

A Simulation Comparison of Estimators of Conditional Extreme Value Index under Right Random Censoring

Richard Minkah^{1,*}, Tertius de Wet² and Ezekiel Nii Noye Nortey¹

¹Department of Statistics and Actuarial Science, University of Ghana, Ghana. ²Department of Statistics and Actuarial Science, Stellenbosch University, South Africa

Received on March 01, 2018; Accepted on June 15, 2018; Published Online on June 18, 2018

Copyright © 2018, African Journal of Applied Statistics (AJAS) and The Statistics and Probability African Society (SPAS). All rights reserved

Abstract. In extreme value analysis, the extreme value index (EVI) plays a vital role as it determines the tail heaviness of the underlying distribution and the primary parameter required for estimating other extreme events. In this paper, we review the estimation of the EVI when observations are subject to right random censoring and the presence of covariate information. In addition, we propose some estimators of the EVI, including a maximum likelihood estimator from a perturbed Pareto distribution. The existing estimators and the proposed ones are compared through a simulation study.

Key words: Extreme value index; random censoring; covariate information; moving window; simulation

AMS 2010 Mathematics Subject Classification : 62F10; 62F12; 62G32; 60F05.

Presented by Dr.Tchilabalo A. Kpanzou, University of
Kara (Togo)
Corresponding Member of the Editorial Board.

*Corresponding author : Richard Minkah: rminkah@ug.edu.gh

Tertius de Wet : tdewet@sun.ac.za

Ezekiel N. N. Nortey : ennnortey@ug.edu.gh

Full Abstract (English) In extreme value analysis, the extreme value index (EVI) plays a vital role as it determines the tail heaviness of the underlying distribution and the primary parameter required for estimating other extreme events. In this paper, we review the estimation of the EVI when observations are subject to right random censoring and the presence of covariate information. In addition, we propose some estimators of the EVI, including a maximum likelihood estimator from a perturbed Pareto distribution. The existing estimators and the proposed ones are compared through a simulation study. The results show that the performance of the estimators depend on the percentage of censoring, the underlying distribution, the size of extreme value index and the number of top order statistics. Overall, we found the proposed estimator from the perturbed Pareto distribution to be more robust to censoring, size of the EVI and the number of top order statistics.

Résumé (French) Dans la théorie des valeurs extrêmes (TVE), l'indice des valeurs extrêmes (IVE) joue un rôle essentiel car il détermine la lourdeur de la queue de la distribution sous-jacente et le paramètre principal requis pour estimer d'autres événements extrêmes. Dans cet article, nous passons en revue l'estimation de l'EVI lorsque les observations sont soumises à une censure aléatoire et à la présence d'informations auxiliaires. En outre, nous proposons quelques estimateurs de l'EVI, y compris un estimateur du maximum de vraisemblance à partir d'une distribution de Pareto perturbée. Les estimateurs existants et ceux proposés sont comparés à l'aide d'une étude de simulation. Les résultats montrent que les performances des estimateurs dépendent du pourcentage de censure, de la distribution sous-jacente, de la taille de l'indice de valeur extrême et du nombre de statistiques d'ordre supérieur. Dans l'ensemble, nous avons trouvé que l'estimateur proposé provenant de la distribution de Pareto perturbée était plus robuste par rapport à la censure, à la taille de l'IVE et au nombre d'observations de queue utilisées.

1. Introduction

The study of extreme events has received much attention in many fields of application due to the nature of their impact. For instance, extreme earth quakes cause many deaths and destruction to properties; large price movements in equities result in huge losses, profits or collapse of financial markets; large insurance claims lead to solvency problems.

Unlike traditional statistical methods that focus on the central part of distributions, statistics of extremes focuses on the tail of the underlying distribution. Interest is then on parameters associated with the tail of the underlying distribution, such as high quantiles and exceedance probabilities. For inference on such parameters, distributional results are needed on the extreme observations. The first such result in extreme value theory was obtained by Fisher and Tippett (1928) and further developed by e.g. Gnedenko (1943) and de Haan (1970). These asymptotic distributions form the basis for carrying out inference in extreme value analysis.

Let Y_1, \dots, Y_n be an independent and identically distributed sample on some random variable Y and let $Y_{1,n} \leq \dots \leq Y_{n,n}$ be the corresponding order statistics. The mentioned results state that, if there exist normalising constants $a_n > 0$ and $b_n \in \mathbb{R}$, and some nondegenerate function Ψ , such that

$$\frac{Y_{n,n} - b_n}{a_n} \xrightarrow{d} \Psi, \tag{1}$$

then the constants can be redefined such that for $\gamma \in \mathbb{R}$,

$$\Psi(y) \equiv \Psi_\gamma(y) = \begin{cases} \exp\left(-(1 + \gamma y)^{-1/\gamma}\right), & 1 + \gamma y > 0, \gamma \neq 0, \\ \exp(-\exp(y)), & y \in \mathbb{R}, \gamma = 0. \end{cases} \tag{2}$$

Here, (2) is the so-called Generalized Extreme Value distribution and γ is the extreme value index (or tail index). The parameter γ is the primary parameter needed in extreme value analysis and determines the tail heaviness of the underlying distribution. If $\gamma > 0$, Ψ_γ belongs to the Pareto domain (heavy tailed); if $\gamma < 0$, Ψ_γ belongs to the Weibull domain (short-tailed); and if $\gamma = 0$, Ψ_γ belongs to the Gumbel domain (light-tailed). An underlying distribution function, F , for which (1) and (2) hold, is said to be in the domain of attraction of Ψ_γ , denoted $F \in D(\Psi_\gamma)$.

The estimation of γ has been addressed in many papers, including Hill (1975); Pickands (1975); de Haan and Peng (1998); Tsourti and Panaretos (2003); Beirlant et al. (2004); Diop and Lo (2006); Gomes et al. (2008); Diop and Lo (2009); Deme et al. (2009); Gomes and Guillou (2014) and Ngom and Lo (2016). When covariate information is available, the focus is to include it in the estimation by modelling the parameters of the extreme value distribution as a function of the covariate(s). For example, Davison and Smith (1990) fitted a Generalised Pareto (GP) distribution with parameters taken as an exponential function of the covariates; Gardes and Girard (2008) used moving-window methodology; Beirlant and Goegebeur (2003) and Wang and Tsai (2009) used a conditional exponential regression model; and Beirlant and Goegebeur (2004) employed repeated fitting of local polynomial maximum likelihood estimation.

In the case of censoring, Beirlant et al. (2007) and Einmahl et al. (2008) proposed an inverse probability-of-censoring weighted method to adapt classical extreme value index estimators to censoring. Similarly, Gomes and Neves (2011) and Brahimi et al. (2013) used this idea to adapt various estimators to censoring. In addition, Beirlant et al. (2010) addressed the issue of censoring, obtaining maximum likelihood estimators by adapting the likelihood function of the generalised Pareto distribution to censoring. Also, Worms and Worms (2014) considered estimators based on Kaplan-Meier integration and censored regression. Furthermore, Amaraoui et al. (2016) estimated the extreme value index from a Bayesian perspectives and Beirlant et al. (2017) proposed a reduced-bias estimator based on an extended Pareto distribution.

In the case of the presence of both covariate information and censoring, Ndao et al. (2014) proposed three estimators for the estimation of the conditional extreme value index and extreme quantiles for heavy-tailed distributions. In particular, the Hill, generalised Hill and moment type estimators were proposed using the moving window method of left-censoring weighted method (Beirlant et al., 2007; Einmahl et al., 2008). Unlike Ndao et al. (2014), Stupfler (2016) proposed a moment estimator valid for all domains of attraction. In addition, Ndao et al. (2016) addressed the estimation of the extreme value index under censoring and the presence of random covariates.

Although quite a number of papers have compared the available estimators of the extreme value index, no paper has compared the estimators of the conditional extreme value index when observations are subject to random censoring. The aim of this paper is two-fold. Firstly, to adapt some classical estimators to the current context, including a reduced-bias maximum likelihood estimator based on a perturbed Pareto distribution in Beirlant et al. (2004). Secondly, we review the available estimators that were proposed in the literature for estimating conditional extreme value index under right random censoring and compare them together with the proposed ones in a simulation study.

