Asymptotic Properties of the Hill Estimator

Jaakko Pere

School of Science

Bachelor's thesis
Espoo 23.8.2018

Supervisor

Ph.D Pauliina Ilmonen

Advisor

M.Sc Matias Heikkilä

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of the bachelor's thesis

Author Jaakko Pere	
Title Asymptotic Properties of the Hill Estimator	
Degree programme Technical Physics and Mathematics	
Major Mathematics and Systems Analysis	**Code of major** SCI3029
Supervisor Ph.D Pauliina Ilmonen	
Advisor M.Sc Matias Heikkilä	
Date 23.8.2018 **Number of pages** 16+10	**Language** English

Abstract

We review the proof of the Hill estimator's consistency. Properties of the regularly varying functions play a crucial role in the theory. Consistency is also studied using simulations. Other properties of the Hill estimator such as asymptotic normality are also reviewed briefly.

Keywords Hill estimator, extreme value theory, asymptotic properties, statistics

Preface

Simulations were performed using computer resources within the Aalto University School of Science "Science-IT" project.

Otaniemi, 23.8.2018

Jaakko Pere

Contents

Symbols and Abbreviations

Symbols

$x^* = \sup\{x : F(x) < 1\}$	the right endpoint of the distribution F
γ	extreme value index
$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$	left-continuous inverse
U	the left-continuous inverse of $\frac{1}{1-F}$
$\mathbb{1}(p) = \begin{cases} 1, \text{if p is true} \\ 0, \text{otherwise} \end{cases}$	indicator function
$X_{i,n}$	i:th order statistic
λ	Lebesque measure
$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$	limit supremum of a sequence of sets A_n
RV_α	the set of regularly varying functions with index α
$D(G_\gamma)$	the maximum domain of attraction of G_γ
$f \sim g$	$\lim\limits_{x\to\infty} f(x)/g(x) = 1$

Abbreviations

cdf	cumulative distribution function
i.d.d.	independent and identically distributed
a.s.	almost surely

1 Introduction

Heavy-tailed distributions appear in many fields of science such as insurance and finance. For example the loss returns of a speculative asset or insurance are often heavy-tailed and in practice the focus is often on risk evaluation. One example of such data set is Danish fire insurance loss data [14] [18]. Further examples are given in [19], [2] and [7].

Analyzing the tail is challenging since using classical methods, such as empirical quantiles, it is impossible to extrapolate beyond the available data. This motivates the use of Extreme Value Theory (EVT), that provides a rigorous mathematical framework to study extremes. EVT characterizes the behavior of the distribution's tail and tail estimation is often of high practical importance.

One of the possible estimators for extreme value index is the Hill estimator, which was introduced by Bruce M. Hill in 1975 [12] and remains one of the most used tail estimators for heavy-tailed distributions. In this article we will review asymptotic properties of the Hill estimator and particularly its consistency.

Regularly varying functions form the mathematical foundation for the theory of heavy-tailed distributions and consequently they play an important role in the theory of the Hill estimator. We carefully review their properties in addition to consistency of the Hill estimator. We also examine consistency of the Hill estimator by performing Monte Carlo simulations. Results for asymptotic normality of the Hill estimator are provided but won't be discussed in detail.

The rest of the article is organized as follows. Section 2 gives requisites for the proofs and theory of the Hill estimator including properties of heavy tailed functions. Section 3 represents the main result, which is consistency and necessary lemmas for the proof of the consistency. Simulations are presented in Section 4 and concluding remarks are given in Section 5. Lastly, all the proofs and figures are found in Appendixes A and B.

2 Backround

2.1 Fisher-Tippett-Gnedenko Theorem

Let $X_1, X_2, ..., X_n$ be i.d.d. random variables with a cdf F. Consider the sample maxima $M_n = \max(X_1, X_2, ..., X_n)$. The limiting behavior of M_n is trivial: Since $X_1, X_2, ..., X_3$ are i.d.d.

$$\begin{aligned} P(\max(X_1, X_2, ..., X_n) \leq x) &= P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x) = F^n(x). \end{aligned}$$

Now,

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 0, x < x^* \\ 1, x \geq x^*. \end{cases}$$

To achieve a nondegerate distribution it is necessary to normalize the sample maxima M_n. The Fisher-Tippett-Gnedenko Theorem states that with a suitable normalization a nondegenate distribution is obtained [8] [10].

Theorem 2.1. *There exists real constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(ax + b), \tag{1}$$

where

$$G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}}), \gamma \neq 0 \\ \exp(-e^{-x}), \gamma = 0, \end{cases}$$

for all x with $1 + \gamma x > 0$ where $\gamma \in \mathbb{R}$.

If F satisfies the Equation (1) for some $\gamma \in \mathbb{R}$ then it is said that F is in the maximum domain of attraction of G_γ, denoted $F \in D(G_\gamma)$.

We are especially interested in heavy-tailed distributions, i.e. the case $F \in D(G_\gamma)$, with $\gamma > 0$. It turns out that F being heavy-tailed is equivalent to the function $1 - F$ being regularly varying with index $-1/\gamma$. In other words, if the tail function $1 - F$ of a distribution is regularly varying then the distribution F is heavy tailed and vice versa. [2]

Theorem 2.2. *Cdf F is in the maximum domain of attraction of the extreme value distribution G_γ with $\gamma > 0$ if and only if $x^* = \infty$ and*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\frac{1}{\gamma}}, x > 0. \tag{2}$$

In addition, (2) can be written in a different form in terms of a left-continuous inverse $U = (1/(1 - F))^{\leftarrow}$ [2].

Corollary 2.3. *Condition* (2) *is equivalent to*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, x > 0. \tag{3}$$

Relation (3) turns out to be easier to work with than (2) and thus it is used in furher calculations.

