"Mircea cel Batran" Naval Academy Scientific Bulletin, Volume XVIII – 2015 – Issue 2
Published by "Mircea cel Batran" Naval Academy Press, Constanta, Romania // The journal is indexed in:
PROQUEST SciTech Journals, PROQUEST Engineering Journals, PROQUEST Illustrata: Technology, PROQUEST Technology Journals, PROQUEST Military Collection PROQUEST Advanced Technologies & Aerospace

A MODIFICATION OF HILL'S TAIL INDEX ESTIMATOR

L. GLAVAŠ[1]
J. JOCKOVIĆ[2]
P. MLADENOVIĆ[3]
[1, 2, 3]University of Belgrade, Faculty of Mathematics, Belgrade, Serbia

Abstract: *In this paper, we study a class of tail index estimators that contains the well-known Hill's estimator. We propose an estimator that, in several cases, has smaller mean squared error than commonly used estimators (obtained by Hill (1975), Pickands (1975), Dekkers, Einmahl and de Haan (1989) and C. de Vries), and confirm these findings in a simulation study.*

Keywords: *Hill's estimator, Mean squared error, Regular variation, Tail index.*

INTRODUCTION

Heavy-tailed distributions often appear to be adequate in statistical modelling of real data. Several classes of such distributions have been introduced and studied. In this paper we focus on the class of distributions with regularly varying (right) tail.

Let $X_1, X_2, \dots, X_n$ be a sample of i.i.d. or possibly dependent random variables, with the same marginal distribution function (cdf) $F(x) = P\{X_k \leq x\}$. We assume that the tail $1 - F(x) = P\{X_k > x\}$ is regularly varying at ∞, that is, for all $x > 0$,

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \qquad (1.1)$$

where $\alpha > 0$ is the index of regular variation. Equality (1.1) can be written equivalently as follows

$$1 - F(x) = x^{-\alpha} L(x), \qquad x > 0, \qquad (1.2)$$

for some slowly varying function $L(x)$.

Distribution functions that allow representation (1.2) for some $\alpha > 0$ have applications in finance, insurance, meteorology, telecommunications, and many other fields (see, e.g. [11], [24], [25], [30]).

Without loss of generality we shall restrict our considerations to the class of cdfs that additionally satisfy the condition $F(0) = 0$.

The problem of statistical estimation of the tail index α in (1.2) has been addressed by many authors, and several estimators have been proposed. One of the best known estimators was introduced by Hill [16] and defined by a portion of extreme order statistics as follows

$$\begin{aligned} H_n^{-1} &= \frac{1}{m} \sum_{k=1}^{m} \ln \frac{X_{(k)}}{X_{(m+1)}} \\ &= \frac{1}{m} \sum_{k=1}^{m} (\ln X_{(k)} - \ln X_{(m+1)}), \end{aligned} \qquad (1.3)$$

where $X_1, X_2, \dots, X_n$ is a size n sample of random variables with the cdf satisfying (1.2), $X_{(n)} \leq \dots \leq X_{(2)} \leq X_{(1)}$ the corresponding order statistics, and $m = m(n)$ is a sequence of positive integers such that $1 \leq m \leq n$ for each integer $n \in \mathbb{N}$, and $m \to \infty$, $m/n \to 0$, as $n \to \infty$. Consistency of Hill's estimator was proved by Mason [19] under conditions $m \to \infty$ and $m/n \to 0$. Strong consistency was obtained by Deheuvels, Häusler and Mason [9] under conditions $m/n \to 0$ and $m/\ln\ln n \to \infty$, as $n \to \infty$. Under certain additional conditions it was proved that Hill's estimator is asymptotically normal, see Davis and Resnick [6], Csörgö, Deheuvels and Mason [5], Häusler and Teugels [13], Goldie and Smith [12], Beirlant and Teugels [1]. Other well-known estimators of the tail index were proposed by Pickands [23], Dekkers, Einmahl and de Haan [7]. The problem of tail index estimation was also addressed by Hall [14], DuMouchel [10]. A very recent study of this problem is given in the paper by Brilhante, Gomes and Pestana [3].

