Weak properties and robustness of t-Hill estimators

Pavlina Jordanova, Zdeněk Fabián, Philipp Hermann, Lubos Strelec, Andrés Rivera, Stéphane Girard, Sebastián Torres, Milan Stehlík

HAL Id: hal-01327002

https://hal.archives-ouvertes.fr/hal-01327002

Submitted on 6 Jun 2016

Weak properties and robustness of t-Hill estimators

Pavlina Jordanova · **Zdeněk Fabián** · **Philipp Hermann** · **Luboš Střelec** · **Andrés Rivera** · **Stéphane Girard** · **Sebastián Torres** · **Milan Stehlík**

Abstract We describe a novel method of heavy tails estimation based on transformed score (t-score). Based on a new score moment method we derive the t-Hill estimator, which estimates the extreme value index of a distribution function with regularly varying tail. t-Hill estimator is distribution sensitive, thus it differs in e.g. Pareto and log-gamma case. Here, we study both forms of the estimator, i.e. t-Hill and t-lgHill. For both estimators we prove weak consistency in moving average settings as well as the asymptotic normality of t-lgHill estimator in iid setting. In cases of contamination with heavier tails than the tail of original sample, t-Hill outperforms several robust

P. Jordanova
Faculty of Mathematics and Informatics, Shumen University, Universitetska Str. 115, 9700 Shumen, Bulgaria

Z. Fabián
Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 18200 Prague, Czech Republic

P. Hermann
Department of Applied Statistics, Johannes Kepler University Linz, Altenbergerstrasse 69, 4040 Linz, Austria

L. Střelec
Department of Statistics and Operation Analysis (FBE), Mendel University, Zemědelská 1, 61300, Brno, Czech Republic

A. Rivera
Laboratorio de Glaciología, Centro de Estudios Científicos (CECs), Arturo Prat 514, Valdivia, Chile / Departamento de Geografía, Universidad de Chile, Portugal 84, Santiago

S. Girard · S. Torres
INRIA Rhône-Alpes, team Mistis, Inovallée, 655, av. de l'Europe, Montbonnot, 38334 Saint-Ismier cedex, France.

M. Stehlík
Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile / Department of Applied Statistics, Johannes Kepler University Linz, Altenbergerstrasse 69, 4040 Linz, Austria
Tel.: +43 732 2468 6806
Fax: +43 732 2468 6800
E-mail: mlnstehlik@gmail.com

tail estimators, especially in small samples. A simulation study emphasizes the fact that the level of contamination is playing a crucial role. The larger the contamination, the better are the t-score moment estimates. The reason for this is the bounded t-score of heavy-tailed distributions (and, consequently, bounded influence functions of the estimators). We illustrate the developed methodology on a small sample data set of stake measurements from Guanaco glacier in Chile.

Keywords Point estimation, Asymptotic properties of estimators, t-Hill estimator, t-lgHill estimator

Mathematics Subject Classification (2000) 62F10 · 62F12

1 Introduction

The aim of this paper is to introduce a novel method of heavy tails estimation based on transformed score. Especially for small or/and contaminated samples such a method can be better than standard maximum likelihood, since it relates to the method of moments for transformed scores (see Fabián (2015); Hosking and Wallis (1987)). Let us denote RV_a the class of regularly varying functions at infinity, with an index of regular variation equal to $a \in \mathbb{R}$, i.e. positive measurable functions $g(\cdot)$ such that for all $x > 0$, $g(tx)/g(t) \to x^a$, as $t \to \infty$.

Here we suppose that $\mathbf{X}_1, X_2, \ldots, X_n$ are possibly dependent copies of X with distribution function (d.f.) F, upper order statistics

$$X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$$

and

$$1 - F \in RV_{-\alpha}, \quad \alpha > 0. \tag{1}$$

We define α to be the tail parameter and $\gamma := 1/\alpha$ to be the extreme value index.

Hill (1975) derived a procedure of Pareto tail estimation by MLE, obtaining the following Hill estimator

$$H_{k_n,n} = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \left\{ \frac{X_{(n-i+1,n)}}{X_{(n-k_n,n)}} \right\}, \quad k_n = 1, 2, \ldots, n-1. \tag{2}$$

Later on, many authors tried to robustify the Hill estimator, but they still relied on maximum likelihood method. Alves (2001) has introduced a new lower bound and Gomes and Oliveira (2003) respectively Li et al. (2010) have introduced powers of original statistics. However, the influence function of Hill estimator is slowly increasing but unbounded, thus, the Hill procedure is not robust. Further approaches of robustifying the original Hill estimator were given in Beran and Schell (2012); Vandewalle et al. (2007). In Fabián (2001) a new score method of score moment estimators has been proposed. It appeared that these score moment estimators are robust for very heavy tailed distributions, see Stehlík et al. (2010a). Jordanova and Pancheva (2012) consider an independent identically distributed (i.i.d.) sample and find the limit distribution of the t-Hill estimator for a fixed number k of the threshold order statistics. They prove that a sample of Pareto distributed observations does not

have to be large in order to receive the corresponding limit distribution for fixed k. In that case under suitable normalizations and a large sample the t-Hill estimator is asymptotically normal for $k(n) \to \infty$. Under the more general conditions, the t-Hill estimator is asymptotically normal for $k(n) = o(n)$. The Hill estimator procedure with the score moment estimator has been investigated in Stehlík et al. (2012) for optimal testing for normality against Pareto tail. Recently, nice generalizations of t-Hill have been published, see Brilhante et al. (2013); Paulauskas and Vaiciulis (2013); Beran et al. (2014).

Resnick and Starica (1993) generalize the Hill estimator for more general settings with possibly dependent data. In this paper we continue these investigations, since dependencies are expected in real data applications. We obtain weak consistency of the t-Hill estimator for a special class of dependent data, the infinite moving average model. Moreover, we provide some examples showing that in contrast to the i.i.d. case the t-Hill and the Hill estimators applied to the moving average model are not robust with respect to large observations. Under the concept of the Hill estimator we understand the successive averaging of ordered values up to given k. In this paper we understand "*The Hill estimator*" as a specific procedure for studying the tail of Pareto like distributions. Instead of implementing "The Hill estimator" procedure, we implement the t-score moment procedure, which can give different estimators for different families. We illustrate t-Hill and we also quantify the robustness and compare efficiency with other competitors.

The paper is organized as follows: in the next section we recall the theory of score function of the distribution introduced firstly by Fabián (2001) and studied later in series of works. In section 3 we provide asymptotical results (asymptotic consistency and asymptotic normality) for t-Hill estimators for Pareto and log-gamma distributions, under iid and moving average settings. In section 4 we compare maximum likelihood, t-Hill, transformed score moment and Hill estimators.We also introduce t-lgHill estimator, which is a transformed score moment estimator for case of loggamma distribution. Contamination of underlying data is controlled by the extreme value index, shifting parameter or the means of transformed score variance of contaminating Pareto distribution. Comparisons show that in cases when the contamination has heavier tail than the original distribution, t-Hill estimator outperforms Hill estimator and several robust tail estimators (Integrated Squared Error Estimator (ISE), Partial Density Component (PDC) estimator (Vandewalle et al. (2007)), Least Squares estimator (LS), moment estimator (ME, see Dekkers et al. (1989)), QQ estimator (Kratz and Resnick (1996)) or Weighted MLE estimator (WMLE, see Dupuis and Morgenthaler (2002); Dupuis and Victoria-Feser (2006)). We investigate robustness of t-Hill estimator for a particular case of moving average sequence. In section 5 we provide an application of t-Hill in comparison to Hill and Zipf's estimator for a small sample data set from Guanaco Glacier. Summary concludes the paper. For reasons of readability we provide all proofs in the Appendix.

2 Transformed Score Function and Transformed Score Moment Estimators

Pearson and Filon (1898), Edgeworth (1908a), Edgeworth (1908b), and Fisher (1925) developed the basic statistical inference function, the so called score function. Their score is the gradient with respect to parameter θ of the logarithm of the likelihood function $S(\theta, X) = \frac{\partial}{\partial \theta} \log L(\theta; X)$, indicating the sensitivity of L (its derivative normalized by its value). The variance of the score is the Fisher information $\mathscr{I}(\theta) = E_\theta(S^2)$.

Fabián (2001, 2007, 2010, 2015) has introduced a more general scalar-valued inference function, in this paper called transformed score function (or t-score) of distribution, reflecting main features of continuous probability distributions and enabling an introduction of their new relevant characteristics.

The t-score of distribution G with support $\mathbb{R}$ and unimodal differentiable density g is a function

$$S_G(y) = -\frac{g'(y)}{g(y)} \tag{3}$$

expressing the relative rate of the change of g. For a distribution G_μ with location parameter $\mu \in \mathbb{R}$, that is, with density in the form $g(x - \mu)$, the generalized t-score function equals the Fisher score for μ as

$$S_G(y - \mu) = \frac{\partial}{\partial \mu} \log g(y - \mu).$$

We consider that t-score in equation (3) is a significant function of parametric distribution $G_\theta, \theta \in \Theta \subseteq \mathbb{R}$, and the mode $y^* : S_G(y; \theta) = 0$ its 'central' point.

Example 1 *The t-score of $\mathscr{N}(\mu, \sigma)$ is $S_{\mathscr{N}}(y) = \frac{y - \mu}{\sigma^2}$ and $y^* = \mu$.*

In general, the t-score of distribution F with arbitrary interval support $\mathscr{X} \subseteq \mathbb{R}$ is defined by the following construction.

Definition 1 *Let $\eta : \mathscr{X} \to \mathbb{R}$ be a strictly increasing smooth mapping. The t-score of distribution F with interval support $\mathscr{X} \subseteq \mathbb{R}$ and (almost surely) two-times differentiable density $f(x)$ is*

$$T_F(x) = -\frac{1}{f(x)} \frac{d}{dx} \left[\frac{1}{\eta'(x)} f(x) \right]. \tag{4}$$

Supposing that the solution x^* of equation

$$T_F(x) = 0 \tag{5}$$

is unique, it is called the t-score mean and function

$$S_F(x) = \eta'(x^*) T_F(x) \tag{6}$$

the generalized t-score function of distribution F.

Definition 1 can be explained as follows: F is a transformed distribution $F(x) = G(\eta(x))$, where $G(y)$ is a 'prototype' with support $\mathbb{R}$. F has density

$$f(x) = g(\eta(x))\eta'(x) \tag{7}$$

where $\eta'(x) = d\eta(x)/dx$ is the Jacobian of the transformation. It was shown in Fabián (2001) that the t-score is actually the transformed generalized t-score function of the prototype, $T_F(x) = S_G(\eta(x))$. The Jacobian of the transformation does not carry any information about the distribution. The term in brackets of equation (4) is the density without the Jacobian.

The justification of expression (6) is the following: consider a location prototype G_μ with density $g(y - \mu)$. The density of the transformed distribution $F_\tau = G_\mu \circ \eta(x)$ is

$$f(x; \tau) = g(\eta(x) - \eta(\tau))\eta'(x),$$

where $\tau = \eta^{-1}(\mu)$ is a so called *transformed location parameter*. It was proven in Fabián (2007) that the generalized t-score function with transformed location parameter is identical with the Fisher score for this parameter,

$$S_F(x; \tau) = \frac{d}{dx}\log f(x; \tau).$$

Generalized t-score function (6) and its parametric form $S_F(x; \theta)$ generalize the concept of the Fisher score for distributions with arbitrary support without a 'central' parameter or without parameters at all. If the prototype G has mode y^*, the score mean of the transformed distribution is the transformed mode of the prototype G, $x^* = \eta^{-1}(y^*)$. The generalized t-score function is actually the (generalized) Fisher score for the score mean, which may not be a parameter of the distribution and is unique if the prototype is unimodal.

The mapping $\eta(x)$ occurring in Definition 1 is often the inner part of $f(x)$, so that $\eta(x)$ and/or $\eta'(x)$ in equation (7) are clearly identifiable. This η is called an innate mapping; Fabián (2015) discusses how to choose a most suitable innate mapping if it is not apparent from the density formula.

Let us consider two examples, related to log-gamma and Pareto distributions studied later in this paper. For other examples, e.g. beta-prime distribution see Fabián (2015).

Example 2 *The log-gamma distribution with support $\mathscr{X} = (1, \infty)$ has density*

$$f(x) = \frac{\alpha^c}{\Gamma(c)}(\log x)^{c-1}\frac{1}{x^{\alpha+1}} = \frac{\alpha^c}{\Gamma(c)}(\log x)^c x^{-\alpha}\frac{1}{x\log x} \tag{8}$$

so that $\eta'(x) = \frac{1}{x\log x}$ and the innate mapping is $\eta(x) = \log\log x$. By (4) one obtains

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[(\log x)^\alpha x^{-\alpha}] = \alpha\log x - c,$$

with score mean $x^* = e^{c/\alpha}$.