2.2 Regularly Varying Functions

In Section 2.1 we saw that $F \in D(G_\gamma)$ with $\gamma > 0$ and U being regularly varying function are equivalent conditions. Regularly varying functions have some useful properties that are needed for the proof of the Hill estimator's consistency. Let's define regularly varying functions properly [2]:

Definition 2.4. *A Lebesque measurable function $f : \mathbb{R}^+ \to \mathbb{R}$ that is eventually positive is regularly varying if for some index $\alpha \in \mathbb{R}$,*

$$\lim_{x\to\infty} \frac{f(tx)}{f(t)} = x^\alpha, x > 0. \tag{4}$$

If a function f is regularly varying with index $\alpha = 0$ then f is said to be slowly varying. One property of regularly varying functions is that all of them can be written in form $f(x) = l(x)x^\alpha$ where $l(x)$ is a slowly varying function. This supports the intuition that heavy tailed distributions behave like Pareto distribution after a high threshold.

Theorem 2.5. *If $f \in RV_\alpha$ then the convergence in the equation* (4) *is uniform .*

$$\lim_{t\to\infty} \sup_{x\in[a,b]} \left| \frac{f(tx)}{f(t)} - x^\alpha \right| = 0,$$

for $0 < a < b < \infty$.

By uniform convergence it can be proved that all the regularly varying functions can be represented in certain form [2]:

Theorem 2.6 (Karamata's representation theorem)**.** *If $f \in RV_\alpha$ then there exists measurable functions $a : \mathbb{R}^+ \to \mathbb{R}$ and $c : \mathbb{R}^+ \to \mathbb{R}$ with*

$$\lim_{t\to\infty} c(t) = c_0 \text{ and } \lim_{t\to\infty} a(t) = \alpha$$

and $t_0 \in \mathbb{R}^+$ such that for $t > t_0$,

$$f(t) = c(t) \exp\left(\int_{t_0}^{t} \frac{a(s)}{s} \,\mathrm{d}s \right). \tag{5}$$

Conversely, if 2.6 holds, then $f \in RV_\alpha$.

For the proof of the above theorem following lemma about the integrals of the regularly varying functions is needed [2].

Lemma 2.7. *Suppose $f \in RV_\alpha$. There exists $t_0 > 0$ such that $f(t)$ is positive and locally bounded for $t \geq t_0$. If $\alpha \geq -1$ then*

$$\lim_{t\to\infty} \frac{tf(t)}{\int_{t_0}^{t} f(s) \,\mathrm{d}s} = \alpha + 1. \tag{6}$$

If $\alpha < -1$ *or* $\alpha = -1$ *and* $\int_0^\infty f(s)\,\mathrm{d}s < \infty$*, then*

$$\lim_{t\to\infty} \frac{tf(t)}{\int_t^\infty f(s)\,\mathrm{d}s} = -\alpha - 1. \tag{7}$$

Conversely, if (6) *holds for* $-1 \leq \alpha < \infty$ *or* (7) *holds for* $-\infty < \alpha < -1$*, then* $f \in RV_\alpha$*.*

The following corollary [16] of the Karamata's representation theorem will play a key role in the proof of the consistency of the Hill estimator.

Corollary 2.8. *Suppose* $f \in RV_\alpha$*. If* $\varepsilon, \delta > 0$ *are arbitrary, there exists* $t_0 = t_0(\varepsilon, \delta)$ *such that for* $t \geq t_0$*,* $tx \geq t_0$*,*

$$(1-\varepsilon)x^{\alpha-\delta} < \frac{f(tx)}{f(t)} < (1+\varepsilon)x^{\alpha+\delta}.$$

3 Hill Estimator

The Hill estimator, proposed in [12], is defined as follows:

$$\hat{\gamma} = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n},$$

where $1 \leq k \leq n-1$ is a free parameter and $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ are order statistics. We proceed by reviewing its asymptotic properties.

3.1 Consistency

The following theorem states that Hill estimator is consistent i.e. estimator converges in probability to the extreme value index. [12]

Theorem 3.1. *Let $X_1, X_2, ...$ be i.d.d. random variables with cdf F_X. Suppose $F_X \in D(G_\gamma)$ with $\gamma > 0$. Then if $k = k(n) \to \infty$, $k/n \to 0$ as $n \to \infty$,*

$$\hat{\gamma} \xrightarrow{p} \gamma.$$

For the proof of the above theorem following lemmas and Corollary 2.8 are needed, firstly the Rényi's representation [17].

Lemma 3.2. *If $E_1, E_2, ...$ are i.d.d. random variables from the standard exponential distribution and $E_{1,n} \leq E_{2,n} \leq ... \leq E_{n,n}$ then for $k \leq n$ we have*

$$\Big(E_{1,n}, E_{2,n}, ..., E_{k,n}\Big) \stackrel{d}{=} \Big(\frac{E_1^*}{n}, \frac{E_1^*}{n} + \frac{E_2^*}{n-1}, ..., \frac{E_1^*}{n} + \frac{E_2^*}{n-1} + ... + \frac{E_k^*}{n-k+1}\Big),$$

where $E_1^*, E_2^*, ...$ are i.d.d. random variables from standard exponential distribution.

With Rényi's representation the complicated joint distribution of dependent order statistics can be represented in terms of independent random variables. Proof of the Rényi's representation is omitted here.

Secondly a lemma about the order statistics of Pareto distribution is necessary [2].

Lemma 3.3. *Let $Y_1, Y_2, ...$ be i.d.d. random variables from Pareto distribution $F_Y(y) = 1 - \frac{1}{y}$, $y \geq 1$ and let $Y_{1,n} \geq Y_{2,n} \geq ... \geq Y_{n,n}$ be the nth order statistics. Then if $k = k(n) \to \infty$, $k/n \to 0$ as $n \to \infty$,*

$$\lim_{n \to \infty} Y_{n-k,n} = \infty \quad a.s.$$

Finally we need the fact that U(Y) is equal in distribution to X, where Y is random variable from Pareto distribution and X is random variable from some distribution F_X.

Lemma 3.4. *Let Y be random variable from Pareto distribution $F_Y = 1 - \frac{1}{y}$, $y \geq 1$. Let X be random variable with cdf F_X then $U(Y) \stackrel{d}{=} X$.*

Notably, there is also converse version of the Theorem 3.1 [13]. Proof is omitted here.