Extreme value distributions in γ-parametrization are given by

$$G_\gamma(x) = \exp\{-(1 + \gamma x)^{-1/\gamma}\}, \qquad (1.4)$$

where $\gamma \in \mathbb{R}$ and $1 + \gamma x > 0$. Here $\gamma = 1/\alpha > 0$ corresponds to the class of Fréchet extreme value distributions, while $\gamma < 0$ corresponds to the class of Weibull extreme value distributions. The Pickands estimator can be used both for positive and negative values of the tail index γ. Asymptotic properties of Pickands estimator were studied by Pickands [23], Smith [31], Dekkers and de Haan [8], Boss [2]. The cases of possibly dependent random variables and missed observations were also considered, see, e.g. Hsing [18], Resnick and Stărică [27], [28], [29], Resnick [26], Mladenović and Piterbarg [22], Hill [17].

FORMULATION OF THE PROBLEM

Although several estimators of the tail index $\alpha > 0$ were researched extensively in the last couple of decades, there were not many conclusions about the optimality of the estimation. De Haan and Peng [15] compared various estimators (including those proposed by Hill [16] and Pickands [23]) by calculating their asymptotic mean squared error after choosing the optimal number of upper order statistics that are involved in definition of these estimators. A detailed comparison of recently obtained modifications of Hill's estimator is given in [3].

The following question is, nevertheless, without the answer. How should the estimator of the tail index α with the minimum mean squared error be defined, in general? Such general problem is probably very difficult to solve. In this paper we focus our attention to a class of estimators that will be defined below.

For simplicity reasons, in the remaining of the paper we work with the γ-parametrization, i.e. discuss estimation of the parameter $\gamma = \frac{1}{\alpha}$.

Hill's estimator given by (1.3) is the average of log-exceedances. We shall consider all convex combinations of these exceedances as possible estimators of the index of regular variation $\alpha > 0$, that is

$$\begin{aligned}\tilde{H}_n &= \sum_{k=1}^{m} c_k \ln \frac{X_{(k)}}{X_{(m+1)}} \\ &= \sum_{k=1}^{m} c_k (\ln X_{(k)} - \ln X_{(m+1)}),\end{aligned} \tag{2.1}$$

where $c_k \geq 0$ for $k \in \{1, 2, \dots, m\}$ and under the constraint $c_1 + c_2 + \cdots + c_m = 1$.

The main question we are interested in is the following one: What are the values of the non-negative coefficients $c_1, c_2, \dots, c_m$, $c_1 + c_2 + \cdots + c_m = 1$, that minimize the mean squared error

$$E\left(\tilde{H}_n - \gamma\right)^2 = E\left(\sum_{k=1}^{m} c_k \ln \frac{X_{(k)}}{X_{(m+1)}} - \gamma\right)^2 \tag{2.2}$$

of the estimator given by (2.1)? Explicit solution of this extremal problem is not easy to obtain and we shall start with a simulation study.

RESULTS OF SIMULATIONS

A. Global Optimization Procedure

Considered from the optimizational point of view, the problem formulated in Section II is, in general case, multimodal (it can possess several local optima). Therefore, in order to obtain an optimal solution, i.e. a set of constants $(c_1, c_2, \dots, c_m)$ such that the value of (2.2) is minimal, we should use a global optimization procedure. Our choice is the metaheuristic procedure called Variable Neighborhood Search (VNS) (Mladenović and Hansen [21]), which is simple and can be modified to deal with various types of problems, including this one.

The basic idea of VNS is to systematically change regions where local search is performed in order to avoid getting stuck in bad local optima. More detailed explanation of the procedure can be found in [21]. There are several variants of this metaheuristic. For example, General VNS (Mladenović et. al. [20]) is used for continuous optimization with bounded feasible region, and Gaussian VNS (Carrizosa et al. [4]) may be used for continuous optimization with unbounded domain.

In order to apply VNS, we have to define the starting point, the random distribution for the shaking step (a way of generating a random point in a current neighborhood) and the stopping criterion. In our case, the starting point is Hill's estimate $\left(c_1 = c_2 = \cdots = c_m = \frac{1}{m}\right)$, the random distribution for the shaking step is Uniform on the appropriate interval, and the stopping criterion is 20 local search calls performed.

All the computations related to this work were performed in MATLAB. As a local procedure, we used MATLAB local search routine fmincon, with stopping criteria set as follows:

(1) MaxFunEvals. The bound on the number of function evaluations is set to 4000.

(2) MaxIter. The bound on the number of solver iterations is set to 4000.

(3) TolFun. The lower bound on the change in the value of the objective function during a step has been set to $1e - 04$.

(4) TolX. The lower bound on the stepsize has been set to $1e - 04$.

