



**HAL**  
open science

# Nonparametric asymptotic confidence intervals for extreme quantiles

Laurent Gardes, Samuel Maistre

► **To cite this version:**

Laurent Gardes, Samuel Maistre. Nonparametric asymptotic confidence intervals for extreme quantiles. *Scandinavian Journal of Statistics*, 2023, 50 (2), pp.825-841. 10.1111/sjos.12610. hal-03382329v2

**HAL Id: hal-03382329**

**<https://hal.science/hal-03382329v2>**

Submitted on 2 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Nonparametric asymptotic confidence intervals for extreme quantiles

Laurent Gardes & Samuel Maistre

Université de Strasbourg & CNRS, IRMA, UMR 7501,

7 rue René Descartes, 67084 Strasbourg Cedex, France

## Abstract

In this paper, we propose new asymptotic confidence intervals for extreme quantiles i.e., for quantiles located outside the range of the available data. We restrict ourselves to the situation where the underlying distribution is heavy-tailed. While asymptotic confidence intervals are mostly constructed around a pivotal quantity, we consider here an alternative approach based on the distribution of order statistics sampled from a uniform distribution. The convergence of the coverage probability to the nominal one is established under a classical second-order condition. The finite sample behavior is also examined and our methodology is applied to a real dataset.

**Keywords:** Extreme quantiles; confidence interval; heavy-tailed distribution.

## 1 Introduction

Let  $X$  be a random variable defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . The distribution of  $X$  is entirely characterized by its cumulative distribution function  $F(\cdot) := \mathbb{P}(X \leq \cdot)$  or equivalently by its quantile function given by

$F^{\leftarrow}(\cdot) = \inf\{x \in \mathcal{S} \mid F(x) \geq \cdot\}$ , where  $\mathcal{S}$  is the support of the distribution. In what follows, we assume that  $F$  is a continuous and strictly increasing function so that  $F^{\leftarrow} = F^{-1}$  is the inverse function of  $F$ .

The question of the estimation of a large quantile  $F^{\leftarrow}(1 - \beta) =: Q(\beta)$  for a level  $\beta$  close to 0 is crucial in actuarial science or in finance in order to be able to compute risk measures such as the Value-at-Risk or the Tail-Value-at-Risk, see Embrechts et al. [10] and Konstantinides [21]. Large quantiles are also useful in climate science to calculate return levels of extreme precipitation or temperature, see for instance Katz et al. [19] and Naveau et al. [22].

There is a large literature dedicated to quantile estimation when  $n$  independent copies  $X_1, \dots, X_n$  of  $X$  are observed. When the level  $\beta = \beta_n$  is not too small in the sense that  $\beta_n \rightarrow 0$  with  $n\beta_n \rightarrow c \in [1, \infty]$ , the quantile  $Q(\beta_n)$  is usually estimated by inverting the empirical cumulative distribution function that is to say by  $\widehat{Q}_n^{(E)}(\beta_n) := X_{n - \lfloor n\beta_n \rfloor, n}$  where  $X_{1,n} \leq \dots \leq X_{n,n}$  is the sample arranged in ascending order. When the level  $\beta_n$  is eventually smaller than  $1/n$ , i.e., when  $n\beta_n \rightarrow c \in [0, 1)$ , the quantile  $Q(\beta_n)$  is located outside the range of the sample  $\{X_1, \dots, X_n\}$ . This kind of quantile is referred to as *extreme quantile* and its estimation requires a priori knowledge of the shape of the tail-distribution. For instance, one can assume that the distribution is heavy-tailed or equivalently that the quantile function  $Q$  is a regularly varying function at 0 with index  $-\xi < 0$  that is to say that for all  $t > 0$ ,

$$\lim_{\alpha \rightarrow 0} \frac{Q(t\alpha)}{Q(\alpha)} = t^{-\xi}.$$

Under this assumption, Weissman [26] proposed to estimate the quantile  $Q(\beta_n)$  by the statistic

$$\widehat{Q}_n^{(W)}(\beta_n) = \widehat{Q}_n^{(E)}(\alpha_n) \left( \frac{\alpha_n}{\beta_n} \right)^{\widehat{\xi}_n},$$

where  $\alpha_n$  is a sequence converging to 0 and such that  $n\alpha_n \rightarrow \infty$  and  $\widehat{\xi}_n$  is a consistent estimator of the tail-index  $\xi$ .

Despite its practical interest, little attention has been paid to the construction of asymptotic confidence intervals for  $Q(\beta_n)$  when  $\beta_n \rightarrow 0$ . Recall that for a preselected probability  $\gamma \in (0, 1)$ , an asymptotic confidence interval for  $Q(\beta_n)$

is a random interval  $[A_{n,\gamma}, B_{n,\gamma}]$  such that,

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ [A_{n,\gamma}, B_{n,\gamma}] \ni Q(\beta_n) \} = \gamma.$$

The common approach to construct asymptotic confidence intervals is to use an asymptotic pivot. This approach was adopted, in the situation where the level  $\beta$  is fixed, by many authors such as Fan and Liu [11], Kocherginsky et al. [20], Parzen et al. [23] among others. For the extreme scenario where  $n\beta_n \rightarrow c \in [0, 1)$ , Buitendag et al. [6] propose a new pivotal quantity leading to various confidence intervals for extreme heavy-tailed quantiles.

Recently, Gardes [13] has adapted a procedure initially introduced by Thomson [25] to construct an asymptotic confidence interval of  $Q(\beta_n)$  when  $\beta_n$  is a sequence converging to 0 with  $n\beta_n \rightarrow \infty$ . To be self contained, let us recall the definition of this confidence interval. First, for  $u \in (0, 1)$ ,  $m \in \mathbb{N} \setminus \{0\}$  and  $\delta \in (0, 1)$ , let us introduce the sets

$$S_{\text{inf}}(u, m, \delta) := \{j \in \{1, \dots, m\} \mid F_{\text{beta}}(u; m - j + 1, j) \leq \delta\} \text{ and}$$

$$S_{\text{sup}}(u, m, \delta) := \{j \in \{1, \dots, m\} \mid F_{\text{beta}}(u; m - j + 1, j) \geq 1 - \delta\},$$

where  $F_{\text{beta}}(\cdot; a, b)$  is the distribution function of a Beta distribution with parameters  $a > 0$  and  $b > 0$ . Since  $\beta_n \in (0, 1)$  is a sequence converging to 0 with  $n\beta_n \rightarrow \infty$ , the sets  $S_{\text{inf}}(\beta_n, n, \delta)$  and  $S_{\text{sup}}(\beta_n, n, \delta)$  are non empty for all  $\delta \in (0, 1)$  (see Lemma 1, i)) allowing us to introduce the integers

$$\mathcal{L}_\gamma(\beta_n) := \max S_{\text{inf}} \left( \beta_n, n, \frac{1 - \gamma}{2} \right) \text{ and } \mathcal{R}_\gamma(\beta_n) := \min S_{\text{sup}} \left( \beta_n, n, \frac{1 - \gamma}{2} \right).$$

The confidence interval proposed in Gardes [13] is given by  $[X_{\mathcal{L}_\gamma(\beta_n), n}, X_{\mathcal{R}_\gamma(\beta_n), n}]$  and we have the following result.

**Proposition 1** *Let  $\beta_n \in (0, 1)$  be a sequence converging to 0 with  $n\beta_n \rightarrow \infty$ . If, as  $n \rightarrow \infty$ ,*

$$\tau_n := \frac{\ln(n)}{(n\beta_n)^{1/2}} \rightarrow 0,$$

*one has  $\mathbb{P} \left( [X_{\mathcal{L}_\gamma(\beta_n), n}, X_{\mathcal{R}_\gamma(\beta_n), n}] \ni Q(\beta_n) \right) = \gamma + o(\tau_n)$ .*

Unfortunately, this confidence interval cannot be used for an extreme quantile.

Indeed, if  $n\beta_n \rightarrow c \in [0, 1)$ ,

$$\lim_{n \rightarrow \infty} \frac{\ln(n)}{(n\beta_n)^{1/2}} = \infty,$$

and thus Proposition 1 does not hold for the extreme quantile  $Q(\beta_n)$ . Moreover, for  $n$  large enough,  $S_{\text{sup}}(\beta_n, n, (1 - \gamma)/2) = \infty$  for all  $\gamma > 1 - 2\exp(-c)$ , see Lemma 1, ii). It thus appears that an extrapolation of the confidence interval bounds beyond the range of the data is required to deal with extreme quantiles.

The aim of this paper is to adapt the construction procedure of quantile confidence intervals introduced in Gardes [13] to the extreme case  $n\beta_n \rightarrow c \in [0, 1)$ . The distribution is assumed to be heavy-tailed and the extrapolation method is inspired by the construction of Weissman's estimator.

The rest of the paper is organized as follows. The new confidence interval for an extreme quantile is given in Section 2 and the convergence of its coverage probability to  $\gamma$  is established. A procedure to reduce the bias of the coverage probability is described in Section 3. The finite sample performance is examined in Section 4 and an application to a real dataset is presented in Section 5. All the proofs are gathered in Appendix.

## 2 Construction of the confidence interval

The following model is considered for the distribution of the random variable  $X$  of interest.

(A.1) There exist positive constants  $d$  and  $\xi$  such that  $\alpha^\xi Q(\alpha) \rightarrow d$  as  $\alpha \rightarrow 0$ .

Under (A.1), the quantile function  $Q$  is a regularly varying function at 0 with index  $-\xi < 0$ . More specifically,  $Q(\alpha) = \alpha^{-\xi} \ell(\alpha^{-1})$  where the function  $\ell$  converges to  $d$  at infinity. Note that  $\ell$  is a slowly varying function at infinity i.e., such that for all  $t > 0$ ,

$$\lim_{x \rightarrow \infty} \frac{\ell(tx)}{\ell(x)} = 1. \quad (1)$$

Condition (A.1) holds for a large number of heavy-tailed distributions such as Fréchet, Burr, student, among others.

Given an estimator  $\hat{\xi}_n$  of the tail-index  $\xi$ , the proposed confidence interval for the extreme quantile  $Q(\beta_n)$  is defined below.