Example 3 *The density*

$$f(x) = \frac{\alpha}{x^{\alpha+1}} \qquad (9)$$

of the Pareto distribution with $\mathscr{X} = (1,\infty)$ does not contain any 'visible' Jacobian of any transformation. By using $\eta'(x)$ from the foregoing example we would receive

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[x\log x f(x)] = \alpha \log x - 1,$$

which is the Fisher score for α. By using $\eta(x) = \log(x-1)$, $\eta'(x) = 1/(x-1)$, the t-score (4) is

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[(x-1)f(x)] = \alpha - (\alpha+1)/x = \alpha(1 - x^*/x), \qquad (10)$$

with score mean $x^* = (\alpha+1)/\alpha$. Since the 'shifted' Pareto with $\mathscr{X} = \mathbb{R}_+$ and density $f(x) = \frac{\alpha}{(x+1)^{\alpha+1}}$ is the particular case of the beta type II distribution with bounded generalized t-score, the latter one is to be taken as the t-score of the Pareto distribution.

The score moments (SM) were introduced for any $k \in \mathscr{N}$ instead of the ordinary moments by

$$M_k(\theta) = ES_F^k(\theta) = \int_{\mathscr{X}} S_F(x;\theta)^k f(x;\theta)\, dx, \qquad (11)$$

existing only if f satisfies the usual regularity requirements. It appeared that the score moments are often expressed by elementary functions of parameters. It is easy to see that $M_1 = 0$. The value $M_2 = ES_F^2$ of distributions with (transformed) location parameter is the Fisher information for (transformed) location parameter. Accordingly, ES_F^2 is the Fisher information for the t-mean. The reciprocal value

$$\omega^2 = \frac{1}{ES_F^2}, \qquad (12)$$

the t-score variance, appeared to be a measure of the variability (dispersion) of the distribution even in cases where the usual variance does not exist (see Fabián (2007)).

Notice, that in case of the Pareto distribution $ET_F^2 = \alpha/(\alpha+2)$, $ES_F^2 = (x^* - 1)^{-2}ET_F^2$ so that

$$\omega_F^2 = \frac{\alpha+2}{\alpha^3}. \qquad (13)$$

Let $X_1,\ldots,X_n$ be i.i.d. random variables according to some F. Assuming F as a member of the model family $\{F_\theta, \theta \in \Theta\}$, $\Theta \in \mathbb{R}^m$, Fabián (2001, 2010) introduced the t-score moment estimate $\hat{\theta}_{SM}$ as the solution of equations

$$\hat{\theta}_{SM}: \quad \frac{1}{n}\sum_{i=1}^n S_F^k(x_i;\theta) = ES_F{}^k(\theta), \qquad k = 1,\ldots,m, \qquad (14)$$

the statistical counterpart of (11). It was shown that $\hat{\theta}_{SM}$ is consistent and asymptotically normal. The score moment estimators take the assumed form of the distribution into account, similarly as the maximum likelihood (ML) ones. However, since data

enter into estimation equations only by means of $S_F(x_i;\theta)$, and transformed scores of heavy-tailed distributions appeared to be bounded, the transformed score moment estimates of all parameters are protected against outliers in cases of heavy-tailed distributions. Let us remark that since t-scores and generalized t-score functions differ only in a constant factor, one can use the t-scores instead of generalized t-score functions in the score moment equations.

The t-score moment equations for log-gamma and Pareto distributions are:

Example 4 *Log-gamma distribution: As $ET_F^2 = c$, equations (14) are*

$$\sum_{i=1}^{n} (\alpha \log x_i - c) = 0 \tag{15}$$

$$\frac{1}{n} \sum_{i=1}^{n} (\alpha \log x_i - c)^2 = c \tag{16}$$

with solution $\alpha = \frac{s_1}{s_2 - s_1^2}, c = \frac{s_1^2}{s_2 - s_1^2}$ where $s_1 = \frac{1}{n}\sum_{i=1}^{n} \log x_i$ and $s_2 = \frac{1}{n}\sum_{i=1}^{n} (\log x_i)^2$.

Example 5 *Pareto distribution: From the first moment equation $\sum_{i=1}^{n}(1 - x^*/x_i) = 0$ we obtain $\hat{x}^* = \bar{x}_H$, where*

$$\bar{x}_H = n/\sum_{i=1}^{n} 1/x_i$$

is the harmonic mean, and

$$\hat{\alpha} = 1/(\bar{x}_H - 1). \tag{17}$$

We use the same notations of order statistics as in section 1, see (1). Hence, due to (17), the t-Hill estimator of $\gamma = \alpha^{-1}$ was suggested in the form

$$H_{k_n,n}^* = \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} \frac{X_{(n-k_n,n)}}{X_{(n-i+1,n)}} \right\}^{-1} - 1, \quad k_n = 1, 2, ..., n-1. \tag{18}$$

This estimator was firstly published in technical report of Fabián and Stehlík (2009). Since it is based on harmonic mean, which is a generalized t-Hill estimator, it is expected to be resistant to large observations to a certain extent so that it could yield more realistic values than the ordinary Hill estimator. Further generalizations have been made by Brilhante et al. (2013); Beran et al. (2014). In particular, the tradeoff between efficiency and robustness has been studied in Beran et al. (2014), a mean of order $p \geq 0$ (MOP) generalization is given by Brilhante et al. (2013).

For the definition of the t-lgHill estimator we use the same notations of order statistics as in section 1, see (1), and

$$M_n^{(j)} = \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\log \frac{X_{(n-i+1,n)}}{X_{(n-k_n,n)}} \right)^j, \quad j = 1, 2, ...$$

Hence the t-lgHill estimator of $\gamma = \alpha^{-1}$ has the form

$$H_{k_n,n}^L = \frac{M_n^{(2)} - (M_n^{(1)})^2}{M_n^{(1)}}. \tag{19}$$

If we understand the t-Hill estimator as an algorithm depending on the assumed distribution, the t-Hill for log-gamma distribution is, according to the solution of (15) and (16)

$$H_{k_n,n}^L = \frac{\sum_{i=1}^{k_n} \log^2\left\{\frac{X_{(n-i+1,n)}}{X_{(n-k_n,n)}}\right\}}{\sum_{i=1}^{k_n} \log\left\{\frac{X_{(n-i+1,n)}}{X_{(n-k_n,n)}}\right\}} - \frac{1}{k_n}\sum_{i=1}^{k_n} \log\left\{\frac{X_{(n-i+1,n)}}{X_{(n-k_n,n)}}\right\} \quad k_n = 1,2,...,n-1. \quad (20)$$

Let us denote $H_{k_n,n}^L$ sequence by t-lgHill. Referring to Section 2, the t-lgHill estimate of the tail index is given by (16) and has a closed-form expression as

$$H_{k_n,n}^L = \frac{s_2}{s_1} - s_1.$$

3 Asymptotical results

3.1 t-Hill Estimator in Case of Moving Average Sequence

Suppose at least one of the real numbers $c_j, j = 0,1,...$ is positive and there exists $\delta \in (0,1)$, $\delta < \alpha$ such that

$$\sum_{j=0}^{\infty} |c_j|^\delta < \infty. \quad (21)$$

Consider the moving average sequence

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}, \quad -\infty < n < \infty, \quad (22)$$

where $Z_i, -\infty < i < \infty$, are non-negative i.i.d. innovations with d.f. G, such that $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$. We will use the t-Hill estimator as introduced before and consider the point measure

$$\mu_{X,k_n,n}(\cdot) := \frac{1}{k_n}\sum_{i=1}^{n} \varepsilon\left\{\frac{X_i}{b(\frac{n}{k_n})} \in (\cdot)\right\},$$

as a random element in the space $\mathbb{E}^+$ of positive Radon measures on $(0,\infty]$ endowed with the vague topology. Here

$$b(t) := F^{\leftarrow}\left(1 - \frac{1}{t}\right) = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(t) \quad (23)$$

is the tail function of F. Let

$$k_n \to \infty, \quad \frac{k_n}{n} \to 0. \quad (24)$$

By Proposition 3.3 of Resnick and Starica (1993)

$$\frac{1}{k_n}\sum_{i=1}^{n} I_{\left[\frac{X_i}{b(\frac{n}{k_n})} > 0\right]} \varepsilon\left\{\frac{X_i}{b(\frac{n}{k_n})} \in (\cdot)\right\} \Rightarrow \mu, \quad n \to \infty,$$

in $\mathbb{E}^+$, where $\mu : \sigma((0,\infty]) \to [0,\infty)$ and $\mu(x,\infty] = x^{-\alpha}$, $x > 0$.

However, for all $n \in N$, $k_n = 1, 2, ..., n$ and $A \in \mathbb{E}^+$

$$\frac{1}{k_n} \sum_{i=1}^n I_{\left[\frac{X_i}{b(\frac{n}{k_n})} > 0\right]} \varepsilon \left\{ \frac{X_i}{b(\frac{n}{k_n})} \in A \right\} = \mu_{X,k_n,n}(A), \tag{25}$$

in distribution, therefore

$$\mu_{X,k_n,n} \Rightarrow \mu, \quad n \to \infty, \tag{26}$$

in $\mathbb{E}^+$.

By Proposition 2.1. in Resnick and Starica (1993), having an intermediate sequence,

$$\frac{X_{(n-k_n,n)}}{b\left(\frac{n}{k_n}\right)} \xrightarrow{\mathbb{P}} 1 \tag{27}$$

and by Proposition 2.2. in Resnick and Starica (1993)

$$\widehat{\mu}_{X,k_n,n}(\cdot) := \frac{1}{k_n} \sum_{i=1}^n \varepsilon \left\{ \frac{X_i}{X_{(n-k_n,n)}} \in (\cdot) \right\} \xrightarrow{\mathbb{P}} \mu(\cdot), \quad n \to \infty, \tag{28}$$

in $\mathbb{E}^+$.

We will use the following Potter's inequality for distribution functions with regularly varying tails. If a function H is regularly varying with exponent $\gamma \in \mathbb{R}$, then for $\varepsilon > 0$ there will exist $t_0(\varepsilon)$, such that for $t > t_0(\varepsilon)$ and $x \geq 1$,

$$(1 - \varepsilon)x^{\gamma - \varepsilon} \leq \frac{H(tx)}{H(t)} \leq (1 + \varepsilon)x^{\gamma + \varepsilon}. \tag{29}$$

Its proof can be found in de Haan (1970).

Lemma 1 *Let $X_1, X_2, ...$ be random variables with d.f. F such that*

$$\overline{F}(x) \in RV_{-\alpha}. \tag{30}$$

Assume that $k_n \to \infty$, $k_n/n \to 0$ and $\mu_{X,k_n,n} \Rightarrow \mu$ as $n \to \infty$. Define

$$\Phi_n = \frac{1}{k_n} \sum_{i=1}^{k_n} \varphi\left(\frac{X_{(n-i+1,n)}}{X_{(n-k_n,n)}} \right),$$

where φ is a differentiable function on $(1, +\infty)$ such that $x^{-\alpha}\varphi(x) \to \infty$ as $x \to \infty$ and $\int_t^\infty x^{-\alpha+\delta_1} |\varphi'(x)|\, dx < +\infty$, for all $t > 1$, $0 < \delta_1 < \alpha$. Then, Φ_n is a weakly consistent estimator of $\varphi(1) + \int_1^{+\infty} x^{-\alpha} \varphi'(x)dx$.

As a consequence of Lemma 1, considering $\varphi(x) = 1/x$ yields that $H^*_{k_n,n}$, defined in (18) is a weakly consistent estimator of $1/\alpha$.

Now we are ready to prove the weak consistency of the t-Hill estimator in case of infinite moving average (MA) sequence.

Proposition 1 *Suppose at least one of the real numbers $c_j, j = 0, 1, \ldots$ is positive and there exists $\delta \in (0,1)$, $\delta < \alpha$ such that condition (21) is satisfied. Consider the moving average sequence (22) where $Z_i, -\infty < i < \infty$, are non-negative i.i.d. innovations with d.f. G, such that $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$. If (24), then $H_{k,n}^*$ is a weakly consistent estimator for α^{-1}.*

3.2 t-lgHill Estimator in Case of Moving Average Sequence

Suppose at least one of the real numbers $c_j, j = 0, 1, \ldots$ is positive and there exists $\delta \in (0,1)$, $\delta < \alpha$ such that (21) is satisfied.

