=1}^n \mathcal{Z}_i.$$

We define, for  $i = 1, \dots, k_n$ ,

$$\mathcal{E}_i := -\frac{1}{c} \log Y_i, \quad \mathcal{Z}_i := \log \frac{\tilde{L}(Y_i(1 - U_{n-k_n, n}))}{\tilde{L}(1 - U_{n-k_n, n})},$$

where  $Y_i$  is defined as in (2.2) and write  $\mathcal{E} = (\mathcal{E}_1, \dots, \mathcal{E}_{k_n})$  and  $\mathcal{Z} = (\mathcal{Z}_1, \dots, \mathcal{Z}_{k_n})$ . Notice that  $c\mathcal{E}$  has the same distribution as the vector of o.s. of  $k_n$  independent standard exponential r.v..

Now, let  $\mathbf{Y}^* := (Y_1^*, \dots, Y_{k_n}^*)$  be a sample of r.v. drawn with replacement from  $(Y_1, \dots, Y_{k_n})$ . The bootstrap versions of  $\mathcal{E}$  and  $\mathcal{Z}$  are then given respectively by

$$\mathcal{E}_i^* := -\frac{1}{c} \log Y_i^*, \quad \mathcal{Z}_i^* := \log \frac{\tilde{L}(Y_i^*(1 - U_{n-k_n, n}))}{\tilde{L}(1 - U_{n-k_n, n})}, \quad i = 1, \dots, k_n.$$

We define similarly the corresponding iterated bootstrap versions  $\mathcal{E}^{**}$  and  $\mathcal{Z}^{**}$ , based on a sample  $(Y_1^{**}, \dots, Y_{k_n}^{**})$  drawn with replacement from the sample  $\mathbf{Y}^*$ .

Let  $\mathcal{Z}_1, \mathcal{Z}_2, \dots$  and  $\mathcal{Z}'_1, \mathcal{Z}'_2, \dots$  be two arbitrary sequences of r.v.. We denote by  $S_n^2(\mathcal{Z})$  the sample variance of  $\mathcal{Z} = (\mathcal{Z}_1, \dots, \mathcal{Z}_n)$  and by  $S_n(\mathcal{Z}, \mathcal{Z}')$  the sample covariance between  $\mathcal{Z}$  and  $\mathcal{Z}' = (\mathcal{Z}'_1, \dots, \mathcal{Z}'_n)$ , that is,

$$S_n^2(\mathcal{Z}) = \frac{1}{n} \sum_{i=1}^n (\mathcal{Z}_i - \overline{\mathcal{Z}})^2$$

and

$$S_n(\mathbf{Z}, \mathbf{Z}') = \frac{1}{n} \sum_{i=1}^n \left( (Z_i - \bar{\mathbf{Z}})(Z'_i - \bar{\mathbf{Z}}') \right).$$

**Lemma 1.** *Assume that  $F$  satisfies condition (1.1) and  $k_n$  is a sequence of positive integers satisfying (1.3). If  $J(k_n) = O_P(1)$ , where  $J(k_n)$  is the r.v. defined in (2.3), then we have*

$$\tilde{R}(k_n) = (S_{k_n}^2(\mathcal{E}))^{-1/2} \left( 1 - \frac{1}{2}A(k_n) + o_P(A(k_n)) \right),$$

where

$$A(k_n) := \frac{S_{k_n}^2(\mathbf{Z}) + 2S_{k_n}(\mathcal{E}, \mathbf{Z})}{S_{k_n}^2(\mathcal{E})}.$$

Moreover,  $k_n^{1/2} A(k_n) = O_P(J(k_n))$ .

*Proof.* Note that

$$\begin{aligned} \tilde{R}(k_n) &= (S_{k_n}^2(\mathcal{E}) + S_{k_n}^2(\mathbf{Z}) + 2S_{k_n}(\mathcal{E}, \mathbf{Z}))^{-1/2} \\ &= (S_{k_n}^2(\mathcal{E}))^{-1/2} (1 + A(k_n))^{-1/2}, \end{aligned}$$

which implies the first result, if we prove that  $A(k_n) = o_P(1)$ . Since  $S_{k_n}^2(c\mathcal{E})$  is the sample variance of a  $k_n$ -sample from a standard exponential distribution, we have  $S_{k_n}^2(\mathcal{E}) = (1/c)^2(1 + o_P(1))$  and thus,  $S_{k_n}(\mathcal{E}, \mathbf{Z}) = O_P(S_{k_n}^2(\mathbf{Z}))$ . Moreover,