**Definition 1** Let  $X_1, \dots, X_n$  be  $n$  independent random variables with common quantile function  $Q$  satisfying (A.1). For a preselected probability  $\gamma \in (0, 1)$ , a sequence  $\beta_n$  such that  $n\beta_n \rightarrow c \in [0, 1)$  and a sequence  $\alpha_n$  converging to 0 with  $n\alpha_n \rightarrow \infty$ , the two-sided confidence interval for the extreme quantile  $Q(\beta_n)$  is

$$I(\gamma, \beta_n, \hat{\xi}_n) := \left[ X_{n-\lfloor n\alpha_n \rfloor, n} (t_{n,L}(\gamma))^{\hat{\xi}_n}, X_{n-\lfloor n\alpha_n \rfloor, n} (t_{n,R}(\gamma))^{\hat{\xi}_n} \right],$$

where, denoting by  $F_{\text{beta}}^{\leftarrow}(\cdot; a, b)$  the quantile function of a Beta distribution of parameters  $a > 0$  and  $b > 0$ ,

$$t_{n,L}(\gamma) = \beta_n^{-1} F_{\text{beta}}^{\leftarrow} \left( \frac{1-\gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right)$$

and  $t_{n,R}(\gamma) = \beta_n^{-1} F_{\text{beta}}^{\leftarrow} \left( \frac{1+\gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right).$

The idea leading to this confidence interval is given hereafter. We start by remarking that  $1 - F(X_{n-\lfloor n\alpha_n \rfloor, n})$  follows a Beta distribution with parameters  $\lfloor n\alpha_n \rfloor + 1$  and  $n - \lfloor n\alpha_n \rfloor$ . Consequently,

$$\mathbb{P} \left( X_{n-\lfloor n\alpha_n \rfloor, n} (t_{n,L}(\gamma))^{\xi} > Q(\beta_n) \right) = F_{\text{beta}}(z_n(\gamma), \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor), \quad (2)$$

where  $z_n(\gamma) = 1 - F((t_{n,L}(\gamma))^{-\xi} Q(\beta_n))$ . Under (A.1),  $x^{1/\xi}(1 - F(x)) \rightarrow d^{1/\xi}$  as  $x \rightarrow \infty$ . Replacing  $x$  by  $Q(u)$ , we obtain that  $u^{\xi} Q(u) \rightarrow d$  as  $u \rightarrow 0$ . Since  $t_{n,L}(\gamma) \sim \alpha_n / \beta_n$  (see equation (16) in the proof of Theorem 1), one has that  $(t_{n,L}(\gamma))^{-\xi} Q(\beta_n) \sim d\alpha_n^{-\xi} \rightarrow \infty$ . Hence, as  $n \rightarrow \infty$ ,

$$\begin{aligned} z_n(\gamma) &\sim d^{1/\xi} t_{n,L}(\gamma) [Q(\beta_n)]^{-1/\xi} \sim t_{n,L}(\gamma) \beta_n \\ &= F_{\text{beta}}^{\leftarrow} \left( \frac{1-\gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right). \end{aligned}$$

Replacing  $z_n(\gamma)$  in (2) by  $F_{\text{beta}}^{\leftarrow} \left( \frac{1-\gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right)$ , we finally obtain

$$\mathbb{P} \left( X_{n-\lfloor n\alpha_n \rfloor, n} (t_{n,L}(\gamma))^{\xi} \leq Q(\beta_n) \right) \approx 1 - \frac{1-\gamma}{2} = \frac{1+\gamma}{2}. \quad (3)$$

A similar approximation can be obtained for  $t_{n,R}(\gamma)$ . Replacing  $\xi$  by a consistent estimator conducts us to the definition of the confidence interval  $I(\gamma, \beta_n, \hat{\xi}_n)$ . It is worth noting that even if  $z_n(\gamma) \sim \alpha_n$ , the sequence  $F_{\text{beta}}(z_n(\gamma), \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor)$  still depends on  $\gamma$  when  $n$  goes to infinity.

We will prove hereafter that the approximation (3) is theoretically justified and more specifically that the coverage probability of the random interval  $I(\gamma, \beta_n, \widehat{\xi}_n)$  converges to  $\gamma$ . To establish this result, one needs to control the rate of convergence in (A.1).

(A.2) There exist a function  $A$  converging to 0 at infinity and  $\rho < 0$  such that

$$\lim_{\alpha \rightarrow 0} \frac{1}{A(\alpha^{-1})} \left( \frac{Q(t\alpha)}{Q(\alpha)} - t^{-\xi} \right) = t^{-\xi} \int_1^{t^{-1}} u^{\rho-1} du.$$

Condition (A.2) is a standard second order condition. For a detailed discussion on this condition, see for instance the monographs of Beirlant *et al.* [4, Section 3.3] and de Haan and Ferreira [16, Section 2.3]. The function  $|A|$  is a regularly varying function at infinity with index  $\rho < 0$ . The second order parameter  $\rho$  controls the rate of convergence of  $\ell$  to  $d$ . The larger the  $\rho$ , the slower the convergence. Note that in (A.2), the worst situation where  $\rho = 0$  is not considered. This is not a strong restriction since a large range of distributions have a negative second-order parameter. This is for instance the case for the Hall class of heavy-tailed distributions introduced by Hall [17].

The asymptotic coverage probability of the confidence interval  $I(\gamma, \beta_n, \widehat{\xi}_n)$  is established in the following result.

**Theorem 1** *Assume that conditions (A.1) and (A.2) hold. Let  $\alpha_n \in (0, 1)$  be a sequence converging to 0 with  $n\alpha_n \rightarrow \infty$  and  $\beta_n \in (0, 1)$  a sequence such that  $n\beta_n \rightarrow c \in [0, 1)$ . Let also  $\widehat{\xi}_n$  be an estimator of the tail-index  $\xi$  for which there exist a non-degenerate distribution  $\mathcal{D}$  and a sequence  $\sigma_n \rightarrow 0$  as  $n \rightarrow \infty$  such that*

$$\sigma_n^{-1} \left( \frac{\widehat{\xi}_n - \xi}{\xi} \right) \xrightarrow{d} \mathcal{D}.$$

*If  $(n\alpha_n)^{1/2} \max(A(\alpha_n^{-1}); \sigma_n \ln(\alpha_n/\beta_n)) \rightarrow 0$ , then*

$$\lim_{n \rightarrow \infty} \mathbb{P}(I(\gamma, \beta_n, \widehat{\xi}_n) \ni Q(\beta_n)) = \gamma.$$

Note that under the conditions of Theorem 1,

$$\mathbb{P} \left( X_{n - \lfloor n\alpha_n \rfloor, n}(t_{n,L}(\gamma))^{\widehat{\xi}_n} \leq Q(\beta_n) \right) \rightarrow \frac{1 + \gamma}{2}.$$

Hence, as a by-product of Theorem 1, we obtain a one-sided confidence interval for  $Q(\beta_n)$  given by

$$\left[ X_{n-\lfloor n\alpha_n \rfloor, n}(t_{n,L}(2\gamma - 1))^{\widehat{\xi}_n}, \infty \right).$$

To estimate  $\xi$ , the natural way is to use the classic Hill estimator [18] defined for a sequence  $k_n \in \{1, \dots, n\}$  with  $k_n \rightarrow \infty$  and  $k_n/n \rightarrow 0$ , by

$$\widehat{\xi}_n^{(H)}(k_n) := \frac{1}{k_n} \sum_{i=1}^{k_n} \log \left( \frac{X_{n-i+1, n}}{X_{n-k_n, n}} \right). \quad (4)$$

According to [16, Theorem 3.2.5], under (A.2) and if  $k_n^{1/2} A(n/k_n) \rightarrow 0$ ,

$$k_n^{1/2} \left( \frac{\widehat{\xi}_n^{(H)}(k_n) - \xi}{\xi} \right) \xrightarrow{d} \mathcal{N}(0, 1). \quad (5)$$

We thus have the following result which is a direct consequence of Theorem 1.

**Corollary 1** *Assume that conditions (A.1) and (A.2) hold. Let  $\alpha_n \in (0, 1)$ ,  $k_n \in \{1, \dots, n\}$  and  $\beta_n \in (0, 1)$  be sequences such that  $\alpha_n, k_n/n \rightarrow 0$ ,  $k_n, n\alpha_n \rightarrow \infty$  and  $n\beta_n \rightarrow c \in [0, 1]$ . If  $(n\alpha_n)^{1/2} \max \left( A(\alpha_n^{-1}); (k_n)^{-1/2} \ln(\alpha_n/\beta_n) \right) \rightarrow 0$  and  $k_n^{1/2} A(n/k_n) \rightarrow 0$  then*

$$\lim_{n \rightarrow \infty} \mathbb{P}(I(\gamma, \beta_n, \widehat{\xi}_n^{(H)}(k_n)) \ni Q(\beta_n)) = \gamma.$$

The choice of the sample fraction  $k_n$  in the Hill estimator is a difficult task and remains an open problem. Several procedures of selection can be found in the literature. Dekkers et de Haan [9] propose to minimize the mean squared-error of  $\widehat{\xi}_n^{(H)}(k_n) - \xi$ . Adaptation of bootstrap methods are investigated by Danielsson *et al.* [8] and Gomes and Oliveira [15]. Graphical based methods are considered by Sousa and Michailidis [24] and Beirlant *et al.* [2, 3]. In the simulation study (see Section 4), we propose an ad-hoc procedure to select the optimal number  $k_n$  of order statistics.

Finally, note that we can obtain similar results to Corollary 1 by using other tail index estimators such as for instance the bias-reduced Hill estimator proposed by Cairo *et al.* [7]. Because of its simplicity, we choose to focus on the classical Hill estimator which provides good finite sample performances (see Section 4).



### 3 A bias reduction procedure

We propose a heuristic method to reduce the bias in the coverage probability of the confidence interval  $I(\gamma, \beta_n, \hat{\xi}_n)$ . For  $i \in \{1, \dots, n\}$ , let  $U_i = 1 - F(X_i)$ . Since  $F$  is a continuous and strictly increasing function,  $U_1, \dots, U_n$  are independent uniform random variables and for all  $k \in \{1, \dots, n\}$  the order statistic  $U_{k,n}$  follows a beta distribution with parameters  $k$  and  $n - k + 1$ . In the proof of Theorem 1, it is shown that

$$\mathbb{P} \left( X_{n - \lfloor n\alpha_n \rfloor, n} (t_{n,L}(\gamma))^{\hat{\xi}_n} \leq Q(\beta_n) \right) = \mathbb{P} \left( U_{\lfloor n\alpha_n \rfloor + 1, n} \geq \beta_n t_{n,L}(\gamma) W_{n,L}(\gamma) \right),$$

and

$$\mathbb{P} \left( X_{n - \lfloor n\alpha_n \rfloor, n} (t_{n,R}(\gamma))^{\hat{\xi}_n} \geq Q(\beta_n) \right) = \mathbb{P} \left( U_{\lfloor n\alpha_n \rfloor + 1, n} \leq \beta_n t_{n,R}(\gamma) W_{n,R}(\gamma) \right),$$

with

$$W_{n,\bullet}(\gamma) = \exp \left( \frac{\hat{\xi}_n - \xi}{\xi} \ln t_{n,\bullet}(\gamma) \right) \times \left( \frac{1 - F \left( Q(\beta_n) (t_{n,\bullet}(\gamma))^{-\hat{\xi}_n} \right)}{\beta_n (t_{n,\bullet}(\gamma))^{\hat{\xi}_n / \xi}} \right).$$

The construction of the confidence interval is based on the fact that, for  $n$  large enough,  $W_{n,\bullet}(\gamma) \approx 1$ . Indeed, since

$$\beta_n t_{n,L}(\gamma) = F_{\text{beta}}^{\leftarrow} \left( \frac{1 - \gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right),$$

we obtain the approximation

$$\mathbb{P} \left( X_{n - \lfloor n\alpha_n \rfloor, n} (t_{n,L}(\gamma))^{\hat{\xi}_n} \leq Q(\beta_n) \right) \approx \mathbb{P} \left( U_{\lfloor n\alpha_n \rfloor + 1, n} \geq \beta_n t_{n,L}(\gamma) \right) = \frac{1 + \gamma}{2},$$

and similarly for the right-bound of the interval.

To reduce the bias in the coverage probability of the confidence interval, we propose to find a more accurate approximation  $\hat{w}_{n,\bullet}(\gamma)$  of  $W_{n,\bullet}(\gamma)$ . In Lemma 4, we show that under the conditions of Theorem 1,

$$W_{n,\bullet}(\gamma) = 1 + \frac{\hat{\xi}_n - \xi}{\xi} \ln t_{n,\bullet}(\gamma) (1 + o_{\mathbb{P}}(1)), \quad (6)$$

where  $o_{\mathbb{P}}(1)$  is a notation for a random variable converging to 0 in probability. It appears that the error in the approximation of  $W_{n,\bullet}(\gamma)$  by 1 is mainly due to the estimation of the tail index. In view of (6), we decide to approximate

$W_{n,\bullet}(\gamma)$  by taking into account the bias of the estimator  $\widehat{\xi}_n$ . A first solution could consists in replacing  $\widehat{\xi}_n$  by a bias-corrected estimator of the tail index, see e.g., Cairo *et al.* [7]. This requires the estimation of the second order parameter which is a quite difficult task in practice. In this paper, we propose an alternative solution.

