

Trimmed extreme value estimators for censored heavy-tailed data*

Martin Bladt
Hansjörg Albrecher

*Department of Actuarial Science,
Faculty of Business and Economics,
University of Lausanne,
CH-1015 Lausanne, Switzerland*

e-mail: martin.bladt@unil.ch; hansjoerg.albrecher@unil.ch

Jan Beirlant

*Department of Mathematics,
KU Leuven,
Celestijnenlaan 200B,
B-3001 Leuven,
Belgium*

e-mail: jan.beirlant@kuleuven.be

Abstract: We consider estimation of the extreme value index and extreme quantiles for heavy-tailed data that are right-censored. We study a general procedure of removing low importance observations in tail estimators. This trimming procedure is applied to the state-of-the-art estimators for randomly right-censored tail estimators. Through an averaging procedure over the amount of trimming we derive new kernel type estimators. Extensive simulation suggests that one of the new considered kernels leads to a highly competitive estimator against virtually any other available alternative in this framework. Moreover we propose an adaptive selection method for the amount of top data used in estimation based on the trimming procedure minimizing the asymptotic mean squared error. We also provide an illustration of this approach to simulated as well as to real-world MTPL insurance data.

Keywords and phrases: Heavy-tailed estimation, right-censored data, asymptotic distributions.

Received January 2021.

Contents

1	Introduction	3113
2	Trimmed estimators for ξ	3116
	2.1 Trimming tail estimators	3116
	2.2 Averaging and kernels	3117
	2.3 Quantile estimation	3119

*The authors acknowledge financial support from the Swiss National Science Foundation Project 200021_191984.

3 Asymptotic representations 3119
 4 Optimal choice of k when estimating ξ 3122
 5 Simulations 3124
 6 Insurance application: censored claims data vs ultimates 3126
 7 Conclusion 3131
 References 3135

1. Introduction

In recent years the problem of tail estimation with right-censored data has received considerable attention starting with Beirlant et al. (2007) and Einmahl et al. (2008), who considered the problem for different domains of attraction. Most efforts however have been dedicated to heavy-tailed distributions, with several papers being motivated by heavy-tailed insurance claim data with long development times of the claims, see e.g. Beirlant et al. (2010, 2016, 2018, 2019); Worms and Worms (2014, 2016, 2018); Ndao et al. (2014). The underlying model assumption here is that the random variable of interest X has a Pareto-type distribution function

$$F(x) = \mathbb{P}(X \leq x) = 1 - x^{-1/\xi} \ell(x), \quad \xi > 0, x > 1, \tag{1}$$

where ℓ is slowly varying at infinity:

$$\lim_{x \rightarrow \infty} \frac{\ell(tx)}{\ell(x)} = 1, \text{ for every } t > 1.$$

Worms and Worms (2019) discuss the analogous problem for Weibull-type distributions. Right-censoring for heavy-tailed distributions in a regression setting was discussed in Ndao et al. (2016); Dierckx et al. (2019); Goegebeur et al. (2019a); Stupfler (2016), while Goegebeur et al. (2019b) is on a bivariate extension. Stupfler (2019) discusses the case of dependent censoring.

Here we revisit the case of heavy-tailed data. More specifically, we consider the random right-censoring model, where the independent and identically distributed (i.i.d.) observations X_1, \dots, X_n of X may be preceded by censoring variables C_1, \dots, C_n , and it is known if that happens. One then observes

$$Z_i = \min\{X_i, C_i\}, \quad e_i = 1\{X_i \leq C_i\}, \quad i = 1, \dots, n,$$

where C_1, \dots, C_n is an i.i.d. sequence of censoring random variables, independent of the observations X_i . Here $1\{X_i \leq C_i\}$ denotes the indicator of the event $\{X_i \leq C_i\}$. In order to obtain non-degenerate and tractable identities, one assumes that also the censoring variables are Pareto-type distributed with distribution function

$$G(x) = \mathbb{P}(C_1 \leq x) = 1 - x^{-1/\xi_c} \ell_c(x), \quad \xi_c > 0, x > 1,$$

with ℓ_c another slowly varying function at infinity. This choice also motivates the asymptotic Pareto behaviour to be introduced in the sequel. Then we have that for $x > 1$,

$$H(x) = \mathbb{P}(Z_1 \leq x) = 1 - x^{-1/\xi_z} \ell_z(x), \quad \xi_z = \frac{\xi \xi_c}{\xi + \xi_c},$$

where $\ell_z(x) = \ell(x)\ell_c(x)$. As explained in Einmahl et al. (2008), the parameter $p = \xi_z/\xi = \frac{\xi_c}{\xi + \xi_c}$ is the limit of $p(z) := \mathbb{P}(e_1 = 1 | Z_1 = z)$ as $z \rightarrow \infty$, and can be interpreted as the non-censoring probability in the limit, or the tail limiting proportion of non-censored data. In the exact Pareto setting (i.e. ℓ and ℓ_c being constant) the censoring indicators e_1, \dots, e_n turn out to be i.i.d. Bernoulli(p) random variables, independent of Z_1, \dots, Z_n .

Within this censoring and regularly varying context, Beirlant et al. (2007) proposed a first estimator of ξ in the spirit of the classical Hill estimator. Concretely, define the order statistics of the observed sample as

$$Z_{1,n} \leq \dots \leq Z_{n,n},$$

and $e_{i,n}$ the corresponding censoring indicators, $i = 1, \dots, n$. Then the Hill estimator adapted for censoring is given by

$$H_k = \frac{\sum_{i=1}^k \log(Z_{n-i+1,n}/Z_{n-k,n})}{\sum_{i=1}^k e_{n-i+1,n}} = \frac{H_k^Z}{pk}, \quad 1 < k < n, \quad (2)$$

with

$$pk = \frac{1}{k} \sum_{i=1}^k e_{n-i+1,n}$$

the fraction of non-censored data in the top k observations, and

$$H_k^Z = \frac{1}{k} \sum_{i=1}^k \log(Z_{n-i+1,n}/Z_{n-k,n})$$

the classical Hill estimator (cf. Hill (1975)) based on the top k observations. Einmahl et al. (2008) showed that, under some regularity assumptions, H_k is consistent whatever the value of $p \in (0, 1)$:

$$H_k = \frac{H_k^Z}{pk} \rightarrow_p \frac{\xi_z}{p} = \xi,$$

as $k, n \rightarrow \infty$ and $k/n \rightarrow 0$. Moreover Einmahl et al. (2008) derived the asymptotic normality of H_k under general conditions.

Worms and Worms (2014) proposed an alternative generalization of the Hill estimator based on the fact that

$$\mathbb{E}(\log(Z/t) | Z > t) = \int_1^\infty \frac{\overline{F}(ut)}{\overline{F}(t)} \frac{1}{u} du \rightarrow \xi \text{ as } t \rightarrow \infty, \quad (3)$$

where $\bar{F}(x) = 1 - F(x)$. In the exact Pareto case, the above limit is an equality. Replacing \bar{F} with the Kaplan-Meier estimator

$$\widehat{\bar{F}}(x) = \prod_{Z_{j,n} \leq x} \left(\frac{n-j}{n-j+1} \right)^{e_{j,n}}, \tag{4}$$

for $\bar{F}(x)$ yields the estimator

$$H_k^W = \sum_{i=1}^k \frac{\widehat{\bar{F}}(Z_{n-i,n})}{\widehat{\bar{F}}(Z_{n-k,n})} \log(Z_{n-i+1,n}/Z_{n-i,n}), \tag{5}$$

which was shown to be consistent in Worms and Worms (2014). Observe that both (2) and (5) reduce to H_k^Z when there is no censoring. Based on simulation studies, see e.g. Beirlant et al. (2018), the estimator H_k^W is known to exhibit superior behaviour in comparison with H_k , especially with respect to bias. The mathematical treatment of this estimator turns out to be difficult and the asymptotic normality of H_k^W has only been derived in Beirlant et al. (2019) under light censoring, i.e. $\xi < \xi_c$ or $p > 1/2$, and some regularity conditions. Here we will propose a novel family of estimators on the basis of a trimming procedure that will exhibit competitive behaviour and for which the mathematical treatment is simpler. This then also yields a generalization to the censoring case of the classical class of kernel estimators introduced in Csörgő et al. (1985).

Before introducing the trimming procedure, we first propose simplified versions of H_k^W , that will be more amenable for the approach in the sequel. To this end, note that one can write

$$H_k^W = \sum_{i=1}^k \left[\prod_{j=i+1}^k (1 - 1/j)^{e_{n-j+1,n}} \right] \log(Z_{n-i+1,n}/Z_{n-i,n}). \tag{6}$$

In the exact Pareto case one has $p(z) = p$ and the term in the square bracket has expectation

$$\begin{aligned} \mathbb{E} \left[\prod_{j=i+1}^k (1 - 1/j)^{e_{n-j+1,n}} \right] &= \mathbb{E} \left[\prod_{j=i+1}^k (1 - 1/j)^{e_j} \right] \\ &= \prod_{j=i+1}^k \mathbb{E} [(1 - 1/j)^{e_j}] = \prod_{j=i+1}^k (1 - p/j), \end{aligned} \tag{7}$$

where we have used the fact that the $e_{n-j+1,n}$ are i.i.d. Bernoulli random variables with the same law as the e_j . Notice that for Pareto-type variables, this only holds asymptotically, as $k, n/k \rightarrow \infty$. The Rényi representation also entails for the exact Pareto case that

$$\mathbb{E}(V_j) = \frac{1}{\sum_{m=i}^k m^{-1}} \mathbb{E}(\log(Z_{n-j+1,n}/Z_{n-k,n})) = \xi_z, \tag{8}$$

with $V_j = j \log \frac{Z_{n-j+1,n}}{Z_{n-j,n}}$. Based on (7) and (8) and using the approximations $1 - p_k/j \approx \exp(-p_k/j)$ and $\sum_{j=i+1}^k j^{-1} \approx \log((k+1)/i)$, we propose the following estimator which is closely related to H_k^W :

$$H_k^A = \frac{1}{k+1} \sum_{i=1}^k \left(\frac{i}{k+1} \right)^{p_k-1} \frac{1}{\log((k+1)/i)} \log(Z_{n-i+1,n}/Z_{n-k,n}), \quad k < n, \quad (9)$$

where the log-spacings in the sum are all taken with respect to the same baseline order statistic $Z_{n-k,n}$. The latter will allow to apply the trimming operation of removing low importance observations in the tail estimation, developed in Bladt et al. (2020b) for the non-censoring case, to the present situation with censoring.