Theorem 3.5. *Let $X_1, X_2, ...$ be i.i.d. variables from cdf F. Suppose that for some sequence of integers $k = k(n) \to \infty$, $k(n)/n \to 0$ and $k(n+1)/k(n) \to 1$ as $n \to \infty$*

$$\hat{\gamma} \to \gamma > 0.$$

Then $F \in D(G_\gamma)$.

The almost sure convergence of the Hill estimator is also proved in [13].

3.2 Asymptotic Normality

In addition to the first-order regular variation 2.4 there is a second-order condition for regular variation, which allows to examine the rates of converge of the first-order condition. Second-order regular variation is needed for the asymptotic normality of the Hill estimator. We define second-order regular variation for case $\gamma > 0$, since it is sufficient for our purposes [2]. For the complete definition see [9].

Definition 3.6. *U is second-order regularly varying for parameters $\gamma > 0$ and $\rho \leq 0$ if for $x > 0$*

$$\lim_{t \to \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho},$$

where A is a one-signed function with $\lim\limits_{t \to \infty} A(t) = 0$, with the right-hand side interpreted as $x^\gamma \log x$ for $\rho = 0$.

The asymptotic normality of $\hat{\gamma}$ can be formulated using the second-order condition [3].

Theorem 3.7. *Suppose that U satisfies the second-order condition 3.6 and $k = k(n)$ with $k(n) \to \infty$, $k(n)/n \to \infty$ when $n \to \infty$. Then*

$$\sqrt{k}(\hat{\gamma} - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right)$$

where

$$\lim_{t \to \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda, \quad \lambda \in \mathbb{R}.$$

From Theorem 3.7 it can be seen that for Hill estimator to be asymptotically normal U of the distribution has to satisfy the second-order regular variation. Furthermore, there is an asymptotic bias that depends on the second-order parameter ρ and λ.

4 Simulations

The consistency of the Hill estimator is studied with a Monte Carlo simulation. We simulate observations from the Pareto distribution with an extreme value index $\gamma = 1/3$ and Cauchy distribution ($\gamma = 1$, as can be checked with Equation (2) or (3)). We simulate $N = 2000$ sets of n_i observations from both distributions where $n_i = 100 + 50(i-1), i = 1, 2, ..., 399$. Hence the sample size is a vector from 100 to 20000 with steps of 50. We denote this vector with $\mathbf{n} = (100, 150, ..., 20000)$. For each n_i we calculate N number of estimates $\hat{\gamma}$ with $k = k(n) = o(n)$. For k we chose $k(n) = \lfloor \sqrt{n} \rfloor$, thus the condition $k = o(n)$ is fulfilled. Results are shown in Figures B1 and B2. Below is a summary of simulation settings.

Table 1: Simulation settings.

Figure	Distribution	γ	Estimator	n	N	$k(n)$
B1	Pareto	1/3	Hill	$\mathbf{n}$	2000	$\lfloor \sqrt{n} \rfloor$
B2	Cauchy	1	Hill	$\mathbf{n}$	2000	$\lfloor \sqrt{n} \rfloor$

Both resulting graphs are constructed in the same manner. Real value of the extreme value index γ is plotted as a dashed line. In both figures we plot the median, the first quartile and the third quartile of the simulated values.

In Figure B1 it seems that with small n convergence is very fast and slows down after a while. This can be seen by looking at the quartiles. At the beginning the quartiles squeeze around γ but eventually after about $n = 12500$ it is hard to see any further convergence.

In Figure B2 we can conclude that convergence is fast with small numbers of observations but slows down eventually, as in Figure B1. It seems that the median is a little further from γ than in Figure B1. This suggests that for Cauchy distribution larger sample size is needed for an accurate estimate for γ than for Pareto distribution. In both Figures B1 and B2 quartiles are equally far away from median which suggests that asymptotic distribution of the Hill estimator might be symmetrical.

5 Conclusion

We reviewed the proof of the Hill estimator's consistency. Properties of the regularly varying functions played a crucial role in the proof of the consistency. Consistency was also studied with simulations, which also showed that Hill estimator has good finite sample properties. Other properties of the Hill estimator such as asymptotic normality and bias are not in the scope of this article. However, these properties are studied in other articles such as [3], [11] and [4].

Finally, we note that the Hill estimator can be extended in various ways. For example, there are estimators similar to Hill estimator for the general case $\gamma \in \mathbb{R}$. Examples of these estimators include the Pickands estimator [15] and the Moment estimator [5]. The Hill estimator can also be extended to the multivariate case [6].

References

[1] K. Athreya and S. Lahiri. *Measure Theory and Probability Theory*. Springer Texts in Statistics. Springer, New York, 2006.

[2] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[3] L. de Haan and L. Peng. Comparison of tail index estimators. *Statistica Neerlandica*, 52(1):60–70, 1998.

[4] L. de Haan and S. Resnick. On asymptotic normality of the hill estimator. *Communications in Statistics. Stochastic Models*, 14(4):849–866, 1998.

[5] A. L. M. Dekkers, J. H. J. Einmahl, and L. D. Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17(4):1833–1855, 1989.

[6] Y. Dominicy, P. Ilmonen, and D. Veredas. Multivariate hill estimators. *International Statistical Review*, 85(1):108–142, 2015.

[7] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Stochastic Modelling and Applied Probability. Springer, Berlin, 1997.

[8] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190, 1928.

[9] J. Geluk, L. de Haan, S. Resnick, and C. Stărică. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications*, 69(2):139 – 159, 1997.

[10] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453, 1943.

[11] E. Haeusler and J. L. Teugels. On asymptotic normality of hill's estimator for the exponent of regular variation. *The Annals of Statistics*, 13(2):743–756, 1985.

[12] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

[13] D. M. Mason. Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10(3):754–764, 1982.

[14] A. J. McNeil. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27(1):117–137, 1997.

[15] J. Pickands III. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131, 01 1975.

[16] H. S. A. Potter. The mean values of certain dirichlet series ii. *Proceedings of the London Mathematical Society*, 47(1):1–19, 1942.

[17] A. Rényi. On the theory of order statistics. *Acta Mathematica Academiae Scientiarum Hungarica*, 4(3):191–231, 1953.

[18] S. I. Resnick. Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin*, 27(1):139–151, 1997.