The remainder of the paper is organised as follows. In Section 2, we set out the framework for the estimation of the parameter of interest i.e. extreme value index. In addition, the existing estimators are reviewed and we present the proposed estimators. In section 3, we conduct a simulation study to assess the performance of these estimators. Lastly, general conclusions from the simulation results are presented in section 4.

2. Estimation of the Extreme Value Index

Consider Y_1, \dots, Y_n as independent copies of a positive random variable, Y , and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be the values of an associated d -dimensional covariate vector, $\mathbf{x} \in \Omega$, where $\Omega \subset \mathbb{R}^d$. Also, in order to incorporate the presence of censoring, let C_1, \dots, C_n be independent copies of another positive random variable C , also associated with the covariate vector \mathbf{x} . We assume that for all $\mathbf{x} \in \Omega$, the random variables, Y and C , are independent. Furthermore, for every $\mathbf{x} \in \Omega$, we assume that the random variables Y and C have respective conditional distribution functions, $F(\cdot; \mathbf{x}) \in D(\Psi_{\gamma_1(\mathbf{x})})$ and $G(\cdot; \mathbf{x}) \in D(\Psi_{\gamma_2(\mathbf{x})})$, where $\gamma_1(\mathbf{x})$ and $\gamma_2(\mathbf{x})$ are real functions. We consider the case where $F(\cdot; \mathbf{x})$ and $G(\cdot; \mathbf{x})$ are in the Pareto domain of attraction i.e. $\gamma_1(\mathbf{x}) > 0$ and $\gamma_2(\mathbf{x}) > 0$. For this domain of attraction, the distribution functions can be represented as

$$1 - F(y; \mathbf{x}) = y^{-\frac{1}{\gamma_1(\mathbf{x})}} \ell_F(y; \mathbf{x}) \text{ and } 1 - G(y; \mathbf{x}) = y^{-\frac{1}{\gamma_2(\mathbf{x})}} \ell_G(y; \mathbf{x}). \quad (3)$$

or equivalently in terms of the tail quantile function,

$$U_F(y; \mathbf{x}) = y^{\gamma_1(\mathbf{x})} \ell_{U_F}(y; \mathbf{x}) \text{ and } U_G(y; \mathbf{x}) = y^{\gamma_2(\mathbf{x})} \ell_{U_G}(y; \mathbf{x}). \quad (4)$$

Here, $\ell_F(y; \mathbf{x})$ (and $\ell_{U_F}(y; \mathbf{x})$) and $\ell_G(y; \mathbf{x})$ (and $\ell_{U_G}(y; \mathbf{x})$) are slowly varying functions associated with F and G respectively and defined as

$$\lim_{y \rightarrow \infty} \frac{\ell_j(by; \mathbf{x})}{\ell_j(y; \mathbf{x})} = 1, \quad b > 0; j \in \{F, G, U_F, U_G\}. \quad (5)$$

In this context, we observe the triplets $\{Z_i, \delta_i, \mathbf{x}_i\}$, $i = 1, \dots, n$ where $Z_i = \min\{Y_i, C_i\}$ and $\delta_i = \mathbb{I}\{Y_i \leq C_i\}$. By the independent assumption of Y and C , the conditional distribution function $H(\cdot; \mathbf{x})$ of the random variable Z_i is related to $F(\cdot; \mathbf{x})$ and $G(\cdot; \mathbf{x})$, as

$$1 - H(\cdot; \mathbf{x}) = (1 - F(\cdot; \mathbf{x})) (1 - G(\cdot; \mathbf{x})). \quad (6)$$

Therefore, $H(\cdot; \mathbf{x})$ is also in the Pareto domain with conditional extreme value index given in Einmahl et al. (2008) as

$$\gamma(\mathbf{x}) = \frac{\gamma_1(\mathbf{x})\gamma_2(\mathbf{x})}{\gamma_1(\mathbf{x}) + \gamma_2(\mathbf{x})}. \quad (7)$$

Before presenting the estimators, we define a ball $B(\mathbf{x}, r)$ in Ω where \mathbf{x} and r ($r > 0$) are the center and radius respectively. Thus,

$$B(\mathbf{x}, r) = \{\mu \in \mathbb{R}^d : d(\mathbf{x}, \mu) \leq r\}. \quad (8)$$

In addition, let $h_{n,\mathbf{x}}$ be a positive integer that approaches 0 as $n \rightarrow \infty$. The estimators of the EVI are based on observations of Z_i for which the corresponding values of \mathbf{x}_i fall within the ball $B(\mathbf{x}, h_{n,\mathbf{x}})$. The proportion of the design points falling within the ball is defined as

$$\phi(h_{n,\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{\mathbf{x}_i \in B(\mathbf{x}, h_{n,\mathbf{x}})\}, \quad (9)$$

where \mathbb{I} is the indicator variable. Relation (9) plays an important role in this procedure as it describes how the points are concentrated around the neighbourhood of \mathbf{x}_i when $h_{n,\mathbf{x}}$ approaches 0 (Gardes and Girard (2008)). The number of nonrandom observations in $(0, \infty) \times B(\mathbf{x}, h_{n,\mathbf{x}})$ is given by $m_{n,\mathbf{x}} = n\phi(h_{n,\mathbf{x}})$.

Let $(W_1(\mathbf{x}), \delta_{(1)}), \dots, (W_{m_{n,\mathbf{x}}}(\mathbf{x}), \delta_{(m_{n,\mathbf{x}})})$ denote the pair, (Z_i, δ_i) , $i = 1, 2, \dots, n$, that have their corresponding \mathbf{x}_i -values falling within the ball as defined in (8). Also, let $W_{1,m_{n,\mathbf{x}}}(\mathbf{x}) \leq \dots \leq W_{m_{n,\mathbf{x}},m_{n,\mathbf{x}}}(\mathbf{x})$ be the corresponding order statistics of W 's and $\delta_{(i)}^{(W)}$, $i = 1, \dots, m_{n,\mathbf{x}}$ be the values of δ 's associated with $W_{i,m_{n,\mathbf{x}}}(\mathbf{x})$, $i = 1, \dots, m_{n,\mathbf{x}}$. The values of $\delta_{(i)}^{(W)}$, $i = 1, \dots, m_{n,\mathbf{x}}$ form the basis for adapting the classical estimators of the conditional extreme value index presented below to censoring.

In what follows, given a sample $\{Z_1, \delta_1, \mathbf{x}_1\}, \dots, \{Z_n, \delta_n, \mathbf{x}_n\}$, we consider the estimation of $\gamma_1(\mathbf{x})$. To do this, we rely on the observations $(W_1(\mathbf{x}), \delta_{(1)}), \dots, (W_{m_{n,\mathbf{x}}}(\mathbf{x}), \delta_{(m_{n,\mathbf{x}})})$ resulting from the moving window approach of Gardes and Girard (2008) described after (9).

2.1. The Existing Estimators

The existing estimators result from the application of the moving window technique (Gardes and Girard, 2008) and the inverse probability-of-censoring weighted method (Beirlant et al., 2007; Einmahl et al., 2008) to adapt classical estimators to censoring. Ndao et al. (2014) used this approach to adapt the Hill, generalised Hill and moment estimators to censoring. In this section, we review these estimators and follow a similar approach to propose other estimators of the extreme value index in the next section.

