TRIMMING AND THRESHOLD SELECTION IN EXTREMES

MARTIN BLADT, HANSJÖRG ALBRECHER, AND JAN BEIRLANT

Abstract. We consider removing lower order statistics from the classical Hill estimator in extreme value statistics, and compensating for it by rescaling the remaining terms. As a function of the extent of trimming, the resulting trajectories are unbiased estimators of the tail index of exact Pareto samples, with lower variance than the Hill estimator when using the same amount of data. For the regularly varying case, the classical threshold selection problem in tail estimation is revisited with this method, both visually via trimmed Hill plots and, for the Hall class, also mathematically via minimizing the expected empirical variance. This leads to a simple threshold selection procedure for the classical Hill estimator, and at the same time also suggests an alternative estimator of the tail index, which assigns more weight to large observations, and works particularly well for relatively lighter tails. A simple ratio statistic routine is suggested to evaluate the goodness of the implied selection of the threshold. We illustrate the performance and potential of the proposed method with simulation studies and real insurance data.

1. Introduction

The use of Pareto-type tails has been shown to be important in different areas of risk management, such as for instance in computer science, insurance and finance. In social sciences and linguistics the model is referred to as Zipf’s law. This model corresponds to the max-domain of attraction of a generalized extreme value distribution with a positive extreme value index (EVI) $\xi$:

$$1 - F(x) = x^{-1/\xi} \ell(x), \quad \xi > 0,$$

where $\ell$ denotes a slowly varying function at infinity:

$$\lim_{x \to \infty} \frac{\ell(ux)}{\ell(x)} = 1, \text{ for every } u > 0.$$

Since the appearance of the paper of Hill (1975) in which the EVI estimator

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n}$$

was proposed with

$$X_{n,n} \geq X_{n-1,n} \geq \cdots \geq X_{n-i+1,n} \geq \cdots \geq X_{1,n}$$

denoting the ordered statistics of a random sample from $F$, the literature on estimation of $\xi > 0$ and other tail quantities such as extreme quantiles and tail probabilities has increased exponentially. We refer to Embrechts et al. (2013), Beirlant et al. (2004), de Haan and Ferreira (2007), and Gomes and Guillou (2015) for detailed discussions and reviews of these estimation problems. Next to the proposal of numerous estimators, focus has gradually shifted to selection methods of $k$ and to the construction of bias-reduced estimators which exhibit plots of estimates which, as a function of $k$,
are as stable as possible. Indeed, plots of estimators of $\xi$ as a function of $k$ that are consistent under the large semi-parametric model (1) are hard to interpret. In case of the Hill estimator some authors refer to Hill horror plots. While it has been frequently suggested to choose a ‘stable’ area (see for instance Drees et al. (2000) and De Sousa and Michailidis (2004)), such a stable part is often absent or hard to find. Sometimes more than one stable section is present, like in some insurance applications as we will discuss later.

The typical available guidelines for the choice of $k$ to be used in the implementation of the EVI estimators depend strongly on the properties of the tail itself, and $k$ needs to be estimated adaptively from the data. This problem can be compared with choosing a bandwidth parameter in density estimation. It is typically suggested that the optimal value of $k$ should be the one that minimizes the mean-squared error (MSE). However, this optimum depends on the sample size, the unknown value of $\xi$ as well as on the nature of $\ell$, as was first described in Hall et al. (1985). Bootstrap methods were proposed in Hall (1990), Draisma et al. (1999), Danielsson et al. (2001), and Gomes and Oliveira (2001). Beirlant et al. (1996, 2002) derived regression diagnostic methods on a Pareto quantile plot. Other selection procedures can be found in Drees and Kaufmann (1998) and Guillou and Hall (2001). Possible heuristic choices are provided in Gomes and Pestana (2007), Gomes et al. (2008) and Beirlant et al. (2011). Almost all authors consider the adaptive choice of $k$ for the Hill estimator, while methods can be adapted to other estimators as well.