Consider the moving average sequence (22), where $Z_i, -\infty < i < \infty$, are non-negative i.i.d. innovations with d.f. G, such that $\overline{G} \in RV_{-\alpha}$, $c = \alpha > 0$.

Assume (24) holds. By weak consistency of Hill estimator we obtain that

$$M_n^{(1)} \xrightarrow{\mathbb{P}} \frac{1}{\alpha}.$$

Now we need to prove that

$$M_n^{(2)} - (M_n^{(1)})^2 \xrightarrow{\mathbb{P}} \frac{1}{\alpha^2},$$

i.e.

$$M_n^{(2)} \xrightarrow{\mathbb{P}} \frac{2}{\alpha^2} \tag{31}$$

and then to use Slutsky arguments. Consider the point measure

$$\mu_{X,k_n,n}(\cdot) := \frac{1}{k_n} \sum_{i=1}^{n} \varepsilon \left\{ \frac{X_i}{b(\frac{n}{k_n})} \in (\cdot) \right\},$$

as a random element in the space $\mathbb{E}^+$ of positive Radon measures on $(0, \infty]$ endowed with the vague topology, where b is defined in (23).

In section 3.1 we proved that under these conditions

$$\mu_{X,k_n,n} \Rightarrow \mu, \quad n \to \infty, \tag{32}$$

in $\mathbb{E}^+$ and (26) and (27) are true.

As a consequence of Lemma 1, letting $\varphi(x) = \ln(x)$ entails that $M_n^{(2)}$ is a weakly consistent estimator for $2/\alpha^2$.

Proposition 2 *Suppose at least one of the real numbers $c_j, j = 0, 1, \ldots$ is positive and there exists $\delta \in (0,1)$, $\delta < \alpha$ such that condition (21) is satisfied. Consider the moving average sequence (22) where $Z_i, -\infty < i < \infty$, are non-negative i.i.d. innovations with d.f. G, such that $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$. If (24), then*

i.) $M_n^{(2)}$ is a weakly consistent estimator for $2\alpha^{-2}$;
ii.) t-lgHill estimator $H_{k,n}^L$ is a weakly consistent estimator for α^{-1}.

3.3 t-lgHill Estimator is Asymptotically Normal in iid case

Definition 2 *If the tail function of a non-negative random variable X is $\bar{F} := 1 - F$ and $\bar{F} : \mathbb{R} \to [0;1]$ satisfies that $\bar{F} \in RV_{-\alpha}$ with $\alpha > 0$. Then $\bar{F}$ is said to be of second-order regular variation with parameter $\rho \leq 0$, if there exists a function $Q(t)$ that ultimately has a constant sign with $\lim_{t \to \infty} Q(t) = 0$ and a constant $c \neq 0$ such that*

$$\lim_{t \to \infty} \frac{\frac{\bar{F}(tx)}{\bar{F}(t)} - x^{-\alpha}}{Q(t)} = H_{\alpha,\rho}(x) = cx^{-\alpha} \int_1^x u^{\rho-1} du, \, x > 0 \qquad (33)$$

Then it is written as $\bar{F} \in 2RV_{-\alpha,\rho}$ and $Q(t)$ is referred to as the auxiliary function of $\bar{F}$.

It is known from de Haan and Stadtmüller (1996) or a more relevant form in Geluk et al. Geluk et al. (1997) that if $H_{\alpha,\rho}(x)$ is not a multiple of $x^{-\alpha}$ then $\rho < 0$ implies that there exists a $c \neq 0$ such that $H_{\alpha,\rho}(x) = cx^{-\alpha}\frac{x^\rho - 1}{\rho}$ and $|A| \in RV_\rho$ and no other choices of ρ are consistent with $Q(t) \to 0$. There are many distributions which satisfy the second order RV condition. These are (see e.g. Drees et al. (2000)): Cauchy $\gamma = 1, \rho = -2$, Fréchet(1) $\gamma = 1, \rho = -1$, Student t(4) $\gamma = 1/4, \rho = -1/2$, t(10) $\gamma = 1/10, \rho = -1/5$ or loggamma $\gamma = 1/3, \rho = 0$.

The most seminal results on this topic could be found in de Haan and Ferreira (2006).

Theorem 1 *Let $X_1, X_2, \ldots$ be iid random variables. Suppose that the second order condition 33 holds with $\alpha > 0$, and auxiliary function Q. If $\sqrt{k_n}Q(n/k_n) \to 0$ as $n \to \infty$ then,*

$$\sqrt{k_n}\left(\alpha H_{k_n,n}^L - 1\right) \xrightarrow{d} N(0,8).$$

4 Empirical study of the effect of contamination on Pareto tail index.

The maximum likelihood estimator $\hat{\alpha}_{ML,n} := \frac{1}{H_{k_n,n}}$ of Pareto tail index is very sensitive to deviations from the theoretical distributions, namely in the heavy-tailed class of distributions, see Stehlík et al. (2010b). It is unbiased, asymptotically consistent and has smallest variance in the set of all unbiased estimators for α. Its variance is equal to α^2/n. However, it is not robust with respect to large contamination because its finite sample upper breakdown point is 0. Therefore, we are looking for an estimator $\hat{\alpha}_n$ with asymptotic relative efficiency

$$\frac{Var\,\hat{\alpha}_{ML,n}}{Var\,\hat{\alpha}_n} = \frac{\alpha^2}{nVar\,\hat{\alpha}_n} \leq 1$$

and close to 1. The latter means that the estimator's variance should be close to α^2/n, although it will be larger than this value. It seems difficult to theoretically compare mean values and variances of Hill and t-Hill estimators for all distribution functions with regularly varying tails. In this section we empirically compare their properties.

A good discussion of this topic can be found in Finkelstein et al. (2006). First they make the Pareto probability integral transform of the initial data and then they use the properties of the Uniform distribution in order to obtain their estimator of α and to examine its properties. The biggest advantage of their approach is that in this way "even infinite contamination has a bounded effect on the transformed data". Robust methods considering contaminated distributions work much better, but sweep out the differences among distributions so that they do not use the complete prior information. The main concepts about the robustness are very well presented e.g. in Huber and Ronchetti (1981); Rousseeuw and Stahel (1986).

In order to compare the quality of the estimators we calculated the percentage relative bias of Hill, t-Hill and t-lgHill estimators (RB). It is defined as

$$RB = \frac{\frac{1}{m} \sum_{i=1}^{m} (\hat{\gamma}_i - \gamma)}{\gamma} 100,$$

where $\hat{\gamma}_i$ is correspondingly estimator of $\gamma = 1/\alpha$ for the $i-th$ simulated sample, $i = 1, 2, ..., m$.

Another characteristic that we use in order to determine the quality of these estimators is their relative root-mean-square error (RRMSE)

$$RRMSE = \frac{\sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{\gamma}_i - \gamma)^2}}{\gamma} 100,$$

see e.g. Brzezinski (2015).

Analogously to the Hill plot we consider the set of points with coordinates

$$\left(k_n, H_{k_n,n}^* \right), \quad k_n \in \{1, 2, ..., n\}.$$

Further on we call this plot 't-Hill plot'. Similarly for t-lgHill plot. The straight line represents the true value of $\gamma = 1/\alpha$ in all following figures in this section.

4.1 Independent observations.

4.1.1 t-Hill versus other robust estimators

In our Monte Carlo simulation we focus on parameter α of the classical (or Type I) Pareto distribution $P(\alpha, \delta)$, which is defined using cumulative distribution function as follows:

$$F(x) = 1 - \left(\frac{\delta}{x} \right)^{\alpha}, \quad x \geq \delta > 0, \tag{34}$$

where δ is a scale parameter and $\alpha > 0$ is a shape parameter, generally known as the Pareto tail index.

The existing literature offers various estimators for the Pareto tail index α. In our simulation study we focus primarily on the following estimators: Hill estimator (maximum likelihood estimator introduced by Hill (1975), which we use as a non-robust benchmark for the Pareto tail index), Integrated Squared Error Estimator (ISE,

estimator introduced by Vandewalle et al. (2007), which is based on the relative excesses of observations above a certain treshold), Partial Density Component estimator (PDC, estimator introduced by Vandewalle et al. (2007), which is also based on the relative excesses of observations above a certain treshold), Least Squares estimator (LS), moment estimator (ME, see Dekkers et al. (1989)), QQ estimator (QQ, estimator based on the Quantile-Quantile approach, see Kratz and Resnick (1996)), and finally Weighted MLE (WMLE, see Dupuis and Morgenthaler (2002); Dupuis and Victoria-Feser (2006)). For the purpose of comparison we use the percentage relative bias (RB) and the percentage relative root-mean-square error (RRMSE) as defined in section 4.

Therefore, the data sets are simulated from Pareto distribution $P(\alpha, 1)$, for various values of α, as well as from contaminated Pareto distribution with function

$$F = (1 - \varepsilon)P(\alpha_1, 1) + \varepsilon P(\alpha_2, 1) \tag{35}$$

where $\varepsilon = 0.05, 0.10$ and $\alpha_1 = 1, 1.7, \ \alpha_2 = 0.3, 0.5, 1$. For all cases we assume the appropriate sample sizes n.

4.1.2 t-Hill versus Hill and t-lgHill in Pareto sample

Table 1 presents the percentage relative bias (RB) and the percentage relative root-mean-square error (RRMSE) for the uncontamined Pareto distribution $P(\alpha, 1)$. If we suppose large sample size $n = 1000$, the results of RB and RRMSE of the t-Hill, the Hill estimator, ISE estimators are comparable. The largest relative bias as well as relative root-mean-square error have the ME and WML estimators.

| | $n = 1000$ | | | | | |
| | $\alpha = 0.5$ | | $\alpha = 1$ | | $\alpha = 3$ | |
	RB	RRMSE	RB	RRMSE	RB	RRMSE
Hill	0.14	3.18	0.10	3.20	0.08	3.20
tHill	0.14	4.27	0.08	3.67	0.06	3.30
t-lgHill	0.43	3.48	0.78	7.03	2.69	21.83
ISE	0.23	5.52	0.16	5.07	0.09	4.72
LS	0.83	4.59	0.79	4.59	0.78	4.54
ME	0.31	3.59	0.48	4.60	1.93	10.73
PDC	1.02	14.07	0.53	9.40	0.25	6.68
QQ	-0.90	4.49	-0.94	4.51	-0.95	4.46
WML	3.48	8.30	3.43	8.20	3.35	8.08

Table 1 Results of the percentage relative bias (RB) and the percentage relative root-mean-square error (RRMSE) for the various estimators of the Pareto tail index α with data drawn from uncontamined Pareto distribution $P(\alpha, 1)$.

The next investigations refine the conclusions made in Fabián and Stehlík (2009). All estimators are calculated by means of their corresponding values in $m = 100$ samples. First we observe samples with size $n = 40$. The means of the Hill, t-lgHill and t-Hill estimators, separately for any fixed $k_n = 1, 2, ..., 39$, together with their averages are given in Figure 1. They show that both estimators have similar properties for small samples, although the t-Hill oscillates stronger. The values of RB and RRMSE are given in Table 2.
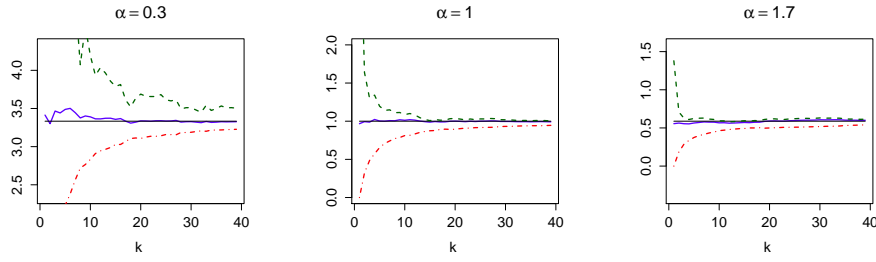
Fig. 1 t-Hill plots (green - -), Hill plots (blue) and t-lgHill (red -.-) for $\alpha = 0.3$ (left), $\alpha = 1$(middle) and $\alpha = 1.7$ (right), $n = 40$.