[19] S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007.

[20] J. S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific Publishing Co., Singapore, second edition, 2006.

A Proofs

A.1 Proof of Theorem 2.5

Proof. For a slowly varying function the limit (4) can be written in different form with function $F = \log f(e^x)$:

$$\lim_{t\to\infty} F(t+x) - F(x) = 0. \tag{A1}$$

The above argument is true, since

$$F(t+x) - F(t) = \log f(e^{t+x}) - \log f(e^t) = \log\left(\underbrace{\frac{f(e^t e^x)}{f(e^t)}}_{\to 1}\right) \to 0$$

as $t \to \infty$.

It can be assumed that $\alpha = 0$. If this isn't the case replace $f(x)$ by $f(x)x^{-\alpha}$. Suppose there exists sequences $t_n \to \infty$, $x_n \to 0$ as $n \to \infty$ such that

$$\left|\frac{f(t_n x_n)}{f(t_n)} - 1\right| > \delta$$

for all $n \in \mathbb{N}$ and some $\delta > 0$. An equivalent condition can be formulated in terms of function $F(x) = \log f(e^x)$:

$$|F(t_n + x_n) - F(t_n)| > \delta \tag{A2}$$

with possibly different x_n, t_n and δ. Let's define sets

$$\begin{aligned}
Y_{1,n} &= \left\{ y \in J : |F(t_n + y) - F(t_n)| > \frac{\delta}{2}\right\}, \\
Y_{2,n} &= \left\{ y \in J : |F(t_n + x_n) - F(t_n + y)| > \frac{\delta}{2}\right\} \quad \text{and} \\
Z_n &= \left\{ z : |F(t_n + x_n) - F(t_n + x_n - z)| > \frac{\delta}{2}, x_n - z \in J\right\} \\
&= \{z : x_n - z \in Y_{2,n}\}
\end{aligned}$$

where $J \subset \mathbb{R}$ is a finite interval. Next we will prove that the Equation (A2) contradicts the pointwise convergence (A1). Pointwise convergence does not hold if some x_0 is included in infinitely many $Y_{1,n}$. This can be noticed by comparing the definition of pointwise convergence and condition in $Y_{1,n}$. Similarly if x_0 is included in infinitely many Z_n then pointwise convergence cannot hold, since the condition in Z_n can be written as

$$\left| F(\underbrace{t_n + x_n}_{=u_n}) - F(\underbrace{t_n + x_n}_{=u_n} \overbrace{-z}^{=x_0}) \right| > \frac{\delta}{2}$$
$$\Leftrightarrow |F(u_n + x_0) - F(u_n)| > \frac{\delta}{2}$$

where $u_n \to \infty$.

Notice that $Y_{1,n} \cup Y_{2,n} = J$, since by the Equation (A2) and triangle inequality we have

$$\begin{aligned} \delta < |F(t_n + x_n) - F(t_n)| &= |(F(t_n + x_n) - F(t_n + y)) + (F(t_n + y) - F(t_n))| \\ &\leq |(F(t_n + x_n) - F(t_n + y))| + |(F(t_n + y) - F(t_n))| \\ \Rightarrow |(F(t_n + x_n) - F(t_n + y))| &> \frac{\delta}{2} \vee |(F(t_n + y) - F(t_n))| > \frac{\delta}{2}. \end{aligned}$$

Additionally $Y_{1,n}$, $Y_{2,n}$ and J are measurable sets. So by subadditivity of the Lebesque measure we have $\lambda(Y_{1,n}) \geq \frac{\lambda(J)}{2} \vee \lambda(Y_{2,n}) \geq \frac{\lambda(J)}{2}$. By the translation property of the Lebesque measure $\lambda(Z_n) = \lambda(Y_{2,n})$ holds. Thus $\lambda(Y_{1,n}) \geq \frac{\lambda(J)}{2} \vee \lambda(Z_n) \geq \frac{\lambda(J)}{2}$ infinitely often. All $Y_{1,n}$ are subsets of finite interval since $Y_{1,n} \subset J$ for all n. Similarly all Z_n are subset of a finite interval since $x_n \to 0$. Hence by Fatou's lemma [1]:

$$\lambda(\limsup Y_{1,n}) \geq \limsup \lambda(Y_{1,n}) \geq \frac{\lambda(J)}{2} \quad \vee$$
$$\lambda(\limsup Z_n) \geq \limsup \lambda(Z_n) \geq \frac{\lambda(J)}{2}.$$

Since at least one of the measures $\lambda(\limsup Y_{1,n})$ or $\lambda(\limsup Z_n)$ is greater than zero, we have some x_0 that is contained in infinitely many $Y_{1,n}$ or Z_n. This completes the proof. □

A.2 Proof of Lemma 2.7

Proof. First we prove the Equation (6). Suppose that $f \in RV_\alpha$. Then by Theorem 2.5 there exists t_0 and c such that $f(tx)/t < c$ when $t \geq t_0$, $x \in [1, 2]$. Then for $t \in [2^n t_0, 2^{n+1} t_0]$ we have

$$\frac{f(t)}{f(t_0)} = \frac{f(t)}{f(2^{-1}t)} \frac{f(2^{-1}t)}{f(2^{-2}t)} \cdots \frac{f(2^{-n}t)}{f(t_0)} < c^{n+1}. \tag{A3}$$

Equation (A3) is true since every fraction is of the form $f(tx)/f(t)$. This implies that for $t \geq t_0$ $f(t)$ is both locally bounded and $\int_{t_0}^{t} f(s)\,\mathrm{d}s < \infty$. Consider a function $F(t) = \int_{t_0}^{t} f(s)\,\mathrm{d}s$. We start by proving that $\lim_{t\to\infty} F(t) = \infty$ when $\alpha > -1$. First notice that $f(2s) \geq 2^{-1} f(s)$ for sufficiently large s. For $n \geq n_0$

$$\int_{2^n}^{2^{n+1}} f(s)\,\mathrm{d}s = 2\int_{2^{n-1}}^{2^n} f(2s)\,\mathrm{d}s \geq \int_{2^{n-1}}^{2^n} f(s)\,\mathrm{d}s \tag{A4}$$

by a change of variables. Then by induction we have

$$\int_{2^n}^{2^{n+1}} f(s)\,\mathrm{d}s \geq \int_{2^{n_0}}^{2^{n_0+1}} f(s)\,\mathrm{d}s = C > 0. \tag{A5}$$