In this paper, we extend the trimming method proposed in Bladt et al. (2020b) to the case of random right censoring, both for H_k and H_k^A . Averaging the trimmed statistics over the amount of trimming then leads to new estimators which belong to a general family of kernel estimators comprising H_k and H_k^A . This family turns out to be closed under the proposed averaging operation after trimming. Then we study the basic asymptotic properties of the kernel estimators in Section 3, even in case of heavy censoring, i.e. $p \leq 1/2$. Trajectories of the trimmed statistics as a function of the amount of trimming turn out to be quite flat near the optimal threshold value minimizing the mean squared error (MSE). Based on this, in Section 4 – for the first time in this setting – an adaptive selection method for the amount of top data used in tail estimation is proposed. In Sections 5 and 6, we show through simulations and a case study from insurance that the new kernel estimators and the threshold selection method exhibit promising properties.

2. Trimmed estimators for ξ

2.1. Trimming tail estimators

In Bladt et al. (2020b), lower trimming of the classical Hill estimator was shown to be an effective strategy to obtain Hill-type plots with lower variance arising from the changes of the baseline order statistic, which aids in the visual selection of a horizontal part of the trajectory. Here, we extend this approach to the censored case, and consider lower trimming of the estimators H_k and H_k^A , deleting the smallest $k-b$ ($b \leq k$) peaks over thresholds $Z_{n-i+1,n}/Z_{n-k,n}$, $i = b+1, \dots, k$: trimming H_k^Z as in Bladt et al. (2020b) one obtains a trimmed version of H_k

$$H_{b,k} = \frac{1}{1 + \sum_{j=b+1}^k j^{-1}} \cdot \frac{\frac{1}{b} \sum_{i=1}^b \log(Z_{n-i+1,n}/Z_{n-k,n})}{p_k}, \quad (10)$$

for $b \leq k \leq n - 1$, while for H_k^A we propose

$$H_{b,k}^A = \frac{1}{b+1} \sum_{i=1}^b \left(\frac{i}{b+1} \right)^{p_k-1} \frac{1}{\log((k+1)/i)} \log(Z_{n-i+1,n}/Z_{n-k,n}), \quad (11)$$

since, when p_k is replaced by the exact value p , using (8) the expected value equals $\frac{\xi_z}{b+1} \sum_{i=1}^b \left(\frac{i}{b+1} \right)^{p-1} \frac{\sum_{m=i}^k m^{-1}}{\log((k+1)/i)} \approx \xi$ for b sufficiently large. Note also that $H_{k,k} = H_k$ and $H_{k,k}^A = H_k^A$.

2.2. Averaging and kernels

The above trimming procedure naturally leads to new estimators when considering the empirical mean of the trimmed estimators across $b = 1, \dots, k$:

$$\frac{1}{k} \sum_{b=1}^k H_{b,k}, \quad \frac{1}{k} \sum_{b=1}^k H_{b,k}^A.$$

For instance, in case of $H_{b,k}^A$ this is asymptotically equivalent to

$$\begin{aligned} \overline{H}_k^A = \frac{1}{k} \left\{ \sum_{i=1}^k \frac{1}{1-p_k} \left(\left(\frac{i}{k+1} \right)^{p_k-1} - 1 \right) \right. \\ \left. \times \frac{1}{\log((k+1)/i)} \log(Z_{n-i+1,n}/Z_{n-k,n}) \right\}, \quad k < n, \end{aligned} \quad (12)$$

as can be seen using partial summation and a simple Riemann sum approximation $\frac{1}{k} \sum_{b=i}^k \left(\frac{b}{k+1} \right)^{-p_k} \approx \frac{1}{1-p_k} \left(1 - \left(\frac{i}{k+1} \right)^{1-p_k} \right)$ for $k \rightarrow \infty$.

In fact H_k , H_k^A and \overline{H}_k^A can all be put into a kernel framework, by defining

$$H_k^{\mathcal{K}} = \frac{1}{k} \sum_{i=1}^k \mathcal{K} \left(\frac{i}{k+1}, p_k \right) \frac{1}{\log((k+1)/i)} \log(Z_{n-i+1,n}/Z_{n-k,n}), \quad (13)$$

where \mathcal{K} is a positive kernel function satisfying

$$\int_0^1 \mathcal{K}(u; p) \, du = \frac{1}{p}, \quad \text{for all } p \in (0, 1].$$

In particular, we get

$$\begin{aligned} H_k &= H_k^{\mathcal{K}_0}, \quad \text{with } \mathcal{K}_0(u, p) = \frac{1}{p} \log \left(\frac{1}{u} \right), \\ H_k^A &= H_k^{\mathcal{K}_1}, \quad \text{with } \mathcal{K}_1(u, p) = u^{p-1}, \\ \overline{H}_k^A &= H_k^{\mathcal{K}_2}, \quad \text{with } \mathcal{K}_2(u, p) = \frac{u^{p-1} - 1}{1-p}. \end{aligned}$$

Note that H_k^W does not fall into this framework, but its simplified version H_k^A does.

Also notice that, when trimming any kernel estimator $H_k^{\mathcal{K}}$ to obtain

$$H_{b,k}^{\mathcal{K}} = \frac{1}{b+1} \sum_{i=1}^b \mathcal{K} \left(\frac{i}{b+1}, p_k \right) \frac{1}{\log((k+1)/i)} \log(Z_{n-i+1,n}/Z_{n-k,n}), \quad (14)$$

the averaging operation $\frac{1}{k} \sum_{b=1}^k H_{b,k}^{\mathcal{K}}$ leads to an associated kernel estimator

$$H_k^{\bar{\mathcal{K}}} = \frac{1}{k} \sum_{i=1}^k \bar{\mathcal{K}} \left(\frac{i}{k+1}, p_k \right) \frac{1}{\log((k+1)/i)} \log(Z_{n-i+1,n}/Z_{n-k,n}), \quad (15)$$

with

$$\bar{\mathcal{K}}(u, p) = \int_u^1 \frac{\mathcal{K}(v, p)}{v} dv,$$

where $\bar{\mathcal{K}}(\frac{i}{k+1}, p)$ is obtained using a Riemann approximation of

$$\frac{1}{k} \sum_{b=i}^k \frac{k}{b+1} \mathcal{K} \left(\frac{i}{k+1} \frac{k+1}{b+1}, p \right),$$

as $k \rightarrow \infty$ for fixed i .

Rewriting the kernel estimators $H_k^{\mathcal{K}}$ from (13) in terms of the random variables V_j ($j = 1, \dots, k$) has some theoretical advantage, since for the exact Pareto case these are independent and exponentially distributed with mean ξ_z thanks to the Rényi representation:

$$\begin{aligned} H_k^{\mathcal{K}} &= \frac{1}{k} \sum_{i=1}^k \mathcal{K} \left(\frac{i}{k+1}, p_k \right) \frac{1}{\log((k+1)/i)} \sum_{j=i}^k \frac{V_j}{j} \\ &= \frac{1}{k} \sum_{j=1}^k \frac{V_j}{j} \sum_{i=1}^j \mathcal{K} \left(\frac{i}{k+1}, p_k \right) \frac{1}{\log((k+1)/i)} \\ &= \frac{1}{k} \sum_{j=1}^k V_j \tilde{\mathcal{K}}_k \left(\frac{j}{k+1}, p_k \right), \end{aligned}$$

with

$$\tilde{\mathcal{K}}_k \left(\frac{j}{k+1}, p_k \right) = \frac{1}{j/(k+1)} \frac{1}{k+1} \sum_{i=1}^j \mathcal{K} \left(\frac{i}{k+1}, p_k \right) \frac{1}{\log((k+1)/i)}.$$

Using a Riemann approximation, one can also propose to use

$$\tilde{H}_k^{\mathcal{K}} = \frac{1}{k} \sum_{j=1}^k \tilde{\mathcal{K}} \left(\frac{j}{k+1}, p_k \right) V_j, \quad (16)$$

with associated kernel

$$\tilde{\mathcal{K}}(u, p) = \frac{1}{u} \int_0^u \frac{\mathcal{K}(v, p)}{\log(1/v)} dv,$$

that also satisfies the norming $\int_0^1 \tilde{\mathcal{K}}(u, p) du = 1/p$. The class of estimators $\tilde{H}_k^{\mathcal{K}}$ can be considered as generalizations of the kernel estimators proposed in Csörgő et al. (1985) from the non-censoring to the censoring case.

2.3. Quantile estimation

Following the approach from Weissman (1978) it is possible to construct quantile estimators as a function of the sample size as follows. Let the quantile function of a regularly varying tail be $Q(p)$. The regular variation property implies that the ratio of increasingly large quantiles satisfies

$$\frac{Q(1-p)}{Q(1-k/n)} \sim \left(\frac{k}{np}\right)^\xi, \quad \text{as } p \downarrow 0, k/n \rightarrow 0, np = o(k), \quad (17)$$

as discussed in Section 4.3 in de Haan and Ferreira (2007). This then leads to a quantile estimator based on k order statistics and the kernel \mathcal{K} as

$$\hat{Q}_k(1-p) = \hat{Q}^{KM}(1-k/n) \cdot \left(\frac{k}{np}\right)^{H_k^{\mathcal{K}}}, \quad (18)$$

where \hat{Q}^{KM} is the quantile function derived from the Kaplan-Meier estimator \hat{F} defined in (4).