α	k_n	estimator	RRMSE	RB
0.3	10	Hill	31.62	0.04
		tHill	102.44	27.88
		t-lgHill	55.89	-18.24
	20	Hill	22.38	0.04
		tHill	47.02	11.08
		t-lgHill	44.12	-9.64
1	10	Hill	31.51	-0.03
		tHill	43.11	7.42
		t-lgHill	55.37	-18.21
	20	Hill	22.29	-0.05
		tHill	27.82	3.41
		t-lgHill	43.67	-9.54
1.7	10	Hill	31.66	-0.04
		tHill	37.46	4.57
		t-lgHill	55.37	-18.38
	20	Hill	22.35	-0.07
		tHill	25.17	2.13
		t-lgHill	43.82	-9.61

α	k_n	estimator	RRMSE	RB
0.3	100	Hill	9.92	-0.07
		tHill	16.27	1.78
		t-lgHill	21.65	-1.94
	200	Hill	7.04	-0.04
		tHill	11.35	0.96
		t-lgHill	15.59	-1.02
1	100	Hill	9.90	0.05
		tHill	11.65	0.73
		t-lgHill	21.68	-1.97
	200	Hill	7.01	-0.01
		tHill	8.14	0.31
		t-lgHill	15.48	-0.97
1.7	100	Hill	10.01	0.10
		tHill	10.83	0.53
		t-lgHill	21.76	-1.74
	200	Hill	7.13	0.00
		tHill	7.66	0.17
		t-lgHill	15.62	-0.78

Table 2 RRMSE and RB of Hill end t-Hill estimators for $n = 40$ i.i.d. data with Pareto distribution, for different α and k_n.

Table 3 RRMSE and RB of Hill end t-Hill estimators for $n = 500$ i.i.d. data with Pareto distribution, for different α and k_n.

We also made analogous observations for samples with size $n = 500$. However, this figure is not provided here. The same conclusions could be made also from the values of RB and RRMSE, given in Table 3.

4.1.3 t-Hill versus Hill and t-lgHill for distribution function with regularly varying tail at a logarithmic rate

Hill, t-lgHill and t-Hill estimators may perform very poorly if the slowly varying function in the tail is faraway from a constant. In (4.16) of Embrechts et al. (1997) the authors consider

$$F^{\leftarrow}(p) = (1-p)^{-1/\alpha}(-\ln(1-p)), \quad p \in (0,1), \tag{36}$$

with respect to the Hill estimator. We simulated samples of $n = 10000$ observations of random variables with quantile functions (36), separately for $\alpha = 0.3, 1, 1.7$ and plotted the Hill, t-lgHill and the t-Hill plots for $k_n = 5, 6, ..., 499$. The rate of convergence

of these estimators can be observed in Figure 2. Notice that the t-lgHill outperforms both Hill and the t-Hill estimators.
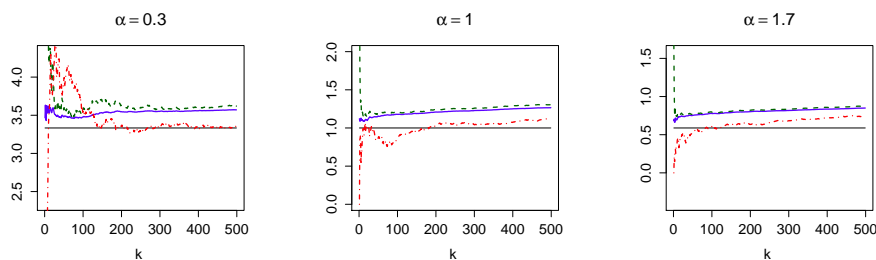
Fig. 2 t-Hill plots (green - -), Hill plots (blue) and t-lgHill (red -.-) for $\alpha = 0.3$ (left), $\alpha = 1$ (middle) and $\alpha = 1.7$ (right), $n = 10000$.

Although Hill estimator seems to be slightly closer to the estimated value, it is clear that three estimators are not good for such c.d.fs, see Table 4, however the t-lgHill has the best performance.

α	k_n	estimator	RRMSE	RB
0.3	100	Hill	11.93	5.41
		tHill	19.32	8.35
		t-lgHill	22.54	1.99
	200	Hill	9.62	6.19
		tHill	14.86	8.31
		t-lgHill	16.39	3.31
1	100	Hill	21.72	18.34
		tHill	25.42	20.79
		t-lgHill	26.00	11.40
	200	Hill	22.61	21.01
		tHill	25.78	23.55
		t-lgHill	21.64	13.69
1.7	100	Hill	33.49	31.00
		tHill	37.00	33.89
		t-lgHill	32.02	20.07
	200	Hill	36.78	35.62
		tHill	40.28	38.82
		t- lgHill	29.63	23.59

Table 4 RRMSE and RB of Hill end t-Hill estimators for $n = 10000$ i.i.d. data with Pareto distribution, for different α and k_n, $m = 10000$.

4.1.4 t-Hill versus t-lgHill and Hill for log-gamma sample

Figure 3 shows the t-lgHill for different α equal to 0.3, 1, 1.7 and log-gamma distribution (8), $c = 1$. Green line is for t-Hill, blue is Hill, red is lgHill.
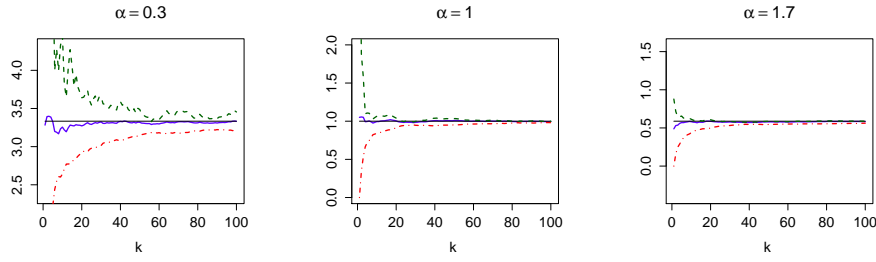
Fig. 3 t-Hill plots (green - -), Hill plots (blue) and t-lgHill (red -.-) for $\alpha = 0.3$ (left), $\alpha = 1$ (middle) and $\alpha = 1.7$ (right), $n = 400$.

4.1.5 t-Hill versus Hill for contaminated data from Pareto sample

Practical examples, see Fabián and Stehlík (2009) among others, have shown that data are expected to be contaminated in specific situations. Frequently there are outliers in the right tail of the distribution. It is known that the Hill estimator is not robust. This is due to the fact that data are entered by their logarithm in its calculation. But what about the t-Hill estimator? In (18) data are involved by their reciprocal value. Therefore, it is not difficult to deduce that these estimators are robust with respect to large values. They are sensitive to the center of the distribution.

Contamination, governed by α_2**.** t-Hill estimator provides the superior results with respect to the relative bias (RB) and the relative root-mean-square error (RRMSE) for the contaminated Pareto distribution with distribution function (35) among all employed estimators ISE, PDC, LS, ME, QQ and WMLE.

Here we simulated small $n = 40$ and bigger samples of $n = 200$ data from contaminated Pareto distribution as given in (35) separately for $\alpha_1 = 0.3, 1$ or 1.7 and $\alpha_2 = 0.5$ or 1. The relative part of the contamination is governed by the parameter ε. We worked with two values 0.1 and 0.05. When considering a sample size with 10% contamination, this would yield $\varepsilon = 0.1$. For sample sizes of 40, see Figure 4.

To obtain proper estimations for $1/\alpha_1$ we need k_n to be large and $k_n < n$. When $\alpha_1 < 1$ and $\alpha_2 \geq 1$ the Hill estimators are better than the t-Hill estimators. When $\alpha_1 < 1$ and $\alpha_2 < 1$, or $\alpha_1 \geq 1$ and $\alpha_2 > 1$ the t-Hill estimators are comparable to the corresponding Hill estimators. Finally if we observe the means of the estimators we can conclude that t-Hill estimators are better than Hill estimators for $\alpha_1 \geq 1$ and $\alpha_2 \leq 1$.

The same conclusions could be made on RB and RRMSE basis of both Hill and t-Hill estimators. They are given in Table 5 for $k = 10$ or 20 and for $k = 100$ or 200, and different values of α_1, α_2 and ε.

Comparison of maximum likelihood with t-score moment estimators

Let us compare maximum likelihood (ML) estimators with t-score moment (SM) estimators for two different cases, on the one hand based on a sample of one distribution and on the other hand in presence of contamination, i.e. distribution of the form

$$F_\varepsilon = (1 - \varepsilon)F(\omega) + \varepsilon F(\omega_1), \tag{37}$$

where $\omega_1 > \omega$ are score variances.
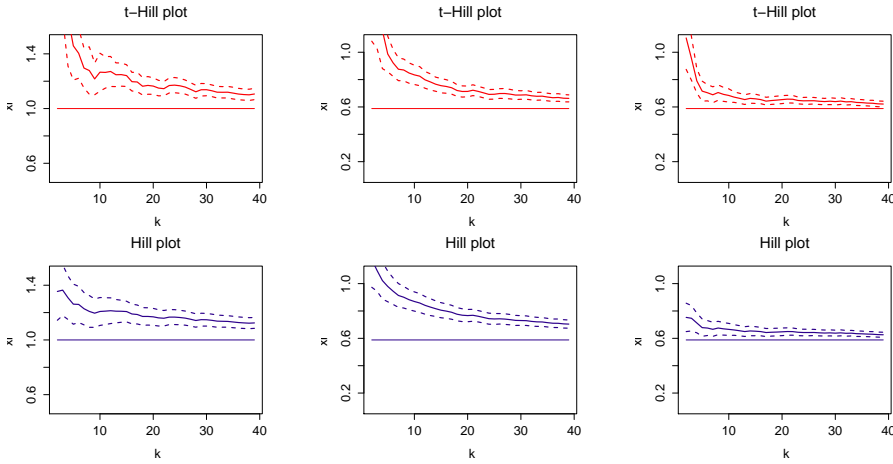
Fig. 4 Rate of convergence of the t-Hill (first line) and Hill (second line) estimators with $\alpha_1 = 1, \alpha_2 = 0.5$ (left), $\alpha_1 = 1.7, \alpha_2 = 0.5$ (middle) and $\alpha_1 = 1.7, \alpha_2 = 1$ (right), $\varepsilon = 0.1$ for $n = 40$.

Note: The score variance of the Pareto distribution is $\omega^2 = (\alpha + 2)/\alpha^3$. If $\omega = 1$, we have $1/\alpha = 0.657$. Pareto distribution with score variance ω^2 will be denoted by $Pa(\omega)$.

Random samples were generated from heavy-tailed distributions, with increasing level ω_1 of contamination and $\alpha = 1$.

Ratios of $r(k) = H_{k_n,n}/\gamma(\omega^*) - 1$ and $r^*(k) = H_{k_n,n}^*/\gamma(\omega^*) - 1$ for $k_n = 100, 250$ and 900 and $n = 1000$ are given in Table 6.

The variances of SM estimates are somewhat higher in accordance with the values of theoretical asymptotic efficiencies, but SM estimates are less biased and thus, do not seem to be much worse than the ML estimates even for not contaminated samples.

However, when data contamination comes into consideration, i.e. $\varepsilon = 0.1$ of $F(10)$, the estimation with t-scores gives better results than standard ML-method. In addition to a smaller bias for very small sample sizes ($n \leq 25$), variances are smaller for SM-method compared with ML-method the smaller the number of observations. Recall that for the two experiments new simulations have been performed such that a comparison of the goodness of the results between contaminated and non-contaminated data is obsolete.

4.1.6 t-Hill versus Hill for data from the log-gamma distribution

It is apparent that both, t-Hill and Hill, show systematic decrease with expecting heavier Pareto tail. Figure 5 shows paths of averages for $H_{k_n,n}, H_{k_n,n}^*$ and $H_{k_n,n}^L$ from 50 experiments with contaminated log-gamma data from (37) with $\omega = 5.546$ and $\varepsilon = 0.05$. The values of ω_1 in plots in Figure 5 are computed from log-gamma distribution.