In this paper we consider trimming of the Hill estimator, omitting some of the lower order statistics in $X_{n-k+1,n}, \ldots, X_{n,n}$, which leads to estimators of the type

$$T_{b,k} = \sum_{i=1}^{b} c_i(b,k) \log \frac{X_{n-i+1,n}}{X_{n-k,n}},$$

for some $1 \leq b \leq k$ and suitable constants $c_i(b,k)$. In Section 2 we derive the respective estimator of $\xi$ that is unbiased when $\ell$ is constant and when the constants $c_i(b,k) = c(b,k)$ do not depend on $i$. We present a novel lower-trimmed Hill plot which provides significant graphical support for the estimation problem of $\xi$, as we illustrate with both simulations and real world data. We also show that, as a function of $b$, the variance of this estimator is lower than the one of the Hill estimator. In Section 3, we examine the asymptotic characteristics of the estimators (4) under the general model (1). The asymptotic expected empirical variance of the trimmed Hill estimator (4) is shown to be less sensitive on the tail parameter $\xi$ than the asymptotic mean-squared error (AMSE) of the usual Hill estimator (3). We identify a link between the optimal $k$-choices of these two estimators which allows to bypass the specification of $\xi$ and other characteristics of the tail behavior for the identification of the optimal threshold in the classical Hill estimate, and the resulting procedure turns out to be simple to implement in practice. Subsequently, we study the estimator $T_k$ obtained by averaging the trimmed Hill estimators over $b = 1, \ldots, k$. This latter estimator naturally assigns more weight to the larger observations, the weights being only moderately changed when increasing $k$. Furthermore, the specification of these weights is independent of the distribution $F$. Note that, in contrast, earlier criteria for reweighting terms in the Hill estimator (such as e.g. Csörgő et al. (1985) in terms of kernel estimates, see also Beirlant et al., 2002 Sec.3) had to heavily rely on the tail parameter $\xi$. We
emphasize that nevertheless our goal here is not the one of optimal reweighting, our method rather suggests, as a by-product, a simple reweighting scheme with already favourable properties. In Section 4 we then present a simple ratio statistic as a tool to evaluate the goodness of selection of \( k \). Section 5 confirms the good performance of the proposed methods using simulations, where \( T_k \) turns out to outperform the classical Hill estimator in almost all cases. Note that our approach eventually suggests a fully automated procedure for the threshold selection, also in the absence of knowledge about, or assumptions on, the tail characteristics. Section 6 favorably illustrates this on a set of real-life motor third party liability insurance data. We would like to emphasize that the approach proposed in this paper suggests a general procedure that can in principle also be applied to other estimators in extreme value analysis.

2. A lower-trimmed Hill estimator

2.1. Derivation. Assume first, for simplicity, that we have independent and identically distributed (i.i.d.) exact Pareto random variables, \( X_1, X_2, \ldots, X_n \), with tail given by

\[
F(x) = (x/\sigma)^{-1/\xi}, \quad x \geq \sigma, \quad \xi, \sigma > 0,
\]

and we are interested in robust estimation of the tail index \( \xi \).

A main tool used throughout the paper is the well-known Rényi representation, which states (in the second distribution equality below), that for the order statistics of a random sample \( X_1, \ldots, X_n \) from the distribution (5), one has, for \( k \leq n \),

\[
\left( \log \left( \frac{X_{n,n}}{X_{n-k,n}} \right), \ldots, \log \left( \frac{X_{n-k+1,n}}{X_{n-k,n}} \right) \right) \overset{d}{=} \left( E_{k,k}, \ldots, E_{1,k} \right) \overset{d}{=} \left( \sum_{j=1}^{k} \frac{E_j^*}{k - j + 1}, \ldots, \frac{E_1^*}{k} \right).
\]

Here, \( E_{k,k} \geq \cdots \geq E_{1,k} \) are the order statistics of an independent i.i.d. exponential sample \( E_1, \ldots, E_k \) with mean \( \xi \), and \( E_1^*, \ldots, E_k^* \) is another independent i.i.d. exponential sample with mean \( \xi \).

Bhattacharya et al. (2017) recently proposed linear estimators of the form

\[
\hat{\xi}_{k_0,k} = \sum_{i=k_0+1}^{k} c_{k_0,k}(i) \log \left( \frac{X_{n-i+1,n}}{X_{n-k,n}} \right), \quad 0 \leq k_0 < k < n,
\]

in order to trim the upper order statistics in outlier-contaminated samples, where the constants \( c_{k_0,k}(i) \) are chosen in a way to ensure that the resulting estimator for \( \xi \) is unbiased. For fixed \( k_0, k \), the problem can then be recast into that of finding suitable weights \( \delta_i \) such that one can write

\[
\hat{\xi}_{k_0,k} = \sum_{i=k_0+1}^{k} c_{k_0,k}(i) E_{k-i+1,k} = \sum_{i=1}^{k-k_0} \delta_i E_{i,k}.
\]

Using the Rényi representation (6) and solving some elementary linear equations, they derived \( \delta_i = \frac{1}{r}, \quad i < r \), and \( \delta_r = (k - r + 1)/r \). This led them to the so-called trimmed Hill estimator

\[
\hat{\xi}_{k_0,k} = \frac{k_0 + 1}{k - k_0} \log \left( \frac{X_{n-k_0,n}}{X_{n-k,n}} \right) + \frac{1}{k - k_0} \sum_{i=k_0+2}^{k} \log \left( \frac{X_{n-i+1,k}}{X_{n-k,n}} \right),
\]
which is shown to be quite useful in outlier detection under (1).