3. Asymptotic representations

In this section we derive the asymptotic distributions of the kernel estimators and their trimmed counterparts as introduced in the preceding section. In Einmahl et al. (2008) the asymptotics for $H_k = H_k^{\mathcal{K}_0}$ was discussed in detail. Beirlant et al. (2019) provided an asymptotic normality result for H_k^W when $p > 1/2$, but that estimator is not in the current kernel framework. Here we provide asymptotic representations for the class of kernel estimators in the form $\tilde{H}_k^{\mathcal{K}}$. To this end, we make use of second-order assumptions which were first proposed in Hall and Welsh (1985) and which have widely been used in papers on the estimation of the extreme value index for Pareto-type distributions both in the non-censoring case such as Csörgő et al. (1985) and the censoring case as in Beirlant et al. (2019):

$$\ell(x) = C(1 + Dx^{-\beta}(1 + o(1))), \quad \ell_c(x) = C_c(1 + D_c x^{-\beta_c}(1 + o(1))), \quad x \rightarrow \infty, \quad (19)$$

where β, β_c, C, C_c are positive constants and D, D_c are real constants. It now follows that

$$\ell_z(x) = C_z(1 + D_z x^{-\beta_z}(1 + o(1))),$$

where

$$C_z = CC_c, \quad \beta_z = \min\{\beta, \beta_c\}, \quad D_z = D \cdot 1_{\beta \leq \beta_c} + D_c \cdot 1_{\beta_c \leq \beta}.$$

Further we set $\rho_z = -\beta_z \xi_z$, $Q_{0,z}(t) = -\xi_z^2 \beta_z D_z C^{\rho_z} t^{\rho_z}$. Concerning the scaled spacings V_j , $j = 1, \dots, k$, one then has the following expansion as $n, k \rightarrow \infty$ and $k/n \rightarrow 0$ as given in Theorem 4.1 in Beirlant et al. (2004):

$$V_j = \left(\xi_z + Q_{0,z}(n/k) \left(\frac{j}{k+1} \right)^{-\rho_z} \right) E_j + R_{j,n} + o_p(Q_{0,z}(n/k)), \quad (20)$$

with E_j standard exponential random variables, independent with each n , and $|\sum_{j=i}^k R_{j,n}/j| = o_p(Q_{0,z}(n/k)) \max(\log((k+1)/i), 1)$. Next, from Einmahl et al. (2008) and Beirlant et al. (2016) one obtains that

$$\sqrt{k}(p_k - p) = \sqrt{p(1-p)}N(1 + o_p(1)) + \sqrt{k}Q_{0,z}(n/k) \frac{\kappa_z}{1 - \rho_z}(1 + o_p(1)), \quad (21)$$

where $N \sim N(0, 1)$ can be chosen appropriately independent of $\{V_1, V_2, \dots\}$, and $\kappa_z = -\frac{(D\xi)_z}{D_z \xi \xi_c}$, with $(D\xi)_z = (D\xi)1_{\beta \leq \beta_c} - (D_c \xi_c)1_{\beta_c \leq \beta}$. Based on (20) and (21) we now derive that

$$\begin{aligned} \tilde{H}_k^{\mathcal{K}} - \xi &= \frac{1}{k} \sum_{j=1}^k \left(\tilde{\mathcal{K}}\left(\frac{j}{k+1}, p_k\right) - \tilde{\mathcal{K}}\left(\frac{j}{k+1}, p\right) \right) V_j \\ &\quad + \left(\frac{1}{k} \sum_{j=1}^k \tilde{\mathcal{K}}\left(\frac{j}{k+1}, p\right) V_j - \xi \right) \\ &=: T_{1,k} + T_{2,k}. \end{aligned} \quad (22)$$

Using the mean value theorem, we have from (21) that

$$\begin{aligned} T_{1,k} &\sim_p \xi_z(p_k - p) \alpha_b^{\tilde{\mathcal{K}}} \\ &= \xi_z \alpha_b^{\tilde{\mathcal{K}}} \left(\sqrt{p(1-p)} \frac{N}{\sqrt{k}} + Q_{0,z}(n/k) \frac{\kappa_z}{1 - \rho_z}(1 + o(1)) \right), \end{aligned} \quad (23)$$

with $\alpha_b^{\tilde{\mathcal{K}}} = \frac{1}{k} \sum_{j=1}^k \frac{\partial \tilde{\mathcal{K}}}{\partial p}\left(\frac{j}{k+1}, p\right)$. Next, using (20),

$$\begin{aligned} T_{2,k} &= \xi_z \left(\frac{1}{k} \sum_{j=1}^k \tilde{\mathcal{K}}_k\left(\frac{j}{k+1}, p\right) - \frac{1}{p} \right) \\ &\quad + \xi_z \frac{1}{k} \sum_{j=1}^k \tilde{\mathcal{K}}_k\left(\frac{j}{k+1}, p\right) (E_j - 1) \end{aligned}$$

$$\begin{aligned}
& + Q_{0,z}(n/k) \frac{1}{k} \sum_{j=1}^k \tilde{\mathcal{K}}_k\left(\frac{j}{k+1}, p\right) \left(\frac{j}{k+1}\right)^{-\rho_z} E_j \\
& + \frac{1}{k} \sum_{j=1}^k R_{j,n} \tilde{\mathcal{K}}_k\left(\frac{j}{k+1}, p\right) + o_p(Q_{0,z}(n/k)). \tag{24}
\end{aligned}$$

Concerning the second last term in (24) we find that

$$\begin{aligned}
\frac{1}{k} \sum_{j=1}^k R_{j,n} \tilde{\mathcal{K}}_k\left(\frac{j}{k+1}, p\right) &= \frac{1}{k} \sum_{j=1}^k \frac{R_{j,n}}{j} \sum_{i=1}^j \frac{\mathcal{K}\left(\frac{i}{k+1}, p\right)}{\log((k+1)/i)} \\
&= \frac{1}{k} \sum_{i=1}^k \frac{\mathcal{K}\left(\frac{i}{k+1}, p\right)}{\log((k+1)/i)} \sum_{j=i}^k \frac{R_{j,n}}{j} = o_p(Q_{0,z}(n/k)).
\end{aligned}$$

From (23) and (24) we can now state an asymptotic expansion for $H_k^{\mathcal{K}} - \xi$.

Theorem 3.1. *Under (19) we have as $k, n \rightarrow \infty$ and $k/n \rightarrow 0$*

$$\begin{aligned}
H_k^{\mathcal{K}} - \xi &= -\sqrt{p(1-p)} \frac{\xi_z}{p^2} \frac{N}{\sqrt{k}} + \xi_z \frac{1}{k} \sum_{j=1}^k \tilde{\mathcal{K}}_k\left(\frac{j}{k+1}, p\right) (E_j - 1) \\
&+ Q_{0,z}(n/k) \left\{ \frac{-\kappa_z \xi_z}{p^2(1-\rho_z)} + \int_0^1 u^{-\rho_z} \tilde{\mathcal{K}}(u, p) du \right\} (1 + o(1)) \\
&+ o_p(Q_{0,z}(n/k)) + \xi_z \left(\frac{1}{k} \sum_{j=1}^k \tilde{\mathcal{K}}_k\left(\frac{j}{k+1}, p\right) - \frac{1}{p} \right), \tag{25}
\end{aligned}$$

where N and $\{E_j, j \geq 1\}$ are introduced respectively in (21) and (20).

In order to select an optimal k , we minimize then the following asymptotic mean squared error of $H_k^{\mathcal{K}}$:

$$\text{AMSE}(H_k^{\mathcal{K}}) = \xi_z^2 v_{k,p} + Q_{0,z}^2\left(\frac{n}{k}\right) b_{k,p}, \tag{26}$$

with

$$\begin{aligned}
v_{k,p} &= \frac{1}{k} \frac{1-p}{p^3} + \frac{1}{k^2} \sum_{j=1}^k \tilde{\mathcal{K}}_k^2\left(\frac{j}{k+1}, p\right), \\
b_{k,p} &= \left\{ \frac{-\kappa_z \xi_z}{p^2(1-\rho_z)} + \int_0^1 u^{-\rho_z} \tilde{\mathcal{K}}(u, p) du \right\}^2.
\end{aligned}$$

Remark 3.2. Note that the first two terms in (25) concern the estimation of p and ξ , respectively. The third term can be regarded as a bias term arising from the second order assumption, which commonly appears in classical extreme value theory. The fourth and final term is proportional to ξ_z and can be regarded as a discretization error term, since the sum inside the parenthesis is a Riemann approximation to $\int_0^1 \tilde{\mathcal{K}}_k(u, p) du = 1/p$. In general, this error is small but non-zero.

Examples. Let us consider the expressions of $v_{k,p}$ and $b_{k,p}$ for some of the kernels \mathcal{K} considered above.

1. For the adapted Hill estimator H_k with $\mathcal{K}_0(u, p) = \log(1/u)/p$ we have that $\tilde{\mathcal{K}}_0(u, p) = 1/p$ and $\frac{1}{k^2} \sum_{j=1}^k \tilde{\mathcal{K}}_{0,k}^2(\frac{j}{k+1}, p) \sim k^{-1} \int_0^1 \tilde{\mathcal{K}}_0^2(v, p) dv = (kp^2)^{-1}$ as $k \rightarrow \infty$ for all $p \in (0, 1)$ so that $v_{k,p} = \frac{1}{k} p^{-3}$. Moreover $\int_0^1 u^{-\rho_z} \tilde{\mathcal{K}}_0(u, p) du = \frac{1}{p} (1 - \rho_z)^{-1}$. In case $\beta \leq \beta_c$, i.e. when the bias is largest compared with the classical Hill estimator in case of no censoring, we have that $-\kappa_z = \xi_c^{-1}$ and then $b_{k,p} = p^{-4} (1 - \rho_z)^{-2}$. Hence, under $\beta \leq \beta_c$,

$$\text{AMSE}(H_k) = \frac{1}{p^4} \left(p \frac{\xi_z^2}{k} + \frac{Q_{0,z}^2(n/k)}{(1 - \rho_z)^2} \right). \tag{27}$$

2. The estimator $\overline{H}_k^A = H_k^{\mathcal{K}_2}$ with $\mathcal{K}_2(u, p) = \frac{u^{p-1}-1}{1-p}$ exhibits excellent finite sample behaviour as shown below through the simulations. Here $\tilde{\mathcal{K}}_2(u, p) \sim \frac{1-p^{-1}u^{p-1}-1}{1-p \log(1/u)}$ as $u \rightarrow 0$.