Figure 5 shows paths of averages for $H_{k_n,n}, H_{k_n,n}^*$ and $H_{k_n,n}^L$ from 50 experiments with contaminated log-gamma data from (37) with $\omega = 5.546$ and $\varepsilon = 0.05$. The values of ω_1 in plots in Figure 5 are computed from log-gamma distribution.

n	α_1	α_2	k_n	ε	estimator	RRMSE	RB
40	1	0.5	10	0.1	Hill	45.39	18.42
					tHill	56.03	22.61
				0.05	Hill	37.89	9.33
					tHill	47.62	14.63
			20	0.1	Hill	31.02	13.800
					tHill	34.57	14.14
				0.05	Hill	25.96	6.63
					tHill	30.74	8.35
	1.7	0.5	10	0.1	Hill	41.01	-6.05
					tHill	37.50	-14.73
				0.05	Hill	39.70	-22.78
					tHill	38.87	-26.63
			20	0.1	Hill	30.76	-18.69
					tHill	32.82	-26.24
				0.05	Hill	35.59	-29.53
					tHill	37.19	-32.92
		1	10	0.1	Hill	40.65	-34.28
					tHill	40.47	-32.34
				0.05	Hill	42.81	-37.59
					tHill	42.26	-35.43
			20	0.1	Hill	38.84	-35.80
					tHill	38.80	-35.26
				0.05	Hill	40.97	-38.44
					tHill	40.64	-37.59
201	1	0.5	100	0.1	Hill	18.49	13.80
					tHill	16.87	10.51
				0.05	Hill	13.34	7.02
					tHill	13.61	5.60
			200	0.1	Hill	13.00	9.99
					tHill	11.57	7.32
				0.05	Hill	9.27	5.00
					tHill	9.35	3.74
	1.7	0.5	100	0.1	Hill	21.55	-18.51
					tHill	29.11	-27.88
				0.05	Hill	31.16	-29.86
					tHill	35.42	-34.63
			200	0.1	Hill	27.79	-26.98
					tHill	33.56	-33.14
				0.05	Hill	34.60	-34.16
					tHill	37.54	-37.22
		1	100	0.1	Hill	36.46	-35.84
					tHill	37.19	-36.53
				0.05	Hill	38.94	-38.43
					tHill	39.20	-38.62
			200	0.1	Hill	37.36	-37.07
					tHill	38.01	-37.70
				0.05	Hill	39.37	-39.12
					tHill	39.66	-39.38

Table 5 RRMSE and RB of Hill end t-Hill estimators for $n = 40$ and 200 i.i.d. data with contaminated Pareto distribution, for different α_1, α_2, k_n and ε.

It is apparent from Figure 5 that for small contamination t-lgHill (red) estimates γ from log-gamma data perform better than both Hill (blue) and t-Hill (green). However, for larger contamination it is definitely not robust, perhaps due to the squared

ε	$\omega*$	$\gamma(\omega*)$	$r(100)$	$r^*(100)$	$r(250)$	$r^*(250)$	$r(900)$	$r^*(900)$
	1.5	0.879	.0363	.0376	.0235	.0187	.0184	.0170
.05	3	1.500	.2027	.1517	.1333	.0982	.0679	.0473
	5	2.165	.4982	.3049	.2792	.1573	.1203	.0656
	1.5	0.879	.0683	.0707	.0502	1.0458	.0339	.0296
.10	3	1.500	.3842	.2828	.2512	.1799	.1340	.0970
	5	2.165	.9298	.6250	.5536	.3352	.2383	.1310

Table 6 Values of $r(k) = H_{k_n,n}/\gamma(\omega^*) - 1$ and $r^*(k) = H^*_{k_n,n}/\gamma(\omega^*) - 1$ are computed for contaminated distributions (P_{cont}) with three different ω^* and $\varepsilon = 0.05$ and 0.1, respectively.

contamination	estimate	$n=12$	$n=25$	$n=50$	$n=75$
	$1/H_{k_n,n}$	1.086	1.041	1.023	1.014
$\varepsilon = 0$	$1/H^L_{k_n,n}$	1.052	1.028	1.010	1.009
	Var $(1/H_{k_n,n})$	0.342	0.215	0.144	0.116
	Var $(1/H^L_{k_n,n})$	0.371	0.241	0.166	0.135
	$1/H_{k_n,n}$	1.298	1.188	1.129	1.127
$\varepsilon = 0.1$	$1/H^L_{k_n,n}$	1.011	1.090	1.198	1.144
	Var $(1/H_{k_n,n})$	0.203	0.073	0.030	0.020
	Var $(1/H^L_{k_n,n})$	0.105	0.061	0.037	0.022

Table 7 Comparison of ML and SM estimators of α (see 8) in addition to their variances with different sample sizes averaged over 5000 experiments without (rows 1-4) and with contaminated Pareto distribution (rows 5-8), i.e. $\varepsilon = 0.1$.
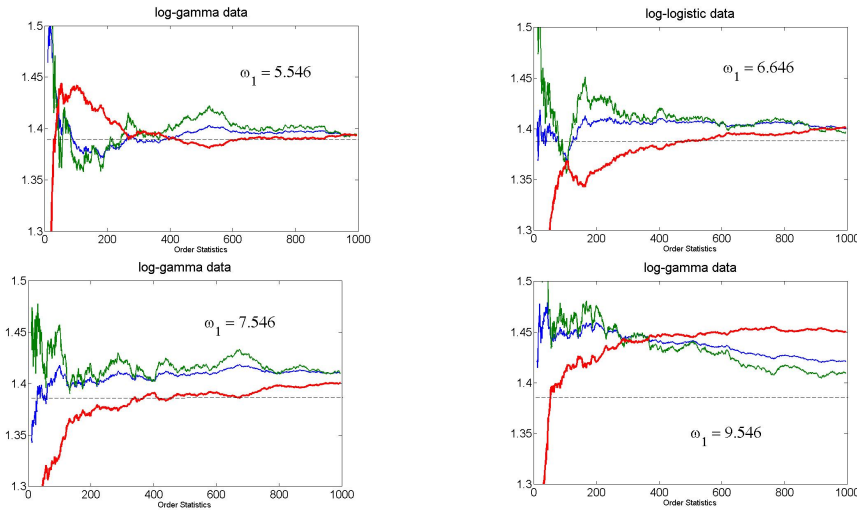
Fig. 5 Graphs of averages for $H_{k_n,n}$ (blue), $H^*_{k_n,n}$ (green) and $H^L_{k_n,n}$ (red) from 50 experiments with contaminated log-gamma data and increasing ω_1, $n = 1000$.

logarithmic term. Notice that t-Hill is more robust than the classical Hill even for non-Pareto data.

## 4.2 Infinite moving average process

In this subsection we consider the infinite moving average process with the following autoregressive form

$$X_i = 1.3X_{i-1} - 0.7X_{i-2} + Z_i, \quad i = 1, 2, ..., n. \tag{38}$$

It is discussed in Resnick and Starica (1993) with respect to the Hill estimator. Here we compare Hill estimator with the corresponding t-Hill estimator.

4.2.1 t-Hill versus Hill and lgHill for Pareto noise

The t-Hill and Hill plots of the averages of the corresponding estimators, calculated on samples of $n = 2000$ observations on moving average sequence (38) with noise Z_i that is $P(\alpha, 1)$ distributed and $k_n = 10, 11, ..., 499$ are given on Figure 6. The corresponding values of α are 0.3 in the first column, 1 in the second column and 1.7 in the third column. At first glance, we see that both estimators have a similar behavior for a fixed number of upper order statistics. However, the values of RRMSE and RB in Table 8 show that in this case the t-Hill estimators are better.
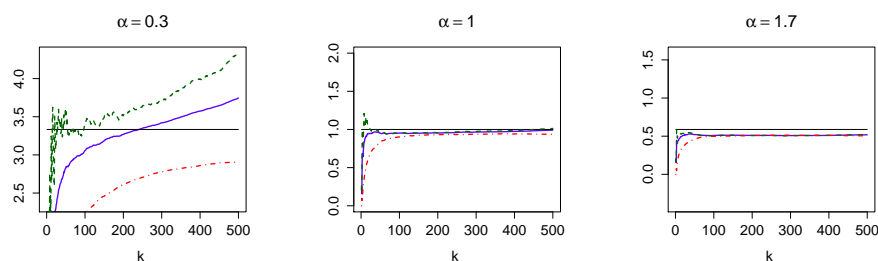
Fig. 6 t-Hill plots (green - -), Hill plots (blue) and t-lgHill (red -.-) for $\alpha = 0.3$ (left), $\alpha = 1$(middle) and $\alpha = 1.7$ (right), $n = 2000$.

4.2.2 t-Hill versus Hill for contaminated Pareto noise

Contamination, governed by ε.

Here we contaminated the distribution of the noise component, calculated the means of the corresponding Hill and t-Hill estimators for $k_n = 10, 11, ..., 9999$ and plotted them in Figure 7 and Figure 8 for relative parts on contamination $\varepsilon = 0.1$ and $\varepsilon = 0.05$. The observations in each sample are $n = 10000$. More precisely the distribution of the noise component Z_i in (38) is (35). The corresponding values of RRMSE and RB are given in Table 9. Although both estimators give almost the same results, in most of these cases the t-Hill estimator is slightly better than the Hill estimator.

α	k_n	estimator	RRMSE	RB
0.3	100	Hill	33.11	-10.26
		tHill	33.87	-0.69
		t-lgHill	45.93	-35.38
	200	Hill	24.45	-4.09
		tHill	23.68	3.38
		t-lgHill	39.80	-23.62
1	100	Hill	20.52	-2.19
		tHill	18.77	0.01
		t-lgHill	36.65	-14.47
	200	Hill	16.03	-1.11
		tHill	15.77	-0.05
		t-lgHill	30.49	-7.75
1.7	100	Hill	20.58	-3.69
		tHill	18.02	-4.98
		t-lgHill	45.27	-1.84
	200	Hill	15.25	-6.56
		tHill	14.27	-8.15
		t-lgHill	37.13	-0.03

Table 8 RRMSE and RB of Hill end t-Hill estimators for $n = 10000$ moving average data with Pareto distribution, for different α and k_n.
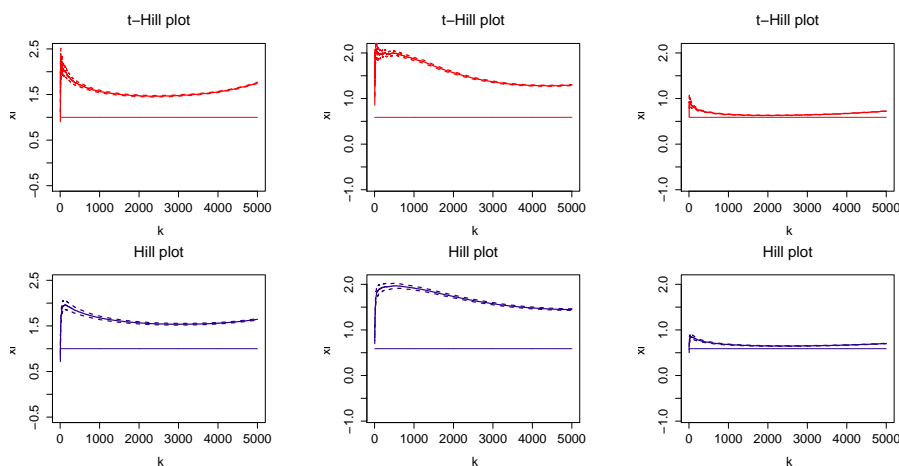
Fig. 7 Rate of convergence of the t-Hill (first line) and Hill (second line) estimators with $\alpha_1 = 1, \alpha_2 = 0.5$ (left), $\alpha_1 = 1.7, \alpha_2 = 0.5$ (middle) and $\alpha_1 = 1.7, \alpha_2 = 1$ (right), $\varepsilon = 0.1, n = 10000$.

4.3 Conclusions of the Simulations

Having in mind the observations in section 4 we can make the following conclusions:

1. when the contamination has heavier tail than the original distribution, t-Hill estimator outperforms Hill estimator and several robust tail estimators (ISE, PDC, LS, ME, QQ and WMLE). In contaminated i.i.d. case, where the severity of the contamination is governed by the extremal index, the Hill estimator has more narrow confidence intervals than t-Hill estimator. When $\alpha_1 < 1$ and $\alpha_2 \geq 1$ the Hill estimators are better than the t-Hill estimators. When $\alpha_1 < 1$ and $\alpha_2 < 1$,
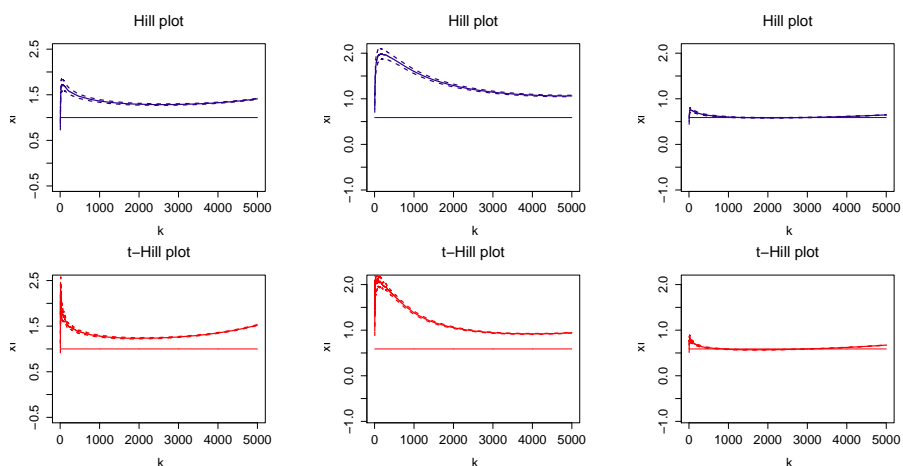
Fig. 8 Rate of convergence of the t-Hill (first line) and Hill (second line) estimators with $\alpha_1 = 1, \alpha_2 = 0.5$ (left), $\alpha_1 = 1.7, \alpha_2 = 0.5$ (middle) and $\alpha_1 = 1.7, \alpha_2 = 1$ (right), $\varepsilon = 0.05, n = 10000$.