In a similar way, but for a different purpose, in this paper we investigate trimming from the left. Concretely, we consider estimators of the form

$$T_{b,k} = \sum_{i=1}^{b} c_i(b,k) \log \left( \frac{X_{n-i+1,n}}{X_{n-k,n}} \right), \quad 0 < b \leq k,$$

where $c_i(b,k)$ are constants to be determined. As above, we would like to find suitable weights $\gamma_i$ such that

$$T_{b,k} = \sum_{i=1}^{b} c_i(b,k) E_{k-i+1,k} = \sum_{i=k-b+1}^{k} \gamma_i E_{i,k}$$

Setting $q = k - b + 1$, the Rényi representation (6) yields

$$T_{b,k} = \sum_{i=q}^{k} \gamma_i E_{i,k} = \sum_{i=q}^{k} \gamma_i \sum_{j=1}^{i} E^*_j \frac{k}{k-j+1} = \sum_{j=1}^{k} E^*_j \sum_{i=j\vee q}^{k} \frac{\gamma_i}{k-j+1} = \sum_{j=1}^{k} \tau_j E^*_j$$

with $\tau_j := \sum_{i=j\vee q}^{k} \frac{\gamma_i}{k-j+1}$. Here we use the notation $j \vee q = \max\{j,q\}$. Unfortunately, the set of equations

$$\tau_j = \frac{1}{k}, \quad j = 1, \ldots, k,$$

has no solution (for $j \leq q$ the left-hand-side cannot remain constant in $j$). Instead, we choose to set

$$\gamma_q = \gamma_{q+1} = \cdots = \gamma_k := \frac{1}{\omega(q,k)}$$

and

$$\mathbb{E}(T_{b,k}) = \xi$$

as the defining equations. The solution of (8) and (9) is given by

$$\omega(q,k) = \sum_{j=1}^{k} \frac{k-j \vee q + 1}{k-j + 1}.$$  

Plugging (10) into (7), we then arrive at the following definition of an unbiased lower-trimmed Hill estimator of $\xi$:

$$T_{b,k} = \sum_{i=q}^{r} \log \left( \frac{X_{n-k+i,n}}{X_{n-k,n}} \right) \omega(q,k) = \sum_{i=1}^{b} \log \left( \frac{X_{n-i+1,n}}{X_{n-k,n}} \right) \omega(k-b+1,k) = \sum_{i=1}^{b} \log \left( \frac{X_{n-i+1,n}}{X_{n-k,n}} \right) b(1 + \sum_{j=b+1}^{k} \gamma_j^{-1}), \quad b = 1, \ldots, k, \quad k < n,$$

where we use the convention $\sum_{j=k+1}^{k} j^{-1} := 0$.

$T_{b,k}$ represents a bivariate family of estimators, such that for each $k$, only $b$ of the $k$ upper order statistics are considered, while pinning down the reference observation to $X_{n-k,n}$. Intuitively, this should reduce the variance for a similar number of summands. Indeed, this can be shown more rigorously:
Proposition 2.1. As a function of the number $b$ of order statistics being used, in the exact Pareto case the estimator $T_{b,k}$ has lower variance than the classical Hill estimator $T_{b,b}$. More precisely,

$$\mathbb{V}(T_{b,b}) = \frac{\xi^2}{b} \quad \text{and} \quad \mathbb{V}(T_{b,k}) \leq \frac{\xi^2}{\sum_{j=1}^{k-b+1} \left( \frac{b}{k-j+1} \right)^2 + b}.$$ 

Proof. Set $q = k - b + 1$. By the Rényi representation,

$$\mathbb{V}(T_{b,k}) = \mathbb{V} \left( \sum_{j=1}^{k} E^* \sum_{i=j \vee q}^{k} \frac{\gamma_i}{k-j+1} \right) = \xi^2 \frac{\sum_{j=1}^{q} \left( \frac{k-q+1}{k-j+1} \right)^2 + k - q + 1}{\left( \sum_{j=1}^{q} \frac{k-q+1}{k-j+1} + k - q + 1 \right)^2}.$$ 

