Michał Stachura
Jan Kochanowski University in Kielce
e-mail: michal.stachura@ujk.edu.pl

ON IMPROVED ESTIMATION OF THE EXTREME VALUE INDEX WITH THE USE OF A SHIFTED HILL'S ESTIMATOR

O UDOSKONALONEJ METODZIE SZACOWANIA INDEKSU EKSTREMALNEGO Z WYKORZYSTANIEM ESTYMATORA HILLA Z PRZESUNIĘCIEM

Summary: The importance of the *heavy-tail phenomenon*, and as well as, limitations of methods used to detect and measure this phenomenon, are sustaining incessant progress in the area of extreme value index estimation (EVI). Therefore, the paper is devoted to illuminate possibility of making the Hill's estimator to be shift-invariant, so that unintended misestimating of the EVI may be reduced. The paper confronts the original Hill's approach with Aban and Meerschaert's [2001] ideas, which are, in addition, adopted, and modified by the author. Thus, the three estimation approaches are compared, and their accuracy is assessed by a simulation study. It occurs that in the Pareto distribution case, the best estimates, amongst the considered, seem to be the ones proposed by the author and based on a proper choice of the sample shift. In general, the present study elucidates clear advantages of the *shift-invariance approach* over the original Hill's estimator.

Keywords: extreme value index, Pareto distribution, shifted Hill's estimator.

Streszczenie: Waga efektu tzw. grubych ogonów, jak i niedoskonałość metod wykorzystywanych do jego wykrywania i pomiaru sprawiają, że w zakresie estymacji indeksu ekstremalnego wciąż trwa nieustanny rozwój. Wobec tego niniejsze opracowanie ma na celu naświetlenie możliwości unieważliwienia estymatora Hilla na translacje próby po to, by w większym stopniu dało się unikać niewłaściwych oszacowań. Artykuł konfrontuje oryginalny pomysł Hilla z ujęciem zaproponowanym w opracowaniu [Aban, Meerschaert 2001], a następnie zmodyfikowanym przez autora. Wszystkie trzy ujęcia estymacji zostały zweryfikowane i porównane dzięki przeprowadzonym badaniom symulacyjnym. Okazuje się, że dla klasy rozkładów Pareta najlepiej spośród rozważonych estymatorów wypada autorski, który odwołuje się do stosownej translacji wartości próby. W pełni ogólności, opisane zagadnienia wskazują na wyraźną przewagę ujęcia dopuszczającego translację próby nad klasycznym ujęciem pochodzącym od Hilla.

Słowa kluczowe: indeks ekstremalny, rozkład Pareta, estymator Hilla z przesunięciem.

1. Introduction

As the *heavy-tail phenomenon* of a sample distribution is becoming more and more widely admitted in a breadth of economic research areas where stochastic modelling is involved, the notions of 'tail thickness' measures are emerging, and their estimates – as well.

The subject matter of the present paper is embedded in the estimation of the *extreme value index* – being one of the most prominent *tail thickness* measures – with use of an amendment of the, probably most common, Hill's estimator which is originally scale-invariant only. A rather new and quite natural approach which is proposed by Aban and Meerschaert [2001] is consisted in making further the estimate to be shift-invariant.

In literature, there is a significant multitude of publications appearing in the recent years, reporting more or less usable, applicative, sophisticated, advanced, theoretical, etc facets of the progress in constructing new estimators of the extreme value index, as well as in improving the known ones. We list arbitrarily ideas presented by Leadbetter at al. [1980], Dekkers at al. [1989], Gomes at al. [2008], de Haan and Ferreira [2006], to mention but a few. This progress is caused doubtlessly by high necessity of practical implementations arising in the wide range of research areas, including, of course, such issues as value at risk estimation, or distribution recognition of a risk posed by *extreme events*. The other reason of the progress is that none of the acknowledged estimators, up to now, seems to be the most adequate and to have an unassailable lead over the others.