(5) TolCon. The upper bound on the magnitude of each constraint function has been fixed as $1e - 04$.

It order to obtain optimal values of the constants $(c_1, c_2, \dots, c_m)$ in (2.2), we have performed the proposed VNS procedure on 100 simulated samples of size 1000, and 100 samples of size 10000 from the generalized Pareto model given by $W_\gamma(x) = 1 + \ln G_\gamma(x)$, where $G_\gamma(x)$ is the extreme value model defined in (1.4), and γ takes value in the set $\{0.1, 0.2, \dots, 1\}$. Number of upper order statistics used in the computation was $m \in \{50, 100, 200\}$ for the samples of size 1000, and $m \in \{50, 100, 200, 500\}$ for the samples of size 10000.

In all of the cases considered in the experiment, the optimization procedure ended successfully. Some of the results obtained, based on samples of size 1000 and m equal to 200 are summarized in Table I.

TABLE I.

Estimated values of the constant c_m and the corresponding maximum values of the remaining constants, based on 1000 samples of size 1000, where $m = 200$

γ	$\max\{c_1, c_2, \dots, c_{m-1}\}$	c_m
0.1	0.0179	0.8837
0.2	0.0178	0.7377
0.3	0.0372	0.6000
0.4	0.0451	0.5931
0.5	0.0381	0.4473

The results lead to the following conclusions:

- The optimal value of c_m is significantly (at least 10 times) greater than the optimal values of $c_1, c_2, \dots, c_{m-1}$ (which are close to zero) in cases in which $\gamma \in \{0.1, 0.2, \dots, 0.5\}$. For other values of γ, no such link is found.
- In all cases considered, the optimal set of constants $(c_1, c_2, \dots, c_m)$ is different from the starting point (Hill's estimator), which means that the estimator defined by this

optimal set has smaller mean squared error than Hill's estimator.

The same conclusion holds for the rest of the cases considered in the experiment.

B. Second Phase of the Optimisation Procedure

Simulation results obtained in Subsection III.A indicate that only the value c_m has a significant impact on the optimization procedure. Therefore, we consider a simplified optimization problem: What are the values of the non-negative coefficients $c_1, c_2, \dots, c_m$, which sum to one, $c_1 = c_2 = \dots = c_{m-1} = \frac{1-c_m}{m-1}$, and which, at the same time, minimize the mean squared error (2.2)? In other words, the question is how to choose c_m in order to minimize (2.2), since the other constants are then uniquely determined. Simulation procedure, analogous to the one in Subsection III.A is performed, and some of the results are presented in Table II. This and the other results of this stage of the experiment show that the optimal values of c_m decrease when γ increases.

TABLE II.

Estimated values of the constant c_m obtained in the second phase of the simulated experiment, based on 100 samples of size 1000, where $m = 200$

γ	c_m
0.1	0.7987
0.2	0.6140
0.3	0.4533
0.4	0.3140
0.5	0.1928
0.6	0.0910
0.7	0.0020
0.8	< 0.0001
0.9	< 0.0001
1.0	< 0.0001

C. New Estimator

Based on the conclusions from Subsection III.A and Subsection III.B, we propose a new estimator, defined as follows

$$\tilde{\pi}_n = \sum_{k=1}^{m} c_k \ln \frac{X_{(k)}}{X_{(m+1)}} = \sum_{k=1}^{m} c_k (\ln X_{(k)} - \ln X_{(m+1)}), \quad (3.1)$$

where

$$c_m = \frac{m}{nH_n} \text{ and } c_k = \frac{1-c_m}{m-1}, \quad (3.2)$$

for $k \in \{1, 2, \dots, m-1\}$.

In this case c_m is defined in terms of Hill's estimator H_n. The intuition behind this choice may be explained in the following way: in the second stage of the simulation experiment, it was observed that the optimal values of c_m decrease when γ increases. Since γ is an unknown parameter, we use H_n as the simplest estimator, to obtain an "approximate" value of γ.