From (6), we have  $W_{n,\bullet}(\gamma) \approx 1 + \text{sgn}(\widehat{\xi}_n - \xi)|\widehat{\xi}_n - \xi|/\xi \ln t_{n,\bullet}(\gamma)$  where  $\text{sgn}$  is the sign function. Using the fact that  $\sigma_n^{-1}|\widehat{\xi}_n - \xi|/\xi \xrightarrow{d} |\mathcal{D}|$  for some sequence  $\sigma_n \rightarrow 0$ , we propose to replace  $|\widehat{\xi}_n - \xi|/\xi$  by  $r\sigma_n$  where  $r > 0$  is any constant in the support of the distribution of  $|\mathcal{D}|$ . For instance,  $r$  can be the expectation of  $|\mathcal{D}|$  or a random value generated from  $|\mathcal{D}|$ . To approximate  $W_{n,\bullet}(\gamma)$ , we also need to estimate  $\text{sgn}(\widehat{\xi}_n - \xi)$ . An estimator  $\widehat{s}_n$  is proposed in the next result.

**Proposition 2** *Let  $\widehat{\xi}_n$  and  $\check{\xi}_n$  be two estimators of  $\xi$  such that*

$$\sigma_n^{-1} \left( \frac{\widehat{\xi}_n - \xi}{\xi} \right) \xrightarrow{d} \mathcal{D} \text{ and } \varsigma_n^{-1} \left( \frac{\check{\xi}_n - \xi}{\xi} \right) \xrightarrow{d} \tilde{\mathcal{D}},$$

*for some non-degenerate distributions  $\mathcal{D}$  and  $\tilde{\mathcal{D}}$  and where  $\sigma_n$  and  $\varsigma_n$  are two sequences converging to 0 and such that  $\sigma_n/\varsigma_n \rightarrow \infty$ . Then, the estimator  $\widehat{s}_n := (\widehat{\xi}_n - \check{\xi}_n)/|\widehat{\xi}_n - \check{\xi}_n|$  is such that*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \widehat{s}_n = \text{sgn}(\widehat{\xi}_n - \xi) \right] = 1.$$

We finally propose the following bias-corrected confidence interval.

**Definition 2** *Let  $X_1, \dots, X_n$  be  $n$  independent random variables with quantile function  $Q$  satisfying (A.1). For a preselected probability  $\gamma \in (0, 1)$ , a sequence  $\beta_n$  such that  $n\beta_n \rightarrow c \in [0, 1)$  and a sequence  $\alpha_n$  converging to 0 with  $n\alpha_n \rightarrow \infty$ , the bias-reduced confidence interval for the extreme quantile  $Q(\beta_n)$  is*

$$I_{\text{BR}}(\gamma, \beta_n, \widehat{\xi}_n) := \left[ X_{n - \lfloor n\alpha_n \rfloor, n} (t_{n,L}(\widehat{\gamma}_{n,L}))^{\widehat{\xi}_n}, X_{n - \lfloor n\alpha_n \rfloor, n} (t_{n,R}(\widehat{\gamma}_{n,R}))^{\widehat{\xi}_n} \right],$$

*where  $\widehat{\gamma}_{n,\bullet}$  is the solution (if it exists) of the equation in  $u \in (0, 1)$*

$$t_{n,\bullet}(u) (1 + r\widehat{s}_n\sigma_n \ln t_{n,\bullet}(u)) = t_{n,\bullet}(\gamma). \quad (7)$$

Conditions for the existence and the uniqueness of the solution of (7) are provided in the particular case where the Hill estimator is used to estimate the tail index, see hereafter.

**Application with the Hill estimator** Let us give the expression of the bias-reduced confidence interval when the Hill estimator  $\widehat{\xi}_n^{(H)}(k_n)$  is used for the estimation of the tail-index  $\xi$ . Recall that under the second-order condition (A.2) and for sequences  $k_n$  and  $\tilde{k}_n$  converging to infinity such that  $k_n^{1/2}A(n/k_n) \rightarrow 0$  and  $\tilde{k}_n^{1/2}A(n/\tilde{k}_n) \rightarrow 0$ , one has

$$k_n^{1/2} \left( \frac{\widehat{\xi}_n^{(H)}(k_n) - \xi}{\xi} \right) \xrightarrow{d} \mathcal{N}(0, 1) \text{ and } \tilde{k}_n^{1/2} \left( \frac{\widehat{\xi}_n^{(H)}(\tilde{k}_n) - \xi}{\xi} \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

If the sequence  $\tilde{k}_n$  is such that  $k_n/\tilde{k}_n \rightarrow 0$ , one can use the result of Proposition 2 to estimate the sign of  $\widehat{\xi}_n^{(H)}(k_n) - \xi$ . This estimator is given by

$$\widehat{s}_n^{(H)} := \frac{\widehat{\xi}_n^{(H)}(k_n) - \widehat{\xi}_n^{(H)}(\tilde{k}_n)}{|\widehat{\xi}_n^{(H)}(k_n) - \widehat{\xi}_n^{(H)}(\tilde{k}_n)|}. \quad (8)$$

An example of sequences  $k_n$  and  $\tilde{k}_n$  satisfying the above conditions is given by  $k_n = d(\ln(n))^2$  for some positive constant  $d$  and  $\tilde{k}_n = k_n \ln(\ln(n))$ . If, for instance, we take for  $r$  the expectation of the folded normal distribution  $|\mathcal{N}(0, 1)|$ , we obtain the bias-reduced confidence interval

$$\left[ X_{n - \lfloor n\alpha_n \rfloor, n} \left( t_{n,L}(\widehat{\gamma}_{n,L}^{(H)}) \widehat{\xi}_n^{(H)}(k_n) \right), X_{n - \lfloor n\alpha_n \rfloor, n} \left( t_{n,R}(\widehat{\gamma}_{n,R}^{(H)}) \widehat{\xi}_n^{(H)}(k_n) \right) \right], \quad (9)$$

where  $\widehat{\gamma}_{n,\bullet}^{(H)}$  is the solution (if it exists) of the equation in  $u \in (0, 1)$

$$t_{n,\bullet}(u) \left( 1 + \sqrt{\frac{2}{k_n \pi}} \widehat{s}_n^{(H)} \ln t_{n,\bullet}(u) \right) = t_{n,\bullet}(\gamma). \quad (10)$$

We close this section by a result dedicated to the existence and the uniqueness of the solution of (10).

**Proposition 3** *If the sequences  $k_n$ ,  $\alpha_n$  and  $\beta_n$  involved in (10) are chosen in such a way that  $t_{n,R}(0) = t_{n,L}(0) =: t_n(0) > 1$ ,*

$$k_n > \frac{2}{\pi} \max \left( \frac{(\ln(\beta_n))^2}{(1 - \beta_n t_{n,R}(\gamma))^2}; [1 + \ln(\beta_n^{-1})]^2 \right), \quad (11)$$

and

$$k_n > \frac{2(\ln(t_n(0)))^2}{\pi} \max \left( \left(1 - \frac{t_{n,L}(\gamma)}{t_n(0)}\right)^{-2} ; \left(1 - \frac{t_{n,R}(\gamma)}{t_n(0)}\right)^{-2} \right), \quad (12)$$

then the solution of (10) exists and is unique.

As an example of sequences satisfying the conditions of Proposition 3, one can take  $\beta_n = c/n$  with  $c > 0$ ,  $\alpha_n = m_n/n$  with  $m_n \rightarrow \infty$  and  $m_n(\ln(m_n))^2 = o(n)$  and  $k_n = d \max((\ln(n))^2, m_n(\ln(m_n))^2)$  for some  $d > 2/\pi$ , see Lemma 5 for a proof. In particular, a concrete choice is given by  $k_n = d(\ln(n))^2$  for some  $d > 2/\pi$ ,  $\beta_n = c/n$  with  $c > 0$  and  $\alpha_n = (\ln(k_n))^a/n$  for some  $a > 0$ . This choice for the sequence  $\alpha_n$  is considered in the simulation study, see Section 4.

## 4 Simulation study

Let  $X_1, \dots, X_n$  be  $n$  independent copies of a random variable  $X$ . We propose to look at the finite sample behavior of the confidence interval for  $Q(\beta_n)$  with  $\beta_n = 1/n$ . Four models are considered for the quantile function of  $X$ .

- **Model 1:** Burr distribution. For  $\kappa > 0$  and  $c > 0$ , the quantile function is given for all  $u \in (0, 1)$  by  $Q(u) = (u^{-1/\kappa} - 1)^{1/c}$ . The tail-index is  $\xi = 1/(c\kappa)$  and the second-order parameter is  $\rho = -1/\kappa$ .
- **Model 2:** Fréchet distribution. The quantile function is given for all  $u \in (0, 1)$  by  $Q(u) = [-\ln(1-u)]^{-\xi}$ . The second-order parameter is  $\rho = -1$ .
- **Model 3:** Absolute Student distribution. For  $\mu > 0$ , the quantile function is given for all  $u \in (0, 1)$  by  $Q(u) = \text{qt}(1 - u/2; \mu)$ , where  $\text{qt}(\cdot; \mu)$  is the quantile function of a Student distribution with  $\mu$  degree of freedom. The tail-index is  $\xi = 1/\mu$  and the second-order parameter is  $\rho = -1$ .
- **Model 4:** Log-Gamma distribution with shape parameter  $a = 2$  and rate parameter  $r = 2$ . The tail-index is given here by  $\xi = 1/r = 1/2$  and the second-order parameter is  $\rho = 0$ .

Note that condition (A.2) is satisfied by models 1 to 3 but not for model 4 for which  $\rho = 0$ . For each model, we consider three sample sizes :  $n \in$

$\{200, 500, 1000\}$ . The tail index is estimated by the Hill estimator  $\widehat{\xi}_n^{(H)}(k_n)$  for some  $k_n \in \{1, \dots, n-1\}$ . The sequence  $\alpha_n$  is taken such as

$$n\alpha_n = \max\left(3; \left\lfloor (\ln(k_n))^{0.85} \right\rfloor\right).$$

This choice is motivated by the remark made after Proposition 3 since for  $k_n$  large enough,  $n\alpha_n \sim (\ln(k_n))^a$  with  $a = 0.85$ . Note that our choice ensures that  $n\alpha_n \geq 3$  whatever the value of  $k_n$ . For the bias-corrected version of the confidence interval given in (9), we take  $\widehat{s}_n^{(H)}$  as in (8) with the sequence

$$\tilde{k}_n = \min(\lfloor k_n \ln(\ln(n)) \rfloor; n-1).$$

This choice corresponds to the example given after the definition of  $\widehat{s}_n^{(H)}$  (equation (8)) with the constraint  $\tilde{k}_n \leq n-1$ . This bias-corrected version is only used when the uniqueness of a solution to (20) has been shown. This corresponds to  $k \geq 26$  for  $n = 200$ ,  $k \geq 34$  for  $n = 500$  and  $k \geq 40$  for  $n = 1000$ . When the uniqueness is not guaranteed, the bias-reduced interval is replaced by the standard one, as presented in Definition 1.

The coverage probability is approximated by generating  $N = 10000$  independent samples of size  $n$  for the four above mentioned models. This approximation is denoted  $\bar{\gamma}_N(k_n)$ . In the same way, we approximate the coverage probabilities

$$\mathbb{P}\left(X_{n-\lfloor n\alpha_n \rfloor, n}(t_{n,L}(\gamma))^{\widehat{\xi}_n} \leq Q(\beta_n)\right) \text{ and } \mathbb{P}\left(X_{n-\lfloor n\alpha_n \rfloor, n}(t_{n,R}(\gamma))^{\widehat{\xi}_n} \geq Q(\beta_n)\right),$$

by  $\bar{\gamma}_{N,L}(k_n)$  and  $\bar{\gamma}_{N,R}(k_n)$ . Of course, it is expected to obtain for  $\bar{\gamma}_N(k_n)$  a value close to  $\gamma$  and for  $\bar{\gamma}_{N,L}(k_n)$  and  $\bar{\gamma}_{N,R}(k_n)$ , values close to  $(1 + \gamma)/2$ .