We focused on an acutely simple perspective of improving the classical Hill's approach. This perspective is still evolving, for instance, see: Fraga [2001], Li at al. [2008, Nuyts [2010]. Therefore, the main goal of the presented study is to illuminate the necessity and possibility of making the Hill's estimator shift-invariant, in order to avoid unintended misestimating of the extreme value index. The paper presents, in turn, the original Hill's idea, a quintessence of Aban and Meerschaert's theoretical considerations, and in addition, author's own proposal of a simple estimation procedure that is based on the appropriate shift detection (Section 2). The three estimation approaches are compared, and their accuracy is assessed by a simulation study concerning the Pareto distribution case (Section 3), which discloses explicit advantages of the *shift-invariance approach*, and indicates the best estimates amongst the considered ones.

2. Theoretical background

In the paper, we assume that $X_1, X_2, X_3, ...$ is a sequence of *independent and identically distributed* (i.i.d.) random variables with a common *cumulative distribution function* (cdf) F. The order statistics of a sample $X_1, ..., X_n$ are denoted by $X_{1:n} \geq \cdots \geq X_{n:n}$ for any fixed size $n \in \mathbb{N}_+$.

The main theorem of the ***extreme value theory*** (EVT) states that if there exist sequences of constants $a_n > 0$, b_n and some non-degenerate distribution function G, such that for all $x \in \mathbb{R}$ holds:

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{X_{1:n} - b_n}{a_n} \le x\right) = G(x), \tag{1}$$

then there exists a constant $\gamma \in \mathbb{R}$, such that the distribution G has the form:

$$G(x) = G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}) & 1 + \gamma x > 0 \quad \gamma \ne 0 \\ \exp(-e^{-x}) & x \in \mathbb{R} \qquad \gamma = 0 \end{cases}, \tag{2}$$

which is unique up to a positive linear transformation.

The parameter γ is called the *extreme value index* (EVI), and it impacts the right tail asymptotics of the common cdf F (for instance see: de Haan and Ferreira [2006]). For the case of Pareto distribution, where $F(x) = 1 - cx^{-\alpha}$ for all $x \ge d$ with some $d > 0$ and $\alpha > 0$, it occurs that the EVI is the inverse of the parameter α, i.e. $\alpha\gamma = 1$.

In 1975, Hill constructed (see: Hill [1975], Gomes et al. [2008]) the estimator:

$$\hat{\gamma}_H^k = \hat{\gamma}_H^k(X_1, \dots, X_n) = \frac{1}{k}\sum_{i=1}^{k} \ln X_{i:n} - \ln X_{k+1:n}, \tag{3}$$

with the use of $k + 1$ largest order statistics, where k is chosen arbitrarily.

Even though the Hill's estimator is derived for the Pareto case only, it is also used in practical applications for the so-called Pareto-like distributions, where $F(x) \approx 1 - cx^{-1/\gamma}$ for $x \to +\infty$. In such circumstances, one should be aware of the following collision of two requirements (for instance, see: Aban and Meerschaert [2001]).

Namely, on one hand, the higher k is, the more information about the sample distribution is involved in the Hill's estimate. On the other hand, the smaller k is, the better approximation of the distributional tail is preserved. Bringing together mentioned requirements is itself quite a rich and still not fully identified question of how to choose the k properly somewhere in-between 1 and n.

In order to avoid an unintended impact of the improper choice of k, we restrict further consideration to the Pareto distribution case only, which additionally conforms to the very Hill's assumptions, and the assumptions, under which a shifted Hill's estimator is introduced by Aban, and Meerschaert [2001].

In general, since the limit distribution G is determined up to a positive linear transformation of sample $X_1, \dots, X_n$, any reliable estimator of the EVI γ should inherit this feature, being both scale-, and shift-invariant. Unfortunately, this is not the classical Hill's estimator case. Straightforward calculation shows that $\hat{\gamma}_H^k(X_1, \dots, X_n) = \hat{\gamma}_H^k(aX_1, \dots, aX_n)$, while $\hat{\gamma}_H^k(X_1, \dots, X_n) \ne \hat{\gamma}_H^k(X_1 + s, \dots, X_n + s)$ for any $a > 0$, and $s \in \mathbb{R}$, $s \ne 0$, provided that both of the equal/unequal sides make sense.