α_1	α_2	k_n	ε	estimator	RRMSE	RB
1	0.5	2000	0.1	Hill	55.92	54.32
				tHill	47.05	46.10
			0.05	Hill	32.79	30.88
				tHill	25.38	24.30
		3000	0.1	Hill	52.85	51.80
				tHill	47.37	46.77
			0.05	Hill	31.15	29.97
				tHill	27.50	26.85
1.7	0.5	2000	0.1	Hill	197.34	195.51
				tHill	166.95	165.59
			0.05	Hill	120.22	117.81
				tHill	79.98	78.63
		3000	0.1	Hill	169.79	168.45
				tHill	133.34	132.47
			0.05	Hill	95.57	93.88
				tHill	61.26	60.42
	1	2000	0.1	Hill	11.52	9.67
				tHill	8.10	6.41
			0.05	Hill	5.64	-0.84
				tHill	5.31	-2.98
		3000	0.1	Hill	11.13	9.96
				tHill	9.41	8.50
			0.05	Hill	4.47	0.44
				tHill	3.67	-0.06

Table 9 RRMSE and RB of Hill end t-Hill estimators for $n = 10000$ moving average data with contaminated Pareto distribution, for different α_1, α_2, k_n and ε.

or $\alpha_1 \geq 1$ and $\alpha_2 > 1$ the t-Hill estimators are comparable to the corresponding Hill estimators. Finally we can conclude that t-Hill estimators are better than Hill estimators for $\alpha_1 \geq 1$ and $\alpha_2 \leq 1$. This effect increases for small samples.

2. Both Hill and t-Hill estimators are asymptotically consistent, however the t-Hill estimator has bigger variance.
3. The reason that the Hill estimator is not robust seems to be the fact that in its formula the data are entered by their logarithm. But what about the t-Hill estimator? In (18) data are involved by their reciprocal value. Therefore, it is not difficult to deduce that these estimators are robust with respect to large values. They are sensitive to the center of the distribution.
4. When the i.i.d. data contamination comes into consideration, i.e. $\varepsilon = 0.1$ of $F(10)$, the estimation with t-scores gives better results than standard ML-method. In addition to a smaller bias for very small sample sizes ($n \leq 25$), variances are smaller for SM-method compared with ML-method, especially for small samples.
5. In moving average case without contamination and with Pareto noise and $\alpha_1 \leq 1$ the t-Hill estimator is better with respect to the unbiasedness when $n \to \infty$ and $k_n/n \to 0$, but it has bigger variance than the Hill estimator.
6. In moving average cases with contaminated Pareto noise the t-Hill estimators are better than Hill only in case $\alpha_1 \geq 1$ and $\alpha_2 \leq 1$ and small relative part of contamination. The smaller the relative part of the contamination, the faster the rate of convergence. Moreover in any cases the estimators converge to $1/min(\alpha_1, \alpha_2)$ for $n \to \infty$, $k_n \to \infty$ and $k_n/n \to 0$.
7. In case when the observed random variable have log-gamma distribution, the t-lgHill estimator is more robust that t-Hill and Hill estimators. This is due to the fact that the t-lgHill estimator is distribution sensitive while the Hill and t-Hill estimators by construction in this case are useful only due to the information about the regular variation of the distribution.
8. From all of the charts it visible that the t-lgHill estimator has relatively small variance. It seems to be almost the same as of the Hill estimator.
9. In Moving Average case with Pareto noise the t-Hill outperforms both Hill and t-lgHill estimators. When alpha increases, the variance of all three estimator decreases.

The larger the number of samples m, e.g. $m = 1000$, the more visible are the conclusions about Hill, t-lgHill and t-Hill estimators, however computational time is very large. We shall recall a general consensus that we shall have a multi-criterial approach in order to estimate the parameters of heavy tail. Consistency is an important property, but we shall also consider importance of approximation of finite sample distribution and robustness.

5 Application to a small sample data from Guanaco glacier

A glacier is a solid ice mass which is fed by solid water (snow, hail or hoarfrost), transforms this solid water in ice and restores it via steam (evaporation/sublimation) or via liquid (water drained by the runoff stream-flow), see Francou and Pouyaud (2004). Such gain and loss of mass can be analyzed as a balance (or budget). Guanaco glacier is located in the III region of Chile (latitude 29S) in the semi-arid Andes, it has a surface area of $1.86[km^2]$ and a maximum thickness of $120[m]$, see Rabatel et al. (2011), the area contributing water to the Pacific Ocean (included in the above

calculation of 23%), is only $1.26[km^2]$, in other words, less than 35% of the total ice area in this zone. The rest of the Guanaco glacier contributes water to the Atlantic ocean. In this sense, the relative importance of glacier meltwater to the river basin as a complete unit is very low compared to the snow and rain contributions. Its geographical location is shown in Figure 9, this glacier (jointly with other minor glaciers and glacierets in the zone) is highly relevant for studies, because it is an important water reservoir in a semi-arid zone of Chile, contributing with up to the 23% of the streamflow in the corresponding valley, see Gascoin et al. (2011). All glaciers in the Pascua Lama region (3.64 km2) can contribute between 3 and 23 % to the very high basin area of the Huasco river. The glacier meltwater contribution goes from 100% at the foot of the glacier, down to near 0 at the sea outlet of this river (Huasco). Given that this zone is one of the world's most richly endowed territories of copper, there is high interest for the occupation of these water resources, resulting in a uneven mining-agricultural competition for water rights, see Oyarzún and Oyarzún (2011).
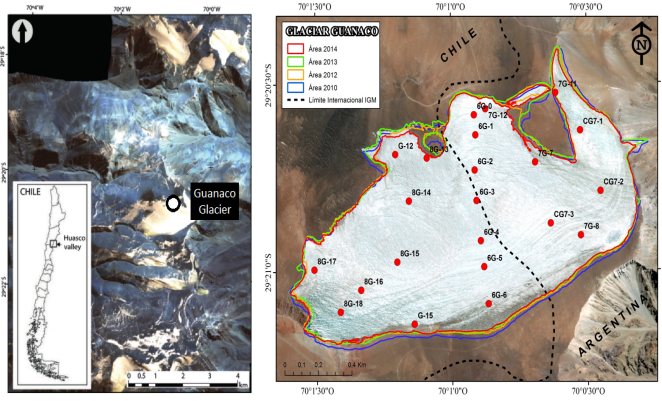

Fig. 9 Geographical location of Guanaco glacier and stakes where mass balance was measured in the period 2011-2014 along with aerial changes of the glacier computed via geodetic methods. Left image by Abermann et al. (2014), right image courtesy of the Center for Scientific Studies (CECs).

The mass balance observations over the glacier consists of annual records for the period 2002-2014, the dataset corresponds to the negated mass balance because we are interested in extremely large losses of volume in the glacier (see Figure 10).

In order to study a threshold model for this data, estimations for the tail index γ are computed using Hill, t-Hill (as previously defined) and Zipf's estimator (see Kratz and Resnick (1996)) which is a form of least square estimator (Schultze and Steinebach (1996)) of the tail index. It follows the fact that if we suspect that $X_{1,n} \leq ... \leq X_{n,n}$ are the order statistics from a Pareto family, then the plot of

$$\left\{ \left(-\log\left(1 - \frac{i}{n+1}\right), \log X_{i,n} \right), 1 \leq i \leq n \right\},$$
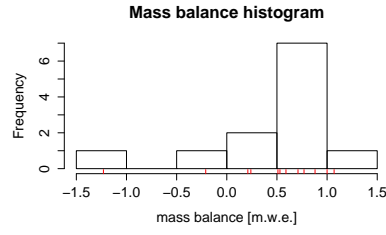
**Fig. 10** Histogram for the dataset of mass balance in Guanaco glacier, $n = 12$.

should be roughly linear with intercept 0 and slope $\hat{\gamma}$. The Zipf's estimator is computed for a given k_n, as

$$\hat{\gamma}_{\mathrm{zipf}}(k_n) = \mathrm{slope}\left(\left\{-\log\left(1 - \frac{i}{k_n + 1}\right), \log X_{n-k_n+i,n}, 1 \le i \le k_n\right\}\right).$$

We used $\hat{\sigma} = X_{n-k_n+1,n}\hat{\gamma}$ as estimator of the scale parameter. The asymptotic standard deviation is computed as $\frac{H_{k_n,n}}{\sqrt{k_n}}$ for Hill's estimator, as $\frac{\hat{\gamma}_{\mathrm{zipf}}(k_n)}{\sqrt{k_n/2}}$ for Zipf's, and for t-Hill as

$$\frac{H^*_{k_n,n}}{\sqrt{k_n}}\left(\frac{1 + H^*_{k_n,n}}{\sqrt{1 + 2H^*_{k_n,n}}}\right).$$

Zipf's estimator plot (see Figure 11) shows linear stability up to $k_n = 7$, which makes this value of k_n a good option for further computations.
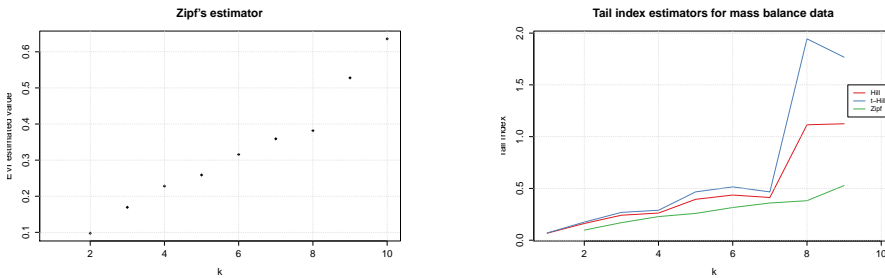
Fig. 11 Left: Zipf's estimator plot for several values of k_n. Right: Extreme value index computed over the dataset using 3 estimators and different values for k_n.

Hill and t-Hill estimators were computed along with Zipf's estimator, taking into account $k_n = 6$ and $k_n = 7$. The stability of the parameter estimation in a small sample context was tested using a parametric bootstrap approach, where 1000 samples of size 10 were generated from GP distribution with shape parameter computed with the three previously discussed estimators. The results are displayed in form of histograms in Figure 12.

Table 10 summarizes the information about the EVI obtained from the simulation:
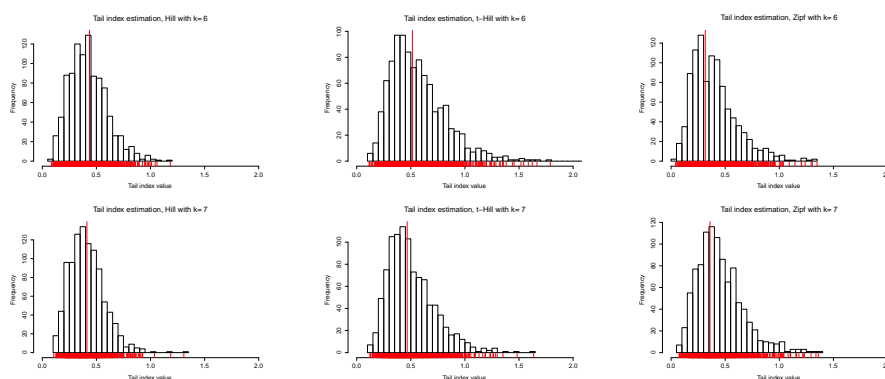
Fig. 12 Histograms showing stability of the estimated tail index using different estimators, bin-width is chosen using the Freedman-Diaconis rule. Estimation over the original sample is shown as a vertical red line. Length of x-axis is fixed.

	Estimator	Initial	Mean	Std.Dev.	Asymp.sd.
	Hill	0.436	0.429	0.171	0.178
$k_n = 6$	t-Hill	0.515	0.541	0.241	0.224
	Zipf	0.315	0.401	0.210	0.257
	Hill	0.412	0.413	0.153	0.156
$k_n = 7$	t-Hill	0.466	0.485	0.198	0.186
	Zipf	0.359	0.441	0.208	0.271

Table 10 The initial value is the obtained using the estimator over the original sample, the mean and standard deviation accounts for the simulated samples.