D. Comparison of the Estimators

In this subsection we compare efficiency of the new estimator and the estimators suggested by:

(1) Hill: H_n

(2) Pickands: $P_n = \frac{1}{\ln 2} \ln \frac{X_{(m)} - X_{(2m)}}{X_{(2m)} - X_{(4m)}}$,

(3) Dekkers, Einmahl and de Haan:
$D_n = H_n + 1 - \frac{1}{2\left(1 - \frac{H_n^2}{M_n}\right)}$,

(4) C. de Vries (according to [15]): $V_n = \frac{M_n}{2H_n}$,

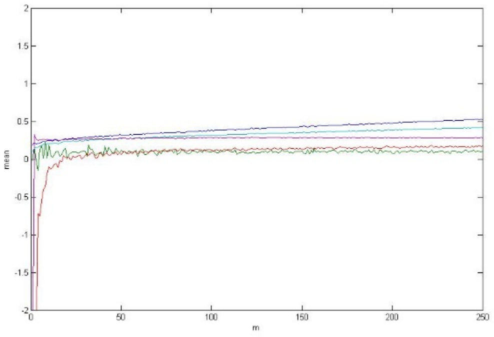
where $M_n = \frac{1}{m} \sum_{k=1}^{m} \left(\ln X_{(k)} - \ln X_{(m+1)}\right)^2$. This is done by simulating the data from several heavy tailed models, and comparing the efficiency of the estimators as functions of (true) tail index, sample size, or number of upper order extremes used in computation (m). Finally, we investigate the behavior of the estimators on one mixture model.

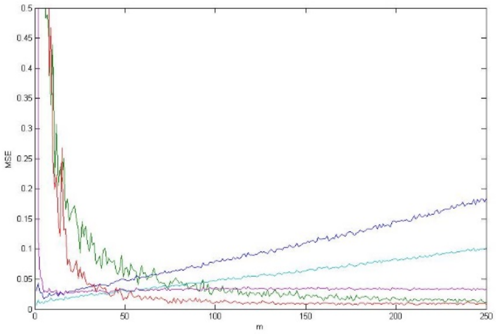
We simulate the following data:

i. 100 samples of size 1000 with $m \in \{50, 100, 200\}$ from each of the underlying models:
   - the extreme value model $G_\gamma(x)$, given by (1.4), for $0.1 \le \gamma \le 1$,
   - the generalized Pareto model given by $W_\gamma(x) = 1 + \ln G_\gamma(x)$, for $0.1 \le \gamma \le 1$,
   - the Student's $t_{1/\gamma}$ model for $0.1 \le \gamma \le 1$,
   - a mixture of 95% of $W_\gamma(x)$ and 5% of $W_{2\gamma}(x)$ distributions, for $0.1 \le \gamma \le 1$;

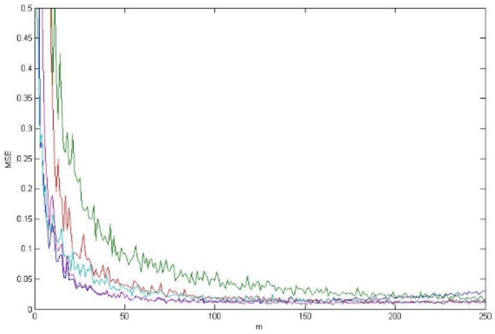
ii. 100 samples of size 10000 with m taking value in the set $\{50, 100, 200, 500\}$ from the same underlying models as in the previous case.

Next we visualize some results obtained by simulations because that is the easiest way to gain an insight into the quality of aforementioned estimators.