In that spirit, one may evaluate quite a wide range of Hill's estimates by substituting different values of s (see: Fig. 1). However, the problem will remain unsolved, as long as one has no criterion of how to make the most appropriate choice of a shift s.
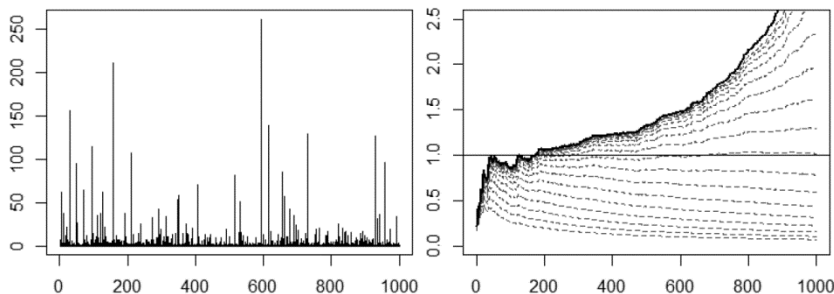
Note: The Hill's plot for zero shift is depicted by the solid line.

Fig. 1. An example of Pareto sample of size $n = 1000$ with $\gamma = 1$ [left panel], and its several Hill's plots for various shifts [right panel]

Source: own study with the use of *actuar R* package [Dutang et al. 2008].

A theoretical approach, proposed by Aban and Meerschaert [2001] yields a collective CML (*conditional maximum likelihood*) estimation of all three parameters of the shifted Pareto distribution:

$$F(x) = 1 - c(x + s)^{-\alpha}, \tag{4}$$

given for any argument $x \geq c^{1/\alpha} - s$ with $\alpha > 0$ (see: Theorem 1 in [Aban, Meerschaert 2001]). The complex, and partially implicit, formulas from therein may be transformed into the explicit function formula[1]:

$$g(s) = \hat{\gamma}_H^k(X_1 + s, \dots, X_n + s) - \frac{k}{(X_{k+1:n}+s)\sum_{i=1}^{k}(X_{k+1:n}+s)^{-1}} + 1 , \tag{5}$$

whose unique (under some assumptions) root $\hat{s} > -X_{k+1:n}$ provides an estimate $\hat{\gamma}_{SH}^k = \hat{\gamma}_H^k(X_1 + \hat{s}, \dots, X_n + \hat{s})$, or $\hat{\alpha}_{SH}^k = 1/\hat{\gamma}_{SH}^k$ equivalently (compare the Formula (7), therein.). Despite desired characteristics, and complete theoretical identification, this approach may raise some nuisance with its implementation in numerical optimisation, which is noticed, and emphasised by the authors. Additionally, they examine their EVI estimator by simulation studies that are surprisingly – in a sense of unintended impact of the improper choice of k – carried out for α-stable distribution case. In the sequel, their estimation approach is simply called a *shifted Hill approach* (SHA).

Because of the complexity appearing in the mentioned estimation procedure, we introduce another approach which is based on such a choice of a shift s of the sample

---

[1] In contrast to the original paper of Aban & Meerschaert [2001], we use equivalently additive, instead of subtractive, notation of shifts.

that stabilises the course of the Hill's plot consisting of all points $\{ (k, \ \hat{\gamma}_{SH}^k(s)) : k \in \{1, 2, \ldots, n-1\} \}$. This approach – henceforth called the *stabilising approach* (STA) – demands an additional criterion that assesses the level of stability. In the sequel, we measure it – probably in the simplest possible way – via slope of a linear regression fitted for k's from range determined arbitrarily by $0.6n \le k \le 0.95n$. Thus, the most preferable shifts are these that yield a slope equal to zero.