Thus

$$\int_{2^{n_0}}^{\infty} f(s)\,\mathrm{d}s = \sum_{n=n_0}^{\infty} \int_{2^n}^{2^{n+1}} f(s)\,\mathrm{d}s \geq \sum_{n=n_0}^{\infty} \int_{2^n_0}^{2^{n_0+1}} f(s)\,\mathrm{d}s = \sum_{n=n_0}^{\infty} C = \infty. \tag{A6}$$

Next we prove that $F \in RV_{\alpha+1}$ for $\alpha > -1$. Let $\varepsilon > 0$ and $t_1 = t_1(\varepsilon)$. Then $f(xt) < (1+\varepsilon)x^{\alpha} f(t)$ for $t > t_1$. Since $\lim_{t\to\infty} F(t) = \infty$,

$$\frac{F(tx)}{F(t)} = \frac{\int_{t_0}^{tx} f(s)\,\mathrm{d}s}{\int_{t_0}^{t} f(t)\,\mathrm{d}s} \sim \frac{\int_{t_1x}^{tx} f(s)\,\mathrm{d}s}{\int_{t_1}^{t} f(t)\,\mathrm{d}s} = \frac{x\int_{t_1}^{t} f(xs)\,\mathrm{d}s}{\int_{t_1}^{t} f(t)\,\mathrm{d}s} < \frac{x\int_{t_1}^{t} (1+\varepsilon)x^{\alpha} f(s)\,\mathrm{d}s}{\int_{t_1}^{t} f(t)\,\mathrm{d}s} = (1+\varepsilon)x^{\alpha+1}$$

by a change of variables. A similar lower bound for $F(tx)/F(t)$ can be derived using $f(xt) < (1-\varepsilon)x^{\alpha} f(t)$ as $t > t_1$. So we have $F \in RV_{\alpha+1}$ for $\alpha > -1$. In the case $\alpha = -1$ and $F(t) \to \infty$ same proof applies. If $\alpha = -1$ and $F(t)$ has a finite limit and $F \in RV_0$. Now for all α

$$\begin{aligned}\frac{F(xt)-F(t)}{tf(t)} &= \frac{1}{tf(t)}\int_t^{tx} f(u)\,\mathrm{d}u = \frac{t}{tf(t)}\int_1^{x} f(ut)\,\mathrm{d}u = \int_1^x \frac{f(ut)}{f(t)}\,\mathrm{d}u \\ &\to \int_1^x u^{\alpha}\,\mathrm{d}u = \frac{x^{\alpha+1}-1}{\alpha+1}, \quad t\to\infty\end{aligned}$$

by the Theorem 2.5 and change of variables. On the other hand

$$\begin{aligned}\frac{F(xt)-F(t)}{tf(t)} = \frac{F(t)}{tf(t)}\left(\underbrace{\frac{F(tx)}{F(t)}}_{\to x^{\alpha+1}} -1\right) &\to \frac{x^{\alpha+1}-1}{\alpha+1} \\ \Rightarrow \lim_{t\to\infty}\frac{tf(t)}{F(t)} &= \alpha+1.\end{aligned}$$

Now we have proven (6). Next we prove Equation (7). Let's define

$$G(t) = \int_t^{\infty} f(s)\,\mathrm{d}s.$$

In the case $\alpha < -1$ there exists $\delta > 0$ such that $f(2s) \leq 2^{-1-\delta} f(s)$ for sufficiently large s. Now we can prove the finiteness of $\lim_{t\to\infty} G(t)$ in a similar way as we proved that $\lim_{t\to\infty} F(t)$ is infinite in Equations (A4), (A5) and (A6). For sufficiently large n_1

$$\begin{aligned}\int_{2^n}^{2^{n+1}} f(s)\,\mathrm{d}s = 2\int_{2^{n-1}}^{2^n} f(s)\,\mathrm{d}s \leq 2^{-\delta}\int_{2^{n-1}}^{2^n} f(s)\,\mathrm{d}s \leq \\ \ldots \leq 2^{-\delta(n-n_1)}\int_{2^{n_1}}^{2^{n_1+1}} f(s)\,\mathrm{d}s = 2^{-\delta(n-n_1)}C'\end{aligned}$$

by induction and a change of variables. Then

$$\int_{2^{n_1}}^{\infty} f(s)\,\mathrm{d}s = \sum_{n=n_1}^{\infty} \int_{2^n}^{2^{n+1}} f(s)\,\mathrm{d}s \leq C' \sum_{n=n_1}^{\infty} 2^{-\delta(n-n_1)}$$
$$= C' \sum_{k=0}^{\infty} \left(\frac{1}{2^\delta}\right)^k = \frac{C'}{1-1/2^\delta} < \infty.$$

Now rest of the proof is analogous. Next we prove the converse results. Suppose that Equation (6) holds. Let's define a function

$$b(t) = t\frac{f(t)}{F(t)}$$

Without loss of generality we may suppose that $f(t) > 0$ and $t > 0$. Integrating both sides of $b(t)/t = f(t)/F(t)$ we obtain for some real c_1 and for all $x > 0$

$$\int_1^x \frac{b(t)}{t}\,\mathrm{d}t = \log F(x) + c_1, \tag{A7}$$

since by a change of variables

$$\int_1^x \frac{f(t)}{F(t)}\,\mathrm{d}t = \int_{F(1)}^{F(x)} \frac{1}{u}\,\mathrm{d}u = \log F(x) + \underbrace{\log F(1)}_{=c_1}.$$

From the Equation (A7) we have

$$F(t) = \exp\left(\int_1^x \frac{b(t)}{t}\,\mathrm{d}t - c_1\right) = \underbrace{\exp(-c_1)}_{=c} \exp\left(\int_1^x \frac{b(t)}{t}\,\mathrm{d}t\right) = c\exp\left(\int_1^x \frac{b(t)}{t}\,\mathrm{d}t\right).$$