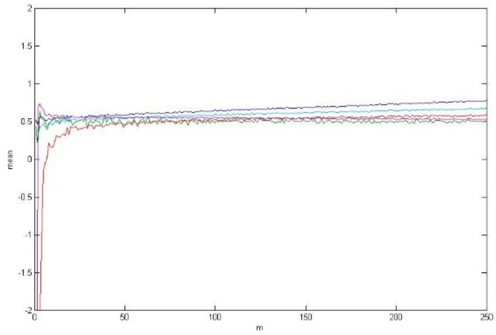


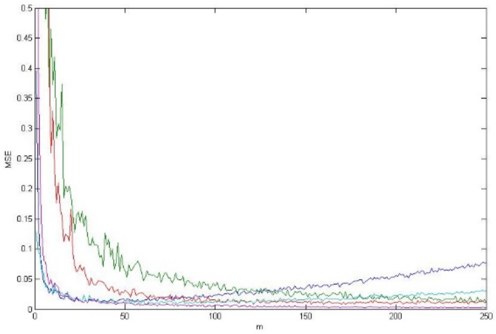


$\gamma = 0.1$



$\gamma = 1$

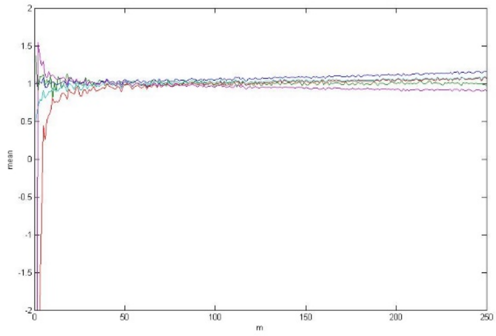


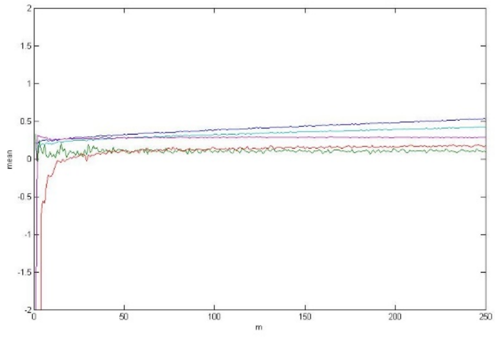


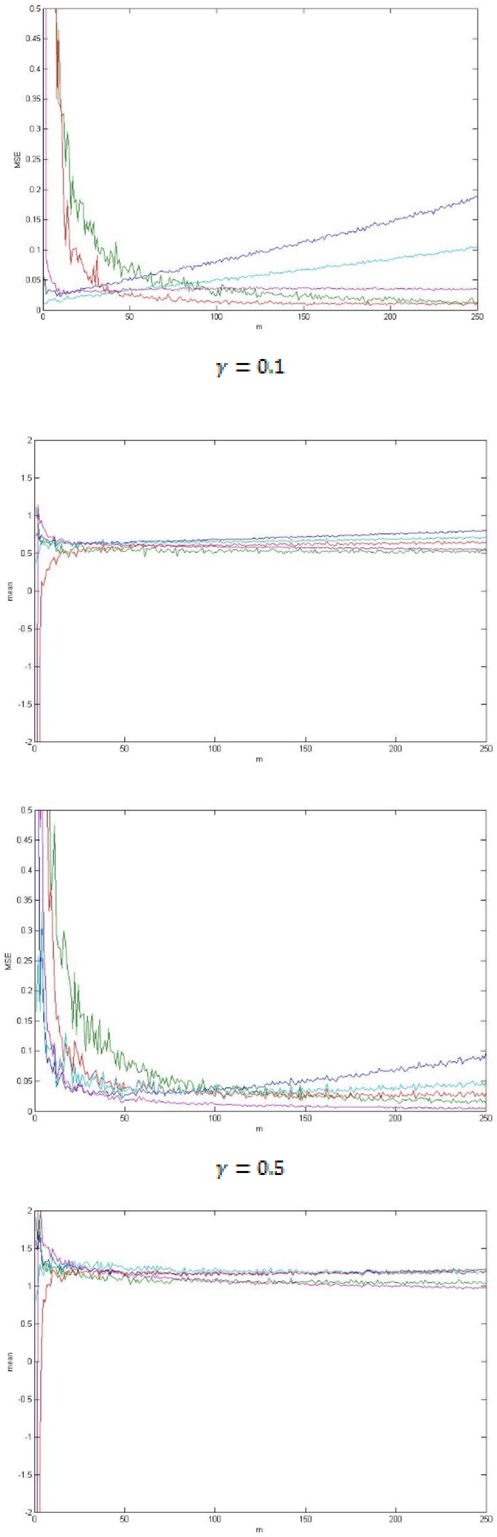
$\gamma = 0.5$

Figure 1. Mean value and mean squared error (MSE) of the estimators suggested by Hill (blue), Pickands (green), Dekkers, Einmahl and de Haan (red), C. de Vries (turquoise), and the new estimator (purple) as a function of m (computation based on 100 samples of size 1000 from $W_\gamma(x)$ model, where $\gamma \in \{0.1, 0.5, 1\}$, respectively, and $1 \leq m \leq 250$)

In Figures 1 and 2 we depict mean value and mean squared error for the estimators H_n, P_n, D_n, V_n, and the new estimator, as a function of m, for $\gamma \in \{0.1, 0.5, 1\}$, based on the samples simulated from the generalized Pareto distributions, respectively. We can conclude that the mean value of the new estimator is relatively stable, and close to the true value of the tail index. The mean squared error pattern indicates that the new estimator outperforms the others for greater values of γ. The same conclusions hold for the other two models (the extreme value distribution and the Student’s distribution).