3. Numerical verification

In order to compare the three mentioned estimation approaches, simulation research is carried out as follows[2]. **Firstly**, for a fixed pair of shifted Pareto distribution parameters[3] $\gamma > 0$ and σ – taken from arbitrarily chosen ranges $\gamma \in \{ 0.05, 0.1, 0.2, 0.4, 0.5, 1, 2, 2.5, 5, 10, 20 \}$, $\sigma \in \{ 1, 10 \}$[4] – a pseudorandom i.i.d. sample of size $n \in \{ 100, 1000 \}$ is generated. **Secondly**,

1. the original Hill's estimators $\hat{\gamma}_H^k$ are calculated for $k \in \{ 0.4n, 0.6n, 0.8n \}$,
2. the shifted Hill's estimators $\hat{\gamma}_{SH}^k$ are calculated for $k \in \{ 0.4n, 0.6n, 0.8n \}$[5],
3. with respect to the best choice of shift due to STA, estimators $\hat{\gamma}_{ST}^k$ are calculated for $k \in \{ 0.4n, 0.6n, 0.8n \}$,
4. additionally, for the best choice of shift in point 3., and for all k's, such that $0.6n \le k \le 0.95n$, a median $\hat{\gamma}_{ST}^M$ of estimators $\hat{\gamma}_{ST}^k$ is computed.

Finally, all the previous steps are replicated 1000 times independently, so that the accuracy of the estimators may be revealed. For this, we start auxiliarily with careful visual examination of boxplots, depicting estimates' location for all the considered combinations of distributional parameters and sample sizes, and in addition for different-scales of graphs (see: Fig. 2, that is drawn in broad common scales for only a few selected combinations of parameters).

Visual examination of all analogue plots discloses generally, that:

- the Hill's estimates $\hat{\gamma}_H^k$ are much more biased than the others, and, moreover, only positive bias is diagnosed in their case. They are also the most dispersed estimates, excluding only the largest values of γ;
- the SHA estimates $\hat{\gamma}_{SH}^k$ have a slight negative bias for small values of γ, and they become the most dispersed ones for large values of γ, and moreover, they are more dispersed than STA $\hat{\gamma}_{ST}^k$ estimates for γ's greater or equal to 5;

---

[2] The simulation research, and additionally all the calculations and plots presented herein, are carried out in R environment.

[3] In R's actuar package, that is employed, we have a family of distributions (called *Pareto Type II*, or *Lomax*), given by $F(x) = 1 - \sigma^\alpha (x + \sigma)^{-\alpha}$ for $x > 0$, $\alpha > 0$, $\sigma > 0$, that is in fact a subfamily of the shifted Pareto distributions, if letting $\sigma = s^\alpha$.

[4] γ's are chosen in order to provide a wide range of positive values, while σ's correspond to two cases: $\sigma = 1$, that provides – irrespective of γ – constant scaling, and $\sigma = 10$, that provides scaling which varies with γ.

[5] The original optimisation problem is solved by implementation of grid search, instead of typical numerical root-finding used by Aban and Meerschaert [2001].
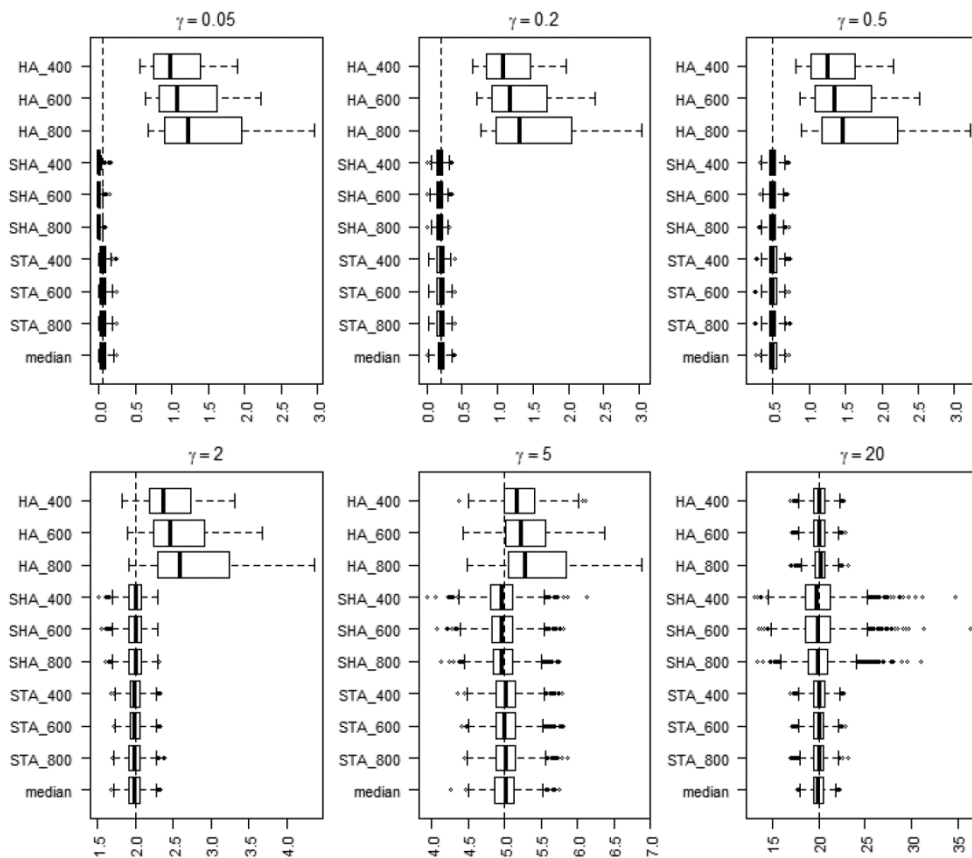
Fig. 2. Boxplots of estimators' sampling distributions for fixed $\sigma = 1$, and $n = 1000$, and selected values of γ