Plugging in $q = 1$ ($b = k$) gives

$$\mathbb{V}(T_{k,k}) = \frac{\xi^2}{k},$$

which corresponds to the usual Hill estimator $T_{k,k}$ and gives the first identity. In the general case,

$$\mathbb{V}(T_{b,k}) = \xi^2 \frac{\sum_{j=1}^{q} \left( \frac{k-q+1}{k-j+1} \right)^2 + k - q + 1}{\left( \sum_{j=1}^{q} \frac{k-q+1}{k-j+1} + k - q + 1 \right)^2}.$$ 

But $j \leq q$ implies $\frac{k-q+1}{k-j+1} \leq 1$, such that

$$\sum_{i=1}^{q} \frac{1}{k-j+1} \geq \sum_{j=1}^{q} \frac{k-q+1}{k-j+1},$$

so

$$\sum_{j=1}^{q} \frac{k-q+1}{k-j+1} + k - q + 1 \geq \sum_{j=1}^{q} \left( \frac{k-q+1}{k-j+1} \right)^2 + k - q + 1.$$ 

Thus

$$\mathbb{V}(T_{b,k}) \leq \xi^2 \frac{\sum_{j=1}^{q} \left( \frac{k-q+1}{k-j+1} \right)^2 + k - q + 1}{\left( \sum_{j=1}^{q} \frac{k-q+1}{k-j+1} + k - q + 1 \right)^2} = \frac{\xi^2}{\sum_{j=1}^{q} \left( \frac{k-q+1}{k-j+1} \right)^2 + k - q + 1},$$

which gives the second identity. \hfill \Box

2.2. A lower-trimmed Hill plot. The family of estimators $T_{b,k}$ defined above is unbiased for any $b, k$, $b \leq k$, by construction. Analogous to the Hill plot, in which $T_{k,k}$ is plotted as a function of $k$, we now exploit the second degree of freedom and plot, for selected values of $k$, $T_{b,k}$ as a function of $b$. That is, the plot is constructed by overlaying the trajectories $(b, T_{b,k})$, $b = 1, \ldots, k$, for a selection of $k$ values. The lower variance of these trajectories comes from the fact that the normalizing order statistic is fixed, and hence a non-constant behaviour is
easier to identify visually than in the classical Hill plot. As a particular consequence, the selection of $k$ that makes the tail resemble a pure Pareto tail is easier to determine, by examining when the trajectories start to be constant.

As an illustration, we now compare the performance of these lower-trimmed Hill (LTH) plots for Pareto, near-Pareto and spliced Pareto distributions. The latter is defined through its cumulative distribution function (c.d.f.)

$$F(x; \xi, r, c) = \frac{(1 - x^{-1/\xi_0 - r}1(x \geq c)) - 1\{x \geq c\}(c^{-1/\xi_0} - c^{-1/\xi_0 - r})}{1 - c^{-1/\xi_0} + c^{-1/\xi_0 - r}}, \ x \geq 1$$ (12)

for $c \geq 1$ and $r > -1/\xi$, which is the c.d.f. of a Pareto random variable with tail index $\xi_0$ up to some splicing point $c$, continuously pasted with the c.d.f. of a Pareto random variable with another tail index $\xi = (1/\xi_0 + r)^{-1}$ thereafter. Splicing models (also sometimes referred to as composite models) are for instance popular in reinsurance modelling, cf. (Albrecher et al., 2017, Ch.4).

Concretely we simulated a sample of size $n = 1000$ from a:

- pure Pareto $\xi = \sigma = 1$ sample, defined in (5).
- spliced Pareto sample, defined in (12), with parameters $\xi = 1$, $\xi_0 = 4$ and splicing point $c = 1.3$.
- spliced Pareto sample, defined in (12), with parameters $\xi = 1$, $\xi_0 = 1/4$ and splicing point $c = 1.3$.
- Burr sample with tail $F(x) = \frac{1}{1+x}, \ x > 0$ (which amounts to a shifted Pareto).
- Loggamma with logshape parameter $3/2$ and lograte parameter 1.