The estimators introduced by Ndao et al. (2014) are presented as follows:

- **The Hill-type estimator:** The Hill estimator (Hill, 1975) is arguably the most common estimator of γ in the Pareto case i.e. $\gamma > 0$. To take into account the available covariate information, the Hill estimator is defined for the $(k_{n,x} + 1)$ -largest order statistics as

$$\hat{\gamma}^{(c,Hill)}(W, k_{n,x}, m_{n,x}) = \frac{1}{k_{n,x}} \sum_{i=1}^{k_{n,x}} i (\log W_{m_{n,x}-i+1, m_{n,x}}(\mathbf{x}) - \log W_{m_{n,x}-i, m_{n,x}}(\mathbf{x})). \tag{10}$$

- **The Moment-type estimator:** Dekkers et al. (1989) introduced the moment estimator as an adaptation of the Hill estimator valid for all domains of attraction. It is defined to take into account the covariate information \mathbf{x} as,

$$\hat{\gamma}^{(c,MOM)}(W, k_{n,x}, m_{n,x}) = M_n^{(1)}(W, k_{n,x}, m_{n,x}) + 1 - \frac{1}{2} \left(1 - \frac{(M_n^{(1)}(W, k_{n,x}, m_{n,x}))^2}{M_n^{(2)}(W, k_{n,x}, m_{n,x})} \right)^{-1} \tag{11}$$

where

$$M_n^{(j)}(W, k_{n,x}, m_{n,x}) = \frac{1}{k_{n,x}} \sum_{i=1}^{k_{n,x}} [\log (W_{m_{n,x}-i+1, m_{n,x}}(\mathbf{x})) - \log (W_{m_{n,x}-k_{n,x}, m_{n,x}}(\mathbf{x}))]^j. \tag{12}$$

- **The generalised Hill Estimator:** Beirlant et al. (1996) proposed the generalised Hill (GH) estimator as an attempt to extend the Hill estimator to the case where $\gamma \in \mathbb{R}$. The GH estimator is obtained as the slope of the ultimately linear part of the generalised Pareto quantile plot of the observations within the defined window as,

$$\hat{\gamma}^{(c,UH)}(W, k_{n,x}, m_{n,x}) = \frac{1}{k_{n,x}} \sum_{j=1}^{k_{n,x}} \log UH_{j, m_{n,x}} - \log UH_{k_{n,x}+1, m_{n,x}}, \tag{13}$$

where

$$UH_{j, m_{n,x}} = W_{m_{n,x}-j, m_{n,x}} \left(\frac{1}{j} \sum_{i=1}^j \log W_{m_{n,x}-i+1, m_{n,x}} - \log W_{m_{n,x}-j, m_{n,x}} \right).$$

The estimators (10), (11) and (13) were adapted to censoring by dividing each estimator by the proportion of noncensored observations in the $k_{n,\mathbf{x}}$ largest order statistics of W 's. Thus, an adapted estimator of the conditional extreme value index is given by

$$\hat{\gamma}^{(c,\cdot)}(W, k_{n,\mathbf{x}}, m_{n,\mathbf{x}}) = \frac{\hat{\gamma}_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(\cdot)}}{\hat{\phi}(\mathbf{x})}, \quad (14)$$

where $\hat{\phi}(\mathbf{x}) = k_{n,\mathbf{x}}^{-1} \sum_{i=1}^{k_{n,\mathbf{x}}} \delta_{m_{n,\mathbf{x}}-i+1, m_{n,\mathbf{x}}}^{(w)}$.

2.2. The Proposed Estimators

In this section, we propose some estimators for the estimation of conditional extreme value index when observations are subject to right random censoring. We follow closely the methodology in Ndao et al. (2014).

Firstly, the estimators are presented to take into account the covariate and are subsequently adapted to censoring.

- (i) **The Zipf estimator:** Kratz and Resnick (1996) derived the Zipf estimator as a smoother version of the Hill estimator through unconstrained least squares fit to the k largest observations on the generalised Pareto quantile plot method of Beirlant et al. (1996). The estimator is valid for $\gamma > 0$ and in the case of a covariate, is given by,

$$\hat{\gamma}^{(c,Zipf)}(W, k_{n,\mathbf{x}}, m_{n,\mathbf{x}}) = \sum_{i=1}^{k_{n,\mathbf{x}}} i \log \left(\frac{W_{m_{n,\mathbf{x}}-i+1, m_{n,\mathbf{x}}}(\mathbf{x})}{W_{m_{n,\mathbf{x}}-i, m_{n,\mathbf{x}}}(\mathbf{x})} \right) \frac{\log(k_{n,\mathbf{x}}/i)}{\sum_{i=1}^{k_{n,\mathbf{x}}} \log(k_{n,\mathbf{x}}/i)} \quad (15)$$

- (ii) **The Moment Ratio:** The Moment Ratio estimator was introduced by Danielsson et al. (1996) as a moment based estimator to reduce bias in the Hill estimator. The moment ratio estimator is valid for the Pareto domain of attraction only. In the case of a covariate, it is given by

$$\hat{\gamma}^{(MomR)}(W, k_{n,\mathbf{x}}, m_{n,\mathbf{x}}) = \frac{1}{2} \frac{M_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(2)}}{M_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(1)}}, \quad (16)$$

where $M_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(j)}$, $j = 1, 2$ is obtained from (12).

- (iii) **The Peng Moment Estimator:** Deheuvels et al. (1997) reports on a variant of the moment estimator for the no covariate case suggested by Liam Peng. This estimator is designed to reduce bias in the moment estimator and it is adapted to the covariate case as

$$\hat{\gamma}^{(PMom)}(W, k_{n,\mathbf{x}}, m_{n,\mathbf{x}}) = \frac{1}{2} \frac{M_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(2)}}{M_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(1)}} + 1 - \frac{1}{2} \left(1 - \frac{(M_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(1)})^2}{M_{W,k_{n,\mathbf{x}},m_{n,\mathbf{x}}}^{(2)}} \right)^{-1}, \quad (17)$$

where $M_{W, k_{n,x}, m_{n,x}}^{(j)}$, $j = 1, 2$ is as before obtained from (12). This estimator is valid for all domains of attraction.

(iv) **The Perturbed Pareto Estimator**

Beirlant et al. (2004) derived the perturbed Pareto estimator as a reduced-biased Hill type estimator by making use of the second-order properties on the underlying distribution function, F . Here, we consider adapting this estimator to censoring and the presence of covariate information. Considering (3) and (5), we can write

$$\lim_{u \rightarrow \infty} \frac{1 - F(uw; \mathbf{x})}{1 - F(u; \mathbf{x})} = w^{-1/\gamma(\mathbf{x})}, \text{ for any } w > 1. \tag{18}$$

Therefore, (18) can be interpreted as

$$1 - F_u(w; \mathbf{x}) = P(W/u > z | W > u) \approx w^{-1/\gamma(\mathbf{x})}, \tag{19}$$

for large u and $w > 1$. Now, consider the relative excesses $V_j = W_i/u$ given $W_i > u$ for a large threshold u , where i is the index of the j th exceedance. Then, it seems natural to consider the strict Pareto distribution as the approximate distribution of the relative excesses, V_j . The maximum likelihood estimator of the parameter results in the Hill estimator in (10). However, if the strict Pareto approximation is poor, the Hill estimator has large bias, and hence, a second-order refinement is needed to address the departure from the strict Pareto distribution (see Beirlant et al., 2004). Assume that the *s.v* function, ℓ , satisfies the second-order assumption:

Assumption I: There exists a real constant $\rho < 0$ and a rate function b satisfying $b(w) \rightarrow 0$ as $w \rightarrow \infty$, such that for all $\lambda \geq 1$,

$$\lim_{w \rightarrow \infty} \frac{\log \ell(\lambda w) - \log \ell(w)}{b(w)} = \kappa_\rho(\lambda) \tag{20}$$

where $\kappa_\rho(\lambda) = \int_1^\lambda u^{\rho-1} du$ (Beirlant and Goegebeur, 2003, page 602).