Then by using the definition of f again

$$\begin{aligned} f(x) = x^{-1}b(x)F(x) &= cb(x)\exp\left(-\int_1^x \frac{1}{t}\,\mathrm{d}t\right)\exp\left(\int_1^x \frac{b(t)}{t}\,\mathrm{d}t\right) \\ &= cb(x)\exp\left(\int_1^x \frac{b(t)-1}{t}\,\mathrm{d}t\right), \end{aligned} \tag{A8}$$

for all $x > 0$. Hence for all $x, t > 0$

$$\begin{aligned} \frac{f(tx)}{f(t)} = \frac{b(tx)\exp\left(\int_1^{tx}\frac{b(s)-1}{s}\,\mathrm{d}s\right)}{b(tx)\exp\left(\int_1^{t}\frac{b(s)-1}{s}\,\mathrm{d}s\right)} &= \frac{b(tx)}{b(t)}\exp\left(\int_1^{tx}\frac{b(s)-1}{s}\,\mathrm{d}s - \int_1^{t}\frac{b(s)-1}{s}\,\mathrm{d}s\right) \\ &= \frac{b(tx)}{b(t)}\exp\left(\int_t^{tx}\frac{b(s)-1}{s}\,\mathrm{d}s\right) = \frac{b(tx)}{b(t)}\exp\left(\int_1^{x}\frac{b(ts)-1}{s}\,\mathrm{d}s\right), \end{aligned}$$

by a change of variables. By the assumption (Equation (6)) $b(t) \to \alpha + 1$ so $b(tx)/b(t) \to 1$. For sufficiently large t

$$\exp\left(\int_1^x \frac{b(ts)-1}{s}\,\mathrm{d}s\right) \approx \exp\left(\int_1^x \frac{\alpha}{s}\,\mathrm{d}s\right) = \exp(\alpha \log x) = x^\alpha.$$

The last statement (Equation (7) implies that $F \in RV_\alpha$) can be proved in a similar way. $\square$

A.3 Proof of Theorem 2.6

Proof. Suppose $f \in RV_\alpha$. The function $t^{-\alpha} f(t)$ is slowly varying and

$$t^{-\alpha} f(t) = cb(t) \exp\left(\int_1^t \frac{b(s)-1}{s}\,\mathrm{d}s\right)$$

by the Equation (A8). Now by Lemma 2.7 $b(t) \to 1$ and function $t^{-\alpha} f(t)$ has the representation as in Theorem 2.6 with $a(t) = b(t) - 1$ and $c(t) = cb(t)$. Then

$$f(t) = c(t) t^\alpha \exp\left(\int_{t_0}^t \frac{a(s)}{s}\,\mathrm{d}s\right)$$

Notice that

$$t^\alpha = \exp\left(\log t^\alpha\right) = \exp\left(\int_1^t \frac{\alpha}{s}\,\mathrm{d}s\right).$$

Then f has the form

$$\begin{aligned}
f(t) = c(t)\exp\left(\int_{t_0}^t \frac{a(s)}{s}\,\mathrm{d}s + \int_1^t \frac{\alpha}{s}\,\mathrm{d}s\right) \\
= c(t)\exp\left(\int_{t_0}^t \frac{a(s)+\alpha}{s}\,\mathrm{d}s + \int_1^{t_0} \frac{\alpha}{s}\,\mathrm{d}s\right) \\
= c(t)\exp\left(\int_{t_0}^t \frac{a(s)+\alpha}{s}\,\mathrm{d}s\right)\exp\left(\log t_0^\alpha\right) \\
= \underbrace{t_0^\alpha c(t)}_{=c'}\exp\left(\int_{t_0}^t \frac{\overbrace{a(s)+\alpha}^{=a'}}{s}\,\mathrm{d}s\right)
\end{aligned} \tag{A9}$$

From the Equation (A9) it can be seen that $f(t)$ has the same representation as in the Theorem 2.6 when a is replaced by a' and c is replaced by c'. □

A.4 Proof of Corollary 2.8

Proof. By the Theorem 2.6

$$\frac{f(tx)}{f(t)} = \frac{c(tx)}{c(t)} \exp\left(\int_1^x \frac{a(st)}{s}\,\mathrm{d}s\right)$$

The function c(t) converges to a constant. Hence $c \in RV_0$ so $c(tx)/c(t) \to 1$ as $t \to \infty$. Furthermore, $a(s) \to \alpha$ as $t \to \infty$. Now we can choose such a t_0 that $\alpha - \delta < a(st) < \alpha - \delta$ and $1 - \varepsilon < \frac{c(tx)}{c(t)} < 1 + \varepsilon$. This implies that

$$\begin{aligned}
(1-\varepsilon)\int_1^x \frac{\alpha-\delta}{s}\,\mathrm{d}s < \frac{f(tx)}{f(t)} < (1+\varepsilon)\int_1^x \frac{\alpha+\delta}{s}\,\mathrm{d}s \\
\Rightarrow (1-\varepsilon)\exp\left(\log\left(x^{\alpha-\delta}\right)\right) < \frac{f(tx)}{f(t)} < (1+\varepsilon)\exp\left(\log\left(x^{\alpha+\delta}\right)\right) \\
\Rightarrow (1-\varepsilon)x^{\alpha-\delta} < \frac{f(tx)}{f(t)} < (1+\varepsilon)x^{\alpha+\delta}
\end{aligned}$$

□

A.5 Proof of Lemma 3.3

Proof. Assume that $Y_{n-k,n} < r$ for some $r > 0$ infinitely often. In other words

$$\frac{k}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i > Y_{n-k,n}) > \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i > r).$$

Now the left side of the equation converges to zero, since

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i > Y_{n-k,n}) = \lim_{n\to\infty} \frac{k}{n} = 0.$$

On the other hand the right side converges to $1/r$ almost surely, since

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i > r) \xrightarrow{a.s} P(Y_i > r) = 1 - F_Y(r) = \frac{1}{r}$$

by the strong law of large numbers [20]. This is a contradiction and

$$P(\lim_{n\to\infty} Y_{n-k,n} = \infty) = 1.$$

□

A.6 Proof of Lemma 3.4

Proof. Consider the condition $U(Y) \le a, a \in \mathbb{R}$.