The LTH plots together with usual Hill plots are shown in the top panels of Figures 1–5. The LTH plots are made for a selection of $k$, from 1 to 1000 by spacings of 50 (1,51,101,...), as a function of the lower trimming $b$. Recall that $b \leq k$, so the lines have different domains on the x-axis. Observe that the right end-point of each of the overlaid lines corresponds to the respective point in the Hill plot.

For the spliced distributions in Figures 2 and 3 observe how the LTH estimator becomes horizontal as a function of $b$ when $k$ is close to the (rank of the) splicing point. For smaller $k$, the plot then looks similar to the exact Pareto case. Loosely speaking, the slope of the lines are a very useful visual tool for detecting the number of upper order statistics $k$ after which a Pareto tail is feasible. This can also be seen in the Burr (Fig.4) and loggamma case (Fig.5), where the regime of a Pareto tail is only reached for high quantiles.

The bottom panels of Figures 1–5 suggest two ways of measuring the aforementioned flatness of the LTH estimator as a function of $b$. The first one computes the empirical variance of $T_{b,k}$, $b = 1, \ldots, k$, while the second one fits a linear model with independent variable $b = 1, \ldots, k$ and response variable $T_{b,k}$, and then plots the magnitude of the resulting slope coefficient.

3. Regularly varying tails

We now move from the simple Pareto sample to a general Fréchet domain of attraction, with tails of the form (1). Denote by $Q$ the quantile function associated to $F$, and
define

\[ U(x) = Q(1 - 1/x), \quad x > 1, \]

such that the condition (1) is equivalent to

\[ \lim_{A \to \infty} \frac{U(Ax)}{U(A)} = x^{-\xi}. \]

Assumptions on the rate of convergence of the above limit make it possible to obtain explicit results concerning asymptotic properties of the lower-trimmed Hill estimator. Hence, we impose the second order condition

\[ \lim_{A \to \infty} \frac{\log U(Ax) - \log U(A) - \xi \log(x)}{Q_0(A)} = \frac{x^p - 1}{p}, \]

for some regularly varying function \( Q_0 \) with index \( p < 0 \).
Theorem 3.1. Under the model (1) and second order condition (13), the lower-
trimmed Hill estimator $T_{b,k}$ as defined in (11) satisfies the following asymptotic distri-
butional identity, for $n, k, n/k \to \infty$,

$$T_{b,k} \overset{d}{=} \xi \frac{\bar{E}_b + \sum_{j=b+1}^{k} E_j/j}{1 + \sum_{j=b+1}^{k} j^{-1}} + \frac{Q_0(n/k)}{p} \frac{(k+1)^p}{1 - p} - \frac{1}{1 + \sum_{j=b+1}^{k} j^{-1}}(1 + o_p(1)),$$

where $E_1, \ldots, E_k$ are i.i.d. standard exponential random variables, and where we use
the notation $\bar{E}_b = b^{-1} \sum_{i=1}^{b} E_i$.

Proof. We first note that

$$T_{b,k} \overset{d}{=} \frac{\sum_{i=1}^{b} \log(U(Y_{n-i+1,n})/U(Y_{n-k,n}))}{b(1 + \sum_{j=b+1}^{k} j^{-1})},$$
where $Y_{1,n} < \cdots < Y_{n,n}$ are the order statistics of a standard Pareto sample (the $\xi = 1$ case). Then, from the second order condition (13) we obtain that for $A = Y_{n-k,n}$ and $x = Y_{n-i+1,n}/Y_{n-k,n}$, as $k,n,n/k \to \infty$,

$$T_{b,k} = \frac{\xi \sum_{i=1}^{b} \log(Y_{n-i+1,n}/Y_{n-k,n}) + \frac{Q_0(Y_{n-k,n})}{p} \sum_{i=1}^{b} ((Y_{n-i+1,n}/Y_{n-k,n})^p - 1)(1 + o_p(1))}{b(1 + \sum_{j=b+1}^k j^{-1})}. $$