Then, from (20), we can write (18) as

$$\lim_{u \rightarrow \infty} \frac{1 - F(uw; \mathbf{x})}{1 - F(u; \mathbf{x})} = w^{-1/\gamma(\mathbf{x})} \left(1 - \frac{b(u)}{\tau(\mathbf{x})} \left(w^{-\tau(\mathbf{x})} - 1 \right) + o(b(u)) \right), \tau(\mathbf{x}) > 0, \tag{21}$$

where b is regularly varying with index $-\tau(\mathbf{x})$. Ignoring the error term, (21) becomes a mixture of two Pareto distributions. The survival function for such a distribution is given by

$$1 - G(w; \mathbf{x}) = (1 - c(\mathbf{x})) w^{-1/\gamma(\mathbf{x})} + c(\mathbf{x}) w^{-(1/\gamma(\mathbf{x}) + \tau(\mathbf{x}))} \tag{22}$$

where $c(\mathbf{x}) \in (-1/\tau(\mathbf{x}), 1)$, $\tau(\mathbf{x}) > 0$ and $w > 1$. In practice, the perturbed Pareto distribution is fitted to the relative excesses, V_j , $j = 1, \dots, k_{n,x}$, and the parameters of the distribution can be estimated through the maximum

likelihood method. The resulting estimator is denoted by $\hat{\gamma}^{(PPD)}(V, k_{n,x}, m_{n,x})$. Similar to Ndao et al. (2014), the estimators of the the extreme value index from (i) through to (iv) are adapted to censoring using (14).

Furthermore, we extend the two estimators of the extreme value index introduced in Worms and Worms (2014) when observations are subject to random censoring to the case where covariate information is available.

(v) The first estimator is given by

$$\hat{\gamma}^{(c, WW, KM)}(W, k_{n,x}, m_{n,x}) := \frac{1}{n \left(1 - \hat{F}(W_{m_{n,x}-k_{n,x}, m_{n,x}})\right)} \times \sum_{j=1}^{k_{n,x}} \frac{\delta_{m_{n,x}-j+1, m_{n,x}}}{1 - \hat{G}(W_{m_{n,x}-j+1, m_{n,x}}^-)} \times \log \left(\frac{W_{m_{n,x}-j+1, m_{n,x}}}{W_{m_{n,x}-k_{n,x}, m_{n,x}}} \right), \quad (23)$$

where \hat{F} and \hat{G} are respectively the Kaplan-Meier estimators for F and G given by

$$1 - \hat{F}(b) = \Pi_{W_{j, m_{n,x}} \leq b} \left(\frac{m_{n,x} - j}{m_{n,x} - j + 1} \right)^{\delta_{j, m_{n,x}}} \quad (24)$$

and

$$1 - \hat{G}(b) = \Pi_{W_{j, m_{n,x}} \leq b} \left(\frac{m_{n,x} - j}{m_{n,x} - j + 1} \right)^{1 - \delta_{j, m_{n,x}}}, \quad (25)$$

for $b < W_{m_{n,x}, m_{n,x}}$. Here, $\hat{G}(W_{m_{n,x}-j+1, m_{n,x}}^-)$ is defined as a function of the form $g(w^-) = \lim_{\nu \rightarrow w} g(\nu)$.

(vi) The second alternative estimator is a weighted version of the Hill-type estimator (23),

$$\hat{\gamma}^{(c, WW, KL)}(W, k_{n,x}, m_{n,x}) := \frac{1}{n \left(1 - \hat{F}(W_{m_{n,x}-k_{n,x}, m_{n,x}})\right)} \times \sum_{j=1}^{k_{n,x}} \frac{1}{1 - \hat{G}(W_{m_{n,x}-j+1, m_{n,x}}^-)} \times j \log \left(\frac{W_{m_{n,x}-j+1, m_{n,x}}}{W_{m_{n,x}-k_{n,x}, m_{n,x}}} \right). \quad (26)$$

In the next section, the performance of the existing and proposed estimators will be compared via a simulation study.

3. Simulation Study

In this section, we present a simulation study to assess the performance of the estimators discussed in the previous section as the asymptotic distribution of most of the estimators are unknown.

3.1. Design

Several sample sizes were considered in the simulation and for simplicity, we report on the simulation for samples of size $n = 2000$ generated from three distributions i.e. Burr, Fréchet and Pareto only. For each distribution, our interest is in the estimation of the conditional extreme value index (EVI) function, $\gamma_1(x) = e^{(\beta_0 + \beta_1 x)}$. Here, β_0 and β_1 were chosen as -0.11 and -2.90 such that the values of $\gamma_1(x)$ for $x \in \text{Uniform}(0, 1)$ are within $(0, 1)$. This range of values of the extreme value index is the most common in extreme value theory literature for simulation studies and practical applications (see for e.g. [Gilli and Këllezli, 2006](#); [Gomes and Neves, 2011](#); [Einmahl et al., 2008](#); [Stupfler, 2016](#)). In particular, we selected values of x equal to, 0.12, 0.37 and 0.75, corresponding respectively to $\gamma_1(x)$ values, 0.63 (large), 0.31 (medium) and 0.10 (small). The choice of parameter functions for each distribution are presented in Table 1.

Table 1. Distributions with parameters as a function of x

Distribution	$1 - F(y; x)$	Parameter function			
		$\tau(x)$	$\lambda(x)$	$\alpha(x)$	$\gamma_1(x)$
Burr	$\left(\frac{\eta(x)}{\eta(x) + y^{\tau(x)}}\right)^{\lambda(x)}$	2	$0.5e^{(0.11 + 2.90x)}$	NA	$e^{-(0.11 + 2.90x)}$
Pareto	$y^{-\alpha(x)}$	NA	1	$e^{(0.11 + 2.90x)}$	$e^{-(0.11 + 2.90x)}$
Fréchet	$1 - \exp\left(-y^{-\alpha(x)}\right)$	NA	NA	$e^{(0.11 + 2.90x)}$	$e^{-(0.11 + 2.90x)}$

Note: $\eta(x)$ the scale parameter was taken as 1. Also, the Pareto distribution is a limiting case of the Burr distribution with $\lambda(x) = 1$.

In addition, the distribution of C is chosen such that the percentage of censoring in the right tail is 10%, 35% and 55%. The performance measures used for examining the estimators of $\gamma_1(x)$ are Mean Square Error (MSE) and median bias (hereafter referred to as bias).

The following algorithm was implemented to obtain the performance measures:

- A1 Generate n ($n = 2000$) random observations from, $x \sim \text{Uniform}(0, 1)$.
- A2 Generate n , random samples from the distributions of Y and C with parameters, $\gamma_1(x)$ and $\gamma_2(x)$, respectively. To maintain an approximately equal percentage of censoring in each sample, γ_2 is chosen as $\gamma_2(x) = \gamma_1(x)\varphi(x)/(1 - \varphi(x))$, where $\varphi(x)$ is the percentage of noncensored observations.
- A3 Let $Z_i = \min\{Y_i, C_i\}$ and $\delta_i = \mathbb{I}\{Y_i \leq C_i\}$, $i = 1, \dots, n$ to obtain the triplets (Z_i, δ_i, x_i) , $i = 1, 2, \dots, n$.

- A4 Choose a covariate value of interest, $x^* \in [0, 1]$, window size, h , and obtain the observations (Z_i, δ_i) , $i = 1, 2, \dots, n^*$, with its x_i values falling within the window $[x^* - h, x^* + h]$, where n^* is the number of observations within the window.
- A5 Compute an estimate of $\gamma_1(x^*)$ using $\hat{\gamma}_1^{(c,\cdot)}(x^*)$, at each number of top order statistics $k \in \{5, \dots, n^*\}$ for the sample in A4.
- A6 Repeat A1-A5 a large number of times, R ($R = 1000$), to obtain $\hat{\gamma}_1^{(c,\cdot)}(x^*) = (\hat{\gamma}_{1,1}^{(c,\cdot)}(x^*), \dots, \hat{\gamma}_{1,R}^{(c,\cdot)}(x^*))'$ at each k .
- A7 At each k value, compute the median bias

$$\text{bias} \left(\hat{\gamma}_1^{(c,\cdot)}(x^*) \right) = \text{median} \left(\hat{\gamma}_1^{(c,\cdot)}(x^*) \right) - \gamma_1(x^*) \quad (27)$$

and the MSE,

$$MSE \left(\hat{\gamma}_1^{(c,\cdot)}(x^*) \right) = \frac{1}{R} \sum_{i=1}^R \left[\hat{\gamma}_{1,i}^{(c,\cdot)}(x^*) - \gamma_1(x^*) \right]^2. \quad (28)$$

3.2. Results and Discussion

The results of the simulation study are presented in this section. For brevity and ease of presentation, we present the results for Burr distribution and leave that of Pareto and Fréchet to the appendix.