For the bias we have $\int_0^1 u^{-\rho_z} \tilde{\mathcal{K}}_2^2(u, p) du = \int_0^1 k_{1-p}(1/v) k_{\rho_z}(1/v) \frac{dv}{\log(1/v)}$ with $k_a(w) = (w^a - 1)/a$ for any $a > 0$. As $k \rightarrow \infty$

$$b_{k,p} \sim b_{2,p} := \left\{ \frac{-\kappa_z \xi_z}{p^2 (1 - \rho_z)} + \int_0^1 k_{1-p}(1/v) k_{\rho_z}(1/v) \frac{dv}{\log(1/v)} \right\}^2.$$

Only in case of weak censoring, i.e. $p \geq 1/2$, $\int_0^1 \tilde{\mathcal{K}}_2^2(u, p) du < \infty$ as $k \rightarrow \infty$, so that then $\frac{1}{k^2} \sum_{j=1}^k \tilde{\mathcal{K}}_{2,k}^2(\frac{j}{k+1}, p) \sim k^{-1} \int_0^1 \tilde{\mathcal{K}}_2^2(u, p) du$. Hence as $\sqrt{k} Q_{0,z}(n/k) \rightarrow \lambda \in \mathbb{R}$

$$\sqrt{k} (\overline{H}_k^A - \xi) \rightarrow_d \mathcal{N}(\lambda b_{2,p}, \sigma_{2,p}^2),$$

with $\sigma_{2,p}^2 = (1-p)/p^3 + \int_0^1 \tilde{\mathcal{K}}_2^2(u, p) du$.

Under heavy censoring, i.e. $p < 1/2$,

$$\lim_{k \rightarrow \infty} k^{2p} (\log k)^2 \left(\frac{1}{k^2} \sum_{j=1}^k \tilde{\mathcal{K}}_{2,k}^2(\frac{j}{k+1}, p) \right) = \sigma_{h,2,p}^2 < \infty,$$

so that, as $k^p (\log k) Q_{0,z}(n/k) \rightarrow \lambda \in \mathbb{R}$,

$$k^p (\log k) (\overline{H}_k^A - \xi) \rightarrow_d \mathcal{N}(\lambda b_{2,p}, \sigma_{h,2,p}^2).$$

4. Optimal choice of k when estimating ξ

Denoting the trimmed version of the Hill estimator in the fully observed case by $H_{b,k}^Z = H_{b,k} p_k$, it was shown in Bladt et al. (2020b) that the value $k_{\text{opt}}(H_k^Z)$ of k minimizing the asymptotic MSE of H_k^Z satisfies

$$k_{\text{opt}}(H_k^Z) = \left(\frac{K}{(1 - \rho_z)^2 f(\rho_z)} \right)^{\frac{-1}{1-2\rho_z}} k_{\text{opt}}(H_{b,k}^Z),$$

for a universal constant K and a specific function f . Here $k_{\text{opt}}(H_{b,k}^Z)$ is the optimal sample fraction minimizing the expectation of the empirical variance $S_k^2 = \frac{1}{k} \sum_{b=1}^k (H_{b,k}^Z - \frac{1}{k} \sum_{b=1}^k H_{b,k}^Z)^2$.

On the other hand, from (27) we obtain that under $\beta \leq \beta_c$ that

$$k_{\text{opt}}(H_k) = \left(\frac{p^{-1} K}{(1 - \rho_z)^2 f(\rho_z)} \right)^{-\frac{1}{1-2\rho_z}} k_{\text{opt}}(H_{b,k}^Z). \quad (28)$$

This means that the optimal k for the estimator H_k with respect to minimization of the AMSE is linked to the optimal k of its trimmed versions for the minimization of the expected empirical variance in the non-censored case. A consequence of the above formula is that

$$k_{\text{opt}}(H_k) = p^{\frac{1}{1-2\rho_z}} k_{\text{opt}}(H_k^Z).$$

That is, a larger percentage of censoring leads to a higher threshold, when compared to the non-censored case. This can already be seen from the expression of the AMSE given in (27), where a smaller p leads to more weight being given to the bias term. From an intuitive point of view, when dealing with censored datasets, two sources of bias have to be accounted for, and hence a smaller sample fraction k is needed to control them.

In practice, for a given sample, one finds an estimate $\hat{k}_0 = \hat{k}_{\text{opt}}(H_{b,k}^Z)$ of $k_{\text{opt}}(H_{b,k}^Z)$ through minimization of S_k^2 over k , from which an adaptive choice of k is found through

$$\left(\frac{p_{\hat{k}_0}^{-1} K}{(1 - \hat{\rho}_z)^2 f(\hat{\rho}_z)} \right)^{-\frac{1}{1-2\hat{\rho}_z}} \hat{k}_0, \quad (29)$$

using an estimate $\hat{\rho}_z$ of ρ_z and replacing p by $p_{\hat{k}_0}$. Estimators of the second-order parameter ρ_z have been proposed for instance in Fraga Alves et al. (2003). Estimators exhibit a high variability and many authors consider the use of a fixed value for ρ_z such as $\rho_z = -1$. In the next section we use the choices $\rho_z = -1, -1/2, -3/2$, but the results are not very sensitive to this parameter, and we here also propose to stick to the choice $\rho_z = -1$.

Remark 4.1. For kernels different from \mathcal{K}_0 , if we restrict to the case $\beta \leq \beta_c$ and $p > 1/2$, and consider the expressions $kv_{k,p} =: \tilde{v}_{k,p}$ and $b_{k,p}$ taken from (26) as constant in k (i.e. converging fast enough to its limit for $k \rightarrow \infty$) we find from (26) that the optimal k_{opt} for any kernel equals

$$\begin{aligned} & \left(-\frac{1}{2\rho_z} \xi_z^2 / M^2 \right)^{1/(1-2\rho_z)} n^{-2\rho_z/(1-\rho_z)} \left(\frac{\tilde{v}_{k,p}}{b_{k,p}} \right)^{1/(1-\rho_z)} \\ & = k_{\text{opt}}(H_k) \left(\frac{\tilde{v}_{k,p}}{b_{k,p}(1-\rho_z)^2} \right)^{1/(1-\rho_z)}, \end{aligned}$$

with $M = -\xi_z^2 \beta_z D_z C^{\rho_z}$, from which one can deduce that

$$k_{\text{opt}}^{\mathcal{K}} = \left(\frac{\tilde{v}_{k,p}}{b_{k,p}} \right)^{1/(1-\rho_z)} \left(\frac{K}{f(\rho_z) p_{\hat{k}_0}} \right)^{-1/(1-\rho_z)} \hat{k}_0.$$

This is then basically the same formula as for H_k , (29), but one has to calculate $\tilde{v}_{k,p}$ and $b_{k,p}$ at every k with p estimated by $p_{\hat{k}_0}$. In heavy censoring cases the rates of convergence of the estimators are different and the approach becomes more involved. This makes the procedure significantly more computer-intensive. In practical applications, however, the difference between the optima across kernels and heavy and light censoring cases is rather small and does not justify the extra calculations.

5. Simulations

We performed simulations using the following distributions.

- Burr distribution with survival function $1 - F(x) = \left(\frac{\theta}{\theta + x^\beta}\right)^\lambda$ with (θ, β, λ) taken as (10,2,1) for X and (10,3,1) for C so that $p < 1/2$, next to (10,3,1) for X and (10,2,1) for C so that $p > 1/2$, and (10,2,1) for both X and C with $p = 1/2$.
- Fréchet distribution with $F(x) = \exp(-x^{-1/\xi})$ with ξ taken as 1/2 for X and 1/4 for C and correspondingly $p < 1/2$, as 1/4 for X and 1/2 for C and correspondingly $p > 1/2$, and finally as $\xi = 1/4$ for both X and C , so that $p = 1/2$.
- Log-gamma distribution with density $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} (\log x)^{\alpha-1} x^{-\lambda-1}$ with (α, λ) taken as (3/2,2) for X and (3/2,4) for C so that $p < 1/2$, as (3/2,4) for X and (3/2,2) for C so that $p > 1/2$, and as (3/2,4) for X and C so that $p = 1/2$. Note that in this case the conditions (19) are not satisfied.

The results are based on 200 simulations of sample size $n = 200$ each. The first-order tail-determining parameters were chosen such that $1/\xi = 3/2, 2, 4$, which seem to be realistic magnitudes for insurance applications, as will be illustrated in the next section. The remaining parameters were chosen in order to satisfy inequalities such as $p > 1/2$ or $p \leq 1/2$ but are otherwise arbitrary.

In Figures 1, 2 and 3 we plot the bias, variance and mean squared error as a function of k of the various estimators considered above. Observe that the plots are in logarithmic scale for display purposes. For the bias term this means that there is an asymptote corresponding to $\lim_{t \downarrow 0} \log(t)$. Note that the MSE characteristics of the estimator $H_k^{\mathcal{K}_2}$ are quite comparable to those of H_k^W in the Burr and Fréchet cases, and are even better for the log-gamma model. The corresponding analysis for the quantile estimators is given in Figures 4, 5 and 6, which are in agreement with the former plots.