$$\begin{aligned} & U(Y) \le a \\ \Leftrightarrow \inf & \Big\{x : \frac{1}{1 - F_X(x)} \ge Y\Big\} \le a \\ \Leftrightarrow \inf & \Big\{x : 1 - \frac{1}{Y} \le F_X(x)\Big\} \le a. \end{aligned} \tag{A10}$$

Let $S = \Big\{x : 1 - \frac{1}{Y} \le F_X(x)\Big\}$ and $b = \inf S$. Notice that F is increasing and right-continuous, since F is a cumulative distribution function and S is an interval of form $[b, \infty)$ or (b, ∞). Define a sequence $x_n = b + \frac{1}{n}, n \in \mathbb{N}$. Notice that $x_n \to b$ and $x_n \in S$ for all n. Now the right-continuity of F implies that $b \in S$ i.e S is an interval $[b, \infty)$. Additionally $a \in S$ since $a \ge b$ so a satisfies the condition $1 - \frac{1}{Y} \le F(a)$. Therefore Equation (A10) implies

$$U(Y) \le a \Leftrightarrow 1 - \frac{1}{Y} \le F_X(a) \Leftrightarrow Y \le \frac{1}{1 - F(a)}.$$

Now from the cdf of U(Y) we have

$$\begin{aligned} F_{U(Y)} = P(U(Y) \le x) = P\Big(Y \le \frac{1}{1 - F_X(x)}\Big) &= F_Y\Big(\frac{1}{1 - F_X(x)}\Big) \\ &= 1 - \Big(\frac{1}{1 - F_X(x)}\Big)^{-1} = F_X(x). \end{aligned}$$

□

A.7 Proof of Theorem 3.1

Proof. $F \in D(G_{\gamma>0})$ is equivalent to $U \in RV_\gamma$ i.e.

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma.$$

From Corollary 2.8 we have that for $x > 1$ and $t \geq t_0$,

$$(1-\varepsilon)x^{\gamma-\delta} < \frac{U(tx)}{U(t)} < (1+\varepsilon)x^{\gamma+\delta},$$

for all $\varepsilon > 0$ and $\delta > 0$. By taking the natural logarithm from both sides of the equation above, it can be written as

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\log(x) &< \log(U(tx)) - \log(U(t)) \\ &< \log(1+\varepsilon) + (\gamma+\delta)\log(x). \end{aligned} \tag{A11}$$

If $Y_1, Y_2, ...$ are i.d.d random variables from Pareto distribution with a cumulative distribution function $F_Y(y) = 1 - \frac{1}{y}$ then $U(Y_i) \stackrel{d}{=} X_i$ as stated in Lemma 3.4. Hence it is sufficient to prove the result for $\hat{\gamma} = \frac{1}{k}\sum_{i=0}^{k-1} \log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n}))$. Substituting $t = Y_{n-k,n}$ and $x = \frac{Y_{n-i,n}}{Y_{n-k,n}}$ Equation (A11) yields

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) &< \log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n})) \\ &< \log(1+\varepsilon) + (\gamma+\delta)\log(\frac{Y_{n-i,n}}{Y_{n-k,n}}). \end{aligned} \tag{A12}$$

Notice that we can replace t with $Y_{n-k,n}$ because we can always find some n_0 such that $Y_{n_0-k,n_0} \geq t_0$ according to lemma 3.3. Furthermore, $Y_{n-i,n}$ is greater than $Y_{n-k,n}$ when $i < k$. Therefore x can be replaced with $\frac{Y_{n-i,n}}{Y_{n-k,n}}$.

Equation (A12) holds for every $i = 0, 1, 2, ..., k-1$. Thus we can write

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) &< \frac{1}{k}\sum_{i=0}^{k-1}\log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n})) \\ &< \log(1+\varepsilon) + (\gamma+\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right). \end{aligned}$$

Notice that this can be written as

$$\log(1-\varepsilon) + (\gamma-\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) < \hat{\gamma}$$

$$< \log(1+\varepsilon) + (\gamma+\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right).$$

Now it is sufficient to prove that

$$\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)\xrightarrow{p}1.$$

The random variable $\log(Y_i)$ has a standard exponential distribution, since

$$F_{\log(Y_i)}(x)=P(\log(Y_i)<x)=P(e^{\log(Y_i)}<e^x)=P(Y_i<e^x)=F_Y(e^x)=1-e^{-x}.$$

Therefore we can write

$$\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)=\frac{1}{k}\sum_{i=0}^{k-1}E_{n-i,n}-E_{n-k,n},$$

where $E_1,E_2,...$ are i.d.d. random variables with a standard exponential distribution. Now Rényi's representation 3.2 implies

$$\begin{aligned}
&\left\{E_{n-i,n}-E_{n-k,n}\right\}_{i=0}^{k-1}\\
&\stackrel{d}{=}\left\{\left(\frac{E_1^*}{n}+\frac{E_2^*}{n-1}+...+\frac{E_{n-(i+1)}^*}{i+2}+\frac{E_{n-i}^*}{i+1}\right)\right.\\
&\qquad\left.-\left(\frac{E_1^*}{n}+\frac{E_2^*}{n-1}+...+\frac{E_{n-k}^*}{k+1}\right)\right\}_{i=0}^{k-1}\\
&=\left\{\frac{E_{n-(k-1)}^*}{k}+\frac{E_{n-(k-2)}^*}{k-1}+...+\frac{E_{n-(i+1)}^*}{i+2}+\frac{E_{n-i}^*}{i+1}\right\}_{i=0}^{k-1}\\
&=\left\{\frac{E_1^*}{k}+\frac{E_2^*}{k-1}+...+\frac{E_{k-(i+1)}^*}{i+2}+\frac{E_{k-i}^*}{i+1}\right\}_{i=0}^{k-1}\\
&\stackrel{d}{=}\left\{E_{k-i,k}\right\}_{i=0}^{k-1}.
\end{aligned}$$

Consequently we have

$$\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)\stackrel{d}{=}\frac{1}{k}\sum_{i=0}^{k-1}E_{k-i,k}=\frac{1}{k}\sum_{i=0}^{k-1}E_i\xrightarrow{p}E[E_i]=1$$

by the weak law of large numbers [20]. Notice that the expected value of a standard exponential is one.