Figures 1, 2 and 3 show the results for estimators of $\gamma_1(x) = 0.63$ ($x = 0.10$), $\gamma_1(x) = 0.31$ ($x = 0.37$) and $\gamma_1(x) = 0.10$ ($x = 0.75$) respectively. From these figures, it can easily be seen that most of the estimators' performance diminish as k increases. This is expected as more intermediate observations are included in the estimation leading to bias. In addition, we observed that the bias and to a larger extent MSE, increases with decreasing value of $\gamma_1(x)$. Furthermore, the performance of the estimators of $\gamma_1(x)$ decreases as the censoring percentage increases. This is in conformity to the findings in Ndao et al. (2014).

We now turn attention to the performance of the individual estimators. Firstly, we found that the Hill estimator has large MSE and bias as k increases. This is in contrast to the simulation results and Corollary 4.2 in Ndao et al. (2014). Thus, we may conclude that the performance of the Hill estimator depends on the choice of parameter function, $\gamma_1(x)$. However, this result is consistent with the performance of the Hill estimator in the case where there is no covariate information nor censoring (see e.g. Beirlant et al., 2004, and references therein).

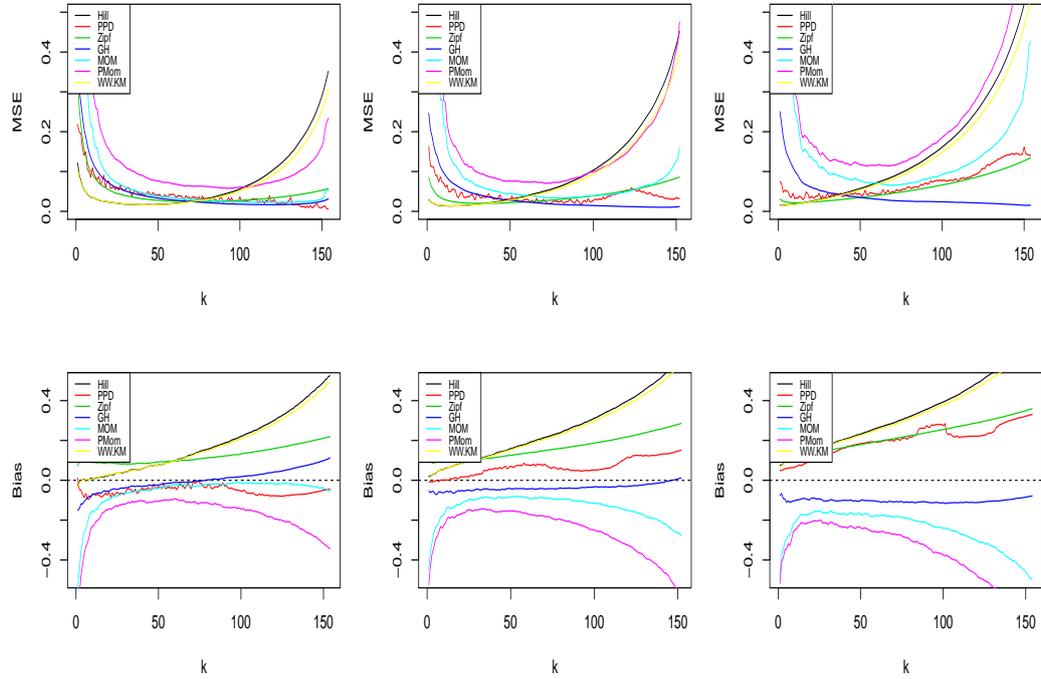


Fig. 1. Results for Burr distribution with $\varphi = 0.1$. Leftmost column: $\gamma_1(x) = 0.63$ ($x = 0.12$); Middlemost column : $\gamma_1(x) = 0.31$ ($x = 0.37$); Rightmost column: $\gamma_1(x) = 0.10$ ($x = 0.75$);

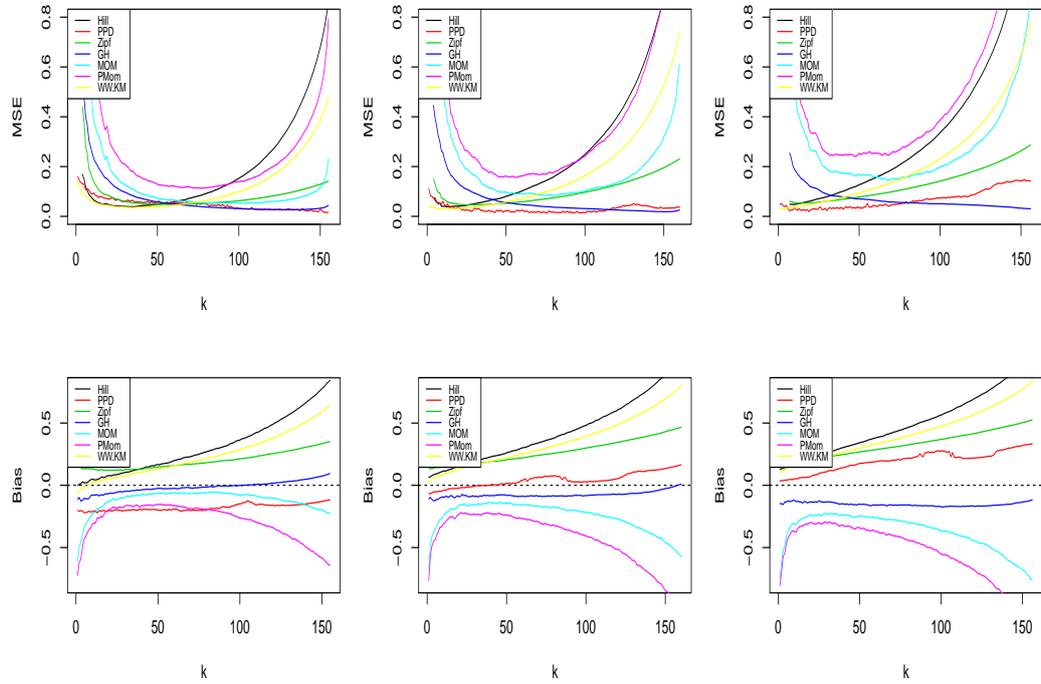


Fig. 2. Results for Burr distribution with $\varphi = 0.35$. Leftmost column: $\gamma_1(x) = 0.63$ ($x = 0.12$); Middlemost column : $\gamma_1(x) = 0.31$ ($x = 0.37$); Rightmost column: $\gamma_1(x) = 0.10$ ($x = 0.75$);