We compare our intervals' performances with the "new normal" ridge confidence interval presented in Buitendag et al. [6]. This interval is based on a ridge regression estimator for the extreme value index, as developed in Buitendag et al. [5], as well as on a new asymptotically normal quantity.

We choose this interval as a competitor among the 12 presented in Buitendag et al. [6] since it ensured accurate coverage in most of their simulations. This new normal ridge confidence interval is based on an estimator

of the second order parameter  $\rho$  from Fraga Alves, Gomes, and de Haan [12]. However, this estimator is not always well-defined. When such problem occurs, we replace the new normal ridge confidence interval by the new normal Hill confidence interval, which is also presented in Buitendag et al. [6] and does not rely on the estimation of  $\rho$ . The coverage comparison of our two confidence intervals (standard and bias reduced) with the new normal ridge is given in Figure 1.

Our confidence intervals provide good coverage probabilities for the distributions we considered, as we can see on the left column of Figure 1.

The bias reduction enlarges the range of values of  $k$  for which the interval has a coverage rate close to the objective. The new normal ridge confidence interval performs especially well for the Student distribution ( $\xi = 1, \rho = -1$ ) and the Fréchet distribution ( $\xi = 0.5, \rho = -1$ ). However, it is not very accurate for the log-gamma distribution ( $\xi = 0.5, \rho = 0$ ), while our bias reduced interval is always close to the nominal level of 0.95 even if the second-order condition (A.2) is not valid in this case.

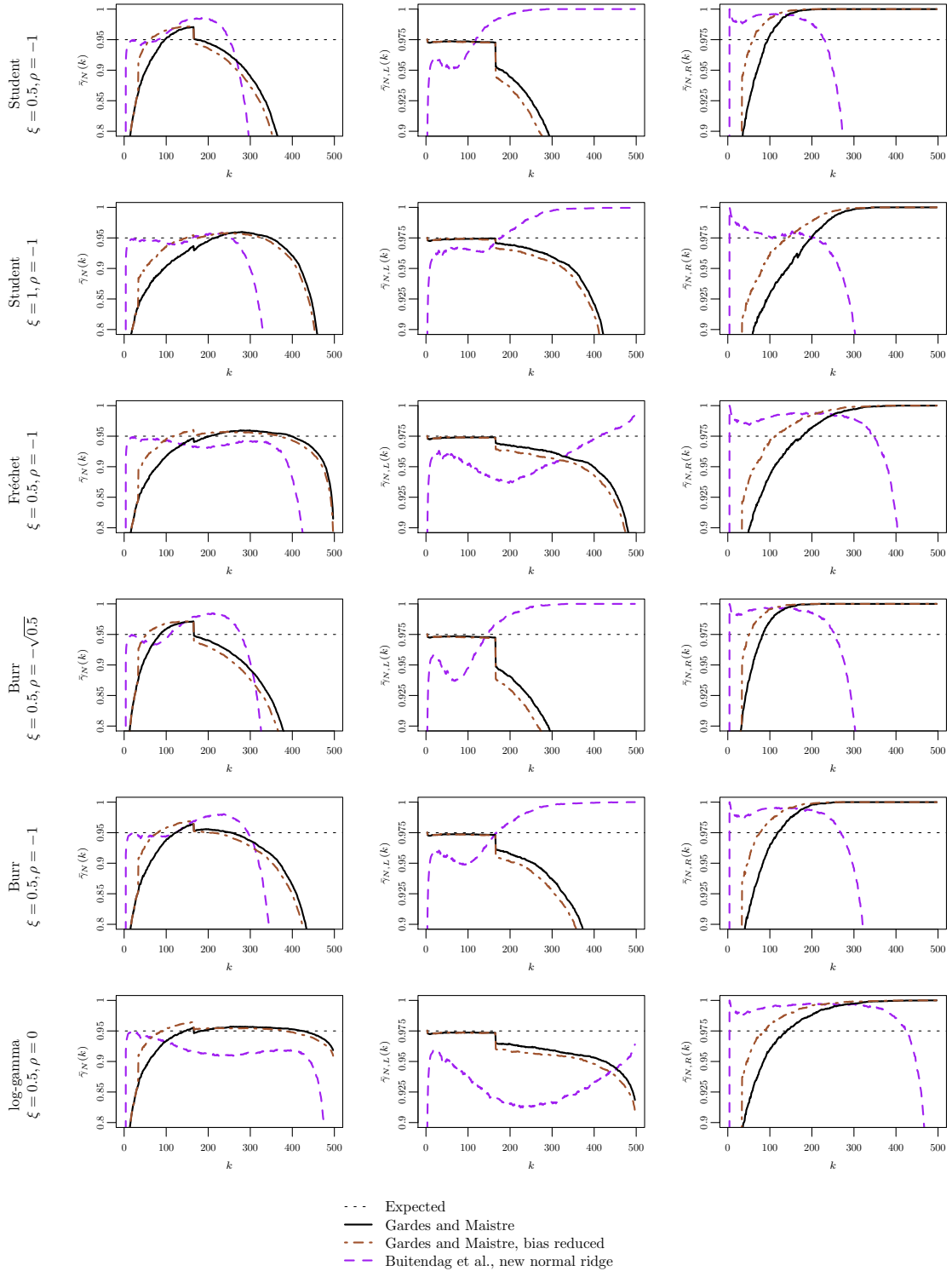
For the left bound, in the central column of Figure 1, our intervals coverage rates are very close to the nominal level of 0.975, as long as  $k$  does not exceed 160. As for the right bound, in the right column of Figure 1, the nominal level is only hit once, for a value of  $k$  below 200, depending on the distribution and the interval considered (bias reduced or not).

However, coverage curves only indicate approximately what will happen if we choose a fixed  $k$ . Usually, we select a value of  $k$  which seems adequate with regard to the Hill plot, as it will be presented in the next paragraph.

**Selection of the sequence  $k_n$**  – To select the number of observations used to estimate the tail-index, we propose for each of the  $N$  samples to pick  $k_n$  in a stability region of the plot  $\{(j, \hat{\xi}_n^{(H)}(j)); j \in J\}$  where  $J := \{\lfloor 0.05n \rfloor, \dots, \lfloor 0.5n \rfloor\}$ . To detect this region, we use the following procedure. Let

$$J^{(H)} := \left[ \min_{j \in J} \hat{\xi}_n^{(H)}(j); \max_{j \in J} \hat{\xi}_n^{(H)}(j) \right],$$

Figure 1: Coverage rate of the interval and of the left and right bounds (from left to right) for absolute Student, Fréchet, Burr and log-gamma distributions and sample size  $n = 500$ .



be the interval of variation of the Hill estimator  $\hat{\xi}_n^{(H)}(k_n)$  when  $k_n$  goes through the interval  $J$ . This interval is divided into  $M$  non-overlapping slices  $\mathcal{S}_1, \dots, \mathcal{S}_M$ . We define the stability region as the longest sequence of consecutive integers for which the Hill estimator belongs to the same slice. More specifically, for each  $m \in \{1, \dots, M\}$ , we denote by  $J_m^{(\text{opt})}$  the longest sequence of consecutive integers such that  $\hat{\xi}_n^{(H)}(j) \in \mathcal{S}_m$  for each  $j \in J_m^{(\text{opt})}$ . Denoting by  $|J|$  the number of elements in an interval  $J$  and letting

$$m_* = \arg \max_{m \in \{1, \dots, M\}} |J_m^{(\text{opt})}|,$$

the stability region is given by  $J_{m_*}^{(\text{opt})}$ . We finally take for  $k_n$  the integer

$$k_n^{(\text{opt})} := \frac{1}{|J_{m_*}^{(\text{opt})}|} \sum_{j \in J_{m_*}^{(\text{opt})}} j = \frac{1}{2} \left( \max J_{m_*}^{(\text{opt})} + \min J_{m_*}^{(\text{opt})} \right).$$

This procedure is applied with  $M = 5$  and the obtained results for the two-sided confidence intervals are shown in Tables 1 to 6. We also give in Tables 7 to 18 (see Appendix), the empirical coverage probabilities for the one-sided confidence intervals (left and right bound).

Since our selection procedure for  $k_n^{(\text{opt})}$  is not tailored for the new normal ridge estimator, we also use a "rule of thumb" with  $k = \lfloor n/5 \rfloor$ . We notice that our bias reduced confidence interval nearly always performs well when applied for  $k_n^{(\text{opt})}$ . The worst case is attained for a sample size of  $n = 200$  with a Student distribution where  $\xi = 1$  and  $\rho = -1$  for which the estimated coverage is 0.9381, see Table 2. The bias reduction is mostly beneficial.

On the contrary, the performances of the new normal ridge confidence interval are often better using  $\lfloor n/5 \rfloor$ , but it only outperforms our confidence intervals when  $n = 200$  for the Student and the Fréchet distributions.

## 5 Application to a real dataset

We apply our procedure to the norwegian fire insurance claims dataset as studied in Beirlant et al. [1]. Data have been corrected for inflation and the claims are expressed in millions of 2012 Norwegian Krone (NOK). The heavy tailness of this data set has already been validated in Gardes and Girard [14].



$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9355	0.9463	0.9732
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9236	0.9580	0.9531
500	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9446	0.9520	0.9764
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9520	0.9653	0.9604
1000	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9510	0.9535	0.9717
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9531	0.9530	0.9713

Table 1: Empirical coverage probabilities for a theoretical coverage probability of  $\gamma = 0.95$  and a Student distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9066	0.9381	0.9506
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.8663	0.9147	0.9484
500	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9240	0.9427	0.9571
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9037	0.9353	0.9415
1000	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9400	0.9512	0.9563
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9191	0.9399	0.9294

Table 2: Empirical coverage probabilities for a theoretical coverage probability of  $\gamma = 0.95$  and a Student distribution ( $\xi = 1$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9203	0.9458	0.9543
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.8823	0.9241	0.9495
500	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9386	0.9500	0.9374
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9171	0.9453	0.9426
1000	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9501	0.9556	0.9300
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9306	0.9483	0.9260

Table 3: Empirical coverage probabilities for a theoretical coverage probability of  $\gamma = 0.95$  and a Fréchet distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9429	0.9519	0.9656
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9308	0.9613	0.9467
500	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9465	0.9475	0.9684
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9588	0.9681	0.9433
1000	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9493	0.9463	0.9707
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9535	0.9498	0.9565

Table 4: Empirical coverage probabilities for a theoretical coverage probability of  $\gamma = 0.95$  and a Burr distribution ( $\xi = 0.5$  and  $\rho = -\sqrt{0.5}$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9319	0.9496	0.9611
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9061	0.9460	0.9493
500	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9410	0.9507	0.9670
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9384	0.9578	0.9441
1000	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9518	0.9548	0.9696
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9458	0.9530	0.9444

Table 5: Empirical coverage probabilities for a theoretical coverage probability of  $\gamma = 0.95$  and a Burr distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9285	0.9503	0.9550
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.8951	0.9316	0.9514
500	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9464	0.9541	0.9200
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9336	0.9530	0.9301
1000	$\bar{\gamma}_N(k_n^{(\text{opt})})$	0.9550	0.9557	0.8846
	$\bar{\gamma}_N(\lfloor n/5 \rfloor)$	0.9441	0.9527	0.8938

Table 6: Empirical coverage probabilities for a theoretical coverage probability of  $\gamma = 0.95$  and a log-Gamma distribution ( $\xi = 0.5$  and  $\rho = 0$ ).