□

B Figures

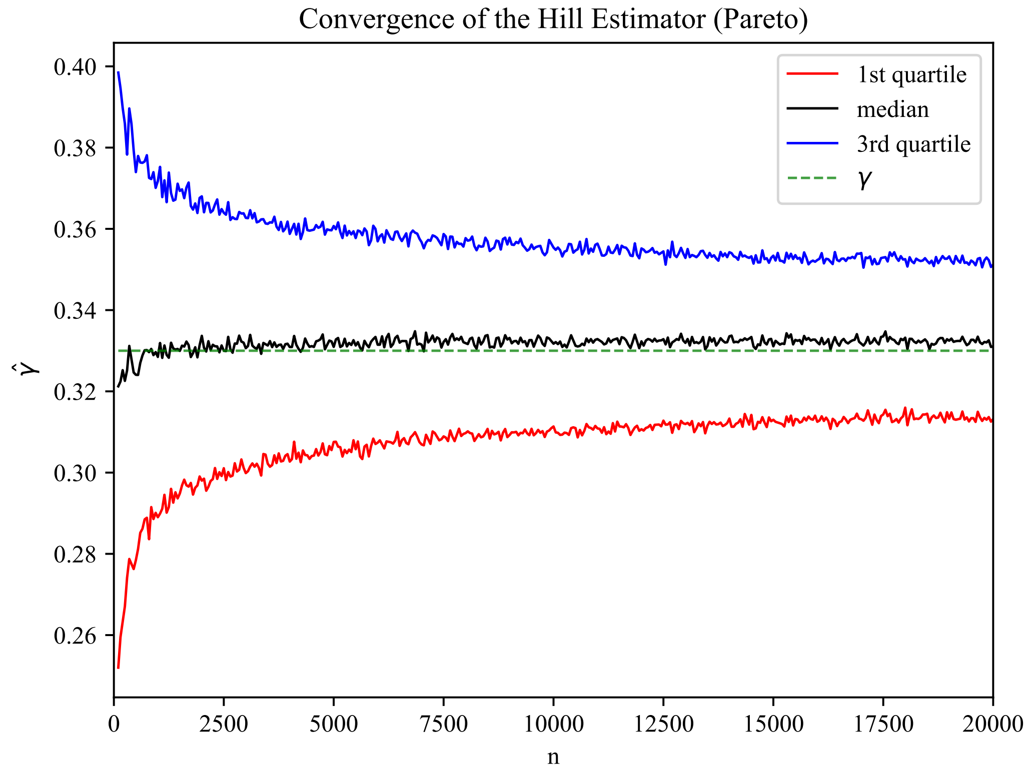

Figure B1: A plot of the median, the first quartile and the third quartile of simulated values of the Hill estimator under the Pareto distribution with extreme value index $\gamma = 1/3$. The observed quantities are plotted against the sample size. The used threshold was $k(n) = \lfloor\sqrt{n}\rfloor$ and each sample size was simulated 2000 times. The correct value of the extreme value index is displayed as a dashed line.

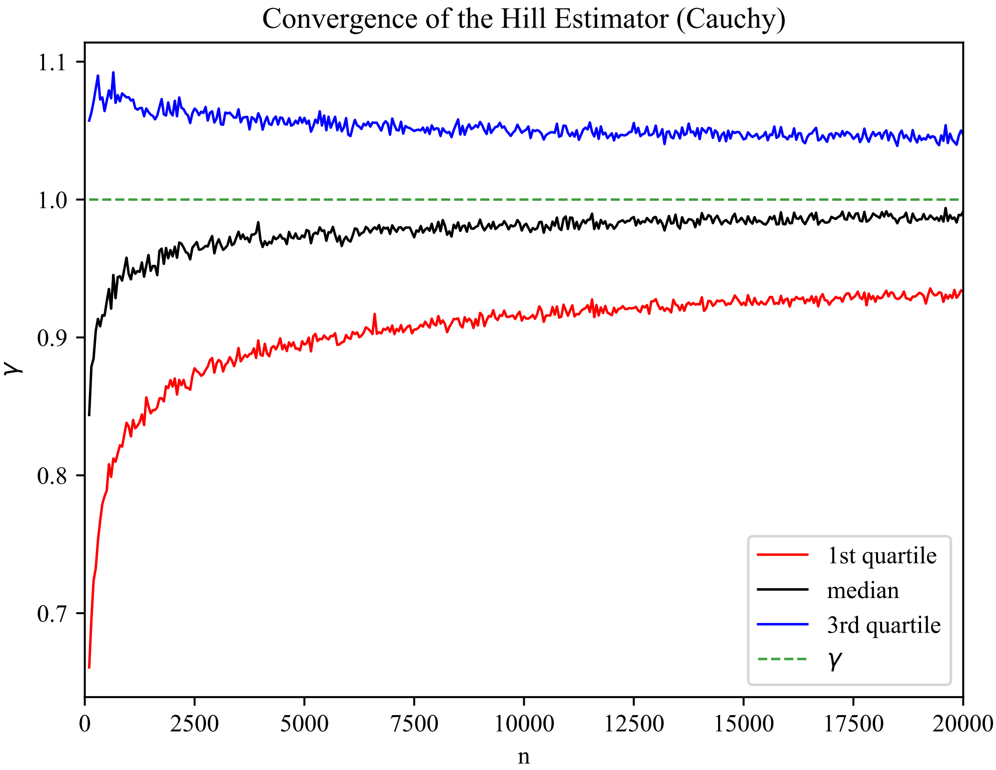

Figure B2: A plot of the median, the first quartile and the third quartile of simulated values of the Hill estimator under the Cauchy distribution. The observed quantities are plotted against the sample size. The used threshold was $k(n) = \lfloor \sqrt{n} \rfloor$ and each sample size was simulated 2000 times. The correct value of the extreme value index is displayed as a dashed line.