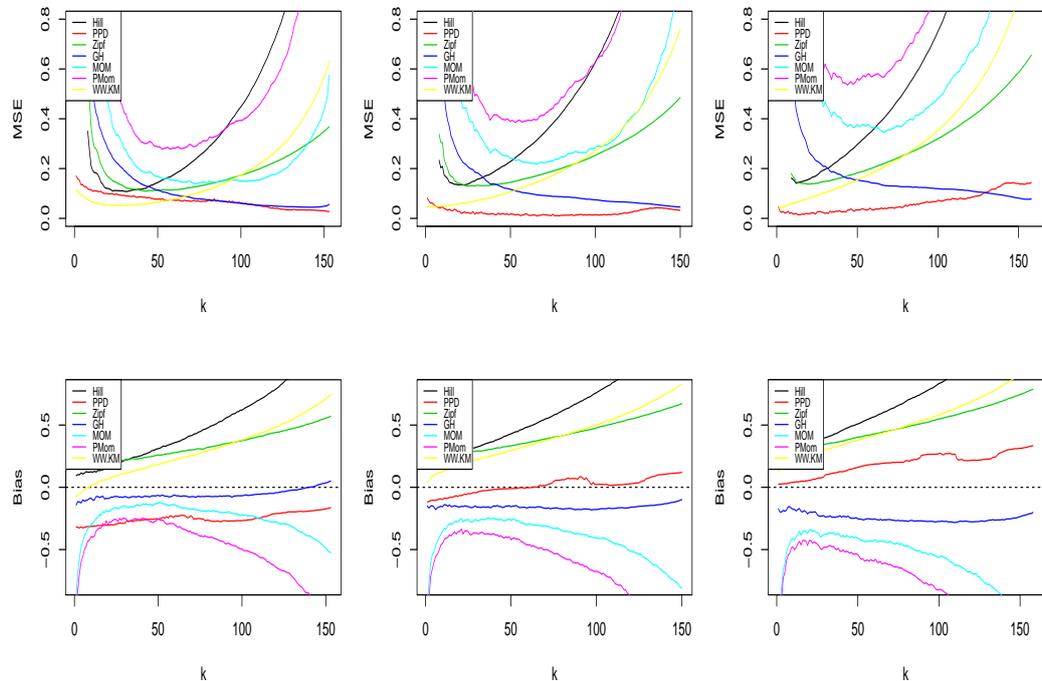


Fig. 3. Results for Burr distribution with $\varphi = 0.55$. Leftmost column: $\gamma_1(x) = 0.63$ ($x = 0.12$); Middlemost column : $\gamma_1(x) = 0.31$ ($x = 0.37$); Rightmost column: $\gamma_1(x) = 0.10$ ($x = 0.75$);

For samples generated from the Burr distribution, the PPD, GH and MOM estimators are the most robust to censoring in most cases. These estimators have smaller bias and MSE compared to the other estimators of $\gamma_1(x)$. In particular, the proposed PPD estimator is seen to have the smallest bias and MSE as the percentage of censoring increases.

With regard to the Pareto distribution, the performance of the estimators are similar to that of the Burr distribution. However, some differences occur which are mentioned. Firstly, the PMom estimator competes with the best three estimators i.e. PPD, MOM and GH in terms of bias and MSE. Secondly, in the case, of large censoring, the PPD estimator has the smallest MSE and relatively good bias.

In the case of samples generated from the Fréchet distribution, all the estimators have good MSE values with the exception of small and large k values where some of the estimators' performance deteriorate. In terms of bias, the performance does not differ significantly from that of MSE. Interestingly, the Hill estimator is seen to compete with the PPD, Zipf and WW.KM in terms of bias and MSE as the

percentage of censoring and k increase.

Overall, we found that the performance of estimators depend on the distribution, the number of top order statistics and the percentage of censoring. Contrary to what was reported in Ndao et al. (2014), we found that the Hill estimator has large bias and MSE except for samples generated from the Fréchet distribution. The proposed PPD estimator is universally competitive in estimating $\gamma_1(x)$ regardless of its size, percentage of censoring and number of top order statistics.

4. Conclusion

In this paper, the central issues were a review and proposals of estimators of conditional extreme value index when observations are subject to right random censoring. In the latter, we proposed adapting some classical extreme value index to censoring and presence of covariate information. The existing and the proposed estimators were compared in a simulation study. Some interesting results were obtained and are outlined as follows:

1. The performance of the estimators depend on the underlying distribution of the sampled data. Thus, no estimator was universally the best under all the simulation conditions considered. However, a closer look at the results reveal that some general conclusions can be reached on some estimators that can be considered as appropriate for estimating $\gamma_1(x)$.
2. The performance of the estimators depend on the size of $\gamma_1(x)$: Bias and MSE values were generally larger for small $\gamma_1(x)$ values and smaller for larger values of $\gamma_1(x)$. Therefore, we recommend that in practice before proceeding to use any estimator for $\gamma_1(x)$, one should assess the potential range of the true value of $\gamma_1(x)$. In such cases, several estimators can be considered and a median or average value used as an estimate to help in the selection of the preferred estimator.
3. The performance of the estimators generally deteriorates as the percentage of censoring increases. However, the estimators that are robust to censoring, to a larger extent, maintain their performance under increased percentage of censoring. These estimators include PPD, MOM and PMom and GH. Conspicuously missing is the Hill estimator which was shown to have large bias and MSE except for samples from the Fréchet distribution.
4. Overall, we found that the proposed PPD estimator mostly has the best MSE and bias behaviour under small and heavy censoring. In addition, it has less bias as the number of top order statistics, k , increases.

Some additional research is needed to establish the consistency and asymptotic normality of the PPD estimator. This will enhance statistical inference and will be considered in future research.

References

- Ameraoui A, Boukhetala K, Dupuy J. F. (2016). Bayesian estimation of the tail index of a heavy tailed distribution under random censoring. *Computational Statistics & Data Analysis*. 104:148–168.

- Beirlant, J. and Goegebeur, Y. (2003). Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis*, 42:595–619.
- Beirlant, J. and Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis*, 89(1):97–118.
- Beirlant, J., Guillou, A., Dierckx, G., and Fils-Villetard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, 10:151–174.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). *Statistics of extremes: Theory and applications*. Wiley, England.
- Beirlant J, Guillou A, Toulemonde G. (2010). Peaks-Over-Threshold modeling under random censoring. *Communications in Statistics - Theory and Methods*. 39(7):1158–1179.
- Beirlant J, Maribe G, Verster A. (2017) Penalized bias reduction in extreme value estimation for censored Pareto-type data , and long-tailed insurance applications. *ArXiv preprint arXiv17050663v1*, 1:1–21.
- Beirlant, J., Vynckier, P., and Teugels, J. L. (1996). Excess functions and estimation of the extreme-value index. *Bernoulli*, 2(4):293–318.
- Brahimi B, Meraghni D, Necir A. (2013). On the asymptotic normality of Hill's estimator of the tail index under random censoring. *ArXiv preprint arXiv13021666v1*, 1:1–11.
- Danielsson, J., Jansen, D. W., and de Vries, C. G. (1996). The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics - Theory and Methods*, 25(4):711–720.
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 52(3):393–442.
- Deheuvels, P., de Haan, L., Peng, L., and Pereira, T. T. (1997). Comparison of extreme value index estimators. Technical Report T400:EUR-09, The Erasmus University Rotterdam, Rotterdam.
- Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855.
- de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. PhD, University of Amsterdam.
- de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52:60–70.
- de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1):60–70.
- Diop, A. and Lo, G. S. (2006). Generalized Hill's Estimator. *Far East J. Theoretical Statistics*. 20(2): 129–149.
- Diop, A. and Lo, G. S. (2009). Ratio of Generalized Hill's Estimator and its asymptotic normality theory. *Math. Method. Statist.* , 18(2): 117–133.
- Dème, E., Lo, G.S. and Diop, A. (2012). On the generalized Hill process for small parameters and applications. *Journal of Statistical Theory and Applications*. 11(4): 397–418.

- Einmahl, J. H. J., Elie, A. M., and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli*, 14(1):207–227.
- Fisher, R. and Tippett, L. (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:80–190.
- Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99(10):2368–2388.
- Gilli, M. and Këllezli, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(1):1–23.
- Gnedenko, B. (1943). Sur la distribution limite du terme Maximum d'une série aléatoire. *Annals of Mathematics*, 44(3):423–453.
- Gomes, M. I. and Guillou, A. (2014). Extreme value theory and statistics of univariate extremes : a review. *International Statistical Review*, pages 1–35.
- Gomes, M. I., Luísa, C. e. C., Fraga Alves, M. I., and Pestana, D. (2008). Statistics of extremes for IID data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions. *Extremes*, 11(1):3–34.
- Gomes, M. I. and Neves, C. (2011). Estimation of the extreme value index for randomly censored data. *Biometrical Letters*, 48(1):1–22.
- Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174.
- Kratz, M. and Resnick, S. I. (1996). The qq-estimator of the index of regular variation. *Communications in Statistics: Stochastic Models*, 12:699–724.
- Mason, D. M. . (1982). Laws of large numbers for sums of extreme values. *Annals of Probability*, 10:754–764.
- Ndao, P., Diop, A., and Dupuy, J. F. (2014). Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. *Computational Statistics and Data Analysis*, 79:63–79.
- 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.
- Ngom, M., Lo, G. S. (2016) A Double-indexed Functional Hill Process and Applications. *Journal of Mathematics Research*: 8(4): 144-165.
- Pickands, J. (1975). Statistical Inference using extreme order statistics. *Annals of Statistics*. 3: 119–131.
- Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, 144:1–24.
- Tsourti, Z. and Panaretos, J. (2003). Extreme Value Index Estimators and Smoothing Alternatives: Review and Simulation Comparison. In Panaretos, J., editor, *Stochastic Musings: Perspectives From the Pioneers of the Late 20th Century*, pages 141–160. Mahwah, New Jersey.
- Wang, H. and Tsai, C.-L. (2009). Tail index regression. *Journal of the American Statistical Association*, 104(487):1233–1240.
- 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.