In Figures 7, 8 and 9 we provide violin plots for the \mathcal{K}_0 and \mathcal{K}_2 -based estimator at the threshold selected according to the automatic procedure given in the previous section. We have taken $\rho_z = -1, -3/2, -1/2$ respectively for the three distributions that we consider. These values were permuted (the resulting plots

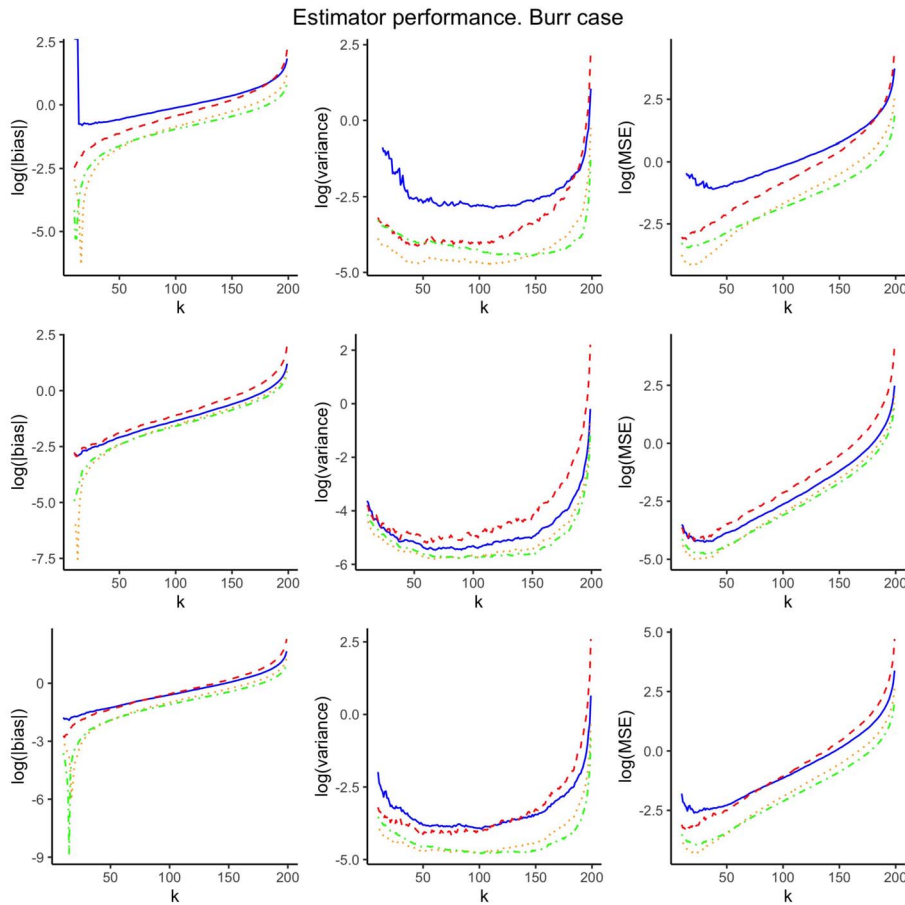


FIG 1. Burr distributions: bias, variance and mean squared error of the kernel estimator ($H_k = H_k^{\mathcal{K}_0}$ in solid blue, $H_k^A = H_k^{\mathcal{K}_1}$ in dashed red, $H_k^{\mathcal{K}_2}$ in dotted orange) and the Worms estimator H_k^W (dashed and dotted green), as a function of k . The top, middle and bottom levels correspond to $2p < 1$, $2p > 1$ and $2p = 1$, respectively.

are omitted) and the results were not very sensitive to the choice of ρ_z . To avoid degeneracies, a cutoff of $1/5$ of the size of the data set was used for the empirical variance estimates. We also add the results of the parameter estimates when taking k fixed at the theoretical optimal value. The latter theoretical optimum is only available in the light-censoring cases and for distributions properly belonging to the Fréchet domain of attraction (not the log-gamma case). A general observation is that the violin plots bundle together close to zero, as is desired. Looking closer, we observe that in the regularly varying setup, the adaptive selection of k together with the use of the kernels \mathcal{K}_0 and \mathcal{K}_2 comes very close to the performance of the Hill estimator evaluated at the theoretical optimum (essentially an oracle estimator, since we input the parameters of the simulated

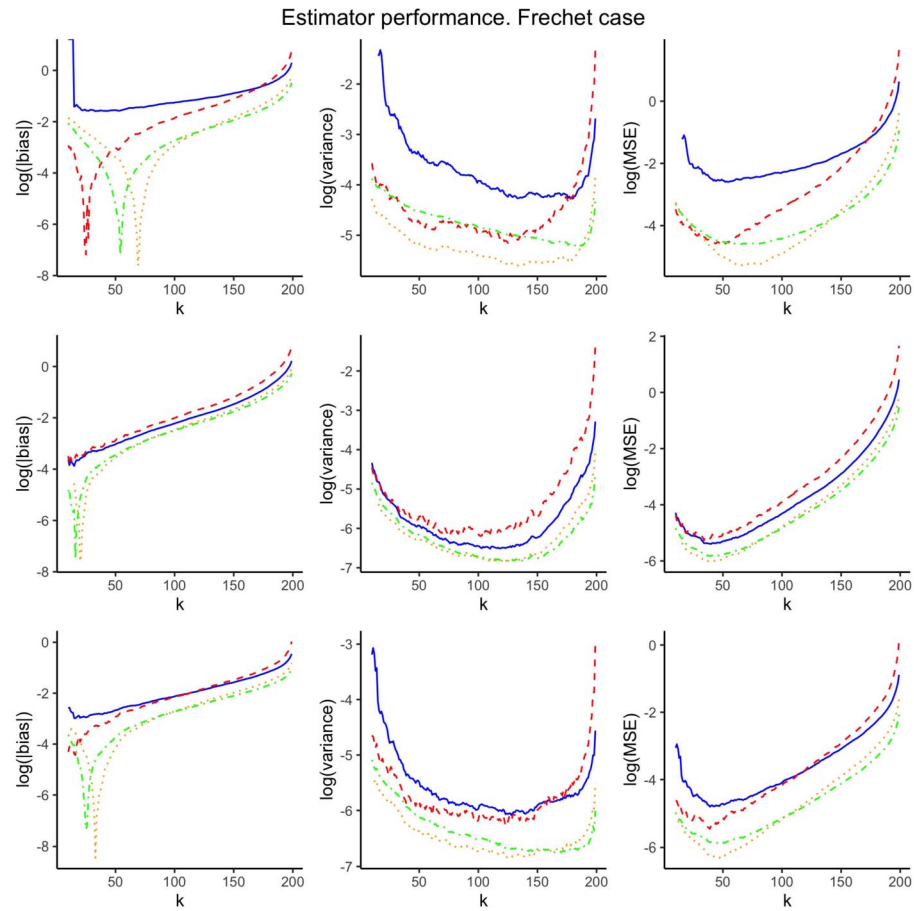


FIG 2. Fréchet distributions: bias, variance and mean squared error of the kernel estimator ($H_k = H_k^{K_0}$ in solid blue, $H_k^A = H_k^{K_1}$ in dashed red, $H_k^{K_2}$ in dotted orange) and the Worms estimator H_K^W (dashed and dotted green), as a function of k . The top, middle and bottom levels correspond to $2p < 1$, $2p > 1$ and $2p = 1$, respectively.

data into this theoretical optimum). As rough guidelines, we observe that the heavy censoring case has significantly worse behaviour than the light censoring case, and that the \mathcal{K}_2 -based estimator has the best behaviour, agreeing with the conclusions from Figures 1, 2 and 3. We believe that this regime-shift between light and heavy censoring is responsible for the increased number of outliers in cases where $p \neq 1/2$.

6. Insurance application: censored claims data vs ultimates

We now proceed to analyze an insurance dataset consisting of 837 motor third-party liability (MTPL) insurance claims from 1995 till 2010. This data set has

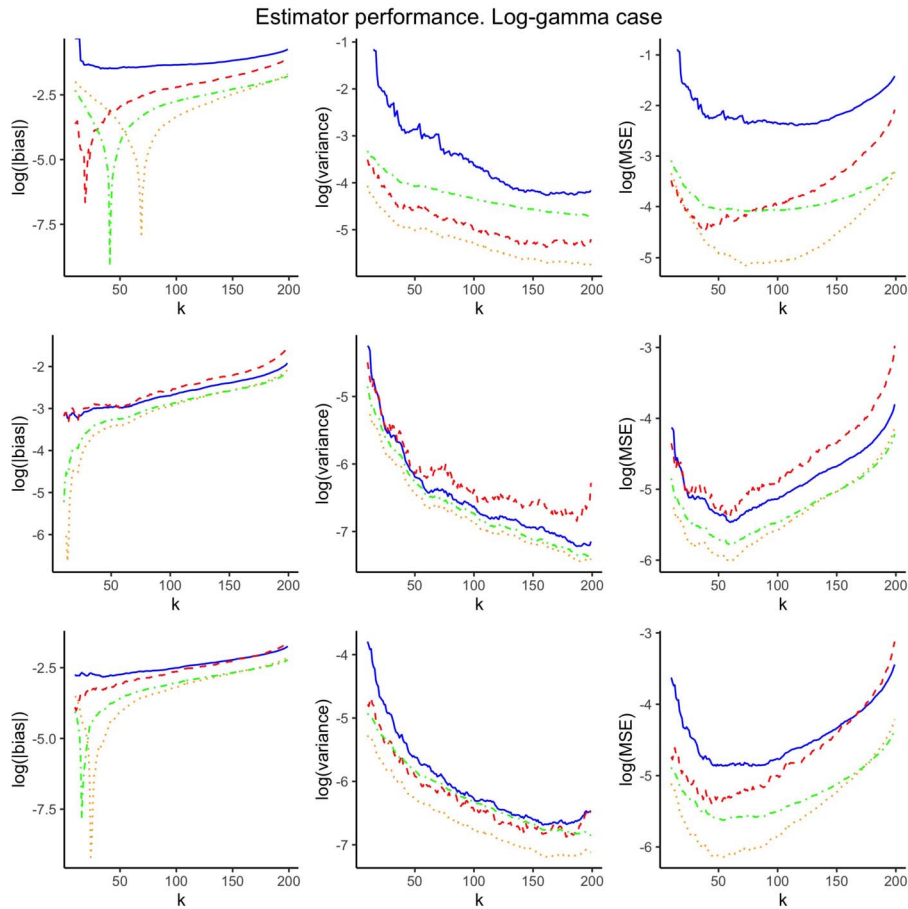


FIG 3. Log-gamma distributions: bias, variance and mean squared error of the kernel estimator ($H_k = H_k^{K_0}$ in solid blue, $H_k^A = H_k^{K_1}$ in dashed red, $H_k^{K_2}$ in dotted orange) and the Worms estimator K_k^W (dashed and dotted green), as a function of k . The top, middle and bottom levels correspond to $2p < 1$, $2p > 1$ and $2p = 1$, respectively.

been previously described and studied in Albrecher et al. (2017), Bladt et al. (2020b) (without censoring, using ultimate values instead) and Bladt et al. (2020a) (using both censoring and ultimate values).