We only use data from years 1985 to 1992, for which the sample sizes are similar, ranging from 607 in 1985 to 827 in 1988. For the quantile level, we take  $\beta = 1/638$  where 638 is the median value of the sample sizes from years 1985 to 1992. For years with a sample size larger than 638, the level  $\beta$  can thus be considered as extreme. For each year, we select  $k_n^{(\text{opt})}$  as described in Section 4. An illustration of the procedure is given in Figure 2 for year 1992, for which  $k_n^{(\text{opt})} = 195$ .

On Figure 3, we check that the procedure is consistent, showing the rough alignment of points

$$\left\{ \left( -\log \left( i/k_n^{(\text{opt})} \right), \log \left( X_{n-i,n}/X_{n-k_n^{(\text{opt}),n}} \right) \right), 1 \leq i < k_n^{(\text{opt})} \right\},$$

on a line of slope  $\widehat{\xi}_n(k_n^{(\text{opt})})$  passing through the origin.

The resulting bias corrected intervals are illustrated in Figure 4.

Note that the larger intervals are obtained for the higher centers of the intervals, which gives an insight on the quantile estimate. This is not surprising since a large quantile is associated with a large tail-index and a high variability in the tail of the distribution. The width of the interval provides us an indication on the tail-variability and in that sense, our results are consistent with those exposed in Gardes and Girard [14]. Note that the confidence interval for year 1985 [177, 850] does not match confidence intervals for years 1990 and 1991 (respectively [46, 158] and [32, 116]).

Figure 2: Selection procedure for  $k_n^{(\text{opt})}$  for year 1992. We represent the estimated extreme value index in terms of  $k$ . The shaded area represent the  $k$  values outside  $[0.05n, 0.5n]$ , the horizontal dashed lines represent the slices and the red part of the curve is the longest part of the curve that belongs to the same slice.

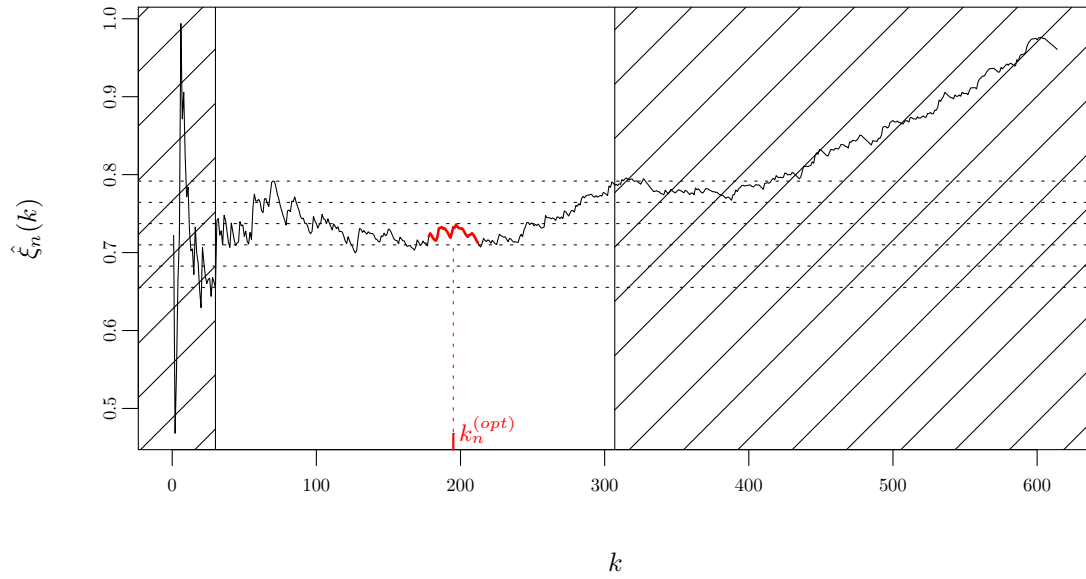


Figure 3: Q-Q plot for year 1992 for the  $k_n^{(\text{opt})} = 195$  largest observations.

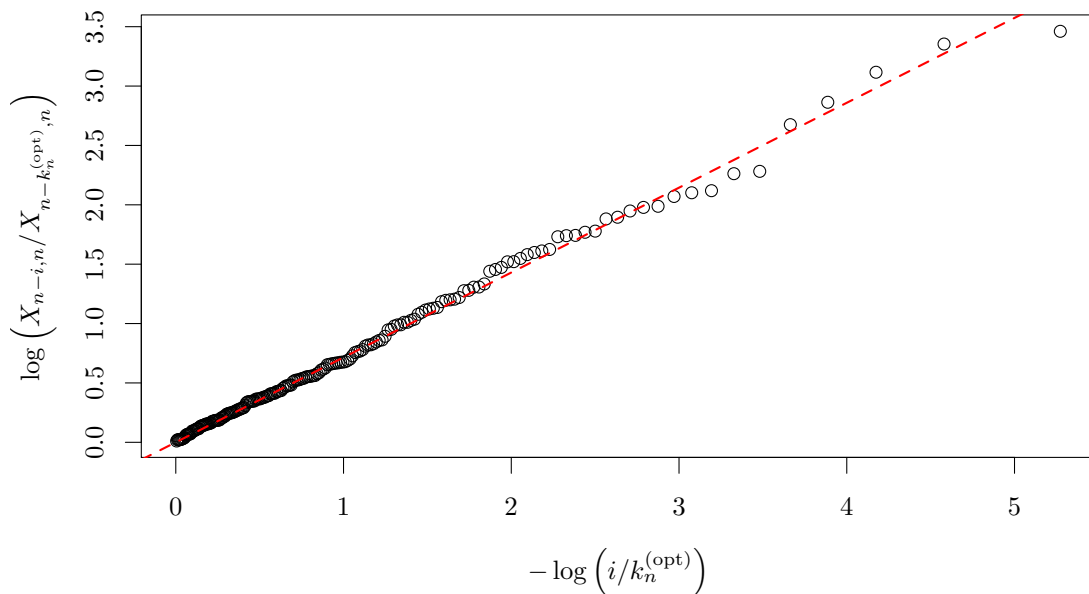
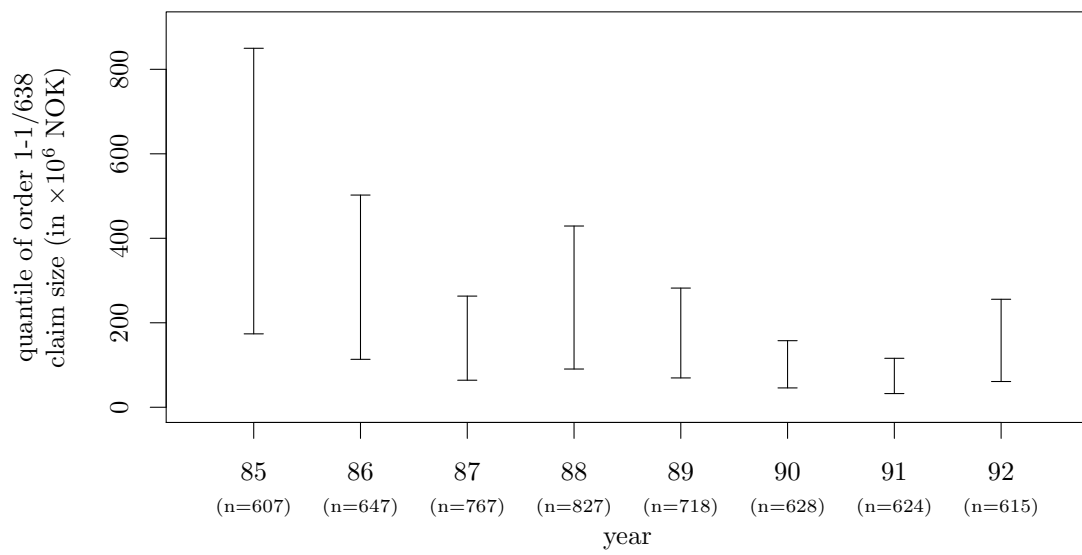


Figure 4: Application to Norwegian Fire Insurance data set for years 1985 to 1992.



## References

- [1] Beirlant J., Teugels, J. and Vynckier, P. (1996). *Practical Analysis of Extreme Values*, Leuven University Press.
- [2] Beirlant, J., Vynckier, P. and Teugels, J. (1996). Tail index estimation, Pareto quantile plots and regression diagnostics, *Journal of the American Statistical Association*, **91**, 1659–1667.
- [3] Beirlant, J., Dierckx, G., Guillou, A. and Starica, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5(2)**, 157–180.
- [4] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004). *Statistics of Extremes - Theory and Applications*, Wiley Series in Probability and Statistics.
- [5] Buitendag, S., Beirlant, J. and de Wet, T. (2019). Ridge regression estimators for the extreme value index, *Extremes*, **22**, 271–292.
- [6] Buitendag, S., Beirlant, J. and de Wet, T. (2020). Confidence intervals for extreme Pareto-type quantiles, *Scandinavian Journal of Statistics*, **47(1)**, 36–55.
- [7] Cairo, F., Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator, *REVSTAT - Statistical Journal*, **3(2)**, 113–136.
- [8] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation, *Journal of Multivariate Analysis*, **76**, 226–248.
- [9] Dekkers, A.L.M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation, *Journal of Multivariate Analysis*, **47(2)**, 173–195.
- [10] Embrechts, P., Kluppelberg, C. and Mikosch, T. (2013). *Modelling Extremal Events for Insurance and Finance*, **Vol. 33**, Springer Science & Business Media.

- [11] Fan, Y. and Liu, R. (2016). A direct approach to inference in nonparametric and semiparametric quantile models, *Journal of Econometrics*, **191**(1), 196–216.
- [12] Fraga Alves, M. I., Gomes, M. I., and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**, 193–213.
- [13] Gardes, L. (2020). Nonparametric confidence interval for conditional quantiles with large-dimensional covariates, *Electronic Journal of Statistics*, **14**(1), 661–701.
- [14] Gardes, L. and Girard, S. (2021). On the estimation of the variability in the distribution tail, *Test Spanish Society of Statistics and Operations Research*, **30**, 884–907.
- [15] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in statistics of extremes - Choice of the optimal sample fraction, *Extremes*, **4**(4), 331–358.
- [16] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering, Springer.
- [17] Hall, P. (1982). On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society, Series B (Methodological)*, **44**(1), 37–42.
- [18] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.
- [19] Katz, R.W., Parlange, M.B. and Naveau, P. (2002). Statistics of extreme in hydrology, *Advances In Water Ressources*, **25**, 1287–1304.
- [20] Kocherginsky, M., He, X. and Mu, Y. (2005). Practical confidence intervals for regression quantiles, *Journal of Computational and Graphical Statistics*, **14**(1), 41–55.
- [21] Konstantinides, D. (2018). *Risk theory. A Heavy Tail Approach*, World Scientific Publishing.



- [22] Naveau, P., Hannart, A. and Ribes, A. (2020). Statistical methods for extreme event attribution in climate science, *Annual Review Of Statistics And Its Application*, **7**, 89–110.
- [23] Parzen, M.I., Wei, L.J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions, *Biometrika*, **81(2)**, 341–350.
- [24] de Sousa, B. and Michailidis, G. (2004). A diagnostic plot for estimating the tail index of a distribution, *Journal of Computational and Graphical Statistics*, **13(4)**, 974–995.
- [25] Thompson, W.R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown distribution form, *Annals of Mathematical Statistics*, **7**, 122–128.
- [26] Weissman, I. (1978). Estimation of parameters and large quantiles based on the  $k$  largest observations, *Journal of the American Statistical Association*, **73**, 812–815.