Source: own study with the use of R package [R Core Team 2015].

- the STA estimates $\hat{\gamma}_{ST}^k$ seem to be the most stable ones – in terms of both bias and dispersion, and this observation is most apparent in the *median estimator's* case ($\hat{\gamma}_{ST}^M$).

Next, we rely fundamentally on properly chosen two positional measures. In that role, we put absolute bias in median (given by $|\text{Med}(\hat{\gamma}) - \gamma|$, and later on referred to as M1), and absolute bias in median enlarged with a half of interquartile range (given by $|\text{Med}(\hat{\gamma}) - \gamma| + \text{IQR}(\hat{\gamma})/2$, and referred to as M2). The latter measure may be regarded as an analogue of the mean square error[6].

---

[6] Positional measures are preferred to classical ones, since they are not limited by their incongruity to any theoretical assumptions, especially if the highest order of existing moment of distribution is unknown. The choice, however justified, may be criticised, as it is rather arbitrary.

With respect to a fixed measure, and a given combination of parameters (including sample size), ranks are appointed to all the estimators (with 0 denoting the best estimator, and 9 indicating the worst one). Further, in order to make an assessment of the considered estimation approaches to be independent of the underlying tail asymptotics[7], we sum up the assigned ranks upon all the values of γ, and then we rank – in the same way – the sums so obtained (see: Table 1, whose values represent ranks for the analysed case of fixed $\sigma = 1$, $n = 1000$).

Table 1. Ranks of estimators for $\sigma = 1$, $n = 1000$, with respect to measure M1

	k	γ											Sums of ranks	Total ranks
		0.05	0.1	0.2	0.4	0.5	1	2	2.5	5	10	20		
HA	400	7	7	7	7	7	7	7	7	7	4	3	70	**7**
	600	8	8	8	8	8	8	8	8	8	5	5	82	**8**
	800	9	9	9	9	9	9	9	9	9	9	8	98	**9**
SHA	400	4	6	5	3	0	5	0	6	6	6	9	50	**5**
	600	5	5	4	2	4	6	1	2	5	7	6	47	**4**
	800	6	4	6	1	6	4	2	3	4	8	7	51	**6**
STA	400	2	0	2	6	3	2	3	1	3	3	1	26	**1.5**
	600	1	2	3	5	5	1	4	4	0	1	0	26	**1.5**
	800	3	1	1	4	2	3	6	5	2	2	4	33	**3**
median		0	3	0	0	1	0	5	0	1	0	2	12	**0**

Source: author's own calculation in *R* package.