$$\begin{aligned} S_{k_n}^2(\mathbf{Z}) &= \frac{1}{k_n} \sum_{i=1}^{k_n} \left( \log \tilde{L}(Y_i(1 - U_{n-k_n, n})) - \frac{1}{k_n} \sum_{i=1}^{k_n} \log \tilde{L}(Y_i(1 - U_{n-k_n, n})) \right)^2 \\ &= \frac{1}{k_n} \sum_{i=1}^{k_n} \left( \log \frac{\tilde{L}(Y_i(1 - U_{n-k_n, n}))}{\tilde{L}(k_n/n)} - \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{\tilde{L}(Y_i(1 - U_{n-k_n, n}))}{\tilde{L}(k_n/n)} \right)^2 \\ &\leq \frac{1}{k_n} \sum_{i=1}^{k_n} \left( \log \frac{\tilde{L}(Y_i(1 - U_{n-k_n, n}))}{\tilde{L}(k_n/n)} \right)^2 \\ &\leq \sup_{Y_{k_n} \leq y \leq 1} \left( \log \frac{\tilde{L}(y(1 - U_{n-k_n, n}))}{\tilde{L}(k_n/n)} \right)^2, \end{aligned}$$

that is,

$$S_{k_n}(\mathbf{Z}) \leq k_n^{-1/2} J(k_n).$$

So,

$$A(k_n) = O_P(k_n^{-1/2})J(k_n).$$

and it follows that  $A(k_n) = o_P(1)$ , completing the proof of the Lemma.  $\square$

We give now a result analogue to Lemma 1 for the bootstrapped version  $\tilde{R}^*(k_n)$ . Write

$$\tilde{R}^*(k_n) = (S_{k_n}^2(\mathcal{E}^*))^{-1/2} (1 + A^*(k_n))^{-1/2},$$

where  $A^*(k_n)$  is defined as  $A(k_n)$  replacing  $\mathcal{E}$  and  $\mathbf{Z}$  by  $\mathcal{E}^*$  and  $\mathbf{Z}^*$ , respectively and denote by  $P^*$  the conditional probability, given  $U_{n-k_n,n}, Y_1, \dots, Y_{k_n}$ .

By Theorem 2.1 of Bickel and Freedman (1981),  $S_{k_n}(\mathcal{E}^*) = (1/c)(1 + o_{P^*}(1))$  almost surely. Using now a similar argument as in the proof of Lemma 1, we obtain:

**Lemma 2.** *Assume that  $F$  satisfies condition (1.1) and  $k_n$  is a sequence of positive integers satisfying (1.3). If  $J(k_n) = O_P(1)$ , where  $J(k_n)$  is the r.v. defined in (2.3), then we have*

$$\tilde{R}^*(k_n) = (S_{k_n}^2(\mathcal{E}^*))^{-1/2} (1 - \frac{1}{2}A^*(k_n) + o_{P^*}(A^*(k_n))).$$

Moreover,  $k_n^{1/2} A^*(k_n) = O_{P^*}(J(k_n))$ .

**Proof of Theorem 1.** By Lemma 2, noting that  $A^*(k_n) = o_{P^*}(1)$ , we may write

$$T_n^{**} = (S_{k_n}^2(c\mathcal{E}^*))^{-1/2} \left( ck_n^{1/2}(\overline{\mathcal{E}^{**}} - \overline{\mathcal{E}^*}) + ck_n^{1/2}(\overline{\mathbf{Z}^{**}} - \overline{\mathbf{Z}^*}) \right) (1 + o_{P^*}(1)).$$

Since  $(Y_1, \dots, Y_{k_n})$  is independent of  $U_{n-k_n,n}$ , we also have that  $(Y_1^*, \dots, Y_{k_n}^*)$  is independent of  $U_{n-k_n,n}$  and thus we have (see e.g. Hall (1992), Section 3.11):

$$\begin{aligned} P[k_n^{1/2}(c\overline{\mathcal{E}^{**}} - c\overline{\mathcal{E}^*}) \leq x | (U_{n-k_n,n}, Y_1^*, \dots, Y_{k_n}^*)] &= \\ P[k_n^{1/2}(c\overline{\mathcal{E}^{**}} - c\overline{\mathcal{E}^*}) \leq x | (Y_1^*, \dots, Y_{k_n}^*)] &\xrightarrow{P^*} \Phi(x) \quad \text{a.s.} \end{aligned} \quad (4.1)$$