But by the Rényi representation of exponential order statistics, the first term is distributed as

$$\sum_{i=1}^{b} \log(Y_{n-i+1,n}/Y_{n-k,n}) \overset{d}{=} \sum_{j=1}^{b} E_j + b \sum_{j=b+1}^{k} E_j/j,$$
where $E_1, E_2, \ldots, E_k$ are i.i.d. standard exponential random variables. For the second term, by convergence to uniform random variables and a Riemann integral approximation, we get

$$\frac{1}{b} \sum_{i=1}^{b} ((Y_{n-i+1,n}/Y_{n-k,n})^p - 1) \approx \frac{1}{b} \sum_{i=1}^{b} (((k + 1)/i)^p - 1) \approx \frac{k+1}{b} \int_{0}^{b/(k+1)} (u^{-p} - 1) \, du = \frac{(k+1/b)^p}{1-p} - 1,$$

and since $(1 - 1/Y_{n-k,n})$ is a uniform order statistic, we further get that

$$\frac{Q_0(Y_{n-k,n})}{Q_0(n/k)} \overset{P}{\to} 1.$$

Putting the three pieces together then establishes (14). \qed
3.1. Distribution of the average. Define the average of the $T_{b,k}$ across $b$ as

$$\bar{T}_k := \frac{1}{k} \sum_{b=1}^{k} T_{b,k},$$

which by the same approximation as above, satisfies the identity

$$\bar{T}_k \overset{d}{=} \xi \sum_{b=1}^{k} \frac{E_b + \sum_{j=b+1}^{k} E_j/j}{1 + \sum_{j=b+1}^{k} j^{-1}} + Q_0(n/k) \sum_{b=1}^{k} \frac{(k+1)/b^p - 1}{1 + \sum_{j=b+1}^{k} j^{-1}} (1 + o_p(1)).$$
We can immediately see that

\[
\mathbb{E}(T_{b,k}) = \xi + \frac{Q_0(n/k)}{p} \left( \frac{(k+1/b)^p - 1}{1 - p} \sum_{j=b+1}^{k} \frac{1}{j} \right) (1 + o_p(1)),
\]

\[
\mathbb{E}(T_k) = \xi + \frac{Q_0(n/k)}{pk} \sum_{b=1}^{k} \frac{(k+1/b)^p - 1}{1 - p} \sum_{j=b+1}^{k} \frac{1}{j} (1 + o_p(1)),
\]

so that the asymptotic bias terms can be recognized directly. To ease notation, let us introduce the constants

\[
\begin{align*}
    c_{b,k,p} := & \frac{1}{p} \cdot \frac{(k+1/b)^p - 1}{1 - p} \approx \frac{1}{p} \cdot \frac{(k+1/b)^p - 1}{1 - p} \\
    \overline{c}_{k,p} := & \frac{1}{pk} \sum_{b=1}^{k} \frac{(k+1/b)^p - 1}{1 - p} \sum_{j=b+1}^{k} \frac{1}{j} \approx \frac{1}{pk} \sum_{b=1}^{k} \frac{(k+1/b)^p - 1}{1 - p} \sum_{j=b+1}^{k} \frac{1}{j}.
\end{align*}
\]

**Theorem 3.2.** The average \( \overline{T}_k \) of the lower-trimmed Hill estimators, (15), under model [1] and second order condition [13] satisfies the following asymptotic distributional identity, for \( n, k, n/k \to \infty \),

\[
\begin{align*}
    \overline{T}_k = & \frac{d}{k-1} \sum_{j=1}^{k} \log(1 + \log(k/j)) + \frac{ek}{j} \mathbb{E}(1 + \log(k/j)) \left(1 + o(1)\right) \\
    & + Q_0(n/k) \left[ \frac{e^{1-p}}{p(1-p)} \mathbb{E}(1-p) - \frac{e}{p} \mathbb{E}(1) \right] \left(1 + o_p(1)\right),
\end{align*}
\]

where

\[
\mathbb{E}(x) := \int_{x}^{\infty} e^{-v/v} \, dv
\]

is the exponential integral.

**Proof.** With the shortened notation, we write

\[
T_{b,k} := \frac{\xi E_b + \sum_{j=b+1}^{k} E_j / j}{1 + \sum_{j=b+1}^{k} j^{-1}} + Q_0(n/k)c_{b,k,p}(1 + o_p(1)),
\]

and by exchange of the order of summation, we can write

\[
T_k = \frac{\xi E_b + \sum_{j=b+1}^{k} E_j / j}{1 + \sum_{j=b+1}^{k} j^{-1}} + Q_0(n/k)\overline{c}_{k,p}(1 + o_p(1))
\]

\[
= \xi \sum_{j=1}^{k} E_j \sum_{b=j}^{k} \frac{1}{b(1 + j^{-1})} + \sum_{j=2}^{k} E_j \sum_{b=1}^{j-1} \frac{1}{b(1 + j^{-1})} + Q_0(n/k)\overline{c}_{k,p}(1 + o_p(1))
\]