Appendix A: Pareto Distribution

For each figure, the following description of the panels apply. Leftmost column: $\gamma_1(x) = 0.63$ ($x = 0.12$); Middlemost column : $\gamma_1(x) = 0.31$ ($x = 0.37$); Rightmost column: $\gamma_1(x) = 0.10$ ($x = 0.75$);

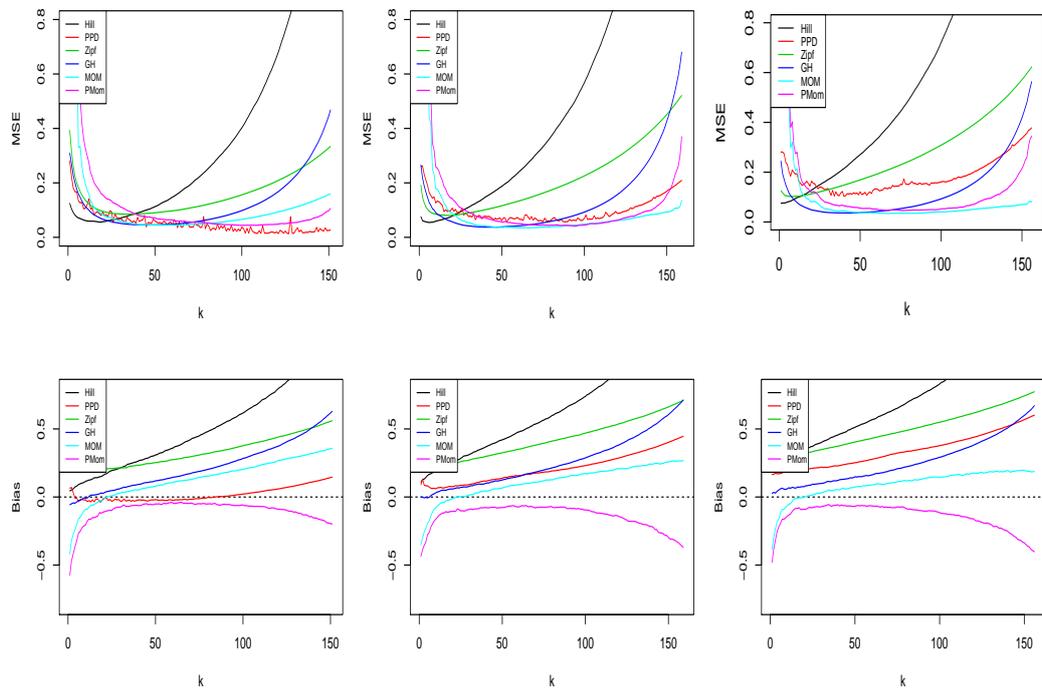


Fig. A1. Results for Pareto distribution with $\varphi = 0.1$.

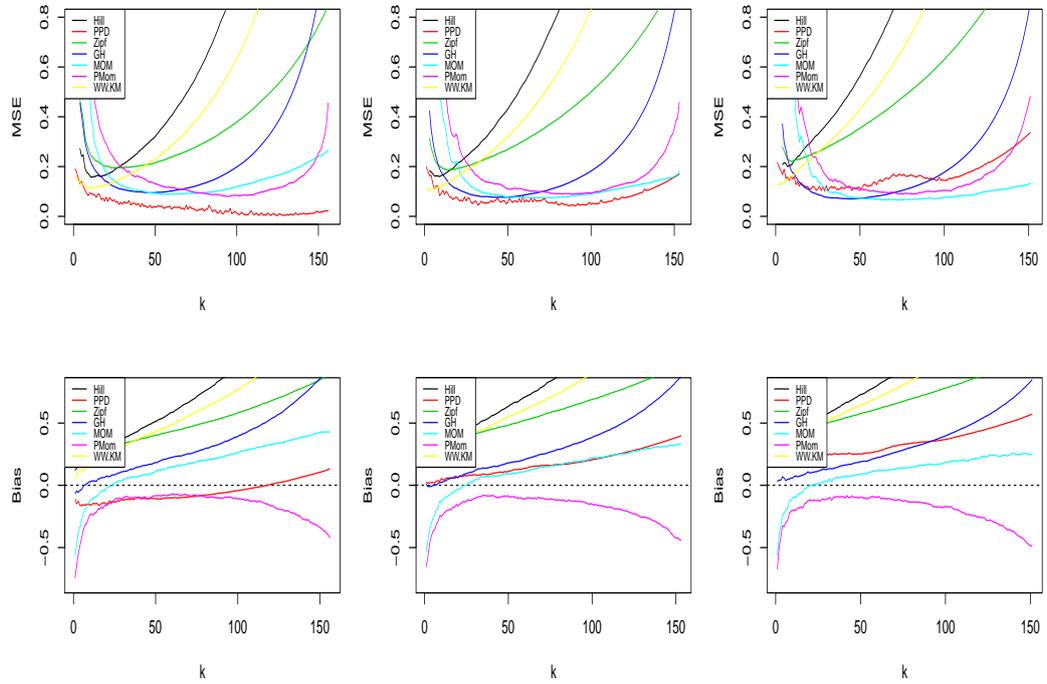


Fig. A2. Results for Pareto distribution with $\phi = 0.35$.

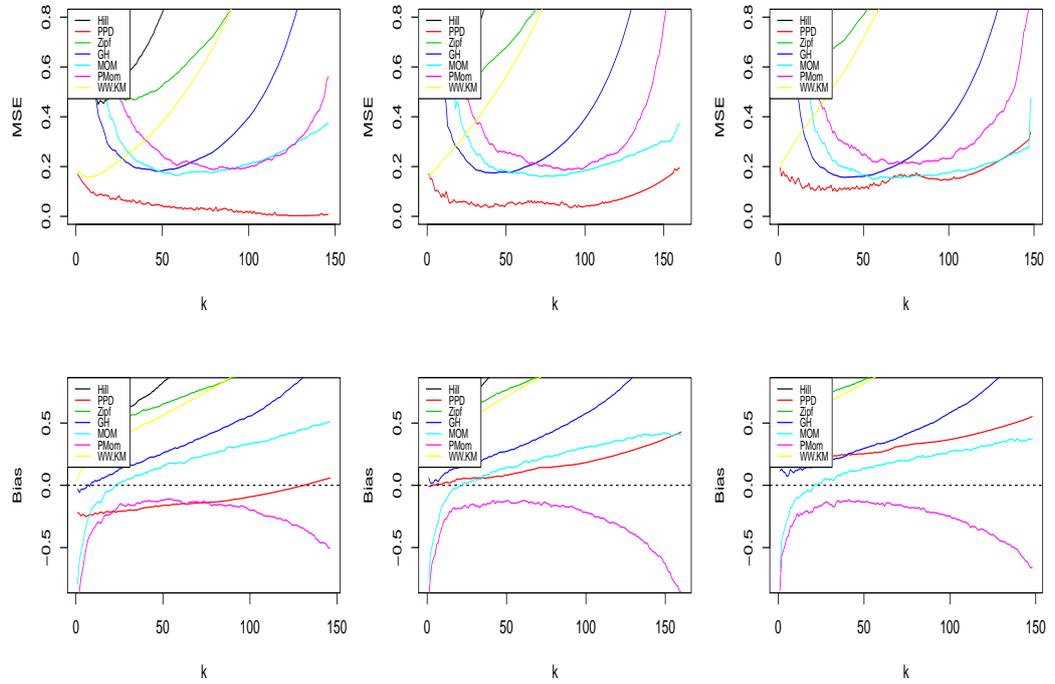


Fig. A3. Results for Pareto distribution with $\phi = 0.55$.