Now, we will show that for every  $\epsilon > 0$ ,

$$P[|ck_n^{1/2}(\overline{\mathbf{Z}^{**}} - \overline{\mathbf{Z}^*})| > \epsilon | (U_{n-k_n,n}, Y_1^*, \dots, Y_{k_n}^*)] = o_{P^*}(1). \quad (4.2)$$

Since  $Y_{1,k_n}^{**} \geq Y_{1,k_n}^* \geq Y_{k_n}$ , we first note that

$$|\overline{\mathbf{Z}^{**}} - \overline{\mathbf{Z}^*}| \leq 2 \sup_{Y_{k_n} \leq y \leq 1} \left| \log \frac{\tilde{L}(y(1 - U_{n-k_n,n}))}{\tilde{L}(\frac{k_n}{n})} \right|.$$

Thus,

$$ck_n^{1/2} |\overline{\mathbf{Z}^{**}} - \overline{\mathbf{Z}^*}| \leq 2cJ(k_n).$$

Consequently,

$$\begin{aligned} &P[|ck_n^{1/2}(\overline{\mathbf{Z}^{**}} - \overline{\mathbf{Z}^*})| > \epsilon | (U_{n-k_n,n}, Y_1^*, \dots, Y_{k_n}^*)] \\ &\leq P[J(k_n) > \epsilon' | (U_{n-k_n,n}, Y_1^*, \dots, Y_{k_n}^*)] \quad (\epsilon' = \epsilon/2c) \\ &= \mathbb{E}(\mathbf{1}_{\{J(k_n) > \epsilon'\}} | (U_{n-k_n,n}, Y_1, \dots, Y_{k_n}, Y_1^*, \dots, Y_{k_n}^*)). \end{aligned} \quad (4.3)$$

Next, choose  $(\lambda_1, \lambda_2) \in ]1, +\infty[ \times ]1, +\infty[$  and consider the event

$$B_n(\lambda_1, \lambda_2) = \left\{ \lambda_1^{-1} \leq \frac{n}{k_n}(1 - U_{n-k_n, n}) \leq \lambda_1, \lambda_2^{-1} \leq k_n Y_{k_n} \leq \lambda_2 \right\}.$$

On  $B_n(\lambda_1, \lambda_2)$  we have

$$\begin{aligned} J(k_n) &= k_n^{1/2} \sup_{Y_{k_n} \leq y \leq 1} \left| \log \frac{\tilde{L}(y(1 - U_{n-k_n, n}))}{\tilde{L}(k_n/n)} \right| \\ &\leq k_n^{1/2} \sup_{\frac{1}{k_n} \leq y \leq 1} \sup_{\frac{1}{\lambda_1 \lambda_2} \leq t \leq \lambda_1} \left| \log \frac{\tilde{L}(y t k_n/n)}{\tilde{L}(k_n/n)} \right|, \end{aligned} \quad (4.4)$$

which converges to zero as  $n \rightarrow \infty$ , by condition (2.4). Since  $P\{B_n(\lambda_1, \lambda_2)\} \rightarrow 1$  as  $n \rightarrow \infty$ ,  $J(k_n) \xrightarrow{P} 0$  and so, using (4.3) we obtain (4.2).

Therefore, by (4.1) and (4.2), we have that

$$P[T_n^{**} \leq x | (U_{n-k_n, n}, Y_1^*, \dots, Y_{k_n}^*)] \xrightarrow{P^*} \Phi(x),$$

which implies that

$$\Delta'_n := |P[T_n^{**} \leq T_n^* | (U_{n-k_n, n}, Y_1^*, \dots, Y_{k_n}^*)] - \Phi(T_n^*)| \xrightarrow{P^*} 0,$$

i.e.,

$$\forall \epsilon > 0 \quad P[\Delta'_n > \epsilon | (U_{n-k_n, n}, Y_1, \dots, Y_{k_n})] \xrightarrow{P} 0. \quad (4.5)$$

Since  $\tilde{R}(k_n) \xrightarrow{P} c$ , by Theorem 1 of Bacro and Brito (1998), we have that

$$P[\Phi(T_n^*) \leq x | (U_{n-k_n, n}, Y_1, \dots, Y_{k_n})] \xrightarrow{P} x.$$

Finally, by (4.5), it follows that

$$P[P[T_n^{**} \leq T_n^* | (U_{n-k_n, n}, Y_1^*, \dots, Y_{k_n}^*)] \leq x | (U_{n-k_n, n}, Y_1, \dots, Y_{k_n})] \xrightarrow{P} x.$$

□

**Proof of Corollary 1.** For the family (2.6),  $\tilde{L}(s) = D^{1/c}\{1 + O(s^{\beta/c})\}$ ,  $0 < s < 1$ .

In this case, we easily obtain that

$$\sup_{1/k_n \leq y \leq 1} \left| \log \left( \frac{\tilde{L}(y t \frac{k_n}{n})}{\tilde{L}(\frac{k_n}{n})} \right) \right| = O \left( \left( \frac{k_n}{n} \right)^{\frac{\beta}{c}} \right).$$

So, condition (2.4) is verified if  $k_n = o(n^{2\beta/(2\beta+c)})$ .