## Appendix - Proofs and additional results

### Preliminaries results

**Lemma 1** *i) If  $\alpha_n$  is a sequence converging to 0 with  $n\alpha_n \rightarrow \infty$  then for all  $\delta \in (0, 1)$  and for  $n$  large enough,  $S_{\text{inf}}(\alpha_n, n, \delta) \neq \emptyset$  and  $S_{\text{sup}}(\alpha_n, n, \delta) \neq \emptyset$ .*

*ii) If  $\beta_n \in (0, 1)$  is a sequence such that  $n\beta_n \rightarrow c \in [0, 1)$  then for  $n$  large enough,  $S_{\text{sup}}(\alpha_n, n, \delta) = \emptyset$  for all  $\delta < \exp(-c)$ .*

**Proof** – i) The function  $j \mapsto F_{\text{beta}}(\alpha_n; n - j + 1, j)$  is increasing. Hence, for all  $j \in \{1, \dots, n\}$ ,

$$\alpha_n^n \leq F_{\text{beta}}(\alpha_n; n - j + 1, j) \leq 1 - (1 - \alpha_n)^n.$$

Since  $\alpha_n^n \rightarrow 0$  and  $1 - (1 - \alpha_n)^n \rightarrow 1$ , the conclusion is straightforward.

ii) As shown before, for all  $j \in \{1, \dots, n\}$ ,

$$0 \leftarrow \beta_n^n \leq F_{\text{beta}}(\beta_n; n - j + 1, j) \leq 1 - (1 - \beta_n)^n \rightarrow 1 - \exp(-c).$$

Hence, if  $1 - \delta > 1 - \exp(-c)$  then, for  $n$  large enough,  $S_{\text{sup}}(\alpha_n, n, \delta) = \emptyset$ . ■

In all what follows,  $U_1, \dots, U_n$  are  $n$  independent standard uniform random variables. The associated order statistics are denoted  $U_{1,n} \leq \dots \leq U_{n,n}$ .

**Lemma 2** *Let  $\alpha_n \in (0, 1)$  be a sequence converging to 0 with  $n\alpha_n \rightarrow \infty$  as  $n \rightarrow \infty$ . If there exist  $\delta \in (0, 1)$  and  $k_n \in \{1, \dots, n\}$  such that  $\mathbb{P}(U_{k_n, n} \leq \alpha_n) \rightarrow \delta$  then  $k_n/(n\alpha_n) \rightarrow 1$ .*

**Proof** – Let us first show that necessarily,  $k_n \rightarrow \infty$ . Suppose that this is not the case. There exists  $A > 0$  such that for all  $N \in \mathbb{N} \setminus \{0\}$ , there exists  $n \geq N$  for which  $k_n < A$ . Now, let  $E_1, \dots$  be independent standard exponential random variables. From Rényi's representation of ordered standard uniform random variables,

$$\mathbb{P}(U_{k_n, n} \leq \alpha_n) = \mathbb{P}\left(\frac{nT_{k_n}}{T_{n+1}} \leq n\alpha_n\right),$$

where for  $i \in \mathbb{N} \setminus \{0\}$ ,  $T_i = E_1 + \dots + E_i$ . By the law of large numbers,  $n/T_{n+1}$  converges in probability to 1 and for all  $N \in \mathbb{N} \setminus \{0\}$ , there exists  $n \geq N$  such that

$$T_{k_n} \leq \sum_{i=1}^{\lfloor A \rfloor + 1} E_i.$$

Hence, for all  $N \in \mathbb{N} \setminus \{0\}$ , there exists  $n \geq N$  such that

$$\mathbb{P}(U_{k_n, n} \leq \alpha_n) \geq \mathbb{P}\left(\frac{n}{T_{n+1}} \sum_{i=1}^{\lfloor A \rfloor + 1} E_i \leq n\alpha_n\right) \rightarrow 1,$$

since  $n\alpha_n \rightarrow \infty$ . Hence,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(U_{k_n, n} \leq \alpha_n) = 1,$$

which is in contradiction with the fact that  $\mathbb{P}(U_{k_n, n} \leq \alpha_n) \rightarrow \delta \in (0, 1)$ .

Now, for all  $\varepsilon > 0$ , let  $A_{n, \varepsilon} = \{nU_{k_n, n}/k_n \in [1 - \varepsilon, 1 + \varepsilon]\}$ . One has

$$\mathbb{P}(U_{k_n, n} \leq \alpha_n) = \mathbb{P}(\{U_{k_n, n} \leq \alpha_n\} \cap A_{n, \varepsilon}) + \mathbb{P}(\{U_{k_n, n} \leq \alpha_n\} \cap A_{n, \varepsilon}^{\text{C}}),$$

where  $A_{n, \varepsilon}^{\text{C}}$  is the complement of the set  $A_{n, \varepsilon}$ . Since  $k_n \rightarrow \infty$ , Rényi's representation of the order statistic  $U_{k_n, n}$ , and the law of large number entail that  $nU_{k_n, n}/k_n$  converges in probability to 1. Hence,

$$\mathbb{P}(\{U_{k_n, n} \leq \alpha_n\} \cap A_{n, \varepsilon}^{\text{C}}) \leq \mathbb{P}(A_{n, \varepsilon}^{\text{C}}) \rightarrow 0.$$

Moreover,

$$\mathbb{P}(\{U_{k_n,n} \leq \alpha_n\} \cap A_{n,\varepsilon}) = \begin{cases} 0 & \text{if } n\alpha_n/k_n < 1 - \varepsilon, \\ \mathbb{P}(A_{n,\varepsilon}) & \text{if } n\alpha_n/k_n > 1 + \varepsilon. \end{cases}$$

Since  $\mathbb{P}(A_{n,\varepsilon}) \rightarrow 1$ , we thus have proven that for all  $\varepsilon > 0$ ,  $\mathbb{P}(U_{k_n,n} \leq \alpha_n) \rightarrow 0$  if  $n\alpha_n/k_n < 1 - \varepsilon$  and  $\mathbb{P}(U_{k_n,n} \leq \alpha_n) \rightarrow 1$  if  $n\alpha_n/k_n > 1 + \varepsilon$ . Since by assumption,  $\mathbb{P}(U_{k_n,n} \leq \alpha_n) \rightarrow \delta \in (0, 1)$ , it appears that for all  $\varepsilon > 0$ , for  $n$  large enough,  $1 - \varepsilon \leq n\alpha_n/k_n \leq 1 + \varepsilon$  and the proof is complete.  $\blacksquare$

Let us now give a technical result which is a consequence of the second-order condition.

**Lemma 3** *Under conditions (A.1) and (A.2), if  $\alpha_n \in (0, 1)$  and  $\beta_n \in (0, 1)$  are sequences converging to 0 with  $\alpha_n/\beta_n \rightarrow \infty$  then, for every sequence  $t_n$  such that  $t_n \sim (\alpha_n/\beta_n)^{-\xi}$ , one has*

$$\lim_{n \rightarrow \infty} \frac{1}{|A(\alpha_n^{-1})|} \left| \frac{1 - F(Q(\beta_n)t_n)}{t_n^{-1/\xi} \beta_n} - 1 \right| = -\frac{1}{\xi\rho}.$$

**Proof** – Condition (A.2) and [16, Theorem 2.3.9] entail that for all  $t > 0$ ,

$$\lim_{x \rightarrow \infty} \frac{1}{A((1 - F(x))^{-1})} \left( \frac{1 - F(tx)}{1 - F(x)} - t^{-1/\xi} \right) = \xi^{-2} t^{-1/\xi} \int_1^t u^{\rho/\xi - 1} du.$$

We can then use [16, Remark B.3.15] to obtain

$$\lim_{x \rightarrow \infty} \frac{1}{A((1 - F(x))^{-1})} \left( t_x^{1/\xi} \frac{1 - F(xt_x)}{1 - F(x)} - 1 \right) = -\frac{1}{\rho\xi},$$

where  $t_x \rightarrow \infty$  as  $x \rightarrow \infty$ . Applying this result with  $x = Q(\beta_n)t_n \sim d\alpha_n^{-\xi} \rightarrow \infty$  and  $t_x = t_n^{-1} \sim (\alpha_n/\beta_n)^\xi \rightarrow \infty$  as  $n \rightarrow \infty$  leads to

$$\frac{\beta_n t_n^{-1/\xi}}{1 - F(Q(\beta_n)t_n)} \frac{1}{A((1 - F(Q(\beta_n)t_n)^{-1})} \times \left( \frac{1 - F(Q(\beta_n)t_n)}{t_n^{-1/\xi} \beta_n} - 1 \right) \rightarrow \frac{1}{\xi\rho},$$

as  $n \rightarrow \infty$ . From (A.1),  $x^{1/\xi}(1 - F(x)) \rightarrow d^{1/\xi}$  as  $x \rightarrow \infty$  and thus,  $1 - F(Q(\beta_n)t_n) \sim \alpha_n$ . Hence, since  $|A|$  is a regularly varying function,  $|A((1 - F(Q(\beta_n)t_n)^{-1})| \sim |A(\alpha_n^{-1})|$  and thus

$$\lim_{n \rightarrow \infty} \frac{1}{|A(\alpha_n^{-1})|} \left| \frac{1 - F(Q(\beta_n)t_n)}{t_n^{-1/\xi} \beta_n} - 1 \right| = -\frac{1}{\xi\rho}. \quad \blacksquare$$

In the next lemma, an asymptotic expansion of the random variable

$$W_{n,\bullet}(\gamma) = \exp\left(\frac{\widehat{\xi}_n - \xi}{\xi} \ln t_{n,\bullet}(\gamma)\right) \times \left(\frac{1 - F\left(Q(\beta_n)(t_{n,\bullet}(\gamma))^{-\widehat{\xi}_n}\right)}{\beta_n(t_{n,\bullet}(\gamma))^{\widehat{\xi}_n/\xi}}\right),$$

is established.

**Lemma 4** *Assume that conditions (A.1) and (A.2) hold. Let  $\alpha_n \in (0, 1)$  be a sequence converging to 0 with  $n\alpha_n \rightarrow \infty$  and  $\beta_n \in (0, 1)$  a sequence such that  $n\beta_n \rightarrow c \in [0, 1)$ . Let also  $\widehat{\xi}_n$  be an estimator of the tail-index  $\xi$  for which there exist a non-degenerate distribution  $\mathcal{D}$  and a sequence  $\sigma_n \rightarrow 0$  as  $n \rightarrow \infty$  such that  $\sigma_n^{-1}(\widehat{\xi}_n - \xi)/\xi \xrightarrow{d} \mathcal{D}$ . If  $(n\alpha_n)^{1/2} \max(A(\alpha_n^{-1}); \sigma_n \ln(\alpha_n/\beta_n)) \rightarrow 0$ , then  $W_{n,\bullet}(\gamma) = 1 + (\widehat{\xi}_n - \xi)/\xi \ln t_{n,\bullet}(\gamma)(1 + o_{\mathbb{P}}(1))$ .*

**Proof** – Recall that  $W_{n,\bullet}(\gamma) = F_{1,n} \times F_{2,n}$  where  $F_{1,n} = \exp(Z_n)$  with  $Z_n := \ln t_{n,\bullet}(\gamma)(\widehat{\xi}_n - \xi)/\xi$  and

$$F_{2,n} := \left(\frac{1 - F\left(Q(\beta_n)(t_{n,\bullet}(\gamma))^{-\widehat{\xi}_n}\right)}{\beta_n(t_{n,\bullet}(\gamma))^{\widehat{\xi}_n/\xi}}\right).$$