Given the results for each estimator, it is clear that $k_n = 7$ is a better choice than $k_n = 6$ because estimations are less biased and all estimators are more close to each other, also, Zipf's estimator should be treated with more care in this context because it shows an important bias in the parametric bootstrap without an improvement in the variance. The main advantage of usage of t-Hill estimator is its robustness and good small sample properties.

In Figure 13 we plot Hill, t-Hill, Zipf and t-lgHill for 40+ data sample of the Pascualama zone.

In Figure 14 we provide a QQ-plot with assumed theoretical distribution of GPD with parameters: $\mu = 0, \sigma = 1, \xi \in \{-0.5, -0.25, 0.25, 0.5\}$, with cdf defined for $\xi \neq 0$ as $G(z) = 1 - [1 + \xi(z - \mu)/\sigma]^{-1/\xi}$.

6 Summary

In this paper we introduce and study properties of t-score estimator under different model settings. In particular, we derived asymptotic normality for t-lgHill estimator
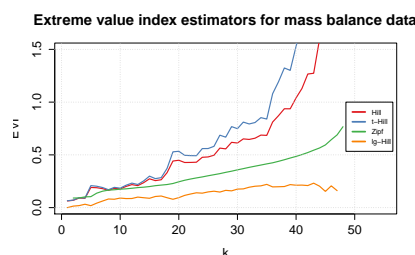
Fig. 13 Hill, t-Hill, Zipf and t-lgHill estimators (right) for 40+ data sample of the Pascualama zone.
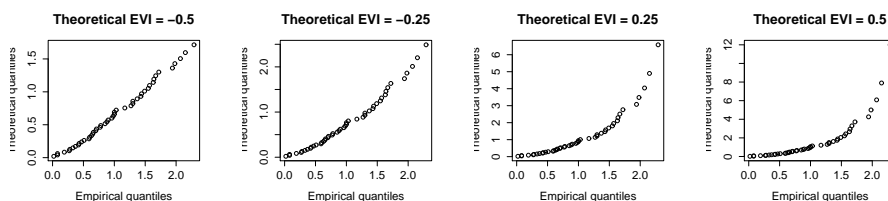

Fig. 14 QQ-Plots with different shapes of -0.5, -0.25, 0.25 and 0.5 plotted from top left to bottom right.

for iid case. We also derived consistency for t-Hill and t-lgHill under i.i.d. and moving average setup. We can conclude that:

i) The t-score is an unbounded function in cases of light-tailed distributions and a bounded function when distributions are heavy-tailed. Therefore, it seems to be extraordinary suitable for the estimation of parameters by the generalized moment method. It secures robustness 'when it is needed by the assumed distribution'. In this sense, the suggested estimates are 'naturally' robust without need of using any further function of robust statistics. In particular robustness of t-Hill can be well applied in cases when existence of first or second moment is questionable. Such an application for LATAM returns, i.e. daily log-returns (in percent) of an equity fund investing in Latin America (LATAM), can be found in Stehlík and Hermann (2015).

ii) It appeared that in cases of very small contamination, the maximum likelihood and t-moment method give similar results. However, the larger the contamination, the better are the t-moment estimates.

iii) It is apparent that an introduction of the t-mean and t-variance for distributions without ordinary mean and variance enables to describe the behavior of the estimates in a unique way and thus, compares results of the estimation for various assumed families of distributions having different parameters.

iv) We propose t-Hill estimator for possibly dependent data and investigate its weak consistency and asymptotic behavior. Particularly we consider the infinite moving average model. Besides the i.i.d. case there are cases where the t-Hill estimator is more robust than the Hill estimator. We applied t-Hill for small sample size real data for Glacier and received satisfactory results despite a small sample size.

References

Abermann J, Kinnard C, MacDonell S (2014) Albedo variations and the impact of clouds on glaciers in the chilean semi-arid andes. Journal of Glaciology 60(219):183–191

Alves MF (2001) A location invariant hill-type estimator. Extremes 4(3):199–217

Beran J, Schell D (2012) On robust tail index estimation. Computational Statistics & Data Analysis 56(11):3430–3443, DOI 10.1016/j.csda.2010.05.028

Beran J, Schell D, Stehlík M (2014) The harmonic moment tail index estimator: asymptotic distribution and robustness. Annals of the Institute of Statistical Mathematics 66:193–220

Billingsley P (2013) Convergence of probability measures. John Wiley & Sons

Brilhante M, Gomes M, Pestana D (2013) A simple generalization of the hill estimator. Computational Statistics & Data Analysis 57:518–535

Brzezinski M (2015) Robust estimation of the pareto tail index: a monte carlo analysis. Empirical Economics DOI 10

Cline D (1983) Infinite series of random variables with regularly varying tails. Tech. Rep. No 83-24, Institute of Applied Mathematics and Statistics, University of British Columbia

Dekkers AL, de Haan L (1989) On the estimation of the extreme-value index and large quantile estimation. Ann Statist 17(4):1795–1832, DOI 10.1214/aos/1176347396

Dekkers AL, Einmahl JH, de Haan L (1989) A moment estimator for the index of an extreme-value distribution. The Annals of Statistics pp 1833–1855

Drees H, de Haan L, Resnick S (2000) How to make a hill plot. The Annals of Statistics 28:254–274

Dupuis DJ, Morgenthaler S (2002) Robust weighted likelihood estimators with an application to bivariate extreme value problems. The Canadian Journal of Statistics/La Revue Canadienne de Statistique pp 17–36

Dupuis DJ, Victoria-Feser MP (2006) A robust prediction error criterion for pareto modelling of upper tails. Canadian Journal of Statistics 34(4):639–658

Edgeworth F (1908a) On the probable errors of frequency constants. Journal of the Royal Statistical Society 71:381–397

Edgeworth F (1908b) On the probable errors of frequency constants (contd). Journal of the Royal Statistical Society 71:499–512

Embrechts P, Klueppelberg C, Mikosch T (1997) Modeling extremal events for insurance and finance. Springer-Verlag

Fabián Z (2001) Induced cores and their use in robust parametric estimation. Communication in Statistics, Theory Methods 30:537–556

Fabián Z (2007) Estimation of simple characteristics of samples from skewed and heavy-tailed distribution. Recent Advances in Stochastic Modeling and Data Analysis pp 43–50

Fabián Z (2010) Score moment estimators. In: Lechavallier Y, Saporta G (eds) Proc. of conference COMPSTAT'2010, Physica-Verlag, Springer

Fabián Z (2015) Score function of distribution and revival of the moment method. Communications in Statistics - Theory and Methods 0(ja):00–00, DOI 10.1080/03610926.2013.857688

Fabián Z, Stehlík M (2009) On robust and distribution sensitive hill like method. Tech. rep., IFAS Reasearch Paper Series 43(4)

Finkelstein M, Tucker HG, Alan Veeh J (2006) Pareto tail index estimation revisited. North American actuarial journal 10(1):1–10

Fisher R (1925) Statistical Methods For Research Workers. Cosmo study guides, Cosmo Publications, URL http://books.google.at/books?id=4bTttAJR5kEC

Francou B, Pouyaud B (2004) Métodos de observación de glaciares en los Andes tropicales, mediciones de campo y procesamiento de datos. Versión 1. Great-Ice, IRD

Gascoin S, Kinnard C, Ponce R, Macdonell S, Lhermitte S, Rabatel A (2011) Glacier contribution to streamflow in two headwaters of the huasco river, dry andes of chile. The Cryosphere (5):1099–1113

Geluk J, de Haan L, Resnick S, Staarica C (1997) Second-order regular variation, convolution and the central limit theorem. Stochastic Process Appl 69:139–159

Gomes MI, Oliveira O (2003) Censoring estimators of a positive tail index. Statistics & Probability Letters 65:147–159

de Haan L (1970) On regular variation and its application to the weak convergence of sample extremes. Mathematisch Centrum

de Haan L, Ferreira A (2006) Extreme Value Theory: An introduction. Springer Series in Operations Research and Financial Engineering. Springer, New York

de Haan L, Stadtmüller U (1996) Generalized regular variation of second order. Journal of the Australian Mathematical Society 61:381–395

Hill B (1975) A simple general approach to inference about the tail of a distribution. Annals of Statistics 3:1163–1174

Hosking JR, Wallis JR (1987) Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29(3):339–349

Huber PJ, Ronchetti E (1981) Robust statistics, series in probability and mathematical statistics

Jordanova P, Pancheva E (2012) Weak asymptotic results for t-hill estimator. Comptes rendus de l'académie bulgare des sciences 65(12):1649–1656

Kratz M, Resnick SI (1996) The qq-estimator and heavy tails. Stochastic Models 12(4):699–724

Li J, Peng Z, Nadarajah S (2010) Asymptotic normality of location invariant heavy tail index estimator. Extremes 13:269–290

Oyarzún J, Oyarzún R (2011) Sustainable development threats, inter-sector conflicts and environmental policy requirements in the arid, mining rich, northern chile territory. Sustainable Development 19(4):263–274

Paulauskas V, Vaiciulis M (2013) On the improvement of hill and some other estimators. Lithuanian Mathematical Journal 53(3):336–355

Pearson K, Filon L (1898) Mathematical contributions to the theory of evolution. iv. on the probable errors of frequency constants and on the influence of random se-

lection on variation and correlation. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 191:229–311, DOI 10.1098/rsta.1898.0007

Rabatel A, Castebrunet H, Favier V, Nicholson L, Kinnard C (2011) Glacier changes in the pascua-lama region, chilean andes (29 s): recent mass balance and 50 yr surface area variations. The Cryosphere 5(4):1029–1041

Resnick S, Starica C (1993) Consistency of hill's estimator for dependent data. Tech. rep., Technical report No. 1077, School of Operational Research and Industrial Engeneering Cornell University Itaca NY 14853

Rousseeuw FHERP, Stahel W (1986) Robust statistics: the approach based on influence functions. J Wiley, New York

Schultze J, Steinebach J (1996) On least squares estimates of an exponential tail coefficient. Statistics & Risk Modeling 14(4):353–372

Stehlík M, Hermann P (2015) Letter to the editor. Annals of Applied Statistics 9:2051

Stehlík M, Potocký R, Fabián Z (2010a) On the favourable estimation of fitting heavy tailed data. Computational Statistics 25:485–503

Stehlík M, Potocký R, Waldl H, Fabián Z (2010b) On the favourable estimation of fitting heavy tailed data. Tech. rep., IFAS res. report Nr. 32 JKU Linz

Stehlík M, Fabián Z, Střelec L (2012) Small sample robust testing for normality against pareto tails. Communications in Statistics - Simulation and Computation 41:1167–1194

Vandewalle B, Beirlant J, Christmann A, Hubert M (2007) A robust estimator for the tail index of pareto-type distributions. Computational Statistics & Data Analysis 51:6252–6268

A Proofs

A.1 Proof of Lemma 1

Proof The result is proved analogously to Resnick and Starica (1993). An integration by parts entails

$$\Phi_n = \int_1^\infty \varphi(x) \left\{ \frac{1}{k_n} \sum_{i=1}^n \varepsilon \left\{ \frac{X_i}{X_{(n-k_n,n)}} \in dx \right\} \right\}$$

$$= \int_1^\infty \varphi(x) \widehat{\mu}_{X,k_n,n}(dx) = -\int_1^\infty \varphi(x) d\, \widehat{\mu}_{X,k_n,n}(x,\infty]$$

$$= \lim_{x\to 1} \varphi(x) \widehat{\mu}_{X,k_n,n}(x,\infty] - \lim_{x\to\infty} \varphi(x) \widehat{\mu}_{X,k_n,n}(x,\infty] + \int_1^\infty \widehat{\mu}_{X,k_n,n}(x,\infty] d\, \varphi(x).$$

The positive measures $\widehat{\mu}_{X,k_n,n}$ are Radon and the measure $\widehat{\mu}_{X,k_n,n}$ converges to μ and $\alpha > 0$. Therefore, in view of (28) and the definition of μ, the Hospital rule yields

$$\lim_{n\to\infty} \lim_{x\to\infty} \varphi(x) \widehat{\mu}_{X,k_n,n}(x,\infty] = \lim_{x\to\infty} \lim_{n\to\infty} \varphi(x) \widehat{\mu}_{X,k_n,n}(x,\infty] = \lim_{x\to\infty} \frac{\varphi(x)}{x^\alpha} = 0,$$

and similarly,

$$\lim_{n\to\infty} \lim_{x\to 1} \varphi(x) \widehat{\mu}_{X,k_n,n}(x,\infty] = \lim_{x\to 1} \lim_{n\to\infty} \varphi(x) \widehat{\mu}_{X,k_n,n}(x,\infty] = \varphi(1).$$