The data exhibit right-censoring, that is, a claim size is partially observed whenever the development of the claim payment is ongoing and the claim is not yet closed. Closed claim sizes are thus considered as observed data points, and open claims are considered as right-censored observations. In Bladt et al. (2020a) it was argued that the assumption of random censoring and heavy-tailedness is adequate. In the top panel of Figure 10 we have the three kinds of data that are available (open claims, closed claims and ultimates), and on the bottom panel the survival function of the open and closed claims using the

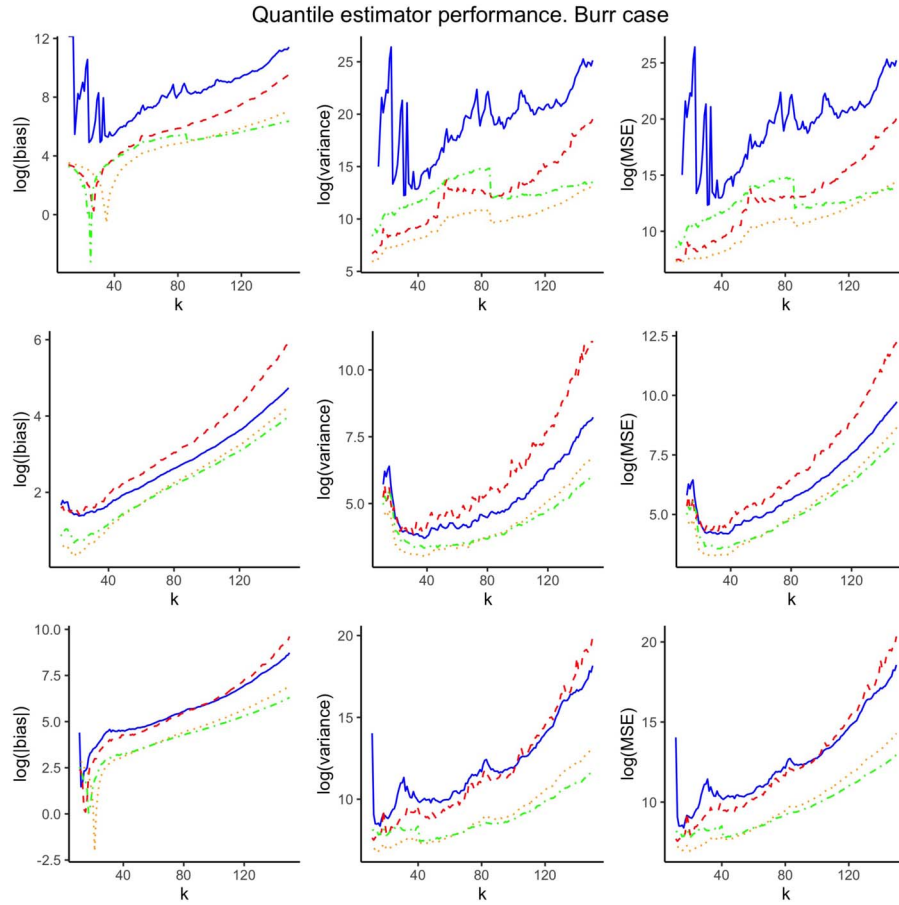


FIG 4. Burr distributions. Bias, variance and mean squared error of the quantile estimator based on: the kernel estimator ($H_k = H_k^{K_0}$ in solid blue, $H_k^A = H_k^{K_1}$ in dashed red, $H_k^{K_2}$ in dotted orange) and the Worms estimator H_k^W (dashed and dotted green), as a function of k . The top, middle and bottom levels correspond to $2p < 1$, $2p > 1$ and $2p = 1$, respectively.

Kaplan-Meier estimator together with the empirical survival function of the ultimates. We observe that the tail of the latter under-estimates the tail index that is suggested by using survival-analysis techniques.

We now apply the censored tail estimators introduced in this paper to the data. Using the same mechanism as for the simulation study (and $\rho_z = -1$) we find that $k = 35$ is the optimal threshold for the estimator using the kernel \mathcal{K}_0 (see Figure 11). As observed in the simulations, it is perfectly reasonable to consider the other estimators ($H_k^{K_1}$, $H_k^{K_2}$ and the Worms estimator H_k^W) at this value as well. This yields the estimates

$$\hat{\xi} = 0.841, 0.761, 0.654, 0.664.$$

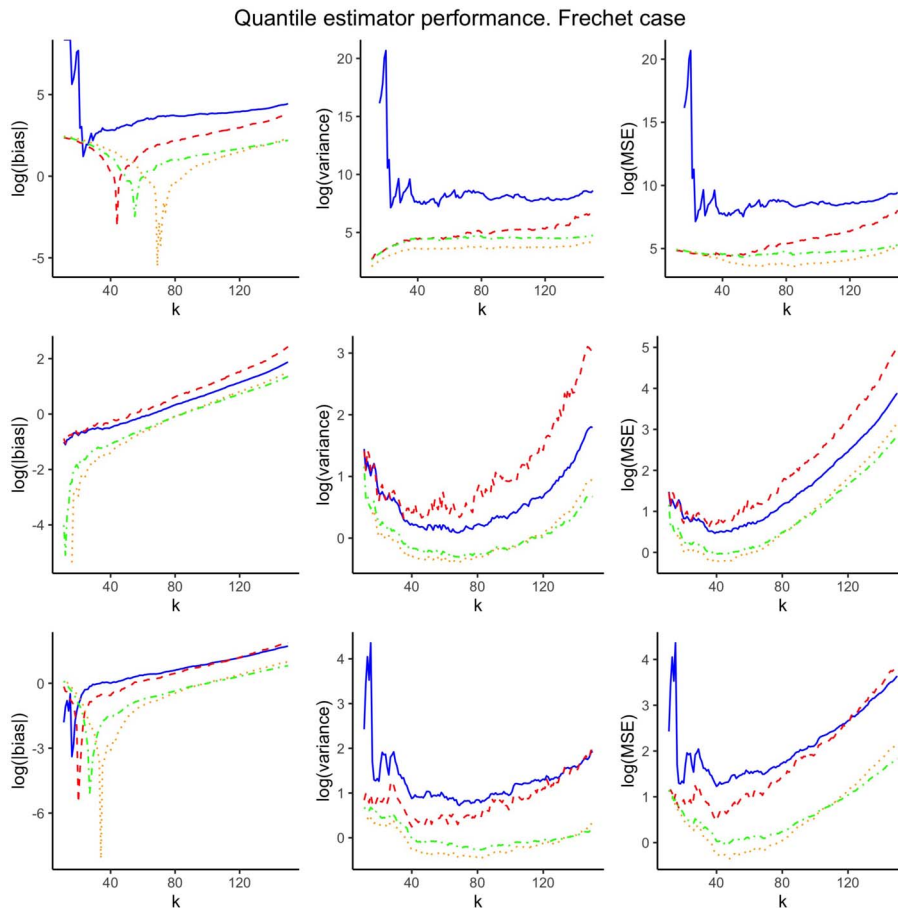


FIG 5. Fréchet distributions. Bias, variance and mean squared error of the quantile estimator based on: the kernel estimator ($H_k = H_k^{K_0}$ in solid blue, $H_k^A = H_k^{K_1}$ in dashed red, $H_k^{K_2}$ in dotted orange) and the Worms estimator H_K^W (dashed and dotted green), as a function of k . The top, middle and bottom levels correspond to $2p < 1$, $2p > 1$ and $2p = 1$, respectively.

The latter two values correspond to the kernel \mathcal{K}_2 and to the Worms estimator H_k^W . Notice that they are quite close, and although the simulations suggest that the two last values are the best performing, the 95% confidence interval for the first of these estimators is given by $[0.415, 1.267]$ which amply accommodates all four estimates. Hence, in practice, with only one sample available, it is not possible to make overly conclusive claims regarding the superiority of these point estimates, since they are not statistically distinguishable. The corresponding 99.5% quantile estimators are given by