\[
= \xi \sum_{j=1}^{k} E_j \sum_{b=j}^{k} \frac{1}{b(1 + \log(k/b))} + \sum_{b=1}^{j-1} \frac{1}{j(1 + \log(k/b))} \left(1 + o(1)\right)
\]

\[
+ Q_0(n/k)\overline{c}_{k,p}(1 + o_p(1)).
\]
Again, by Riemann integration we have that
\[
\frac{1}{k} \sum_{b=1}^{k} \frac{1}{(b/k)(1 + \log(k/b))} \approx \int_{j/k}^{1} \frac{du}{u(1 - \log(u))} = \log(1 + \log(k/j)),
\]
and
\[
\sum_{b=1}^{j-1} \frac{1}{j(1 + \log(k/b))} \approx \frac{k}{j} \int_{0}^{j/k} \frac{du}{1 - \log(u)} = \frac{ek}{j} \mathbb{E}(1 + \log(k/j)).
\]
Similarly,
\[
\bar{c}_{k,p} \approx \frac{1}{p} \int_{0}^{1} \frac{(1 - p)u^{-p}}{1 - \log(u)} \, du - \frac{1}{p} \int_{0}^{1} \frac{du}{1 - \log(u)}
\]
(19)
\[
= \frac{e^{1-p}}{p(1-p)} \mathbb{E}(1-p) - \frac{e}{p} \mathbb{E}(1).
\]
Putting the pieces together then indeed yields (17). □

3.2. Mean of the empirical variance. Equipped with the representations in terms of exponential variables that we obtained in Theorems 3.1 and 3.2 we set on to analyze the mean of the empirical variance of \(T_{b,k}\) as a function of \(b\).

**Theorem 3.3.** The mean of the empirical variance of the lower-trimmed Hill estimator \(T_{b,k}\), under model (1) and second order condition (13), satisfies the following asymptotic identity, for \(n, k, n/k \to \infty\),
\[
\mathbb{E} \left[ \frac{1}{k} \sum_{b=1}^{k} (T_{b,k} - T_{k})^2 \right] = \frac{C}{k} \xi^2 (1 + o(1)) + Q_0^2(n/k)f(p)(1 + o_p(1))
\]
where \(C = 0.502727\) and
\[
f(p) := \frac{1 - e^{-2p}(1 - 2p)}{p^2(1-p)^2} \mathbb{E}(1-2p) - e^{-2p} \mathbb{E}^2(1-p) \\
+ 2 \frac{e^{-p} \mathbb{E}(1-p)}{p^2(1-p)} \mathbb{E}(1-1) - 1 + e^{-p}(1-p) \mathbb{E}(1-p) \\
+ \frac{1 - e \mathbb{E}(1) - e^2 \mathbb{E}^2(1)}{p^2} > 0.
\]

**Proof.** Let us first decompose each summand by writing
\[
\mathbb{E}((T_{b,k} - T_{k})^2) = \mathbb{E}((T_{b,k} - \xi)^2) + \mathbb{E}((T_{k} - \xi)^2) - 2 \mathbb{E}((T_{b,k} - \xi)(T_{k} - \xi)),
\]
and subsequently consider each term separately. From (18) we have that
\[
\mathbb{E}((T_{b,k} - \xi)^2) = \mathbb{V}(T_{b,k}) + \text{Bias}^2(T_{b,k}) = \xi^2 \frac{1}{k} + \frac{\sum_{j=b+1}^{k} 1/j^2}{(1 + \sum_{j=b+1}^{k} j^{-1})^2} + Q_0^2(n/k)c_{b,k,p}^2(1 + o_p(1)).
\]
On the other hand, \([17]\) gives
\[
\mathbb{E}((T_k - \xi)^2) = \mathbb{V}(T_k) + \text{Bias}^2(T_k)
\]
\[
= \frac{\xi^2}{k^2} \sum_{j=1}^{k} \left[ \log(1 + \log(k/j)) + \frac{e^{1-p}}{j} \mathbb{E}(1 + \log(k/j)) \right]^2 (1 + o(1))
\]
\[+ \mathcal{Q}_0(n/k) \left[ \frac{e^{1-p}}{p(1-p)} \mathbb{E}(1-p) - \frac{e}{p} \mathbb{E}(1) \right] (1 + o_p(1)).
\]
The third term can be analyzed using both \([17]\) and \([18]\) as follows
\[
\mathbb{E}((T_{b,k} - \xi)(T_k - \xi)) = \mathbb{E}((T_{b,k} - \mathbb{E}(T_{b,k}))(T_k - \mathbb{E}(T_k))) + \mathcal{Q}_0(n/k)c_{b,k,p}c_{k,p}
\]
\[
= \xi^2 \mathbb{E} \left[ \left( \frac{1}{b} \sum_{j=1}^{b} (E_j - 1) + \frac{k}{1 + \sum_{j=b+1}^{k-1} j^{-1}} \right) \left( \sum_{i=1}^{k} E_i - \frac{1}{k} S(i,k)(1 + o(1)) \right) \right]
\]
\[+ \mathcal{Q}_0(n/k)c_{b,k,p}c_{k,p}
\]
where \(S(j,k) := \log(1 + \log(k/j)) + \frac{e^{1-p}}{j} \mathbb{E}(1 + \log(k/j))\).