Then, for all $t \geq 1$,

$$\lim_{n\to\infty} \Phi_n = \lim_{n\to\infty} \left\{ \int_1^t \widehat{\mu}_{X,k_n,n}(x,\infty] \varphi'(x)\, dx + \int_t^\infty \widehat{\mu}_{X,k_n,n}(x,\infty] \varphi'(x)\, dx \right\}.$$

Let us first remark that

$$\lim_{n\to\infty}\int_1^t \widehat{\mu}_{X,k_n,n}(x,\infty]\,\varphi'(x)\,dx = \int_1^t \mu(x,\infty]\,\varphi'(x)\,dx = \int_1^t x^{-\alpha}\varphi'(x)\,dx$$

and therefore,

$$\lim_{t\to\infty}\lim_{n\to\infty}\int_1^t \widehat{\mu}_{X,k_n,n}(x,\infty]\,\varphi'(x)\,dx = \int_1^{+\infty} x^{-\alpha}\varphi'(x)\,dx.$$

Second, in order to apply Theorem 4.2. of Billingsley (2013), we have to check that

$$\lim_{t\to\infty}\lim_{n\to\infty}P\left\{\int_t^\infty \widehat{\mu}_{X,k_n,n}(x,\infty]\,|\varphi'(x)|\,dx > \varepsilon\right\} = 0. \tag{39}$$

Our aim is to replace $\widehat{\mu}_{X,k_n,n}$ with $\mu_{X,k_n,n}$. To this aim, we choose $\delta > 0$ and consider

$$P\left\{\int_t^\infty \widehat{\mu}_{X,k_n,n}(x,\infty]\,|\varphi'(x)|\,dx > \varepsilon\right\}$$
$$= P\left\{\int_t^\infty \mu_{X,k_n,n}\left(\frac{X_{(n-k_n,n)}}{b(n/k_n)}x,\infty\right]|\varphi'(x)|\,dx > \varepsilon\right\}$$
$$= P\left\{\int_t^\infty \mu_{X,k_n,n}\left(\frac{X_{(n-k_n,n)}}{b(n/k_n)}x,\infty\right]|\varphi'(x)|\,dx > \varepsilon,\left|\frac{X_{(n-k_n,n)}}{b(n/k_n)}-1\right|<\delta\right\}$$
$$+ P\left\{\int_t^\infty \mu_{X,k_n,n}\left(\frac{X_{(n-k_n,n)}}{b(n/k_n)}x,\infty\right]|\varphi'(x)|\,dx > \varepsilon,\left|\frac{X_{(n-k_n,n)}}{b(n/k_n)}-1\right|\geq\delta\right\}.$$

Clearly,

$$0 \leq P\left\{\int_t^\infty \mu_{X,k_n,n}\left(\frac{X_{(n-k_n,n)}}{b(n/k_n)}x,\infty\right]|\varphi'(x)|\,dx > \varepsilon,\left|\frac{X_{(n-k_n,n)}}{b(n/k_n)}-1\right|\geq\delta\right\}$$
$$\leq P\left\{\left|\frac{X_{(n-k_n,n)}}{b(n/k_n)}-1\right|\geq\delta\right\}$$

which, from (27), converges to zero when $n\to\infty$. In view of (27), for any $0<\delta_0<1+\alpha$ there exists n_{δ_0} large enough such that for all $t>1$ and $n>n_{\delta_0}$,

$$P\left\{\int_t^\infty \mu_{X,k_n,n}\left(\frac{X_{(n-k_n,n)}}{b(n/k_n)}x,\infty\right]|\varphi'(x)|\,dx > \varepsilon,\left|\frac{X_{(n-k_n,n)}}{b(n/k_n)}-1\right|<\delta\right\}$$
$$\leq P\left\{\int_t^\infty \mu_{X,k_n,n}((1-\delta_0)x,\infty]\,|\varphi'(x)|\,dx > \varepsilon,\left|\frac{X_{(n-k_n,n)}}{b(n/k_n)}-1\right|<\delta\right\}$$
$$\leq P\left\{\int_t^\infty \mu_{X,k_n,n}((1-\delta_0)x,\infty]\,|\varphi'(x)|\,dx > \varepsilon\right\}.$$

Chebishev's inequality guarantees that

$$P\left\{\int_t^\infty \mu_{X,k_n,n}((1-\delta_0)x,\infty]\,|\varphi'(x)|\,dx > \varepsilon\right\} \leq \frac{1}{\varepsilon}E\int_t^\infty \mu_{X,k_n,n}((1-\delta_0)x,\infty]\,|\varphi'(x)|\,dx.$$

Since $X_1,X_2,\ldots,X_n$ are indentically distributed, it follows that

$$E\mu_{X,k_n,n}((1-\delta_0)x,\infty] = \frac{1}{k_n}\sum_{i=1}^n P\left\{\frac{X_i}{b(n/k_n)}\in((1-\delta_0)x,\infty]\right\}$$
$$= \frac{n}{k_n}P\left\{\frac{X_1}{b(n/k_n)}\in((1-\delta_0)x,\infty]\right\}$$
$$= \frac{n}{k_n}\overline{F}((1-\delta_0)xb(n/k_n))$$
$$= \frac{n}{k_n}\overline{F}(b(n/k_n))\frac{\overline{F}((1-\delta_0)xb(n/k_n))}{\overline{F}(b(n/k_n))}.$$

From the definition of b, $n/k_n \to \infty$ implies $b(n/k_n) \to \infty$ and

$$\frac{n}{k_n}\overline{F}(b(n/k_n)) \to 1$$

as $n \to \infty$. Thus, in view of (30) and Potter's inequality for regularly varying functions, for any $0 < \delta_1 < \alpha$ there exist x_0 and n_{δ_1} large enough such that, for all $x > \max(1, x_0)$ and $n > n_{\delta_1}$,

$$E\mu_{X,k_n,n}((1-\delta_0)x,\infty] \le (1+\delta_1)((1-\delta_0)x)^{-\alpha+\delta_1},$$

and for all $t > 1$,

$$\int_t^\infty E\mu_{X,k_n,n}((1-\delta_0)x,\infty]\,|\varphi'(x)|\,dx \le (1+\delta_1)(1-\delta_0)^{-\alpha+\delta_1}\int_t^\infty x^{-\alpha+\delta_1}\,|\varphi'(x)|\;dx.$$

We apply Fubini's theorem and obtain that for all $t > 1$ and $n > n_\delta$,

$$P\left\{\int_t^\infty \mu_{X,k_n,n}((1-\delta_0)x,\infty]\,|\varphi'(x)|\,dx > \varepsilon\right\}$$

$$\le \frac{1}{\varepsilon}\int_t^\infty E\mu_{X,k_n,n}((1-\delta_0)x,\infty]\,|\varphi'(x)|\,dx$$

$$\le \frac{1}{\varepsilon}(1+\delta_1)(1-\delta_0)^{-\alpha+\delta_1}\int_t^\infty x^{-\alpha+\delta_1}\,|\varphi'(x)|\;dx.$$

As a consequence,

$$\lim_{t\to\infty}\lim_{n\to\infty}P\left\{\int_t^\infty \mu_{X,k_n,n}\left(\frac{X_{(n-k_n,n)}}{b(n/k_n)}x,\infty\right]\,|\varphi'(x)|\,dx > \varepsilon, \left|\frac{X_{(n-k_n,n)}}{b(n/k_n)}-1\right| < \delta\right\} = 0$$

and (39) is satisfied. $\qquad\square$

A.2 Proof of Proposition 1

Proof The random variables X_n, $-\infty < n < \infty$, are identically distributed. Cline (1983) proves that under these settings

$$\overline{F}(x) = P\left(\sum_{j=0}^\infty c_j Z_j > x\right) \sim \sum_{j:c_j>0}^\infty c_j^\alpha \overline{G}(x) \in RV_{-\alpha}.$$

We use (26), apply the above Lemma 1 and complete the proof. $\qquad\square$

A.3 Proof of Proposition 2

Proof i.) The random variables X_n, $-\infty < n < \infty$, are identically distributed. Cline (1983) proves that under these settings

$$\overline{F}(x) = P\left(\sum_{j=0}^\infty c_j Z_j > x\right) \sim \sum_{j:c_j>0}^\infty c_j^\alpha \overline{G}(x) \in RV_{-\alpha}.$$

We use (32), apply Lemma 1 and complete the proof.

ii.) We use Slutsky's arguments and obtain weak consistency of the t-lgHill estimator in case of infinite moving average (MA) sequence. $\qquad\square$

A.4 Proof of Theorem 1

Proof Recall that the t-lgHill estimator $H_{k_n,n}^L$ of $\gamma = \alpha^{-1}$ can be written as

$$H_{k_n,n}^L = \frac{M_n^{(2)} - (M_n^{(1)})^2}{M_n^{(1)}}, \tag{40}$$

while the Moment estimator is defined as

$$\widehat{\gamma}_M = M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1}, \tag{41}$$

see (3.5.9), p.102, in de Haan and Ferreira (2006), also (1.7) in page 1834 of Dekkers and de Haan (1989)). The difference between the two estimators appears clearly when comparing (40) and (41). In the following, we shall also make use of $\widehat{\gamma}_- := \widehat{\gamma}_M - M_n^{(1)}$, see Remark 3.5.7 in de Haan and Ferreira (2006)), which is an estimator of $\gamma_- := \min(0, \gamma)$. It is thus clear that the t-lgHill estimator (40) c an be rewritten as

$$H_{k_n,n}^L = \frac{M_n^{(1)}}{1 - 2\widehat{\gamma}_-}. \tag{42}$$

Corollary 3.5.6 of de Haan and Ferreira (2006) states that

$$\sqrt{k_n}\left\{1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1} - \gamma_-\right\} \xrightarrow{d} (1 - 2\gamma_-)(1 - \gamma_-)^2\left\{\left(\frac{1}{2} - \gamma_-\right)Q - 2P\right\},$$

where Q, P are Gaussian distributions given in Lemma 3.5.5 of de Haan and Ferreira (2006). Using the previous notations and remarking that $\gamma_- = 0$, this result can be simplified as $\xi_n := \sqrt{k_n}\widehat{\gamma}_-$ converges in distribution to $\frac{1}{2}Q - 2P$.

Similarly, let $\xi_n' := \sqrt{k_n}(M_n^{(1)} - \gamma_+)$ where $\gamma_+ := \max(0, \gamma) = \gamma = 1/\alpha$. It is easily seen that ξ_n' can be expanded as $\xi_n' = \sqrt{k_n}(\widehat{\gamma}_M - \gamma) - \sqrt{k_n}\widehat{\gamma}_-$. The limit distribution of the first term is established in (3.5.24), page 109 of de Haan and Ferreira (2006):

$$\sqrt{k_n}(\widehat{\gamma}_M - \gamma) \xrightarrow{d} \frac{1}{2}Q - (\gamma - 2)P.$$

The limit distribution of the second term has already been established to be $\frac{1}{2}Q - 2P$ and therefore ξ_n' converges in distribution to γP. Plugging ξ_n and ξ_n' in (42) yields

$$\begin{aligned}
H_{k_n,n}^L &= \frac{\gamma + k_n^{-1/2}\xi_n'}{1 - 2k_n^{-1/2}\xi_n} \\
&= (\gamma + k_n^{-1/2}\xi_n')(1 + 2k_n^{-1/2}\xi_n(1 + o_p(1))) \\
&= \gamma + k_n^{-1/2}(\xi_n' + 2\gamma\xi_n) + o_p(k_n^{-1/2}).
\end{aligned}$$

Thus, $\sqrt{k_n}(H_{k_n,n}^L - \gamma) = \xi_n' + 2\gamma\xi_n + o_p(1) \xrightarrow{d} \gamma P + 2\gamma(\frac{1}{2}Q - 2P) = \gamma(Q - 2P)$. According to Lemma 3.5.5 of de Haan and Ferreira (2006), (P, Q) is a bivariate centered Gaussian random vector with covariance matrix $\begin{pmatrix} 1 & 4 \\ 4 & 20 \end{pmatrix}$ The asymptotic variance in such case is $var(Q - 2P) = var(Q) + 4var(P) - 4cov(P, Q) = 8$. As a conclusion, we have

$$\sqrt{k_n}\left(\frac{H_{k_n,n}^L}{\gamma} - 1\right) \xrightarrow{d} N(0, 8).$$

$\square$