$$14957214, 12093195, 9129021, 9355831,$$

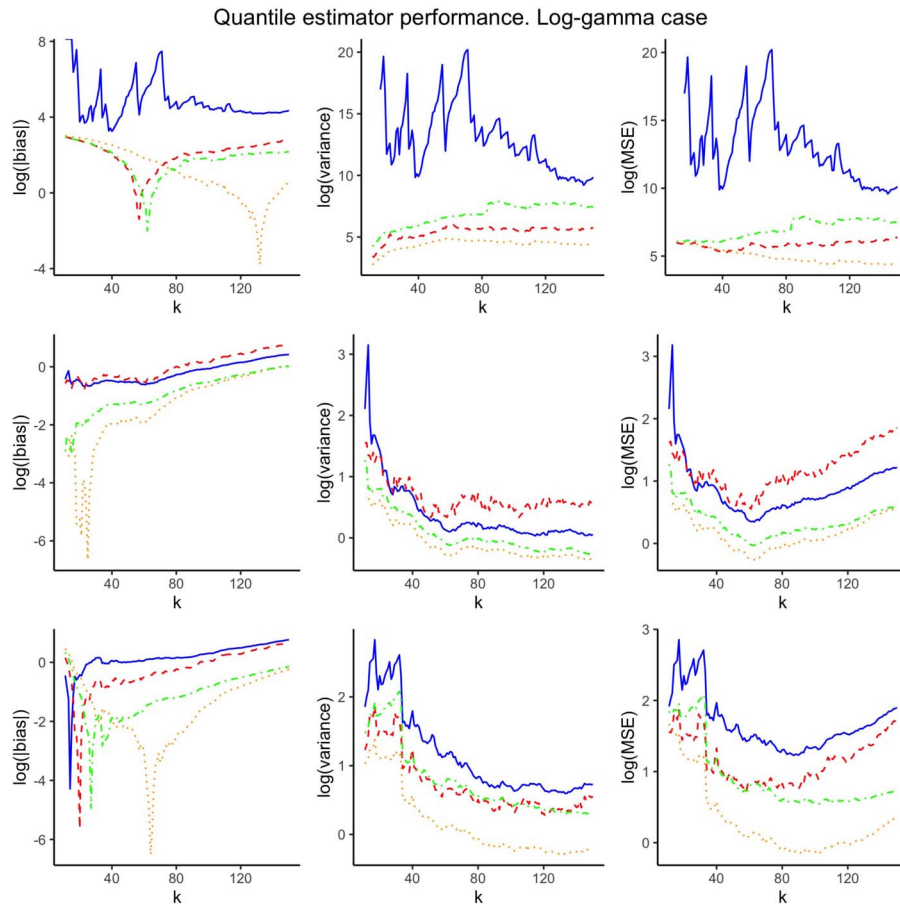


FIG 6. Log-gamma distributions. Bias, variance and mean squared error of the quantile estimator based on: the kernel estimator ($H_k = H_k^{K_0}$ in solid blue, $H_k^A = H_k^{K_1}$ in dashed red, $H_k^{K_2}$ in dotted orange) and the Worms estimator K_k^W (dashed and dotted green), as a function of k . The top, middle and bottom levels correspond to $2p < 1$, $2p > 1$ and $2p = 1$, respectively.

illustrating how small changes in tail estimation can lead to large differences in the quantile scale. Note however that the quantile estimates based on H_k^W and $H_k^{K_2}$ are quite stable for $k \leq 80$.

A more refined analysis of ρ_z is known to be unstable, but can be routinely applied (for instance using the `mop` function from the R package `evt0`). For the Hill estimator of the Z_i variates (ignoring censoring) this gives the estimate $\hat{\rho}_z = -0.616$, which is relatively close to our choice of -1 , given the high variability of second-order parameter estimators. Repeating the above analysis with this value has small quantitative influence and no additional qualitative insight, and is thus omitted.

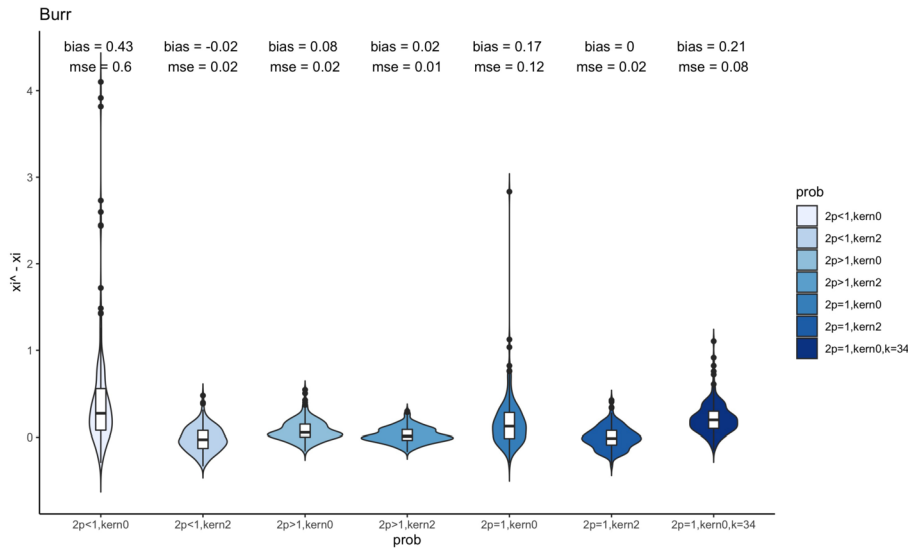


FIG 7. Violin plots for the simulation results in case of Burr distributions under different non-censoring asymptotic probabilities. We take the difference between estimated and theoretical ξ . The cases $2p < 1$ and $2p \geq 1$ correspond to heavy and light censoring, respectively. The labels kern0 and kern2 correspond to the use of the estimators $H_k^{\mathcal{K}_0}$ and $H_k^{\mathcal{K}_2}$. The specified $k = 34$ is the theoretical optimal sample fraction.

Previous studies, using the ultimate values (cf. Bladt et al. (2020b), with subsequent agreement in Albrecher et al. (2019)), that is, internal projected values of the claim sizes at closure, suggested a tail index of about 0.48. In Bladt et al. (2020a), combining this expert information with the estimator corresponding to \mathcal{K}_0 , intermediate values between the purely statistical 0.87 and the purely expert information 0.48 were suggested. The present value of 0.66 is an interesting intermediate value that arises from a purely statistical procedure.

7. Conclusion

In this paper we developed novel extreme value estimators under right-censoring in a kernel framework. The latter class is closed (in the asymptotic sense) under the averaging operation of their trimmed versions, by a simple replacement of kernel. The asymptotic behaviour is given for arbitrary kernels, which allows us to compute, for instance, the expression for the MSE as a function of k . The choice of the optimal threshold with respect to MSE is explored in connection with the empirical variance of the trimmed trajectories, which leads to an automated way of selecting a threshold. As for the non-censored case, the idea of selecting a threshold by exploiting this link, circumvents the usual estimation difficulties and instabilities which arise in previous approaches in the literature which typically require the estimation of the second-order parameter D , and

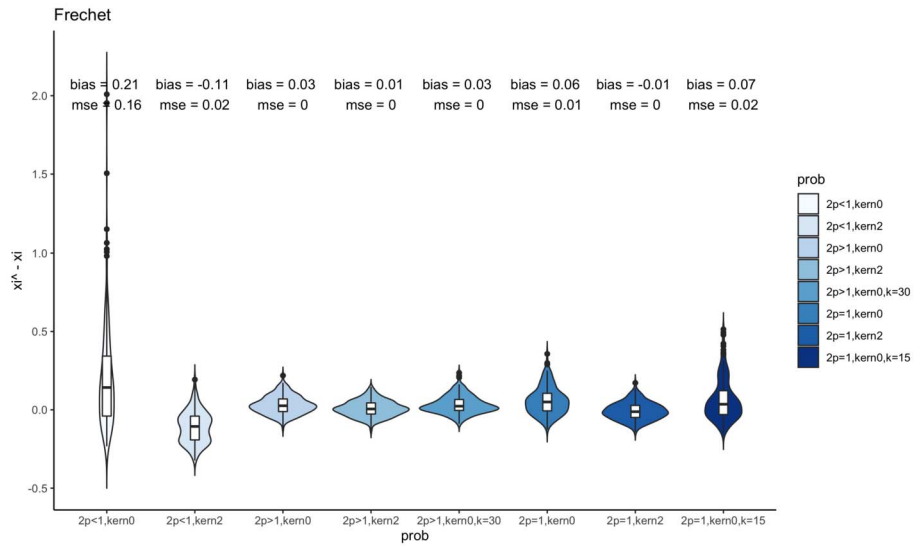


FIG 8. Violin plots for the simulation results in case of Fréchet distributions under different non-censoring asymptotic probabilities. We take the difference between estimated and theoretical ξ . The cases $2p < 1$ and $2p \geq 1$ correspond to heavy and light censoring, respectively. The labels kern0 and kern2 correspond to the use of the estimators $H_k^{K_0}$ and $H_k^{K_2}$. The specified $k = 30, 15$ are the theoretical optimal sample fractions.

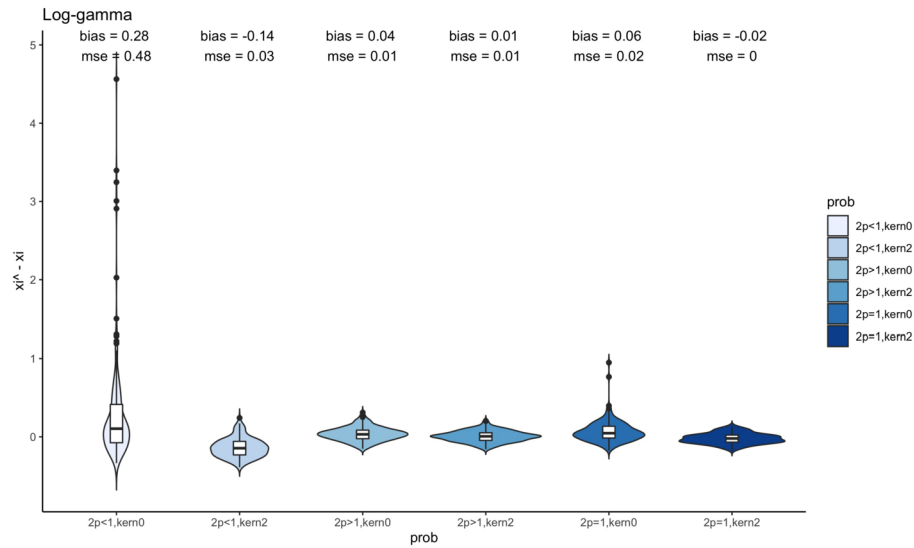


FIG 9. Violin plots for the simulation results in case of log-gamma distributions under different non-censoring asymptotic probabilities. We take the difference between estimated and theoretical ξ . The cases $2p < 1$ and $2p \geq 1$ correspond to heavy and light censoring, respectively. The labels kern0 and kern2 correspond to the use of the estimators $H_k^{K_0}$ and $H_k^{K_2}$.