□

## ACKNOWLEDGEMENTS

ACMF was partially supported by FCT grant SFRH/BPD/66174/2009. JMF was partially supported by FCT grant SFRH/BPD/66040/2009. ACMF and JMF are partially supported by FCT project FAPESP/19805/2014. All three authors were supported by PTDC/MAT/120346/2010, which is financed by national and European Community structural funds through the programs FEDER and COMPETE. All authors are partially supported by CMUP (UID/MAT/00144/2013), which is funded by FCT with national (MEC) and European structural funds through the programs FEDER, under the partnership agreement PT2020.

## References

- Bacro, J.N. and Brito, M. (1998). A tail bootstrap procedure for estimating the tail Pareto index. *Journal of Statistical Planning and Inference* 71, 245-260.
- Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* 74, 457-468.
- Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* 9, 1196-1217.
- Brito, M., Cavalcante, L. and Freitas, A.C.M. (2015). Modelling of extremal earthquakes. In: “Mathematics of Energy and Climate Change”, eds. Bourguignon, J.P., Jeltsch R., Pinto A., Viana M., 39-60, Berlin: Springer.
- Brito, M. and Freitas, A.C.M. (2003). Limiting behaviour of a geometric estimator for tail indices, *Insurance: Mathematics and Economics*, 33, 221-226.
- Brito, M. and Freitas, A.C.M. (2006). Weak convergence of a bootstrap geometric-type estimator with applications to risk theory. *Insurance: Mathematics and Economics* 38, 571-584.
- Brito, M. and Freitas, A.C.M. (2008). Edgeworth expansion for an estimator of the adjustment coefficient. *Insurance: Mathematics and Economics* 43, 203-208.
- Carreau, J., Naveau, P. and Sauquet, E. (2009). A statistical rainfall-runoff mixture model with heavy-tailed components. *Water resources research* 45: W10437.

- Csörgő, M. and Steinebach, J. (1991). On the estimation of the adjustment coefficient in risk theory via intermediate order statistics. *Insurance: Mathematics and Economics* 10, 37-50.
- Deheuvels, P., Haeusler, E. and Mason, D.M. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society* 104, 371-381.
- Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *The Annals of Statistics* 7, 1-26.
- Freitas, A.C.M., Freitas, J.M. and Todd, M. (2010). Hitting time statistics and extreme value theory. *Probab. Theory Related Fields* 147, 675-710.
- de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Math. Centre Tract 32, Mathematisch Centrum, Amsterdam.
- Haeusler, E. and Teugels, J.L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics* 13, 743-756.
- Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B* 44, 37-42.
- Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer-Verlag, New York.
- Hill, B.M. (1975). A simple approach to inference about the tail of a distribution. *The Annals of Statistics* 3, 1163-1174.

Lucarini, V., Faranda, D., Turchetti, G. Vaienti, S. (2012). Extreme value theory for singular measures. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 22: 023135.

MARGARIDA BRITO, CENTRO DE MATEMÁTICA & DEPARTAMENTO DE MATEMÁTICA, FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO, RUA DO CAMPO ALEGRE 687, 4169-007 PORTO, PORTUGAL

*E-mail address:* `mabrito@fc.up.pt`

ANA CRISTINA MOREIRA FREITAS, CENTRO DE MATEMÁTICA & FACULDADE DE ECONOMIA DA UNIVERSIDADE DO PORTO, RUA DR. ROBERTO FRIAS, 4200-464 PORTO, PORTUGAL

*E-mail address:* `amoreira@fep.up.pt`

CORRESPONDING AUTHOR: JORGE MILHAZES FREITAS, CENTRO DE MATEMÁTICA DA UNIVERSIDADE DO PORTO, RUA DO CAMPO ALEGRE 687, 4169-007 PORTO, PORTUGAL

*E-mail address:* `jmfreira@fc.up.pt`