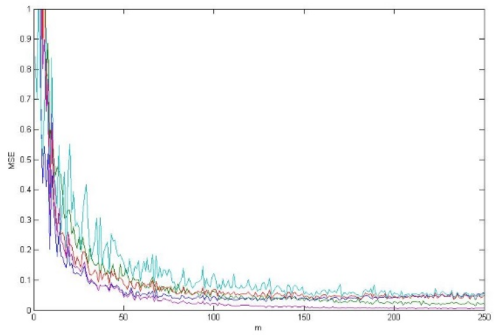






$\gamma = 0.1$

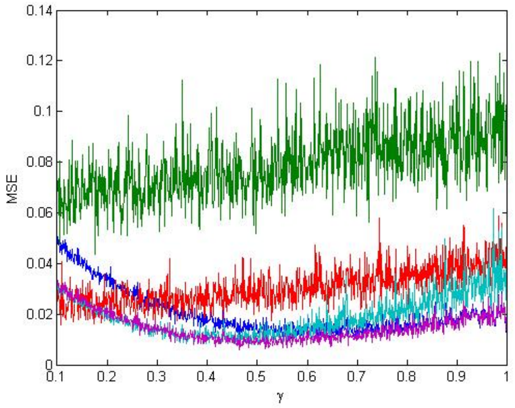
$\gamma = 0.5$



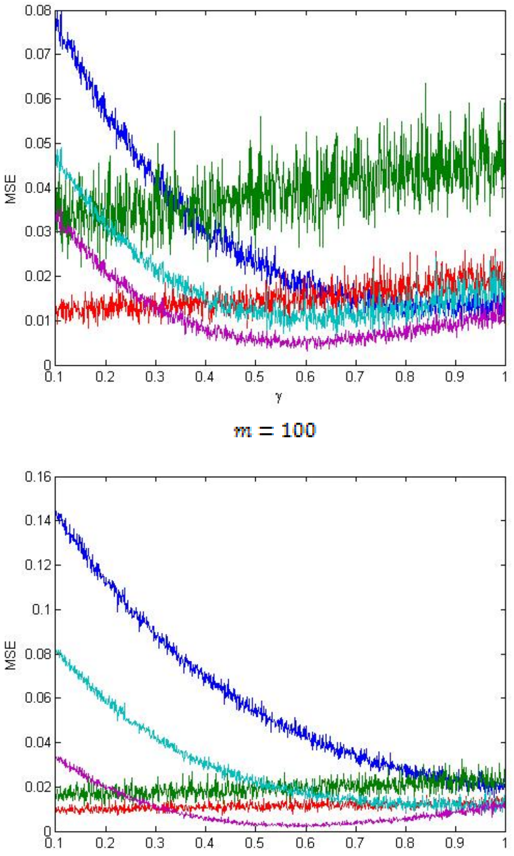
$\gamma = 1$

Figure 2. Mean value and mean squared error (MSE) of the estimators suggested by Hill (blue), Pickands (green), Dekkers, Einmahl and de Haan (red), C. de Vries (turquoise), and the new estimator (purple) as a function of m (computation based on 100 samples of size 1000 from mixture of $W_\gamma(x)$ and $W_{2\gamma}(x)$ models, where $\gamma \in \{0.1, 0.5, 1\}$, respectively, and $1 \le m \le 250$)

In Figure 3, we depict mean squared error as a function of γ, for $0.1 \le \gamma \le 1$, for the estimators H_n, P_n, D_n, V_n, and the new estimator, based on the samples simulated from the generalized Pareto distribution. We can conclude that the new estimator competes well with the other estimators, and outperforms all of them for grater values of γ, $0.1 \le \gamma \le 1$.