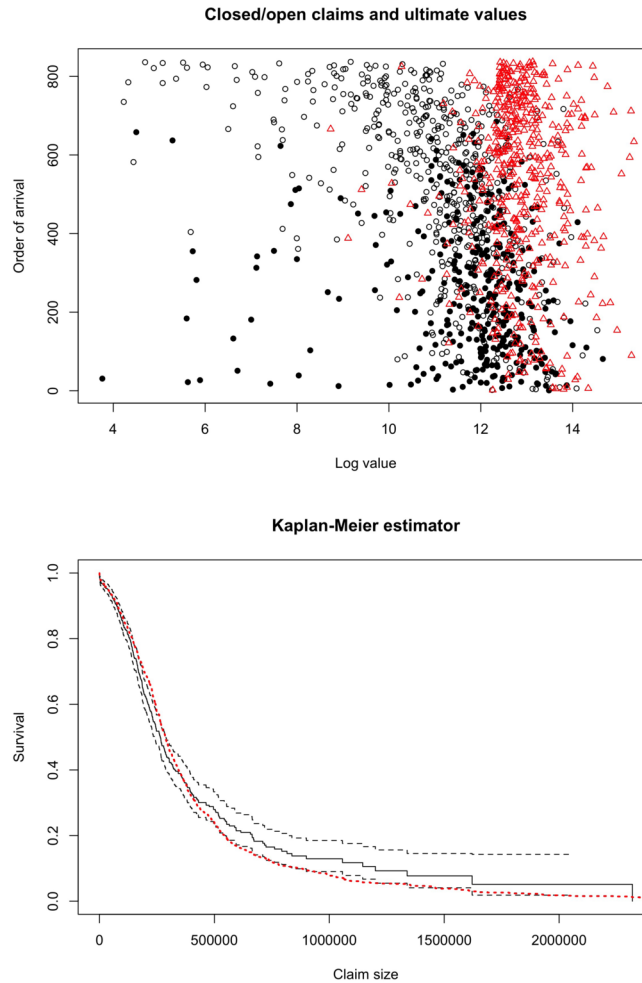


FIG 10. *MTPL insurance claims data. Top: open (empty circle) and closed (full circle) claims, together with the ultimate values (triangle) for the open claims. Bottom: Kaplan-Meier survival estimate for the open and closed claims, together with the empirical survival function of the ultimates (dotted).*

of ξ itself. Despite its simplicity, simulation studies suggest that the method is also efficient. In fact, when compared with the theoretically optimal value, the latter sometimes is too small to be of any practical relevance, and then our adaptive estimator is superior. In the other cases, when the theoretically optimal value is sensible, our estimator also performs well against it. We finally apply the procedure to a well-understood insurance dataset, and the simulation studies suggest that the instances where \mathcal{K}_0 has been used in the literature (either alone, or in combination with expert information) to analyze these data could very possibly be improved by considering \mathcal{K}_2 instead. Interesting directions for

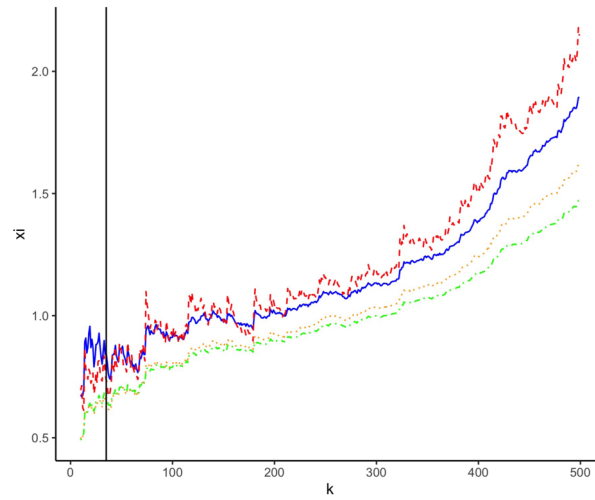


FIG 11. Estimates of ξ for the MTPL insurance claim size data: $H_k = H_k^{K_0}$ in solid blue, $H_k^A = H_k^{K_1}$ in dashed red, $H_k^{K_2}$ in dotted orange and the Worms estimator H_k^W in dashed and dotted green. The vertical line is at the estimated optimal k for $H_k = H_k^{K_0}$.

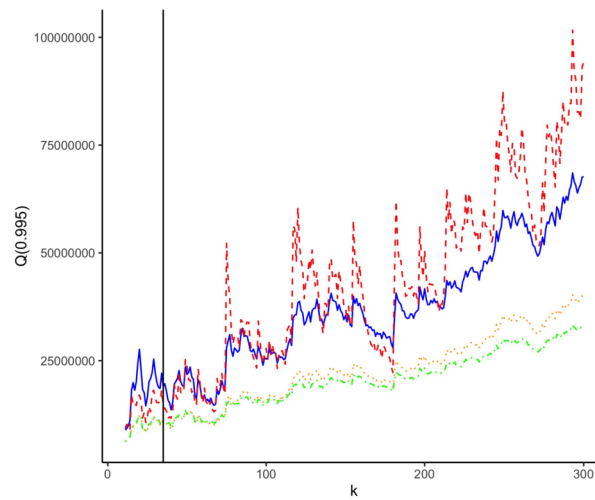


FIG 12. Estimates of the 99.5% quantile for the MTPL insurance claim size data based on the tail estimators: $H_k = H_k^{K_0}$ in solid blue, $H_k^A = H_k^{K_1}$ in dashed red, $H_k^{K_2}$ in dotted orange and the Worms estimator H_k^W in dashed and dotted green. The vertical line is at the estimated optimal k for $H_k = H_k^{K_0}$.

further research include trimming the kernel estimators from above, to remove outliers from data, and to apply combined tail information using censored data and expert information with the new kernels, improving the previous methods.

Finally, it will be interesting to consider optimality criteria for the choice of k for any kernel, and to work out criteria for the selection of the optimal kernel from a purely mathematical point of view.

References

- Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). *Reinsurance: Actuarial and Statistical Aspects*. John Wiley & Sons, Chichester. [MR3791478](#)
- Albrecher, H., Bladt, M., and Bladt, M. (2019). Matrix Mittag-Leffler distributions and modeling heavy-tailed risks. *arXiv preprint arXiv:1906.05316*. [MR4129559](#)
- Beirlant, J., Bardoutsos, A., de Wet, T., and Gijbels, I. (2016). Bias reduced tail estimation for censored Pareto type distributions. *Statistics & Probability Letters*, 109:78–88. [MR3434963](#)
- Beirlant, J., Dierckx, G., Guillou, A., and Fils-Villetard, A. (2007). Bias reduced tail estimation for censored Pareto type distributions. *Extremes*, 10:151–174. [MR2415636](#)
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). *Statistics of Extremes: Theory and Practice*. John Wiley & Sons, Chichester. [MR2108013](#)
- Beirlant, J., Guillou, A., and Toulemonde, G. (2010). Peaks-over-threshold modeling under random censoring. *Communications in Statistics – Theory and Methods*, 39:1158–1179. [MR2745371](#)
- Beirlant, J., Maribe, G., and Verster, A. (2018). Penalized bias reduction in extreme value estimation for censored Pareto-type data, and long-tailed insurance applications. *Insurance: Mathematics and Economics*, 78:114–122. [MR3761071](#)
- Beirlant, J., Worms, J., and Worms, R. (2019). Estimation of the extreme value index in a censorship framework: Asymptotic and finite sample behavior. *Journal of Statistical Planning and Inference*, 202:31–56. [MR3926762](#)
- Bladt, M., Albrecher, H., and Beirlant, J. (2020a). Combined tail estimation using censored data and expert information. *Scandinavian Actuarial Journal*, (6):503–525. [MR4130455](#)
- Bladt, M., Albrecher, H., and Beirlant, J. (2020b). Trimming and threshold selection in extremes. *Extremes*, 23(4):629–665. [MR4165035](#)
- Csörgő, S., Deheuvels, P., and Mason, D. (1985). Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, 13:1050–1077. [MR0803758](#)
- de Haan, L., and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.
- Hall, P. and Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13:331–341.
- Dierckx, G., Goegebeur, Y., and Guillou, A. (2019). Local robust estimation of pareto-type tails with random right censoring. *Sankhya a*, pages 1–39. [MR4227205](#)
- Einmahl, J. H., Fils-Villetard, A., and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli*, 14(1):207–227. [MR2401660](#)

- Fraga Alves, M., Gomes, M., and de Haan, L. (2003). A new class of semiparametric estimators of the second order parameter. *Portugaliae Mathematica*, 60:193–213. [MR1984031](#)
- Goegebeur, Y., Guillou, A., and Qin, J. (2019a). Bias-corrected estimation for conditional pareto-type distributions with random censoring. *Extremes*, 22:459–498. [MR3984358](#)
- Goegebeur, Y., Guillou, A., and Qin, J. (2019b). Robust estimation of the pickands dependence function under random right censoring. *Insurance: Mathematics and Economics*, 87:101–114. [MR3945645](#)
- Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution *The Annals of Statistics*, 3:1163–1174.
- Ndao, P., Diop, A., and Dupuy, J.-F. (2014). Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. *Computational Statistics & Data Analysis*, 79:63–79. [MR3227987](#)
- Ndao, P., Diop, A., and Dupuy, J.-F. (2016). Nonparametric estimation of the conditional extreme-value index with random covariates and censoring. *Journal of Statistical Planning and Inference*, 168:20–37. [MR3412219](#)
- Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, 144:1–24. [MR3434936](#)
- Stupfler, G. (2019). On the study of extremes with dependent random right-censoring. *Extremes*, 22:97–129. [MR3919740](#)
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815. [MR0521329](#)
- Worms, J. and Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes*, 17(2):337–358. [MR3207720](#)
- Worms, J. and Worms, R. (2016). A Lynden-Bell integral estimator for extremes of randomly truncated data. *Statistics & Probability Letters*, 109:106–117. [MR3434966](#)
- Worms, J. and Worms, R. (2018). Extreme value statistics for censored data with heavy tails under competing risks. *Metrika*, 81(7):849–889. [MR3849981](#)
- Worms, J. and Worms, R. (2019). Estimation of extremes for Weibull-tail distributions in the presence of random censoring. *Extremes*, 22:667–704. [MR4031853](#)