Table 2. Total ranks of estimators for all σ's, and n's, with respect to both the measures

	k	Measure (M1)				Measure (M2)			
		$\sigma = 1$ $n = 1000$	$\sigma = 1$ $n = 100$	$\sigma = 10$ $n = 1000$	$\sigma = 10$ $n = 100$	$\sigma = 1$ $n = 1000$	$\sigma = 1$ $n = 100$	$\sigma = 10$ $n = 1000$	$\sigma = 10$ $n = 100$
HA	400	7	7	7	7	7	7	7	7
	600	8	8	8	8	8	8	8	8
	800	9	9	9	9	9	9	9	9
SHA	400	5	6	6	3	6	6	6	6
	600	4	5	5	4	5	5	4.5	1
	800	6	4	4	6	4	4	1	4
STA	400	1.5	3	2.5	2	2	1	2	2
	600	1.5	2	1	1	1	2.5	3	3
	800	3	1	2.5	5	3	2.5	4.5	5
median		0	0	0	0	0	0	0	0

Source: author's own study.

---

[7] Such a formulation of assessment procedure seems to be very reasonable from an empirical study viewpoint, as if we examine a sample distribution, we must face with *a priori* unknown parameter γ.

Lastly in Table 2, we gather total ranks obtained similarly for all combinations of sample sizes n, and scaling parameters σ.

Casting an eye[8] over the results provided in Table 2 shows lucidly that STA estimators exhibit better properties than SHA ones, and further that STA and SHA estimators are definitely preferable than the original Hill's estimator.

4. Conclusion

The presentation of accommodating the Hill's estimator to the *shift-invariance* postulate clearly gives evidence that the Hill's approach may be effectively improved by means of a simple ploy that involves algebraic translation of an analysed sample, albeit with some possible difficulties in optimising such a translation.

Moreover, computer simulation research, that constitutes an integral part of an assessment of both SHA and STA against the backdrop of the Hill's estimator, not only compliments the presented theoretical consideration, but also reveals both methodological framework simplicity and accuracy of the introduced approaches.

However, in addition it should be pointed out that reproduction attempt of the technical solutions for the described estimation procedures, from Pareto's to other distributions' cases may not be conducted straightforwardly, since an appropriate choice of number of the highest order statistics that are to be taken into account is acutely dependent on the underlying law of a sample, and further the question of such a choice remains quite difficult, and still unsolved.

References

Aban I.B., Meerschaert M., 2001, *Shifted Hill's Estimator for Heavy Tails*, Communications in Statistics – Simulation and Computation, no. 30, vol. 4, p. 949-962.

Dekkers A.L.M., Einmahl J.H.J., de Haan L., 1989, *A moment estimator for the index of an extreme-value distribution*, Annals of Statistics, no. 4, p. 1833-1855.

Dutang C., Goulet V., Pigeon M., 2008, *Actuar: An R Package for Actuarial Science*, Journal of Statistical Software, vol. 25, no. 7, p. 1-37.

Fraga Alves M.I., 2001, *A Location Invariant Hill-Type Estimator*, Extremes, no. 4, p. 199-217.

Gomes M.I., e Castro L.C., Fraga Alves M.I., 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, no. 11, p. 3-34.

de Haan L., Ferreira A., 2006, *Extreme Value Theory. An Introduction*, Springer, New York.

Hill B., 1975, *A simple general approach to inference about the tail of a distribution*, Annals of Statistics, no. 3, vol. 5, p.1163-1174.

---

[8] We do not decide to calculate, and analyse then, parallel 'total ranks of total ranks', since it is not so natural, and obvious how to indicate the scope, over which such ranks should be evaluated (even if in a facet of joint, or separate reference to both measures). Additionally, the amount of different values of parameters, to be regarded now in a view of a scope limitation, is incomparably smaller than before.

Leadbetter M., Lindgren G., Rootzén H., 1980, *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York.

Li J., Peng Z., Nadarajah S., 2008, *A class of unbiased location invariant Hill-type estimators for heavy tailed distributions*, Electronic Journal of Statistics, vol. 2, DOI: 10.1214/08-EJS276 (12.09.2016).

Nuyts J., 2010, *Inference about the tail of a distribution: improvement on the Hill estimator*, International Journal of Mathematics and Mathematical Sciences, no. 2010, http://dx.doi.org/10.1155/2010/924013 (12.09.2016).

R Core Team, 2015, R*: A language and environment for statistical computing. R Foundation for Statistical Computing*, http://www.R-project.org (12.09.2016).