$m = 50$

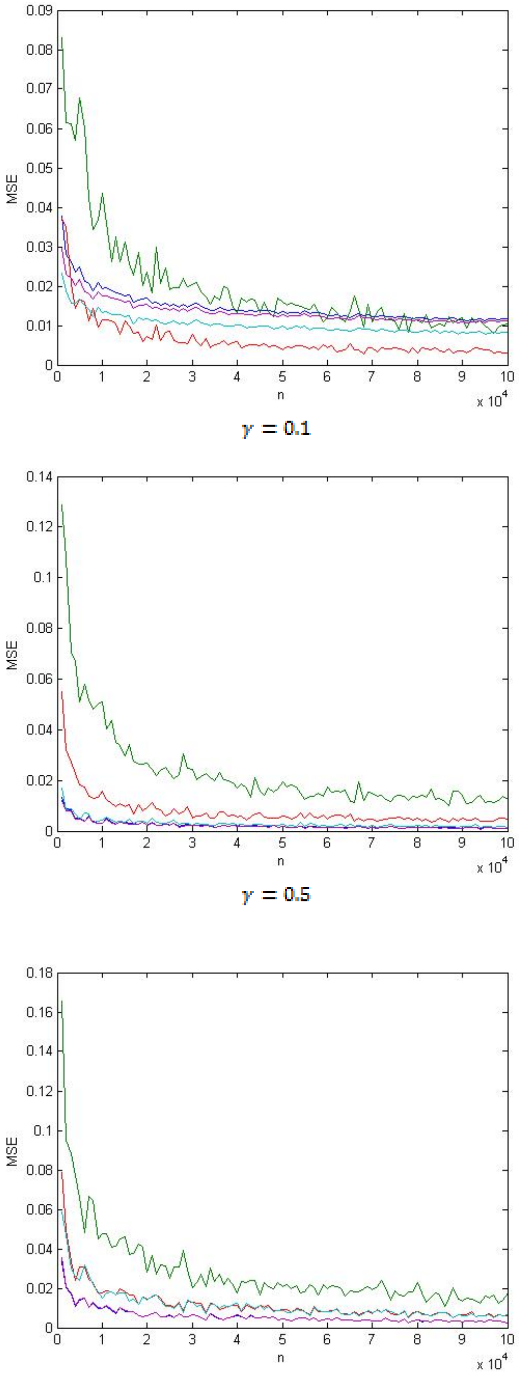


$m = 100$

$m = 200$

Figure 3. Mean squared error (MSE) of the estimators suggested by Hill (blue), Pickands (green), Dekkers, Einmahl and de Haan (red), C. de Vries (turquoise), and the new estimator (purple) as a function of γ (computation based on 100 samples of size 1000, where $m \in \{50, 100, 200\}$, respectively, and $0.1 \leq \gamma \leq 1$, with step 0.001)

In Figure 4, we describe the behavior of the estimators when the sample size changes. Precisely, we depict mean squared errors of the estimators H_n, P_n, D_n, V_n, and the new estimator, as a function of sample size, n, for $1000 \leq n \leq 10000$, and several values of γ, based on the samples simulated from the generalized Pareto distribution. In this case, the new estimator competes well with the other estimators, and outperforms most of them for $\gamma > 0.5$.



$\gamma = 0.1$

$\gamma = 0.5$

$\gamma = 1$

Figure 4. Mean squared error (MSE) of the estimators suggested by Hill (blue), Pickands (green), Dekkers, Einmahl and de Haan (red), C. de Vries (turquoise), and the new estimator (purple) as a function of sample size (computation based on 100 samples of size 1000, where $\gamma \in \{0.1, 0.5, 1\}$, respectively, and $1000 \leq n \leq 10000$, with step 1000, where $m = n^{0.5+eps}$, $eps = 10^{-8}$)

CONCLUSION AND FUTURE WORK

In this paper we propose a way to modify Hill's tail index estimator in order to minimize the mean squared error. When compared to several other estimators, the new estimator outperforms most of them for $0.5 \leq \gamma \leq 1$, which are the values commonly found in practice. Future work will be devoted to obtaining asymptotic properties of the proposed estimator (consistency, normality), which would allow more detailed comparison with other estimators (as it is done in articles [15], [3]).

BIBLIOGRAPHY:

[1] J. Beirlant and J. L. Teugels "Asymptotic normality of Hill's estimator", in: J. Hüsler, R. D. Reiss (Eds.) *Lecture notes in Statistics*, vol. 51. Springer, pp. 148-155.

[2] D. D. Boss, "Using extreme value theory to estimate large percentiles", *Technometrics*, vol. 26, pp. 33-39, 1984.

[3] M. F. Brilhante, M. I. Gomes and D. Pestana, "A simple generalization of the Hill's estimator", *Comput. Statist. Data Anal*, vol. 57, pp. 518-535, 2013.

[4] E. Carrizosa, M. Dražić, Z. Dražić and N. Mladenović, "Gaussian variable neighborhood search for continuous optimization", *Comput. Oper. Res*, vol. 39, pp. 2206-2213, 2012.