Appendix B: Fréchet Distribution

For each figure, the following description of the panels apply. Leftmost column: $\gamma_1(x) = 0.63$ ($x = 0.12$); Middlemost column : $\gamma_1(x) = 0.31$ ($x = 0.37$); Rightmost column: $\gamma_1(x) = 0.10$ ($x = 0.75$);

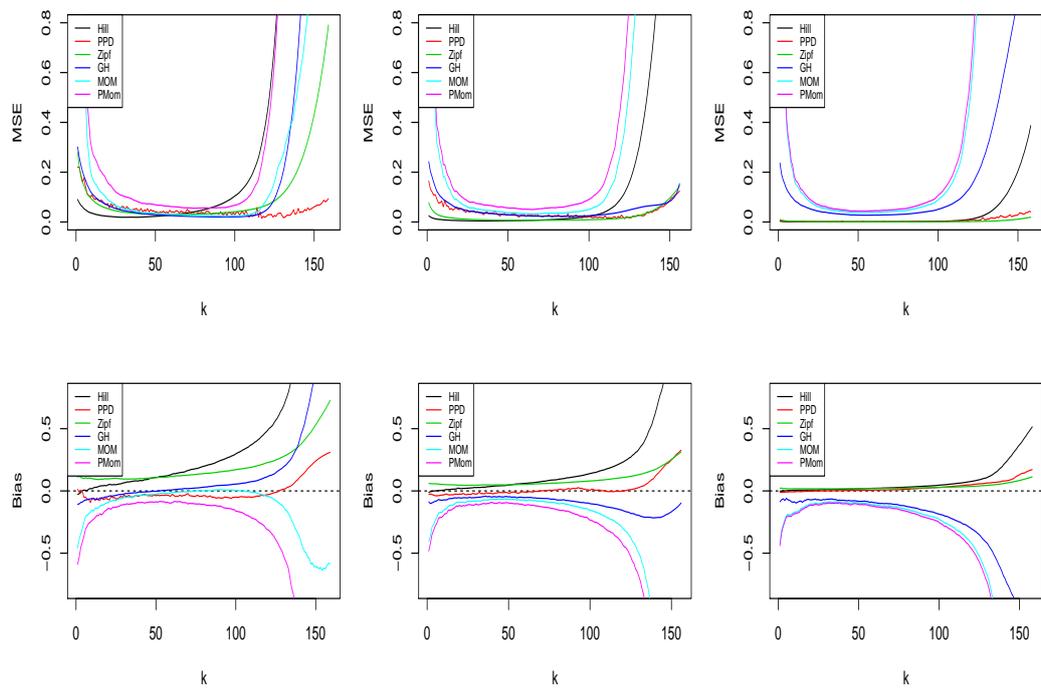


Fig. B1. Results for Fréchet distribution with $\varphi = 0.1$.

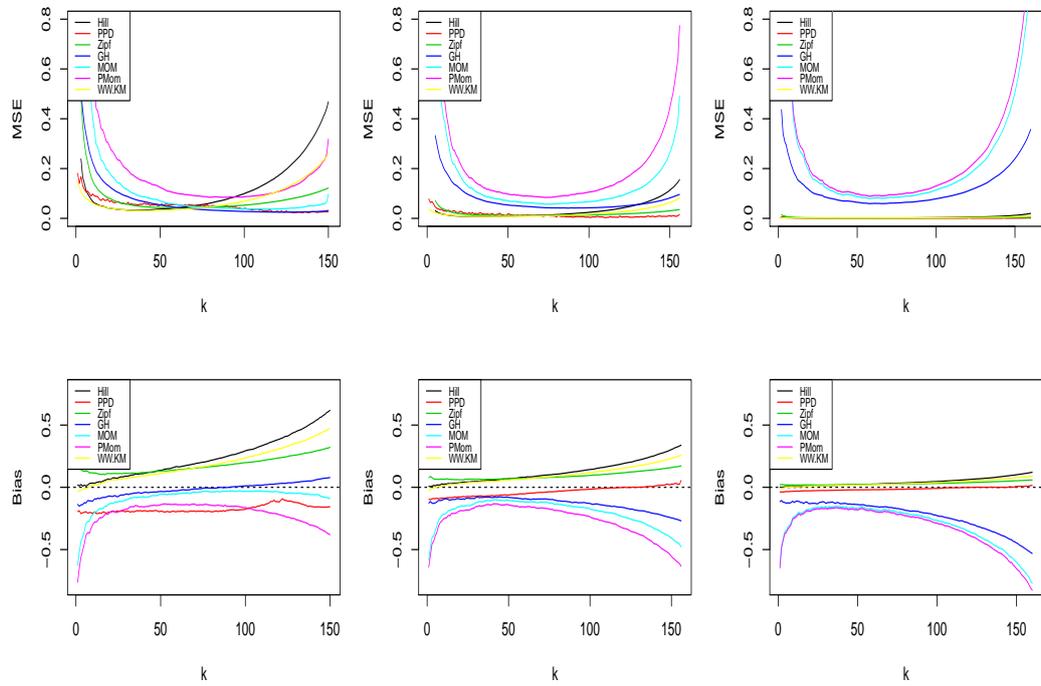


Fig. B2. Results for Fréchet distribution with $\phi = 0.35$.

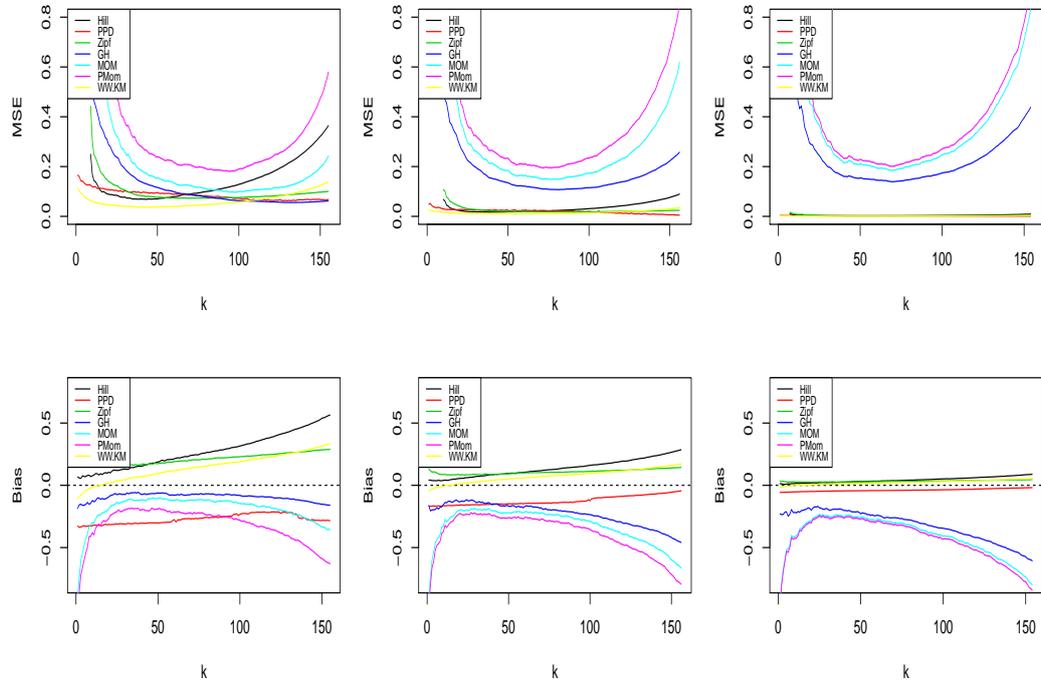


Fig. B3. Results for Fréchet distribution with $\phi = 0.55$.