Let us first focus on the factor  $F_{1,n}$ . Since  $\ln t_{n,\bullet} \sim \ln(\alpha_n/\beta_n)$  and, by assumption,  $\sigma_n^{-1}(\widehat{\xi}_n - \xi)/\xi \xrightarrow{d} \mathcal{D}$ , one has that  $Z_n/[\sigma_n \ln(\alpha_n/\beta_n)] = \mathcal{O}_{\mathbb{P}}(1)$ , where  $\mathcal{O}_{\mathbb{P}}(1)$  represents a stochastically bounded random variable. Since  $\sigma_n \ln(\alpha_n/\beta_n) \rightarrow 0$ , we thus have that  $Z_n = \mathcal{O}_{\mathbb{P}}(\sigma_n \ln(\alpha_n/\beta_n)) = o_{\mathbb{P}}(1)$ . A Taylor expansion of the exponential function leads to

$$F_{1,n} = 1 + Z_n(1 + o_{\mathbb{P}}(1)). \quad (13)$$

Let us now deal with the factor  $F_{2,n}$ . Applying Lemma 3 with  $t_n = (\alpha_n/\beta_n)^{-\widehat{\xi}_n}$  leads to

$$F_{2,n} = 1 - \frac{1}{\xi\rho} A(\alpha_n^{-1})(1 + o(1)), \quad (14)$$

where  $o(1)$  is the generic notation for a sequence converging to 0. Multiplying (13) and (14), we obtain

$$W_{n,\bullet}(\gamma) = 1 + Z_n \left(1 - \frac{1}{\xi\rho} Z_n^{-1} A(\alpha_n^{-1})(1 + o(1)) + o_{\mathbb{P}}(1)\right).$$

Using again the facts that  $\ln t_{n,\bullet} \sim \ln(\alpha_n/\beta_n)$  and  $\sigma_n^{-1}(\widehat{\xi}_n - \xi)/\xi \xrightarrow{d} \mathcal{D}$ , one has that  $Z_n^{-1} = \mathcal{O}_{\mathbb{P}}(\sigma_n^{-1}(\ln(\alpha_n/\beta_n))^{-1})$ . Remarking that  $A(\alpha_n^{-1})/\sigma_n \ln(\alpha_n/\beta_n) \rightarrow 0$ , we have  $Z_n^{-1} A(\alpha_n^{-1}) = o_{\mathbb{P}}(1)$  and the proof is complete.  $\blacksquare$

In the last lemma, we propose an example of sequences  $\alpha_n$ ,  $\beta_n$  and  $k_n$  satisfying the conditions of Proposition 3 and that can be used to construct the bias-reduced confidence interval (9).

**Lemma 5** *Let  $m_n$  be a positive sequence such that  $m_n \rightarrow \infty$  and  $m_n(\ln(m_n))^2 = o(n)$ . The conditions of Proposition 3 are satisfied by taking  $\beta_n = c/n$  with  $c > 0$ ,  $\alpha_n = \delta_n/n$  and  $k_n = d \max((\ln(n))^2, m_n(\ln(m_n))^2)$  for some  $d > 2/\pi$ .*

**Proof** – First note that  $\alpha_n \rightarrow 0$ ,  $n\alpha_n = m_n \rightarrow \infty$  and that  $\beta_n/\alpha_n \rightarrow \infty$ . Since  $t_n(0) \sim \alpha_n/\beta_n \rightarrow \infty$ , one has for  $n$  large enough that  $t_n(0) > 1$ . Next, since  $\ln(\beta_n) \sim \ln(n)$ ,  $[1 - \ln(\beta_n)]^2 \sim (\ln(n))^2$  and  $1 - \beta_n t_{n,R}(\gamma) \rightarrow 1$ , it is readily seen that condition (11) is satisfied since

$$k_n / \max \left( \frac{(\ln(\beta_n))^2}{(1 - \beta_n t_{n,R}(\gamma))^2} ; [1 + \ln(\beta_n^{-1})]^2 \right) \sim \frac{k_n}{(\ln(n))^2} \geq d > \frac{2}{\pi}.$$

Let us now focus on condition (12). One has that  $\ln t_n(0) \sim \ln(\alpha_n/\beta_n) = \ln(m_n) - \ln(c) \sim \ln(m_n)$ . Moreover, a Taylor expansion together with the use of [13, Lemma 1] entail that for  $n$  large enough, there are constants  $0 < c_1 < c_2$  such that

$$m_n^{1/2} \left| \frac{t_{n,\bullet}(\gamma)}{t_n(0)} - 1 \right| \in [c_1, c_2].$$

Hence, for  $n$  large enough, there exists constants  $0 < c_3 < c_4$  such that

$$m_n(\ln(m_n))^2 \frac{\ln t_n(0)}{1 - t_{n,\bullet}(\gamma)/t_n(0)} \in [c_3, c_4].$$

Since  $k_n/[m_n(\ln(m_n))^2] \geq d > 2/\pi$ , condition (12) is checked and the proof is complete. ■

## Proof of main results

**Proof of Theorem 1** – Let us prove that

$$\mathbb{P} \left( X_{n-\lfloor n\alpha_n \rfloor, n} (t_{n,L}(\gamma))^{\hat{\xi}_n} \leq Q(\beta_n) \right) \rightarrow \frac{1+\gamma}{2},$$

the proof of

$$\mathbb{P} \left( X_{n-\lfloor n\alpha_n \rfloor, n} (t_{n,R}(\gamma))^{\hat{\xi}_n} \geq Q(\beta_n) \right) \rightarrow \frac{1-\gamma}{2},$$

being similar. Let  $U_i := 1 - F(X_i)$  for  $i \in \{1, \dots, n\}$ . Since  $F$  is a continuous and strictly increasing function,

$$\mathbb{P}\left(X_{n-\lfloor n\alpha_n \rfloor, n}(t_{n,L}(\gamma))^{\widehat{\xi}_n} \leq Q(\beta_n)\right) = \mathbb{P}\left(U_{\lfloor n\alpha_n \rfloor+1, n} \geq Z_n(\widehat{\xi}_n)\right),$$

where

$$Z_n(\widehat{\xi}_n) = 1 - F\left(Q(\beta_n)(t_{n,L}(\gamma))^{-\widehat{\xi}_n}\right).$$

By assumption,  $\sigma_n^{-1}(\widehat{\xi}_n - \xi) = \mathcal{O}_{\mathbb{P}}(1)$  that is to say that for all  $\varepsilon > 0$ , there exists  $c_\varepsilon > 0$  such that for  $n$  large enough,  $\mathbb{P}(\sigma_n^{-1}|\widehat{\xi}_n - \xi| > c_\varepsilon) < \varepsilon$ . We have the decomposition

$$\begin{aligned} \mathbb{P}\left(U_{\lfloor n\alpha_n \rfloor+1, n} \geq Z_n(\widehat{\xi}_n)\right) &= \mathbb{P}\left(\{U_{\lfloor n\alpha_n \rfloor+1, n} \geq Z_n(\widehat{\xi}_n)\} \cap \{|\widehat{\xi}_n - \xi| > \sigma_n c_\varepsilon\}\right) \\ &+ \mathbb{P}\left(\{U_{\lfloor n\alpha_n \rfloor+1, n} \geq Z_n(\widehat{\xi}_n)\} \cap \{|\widehat{\xi}_n - \xi| \leq \sigma_n c_\varepsilon\}\right) \\ &=: T_{1,n} + T_{2,n}. \end{aligned}$$

For  $n$  large enough,

$$0 < T_{1,n} < \varepsilon. \quad (15)$$

To deal with the second term, the definition of  $t_{n,L}(\gamma)$  and the fact that  $U_{\lfloor n\alpha_n \rfloor+1, n}$  is distributed as a Beta distribution with parameters  $\lfloor n\alpha_n \rfloor + 1$  and  $n - \lfloor n\alpha_n \rfloor$ , entail that

$$\mathbb{P}(U_{\lfloor n\alpha_n \rfloor+1, n} \geq \beta_n t_{n,L}(\gamma)) = \frac{1 - \gamma}{2}.$$

Hence, from Lemma 2,

$$t_{n,L}(\gamma) \sim \alpha_n / \beta_n, \quad (16)$$

and thus, for  $n$  large enough,

$$(1 - \varepsilon) \ln\left(\frac{\alpha_n}{\beta_n}\right) \leq \ln(t_{n,L}(\gamma)) \leq (1 + \varepsilon) \ln\left(\frac{\alpha_n}{\beta_n}\right).$$

As a consequence, when the event  $\{|\widehat{\xi}_n - \xi| \leq \sigma_n c_\varepsilon\}$  occurs,

$$\exp\left(-c_\varepsilon(1 - \varepsilon)\sigma_n \ln\left(\frac{\alpha_n}{\beta_n}\right)\right) \leq (t_{n,L}(\gamma))^{\xi - \widehat{\xi}_n} \leq \exp\left(c_\varepsilon(1 + \varepsilon)\sigma_n \ln\left(\frac{\alpha_n}{\beta_n}\right)\right).$$

Since by assumption,  $\sigma_n \ln(\alpha_n / \beta_n) \rightarrow 0$ , one has for  $n$  large enough

$$1 - 2c_\varepsilon\sigma_n \ln\left(\frac{\alpha_n}{\beta_n}\right) \leq (t_{n,L}(\gamma))^{\xi - \widehat{\xi}_n} \leq 1 + 2c_\varepsilon\sigma_n \ln\left(\frac{\alpha_n}{\beta_n}\right),$$

and thus  $1 - F(Q(\beta_n)t_{n,+}(\xi)) \leq Z_n(\widehat{\xi}_n) \leq 1 - F(Q(\beta_n)t_{n,-}(\xi))$ , where

$$t_{n,\pm}(\xi) := (t_{n,L}(\gamma))^{-\xi} \left( 1 \pm 2c_\varepsilon \sigma_n \ln \left( \frac{\alpha_n}{\beta_n} \right) \right).$$

Applying Lemma 3 with  $t_n = t_{n,\pm}(\xi)$  leads to

$$\lim_{n \rightarrow \infty} \frac{1}{|A(\alpha_n^{-1})|} \left| \frac{1 - F(Q(\beta_n)t_{n,\pm}(\xi))}{(t_{n,\pm}(\xi))^{-1/\xi} \beta_n} - 1 \right| = -\frac{1}{\xi \rho}.$$

Using this convergence, when the event  $\{|\widehat{\xi}_n - \xi| \leq \sigma_n c_\varepsilon\}$  occurs, one has for  $n$  large enough that

$$\beta_n [t_{n,+}(\xi)]^{-1/\xi} \left( 1 + |A(\alpha_n^{-1})| \frac{1+\varepsilon}{\rho \xi} \right) \leq Z_n(\widehat{\xi}_n) \leq \beta_n [t_{n,-}(\xi)]^{-1/\xi} \left( 1 - |A(\alpha_n^{-1})| \frac{1+\varepsilon}{\rho \xi} \right).$$

As a consequence, for  $n$  large enough,

$$p_{n,-} - \varepsilon \leq T_{2,n} \leq p_{n,+}, \quad (17)$$

where

$$p_{n,\pm} := \mathbb{P} \left( U_{\lfloor n\alpha_n \rfloor + 1, n} \geq \beta_n [t_{n,\pm}(\xi)]^{-1/\xi} \left( 1 \pm |A(\alpha_n^{-1})| \frac{1+\varepsilon}{\rho \xi} \right) \right).$$

Since  $U_{\lfloor n\alpha_n \rfloor + 1, n}$  is distributed as a Beta distribution with parameters  $\lfloor n\alpha_n \rfloor + 1$  and  $n - \lfloor n\alpha_n \rfloor$  and since

$$\begin{aligned} & \beta_n [t_{n,\pm}(\xi)]^{-1/\xi} \left( 1 \pm |A(\alpha_n^{-1})| \frac{1+\varepsilon}{\rho \xi} \right) \\ &= F_{\text{beta}}^{\leftarrow} \left( \frac{1-\gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right) \left( 1 \pm 2c_\varepsilon \sigma_n \ln \left( \frac{\alpha_n}{\beta_n} \right) \right) \left( 1 \pm |A(\alpha_n^{-1})| \frac{1+\varepsilon}{\rho \xi} \right), \end{aligned}$$

one has

$$p_{n,\pm} = 1 - F_{\text{beta}} \left( F_{\text{beta}}^{\leftarrow} \left( \frac{1-\gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right) (1 + \delta_{n,\pm}); \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right),$$

where

$$\delta_{n,\pm} = \left( 1 \pm 2c_\varepsilon \sigma_n \ln \left( \frac{\alpha_n}{\beta_n} \right) \right) \left( 1 \pm |A(\alpha_n^{-1})| \frac{1+\varepsilon}{\rho \xi} \right) - 1 = \mathcal{O} \left( \sigma_n \log \left( \frac{\alpha_n}{\beta_n} \right) \right) + \mathcal{O} \left( A(\alpha_n^{-1}) \right).$$