[5] S. Csörgö, P. Deheuvels and D. M. Mason, "Kernel estimates of the tail index of a distribution", *Ann. Statist*, vol. 13, pp. 1050-1077, 1985.

[6] R. A. Davis and S. T. Resnick, "Tail estimates motivated by extreme value theory", *Ann. Statist*, vol. 12, pp. 1467-1487, 1984.

[7] A. L. M. Dekkers, J. H. J. Einmahl and L. de Haan, "A moment estimator for the index of an extreme-value distribution", *Ann. Statist*, vol. 17, pp. 1833-1855, 1989.

[8] A. L. M. Dekkers and L. de Haan, "On the estimation of extreme value index and large quantile estimation", Ann. Statist, vol. 17, pp. 1795-1832, 1989.

[9] P. Deheuvels, E. Häusler and D. M. Mason, "Almost sure convergence of the Hill estimator", *Math. Proc. Cambridge Philos. Soc*, vol. 104, pp. 371-381,

[10] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: A critique", *Ann. Statist*, vol. 11, pp. 1019-1031, 1983.

[11] P. Embrechts, C. Klüppelberg and T. Mikosch, *Modelling extremal events – for insurance and finance*, Springer-Verlag, 1997.

[12] C. M. Goldie and R. L. Smith, "Slow variation with remainder: Theory and applications", *Quart. J. Math. Oxford*, vol. 38, pp. 45-71, 1987.

[13] E. Häusler and J. L. Teugels, "On asymptotic normality of Hill's estimator for the exponent of regular variation", *Ann. Statist*, vol. 13, pp. 743-756, 1985.

[14] P. Hall, "On some simple estimates of an exponent of regular variation", J. R. Statist. Soc. B, vol. 44, pp. 37-42, 1982.

[15] L. de Haan and L. Peng, "Comparison of tail index estimators", *Stat. Neerl*, vol. 52, pp. 60-70, 1998.

[16] B. M. Hill, "A simple general approach to inference about the tail of a distribution", *Ann. Statist*, vol. 3, pp. 1163-1174, 1975.

[17] B. M. Hill, "On tail index estimation for dependent, heterogeneous data", *Econom. Theory*, vol. 26, pp. 1398-1436, 2010.

[18] T. Hsing, "On tail index estimation using dependent data", *Ann. Statist*, vol. 19, pp. 1547-1569, 1991.

[19] D. M. Mason, "Laws of large numbers for sums of extreme values", *Ann. Probab*, vol. 10, pp. 754-764, 1982.

[20] N. Mladenović, M. Dražić, V. Kovačević Vujčić and M. Čangalović, "General variable neighborhood search for continuous optimization", *European J. Oper. Res*, vol. 191, pp. 753-770, 2008.

[21] N. Mladenović and P. Hansen, "Variable neighborhood search", *Comput. Oper. Res*, vol. 24, pp. 1097-1100, 1997.

[22] P. Mladenović and V. Piterbarg, "On estimation of the exponent of regular variation using a sample with missing observations", *Statist. Probab. Lett*, vol. 78, pp. 327-335, 2008.

[23] J. Pickands, "Statistical inference using extreme order statistics", *Ann. Statist*, vol. 3, pp. 119-131, 1975.

[24] V. Pisarenko and M. Rodkin, "Heavy-tailed distributions in disaster analysis", Springer-Verlag, 2010.

[25] S. I. Resnick, "Heavy tail modelling and teletraffic data", *Ann. Statist*, vol. 25, pp. 1805-1869, 1997.

[26] S. I. Resnick, *Heavy-tail phenomena: Probabilistic and statistical modelling*, Springer, 2006.

[27] S. I. Resnick and C. Stărică, "Consistency of Hill's estimator for dependent data", *J. Appl. Probab*, vol. 32, pp. 139-167, 1995.

[28] S. I. Resnick and C. Stărică, "Asymptotic behavior of Hill's estimator for autoregressive data", *Stoch. Models*, vol. 13, pp. 703-723, 1997.

[29] S. I. Resnick and C. Stărică, "Tail index estimation for dependent data", Ann. Appl. Probab, vol. 8, p. 1156-1183, 1998.

[30] G. Salvadori, C. De Michele, N. T. Kottegoda and R. Rosso, *Extremes in Nature – An approach using copulas*, Springer, 2007.

[31] R. L. Smith, "Estimating tails of probability distributions", Ann. Statist, vol. 15, pp. 1174-1207, 1987.