The notation  $\mathcal{O}(1)$  stands for an asymptotically bounded sequence. From the mean value theorem,

$$\begin{aligned} p_{n,\pm} &= \frac{1+\gamma}{2} + \delta_{n,\pm} F_{\text{beta}}^{\leftarrow} \left( \frac{1-\gamma}{2}, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor \right) \\ &\quad \times f_{\text{beta}}(\zeta_n, \lfloor n\alpha_n \rfloor + 1, n - \lfloor n\alpha_n \rfloor), \end{aligned}$$

where

$$\zeta_n = F_{\text{beta}}^{\leftarrow} \left( \frac{1-\gamma}{2}, [n\alpha_n] + 1, n - [n\alpha_n] \right) (1 + \theta\delta_{n,\pm}) \sim \alpha_n,$$

with  $\theta \in (0, 1)$ . From [13, Lemma 1]

$$f_{\text{beta}}(\zeta_n, [n\alpha_n] + 1, n - [n\alpha_n]) = \mathcal{O} \left( \frac{n^{1/2}}{\alpha_n^{1/2}} \right).$$

Since  $F_{\text{beta}}^{\leftarrow}((1-\gamma)/2, [n\alpha_n] + 1, n - [n\alpha_n]) \sim \alpha_n$ , we have  $p_{n,\pm} = (1+\gamma)/2 + \mathcal{O}(\delta_{n,\pm}(n\alpha_n)^{1/2})$ . Since by assumption

$$\delta_{n,\pm}(n\alpha_n)^{1/2} = \mathcal{O} \left( (n\alpha_n)^{1/2} \max \left( A(\alpha_n^{-1}); \sigma_n \ln \left( \frac{\alpha_n}{\beta_n} \right) \right) \right) = o(1),$$

we finally obtained

$$p_{n,\pm} = \frac{1+\gamma}{2} + o(1). \quad (18)$$

Gathering (17) and (18) entail that for  $n$  large enough,

$$\frac{1+\gamma}{2} - 2\varepsilon \leq T_{2,n} \leq \frac{1+\gamma}{2} + \varepsilon. \quad (19)$$

Collecting (15) and (19) conclude the proof.  $\blacksquare$

**Proof of Proposition 2** – Let us introduce the notations

$$\zeta_{1,n} := \sigma_n^{-1} (\widehat{\xi}_n - \xi) \quad \text{and} \quad \zeta_{2,n} := \varsigma_n^{-1} (\check{\xi}_n - \xi).$$

One has  $\widehat{s}_n = \text{sgn}(\widehat{\xi}_n - \xi) \text{sgn}(1 - \Delta_n)$ , where

$$\Delta_n := \frac{\varsigma_n \zeta_{2,n}}{\sigma_n \zeta_{1,n}}.$$

As a consequence,

$$\mathbb{P} \left( \widehat{s}_n = \text{sgn}(\widehat{\xi}_n - \xi) \right) = \mathbb{P}(\text{sgn}(1 - \Delta_n) = 1) = \mathbb{P}(\Delta_n < 1).$$

The conclusion is straightforward since, under the assumptions of Proposition 2,

$$\frac{\zeta_{2,n}}{\zeta_{1,n}} = \mathcal{O}_{\mathbb{P}}(1) \quad \text{and} \quad \frac{\varsigma_n}{\sigma_n} \rightarrow 0.$$

$\blacksquare$



**Proof of Proposition 3** – We focus on the equation (10) with  $\bullet = R$ . The case  $\bullet = L$  is very similar and is omitted for the sake of conciseness. For  $s \in \{-1, 1\}$  and  $d > 0$ , let us introduce the function  $f_{s,d} : [0, \infty) \rightarrow \mathbb{R}$  defined for all  $x \geq 0$  by  $f_{s,d}(x) = x(1 + sd \ln x)$ . Let us also introduce the sequence  $d_n := k_n^{-1/2} \sqrt{2/\pi}$  and remark that when  $u \in (0, 1)$ ,  $t_{n,R}(u) \in (t_n(0), \beta_n^{-1})$ . If we prove the existence and the uniqueness of the solution of the equation

$$f_{\widehat{s}_n^{(H)}, d_n}(x) = t_{n,R}(\gamma), \quad (20)$$

when  $x \in (t_n(0), \beta_n^{-1})$ , then necessarily equation (10) will also admit an unique solution given by  $\widehat{\gamma}_{n,R}^{(H)} = 2F_{\text{beta}}(\beta_n x_n^*, [n\alpha_n] + 1, n - [n\alpha_n]) - 1$ , where  $x_n^*$  is the solution of (20).

Let us start with the study of the function  $f_{s,d}$ . When  $s = 1$ , the function  $f_{s,d}$  is decreasing on  $(0, \exp(-(1 + d^{-1})))$  and increasing elsewhere. We also remark that  $f_{1,d}(x) \leq 0$  for all  $x \leq \exp(-(1 + d^{-1}))$  and that  $f_{s,d}(x) \rightarrow +\infty$  as  $x \rightarrow \infty$ . When  $s = -1$ , the function is increasing on  $(0, \exp(d^{-1} - 1))$  and decreasing elsewhere. Moreover,  $0 < f_{s,d}(x) \leq d \exp(d^{-1} - 1)$  for  $x \leq \exp(d^{-1} - 1)$  and  $f_{s,d}(x) \rightarrow -\infty$  as  $x \rightarrow \infty$ .

First, let us consider the situation when  $\widehat{s}_n^{(H)} = 1$  in equation (20). The inequality

$$k_n > \frac{2}{\pi} \frac{(\ln(t_n(0)))^2}{(1 - t_{n,R}(\gamma)/t_n(0))^2},$$

entails that  $f_{1,d_n}(t_n(0)) < t_{n,R}(\gamma)$ . Moreover, since  $t_n(0) > 1$ , one has  $f_{1,d_n}(\beta_n^{-1}) > \beta_n^{-1} > t_{n,R}(\gamma)$ . Hence, in view of the above study of the function  $f_{1,d}$ , the existence and uniqueness of the solution is proved.

Now, let us focus on the case  $\widehat{s}_n^{(H)} = -1$ . The inequality  $k_n > (2/\pi)[1 + \ln(\beta_n^{-1})]^2$  yields to  $\beta_n^{-1} < \exp(d_n^{-1} - 1)$ . Hence, the function  $f_{-1,d_n}$  is increasing on the interval  $(t_n(0), \beta_n^{-1})$ . Moreover, since  $t_n(0) > 1$ ,  $f_{-1,d_n}(t_n(0)) < t_n(0) < t_{n,R}(\gamma)$  and finally, the inequality

$$k_n > \frac{2}{\pi} \frac{(\ln(\beta_n))^{-2}}{(1 - \beta_n t_{n,R}(\gamma))^2},$$

entails that  $f_{-1,d_n}(\beta_n^{-1}) > t_{n,R}(\gamma)$  and the proof is complete.  $\blacksquare$

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9747	0.9733	0.9749
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9757	0.9746	0.9557
500	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9628	0.9587	0.9803
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9736	0.9727	0.9647
1000	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9610	0.9569	0.9789
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9631	0.9572	0.9770

Table 7: Empirical coverage probabilities for the left bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Student distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

## Additional results

Additional results on the one-sided confidence intervals are given below.

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9608	0.9730	0.9983
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9479	0.9834	0.9974
500	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9818	0.9933	0.9961
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9784	0.9926	0.9957
1000	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9900	0.9966	0.9928
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9900	0.9958	0.9943

Table 8: Empirical coverage probabilities for the right bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Student distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9770	0.9763	0.9643
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9771	0.9770	0.9644
500	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9697	0.9664	0.9735
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9742	0.9743	0.9657
1000	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9707	0.9677	0.9740
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9719	0.9699	0.9592

Table 9: Empirical coverage probabilities for the left bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Student distribution ( $\xi = 1$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9296	0.9618	0.9863
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.8892	0.9377	0.9840
500	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9543	0.9763	0.9836
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9295	0.9610	0.9758
1000	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9693	0.9835	0.9823
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9472	0.9700	0.9702

Table 10: Empirical coverage probabilities for the right bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Student distribution ( $\xi = 1$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9769	0.9761	0.9590
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9767	0.9767	0.9582
500	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9693	0.9664	0.9431
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9740	0.9741	0.9518
1000	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9683	0.9644	0.9355
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9701	0.9676	0.9370

Table 11: Empirical coverage probabilities for the left bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Fréchet distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9434	0.9697	0.9953
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9056	0.9474	0.9913
500	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9693	0.9836	0.9943
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9431	0.9712	0.9908
1000	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9818	0.9912	0.9945
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9605	0.9807	0.9890

Table 12: Empirical coverage probabilities for the right bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Fréchet distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9745	0.9731	0.9666
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9754	0.9743	0.9487
500	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9582	0.9526	0.9719
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9734	0.9725	0.9472
1000	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9538	0.9483	0.9756
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9587	0.9523	0.9608

Table 13: Empirical coverage probabilities for the left bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Burr distribution ( $\xi = 0.5$  and  $\rho = -\sqrt{0.5}$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9684	0.9788	0.9990
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9554	0.9870	0.9980
500	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9883	0.9949	0.9965
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9854	0.9956	0.9961
1000	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9955	0.9980	0.9951
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9948	0.9975	0.9957

Table 14: Empirical coverage probabilities for the right bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Burr distribution ( $\xi = 0.5$  and  $\rho = -\sqrt{0.5}$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9758	0.9744	0.9645
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9760	0.9751	0.9538
500	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9645	0.9609	0.9714
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9737	0.9734	0.9490
1000	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9639	0.9600	0.9754
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9663	0.9624	0.9512

Table 15: Empirical coverage probabilities for the left bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Burr distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9561	0.9752	0.9966
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9301	0.9709	0.9955
500	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9765	0.9898	0.9956
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9647	0.9844	0.9951
1000	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9879	0.9948	0.9942
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9795	0.9906	0.9932

Table 16: Empirical coverage probabilities for the right bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a Burr distribution ( $\xi = 0.5$  and  $\rho = -1$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9765	0.9755	0.9576
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9761	0.9758	0.9567
500	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9659	0.9622	0.9224
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9737	0.9736	0.9355
1000	$\bar{\gamma}_{N,L}(k_n^{(\text{opt})})$	0.9632	0.9594	0.8874
	$\bar{\gamma}_{N,L}(\lfloor n/5 \rfloor)$	0.9656	0.9631	0.8994

Table 17: Empirical coverage probabilities for the left bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a log-Gamma distribution ( $\xi = 0.5$  and  $\rho = 0$ ).

$n$	Coverage probability	Gardes and Maistre	Gardes and Maistre, bias reduced	Buitendag et al., new normal ridge
200	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9520	0.9748	0.9974
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9190	0.9558	0.9947
500	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9805	0.9919	0.9976
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9599	0.9794	0.9946
1000	$\bar{\gamma}_{N,R}(k_n^{(\text{opt})})$	0.9918	0.9963	0.9972
	$\bar{\gamma}_{N,R}(\lfloor n/5 \rfloor)$	0.9785	0.9896	0.9944

Table 18: Empirical coverage probabilities for the right bound (with an objective of  $(1 + \gamma)/2 = 0.975$ ) and a log-Gamma distribution ( $\xi = 0.5$  and  $\rho = 0$ ).