We now proceed to add the \(k\) summands of the expected variance. To this end, some preparatory calculations will be helpful. By \([16]\) and Riemann approximation we have
\[
\frac{1}{k} \sum_{b=1}^{k} c_{b,k,p}^2 \approx \frac{1}{k} \sum_{b=1}^{k} \frac{1}{p^2} \frac{(k+1/b)^{2p} - 2(k+1/b)^{p}}{(1 + \log((k+1)/b))^2} + 1
\]
\[
\approx \frac{1}{p^2(1-p)^2} \left[ 1 - e^{1-2p}(1-2p) \mathbb{E}(1-2p) \right]
\]
\[- \frac{2}{(1-p)^2} \left[ 1 - e^{-p}(1-p) \mathbb{E}(1-p) \right]
\]
\[+ \frac{1}{p^2} \left[ 1 - e \mathbb{E}(1) \right].
\]
By virtue of \([19]\),
\[
\mathcal{C}_{k,p}^2 \approx \frac{e^{2(1-p)}}{p^2(1-p)^2} \mathbb{E}^2(1-p) - \frac{e^{2-p}}{p^2(1-p)} \mathbb{E}(1-p) \mathbb{E}(1) + \frac{e^2}{p^2} \mathbb{E}^2(1),
\]
from which we deduce that as \(k \to \infty\),
\[
\frac{1}{k} \sum_{b=1}^{k} c_{b,k,p}^2 - \mathcal{C}_{k,p}^2 \to f(p),
\]
where \(f(p)\) is given by \([20]\).

Observe that
\[
\frac{1}{k} \sum_{b=1}^{k} \frac{1}{b} + \sum_{j=b+1}^{k} 1/j^2 \approx \frac{2}{k} \int_{0}^{1} \frac{du}{u(1-\log(u))^2} - \frac{1}{k} \int_{0}^{1} \frac{du}{(1-\log(u))^2}
\]
\[= \frac{1 + e \mathbb{E}(1)}{k}.
\]
Next,

\[
\frac{1}{ek} \sum_{b=1}^{k} \sum_{j=1}^{b} \frac{b^{-1} \log(1 + \log(k/j)) + \frac{ek}{j} E_1(1 + \log(k/j))}{1 + \log(k/b)} \\
\approx \int_{0}^{1} \frac{1}{z \log(e/z)} \left( \int_{0}^{\infty} \frac{1}{u} \left( \int_{\log(e/u)}^{\infty} \log(v) e^{-v} \, dv \right) \, du \right) \, dz \\
\approx 0.266 =: I_1
\]

and

\[
\frac{1}{ek} \sum_{j=b+1}^{k} j^{-1} \log(1 + \log(k/j)) + \frac{ek}{j} E_1(1 + \log(k/j)) \\
\approx \int_{0}^{1} \frac{1}{\log(e/z)} \left( \int_{z}^{1} \frac{1}{u^2} \left( \int_{\log(e/u)}^{\infty} \log(v) e^{-v} \, dv \right) \, du \right) \, dz \\
\approx 0.135746 =: I_2
\]

Finally,

\[
\frac{1}{k} \sum_{j=1}^{k} (k/j)^2 \left( \int_{\log(e/u)}^{\infty} \log(v) e^{-v} \, dv \right)^2 \\
\approx \int_{0}^{1} u^{-2} \left( \int_{\log(e/u)}^{\infty} \log(v) e^{-v} \, dv \right)^2 \, du \\
\approx 0.148005 =: I_3
\]

Altogether we hence obtain

\[
\mathbb{E} \left[ \frac{1}{k} \sum_{b=1}^{k} (T_{b,k} - T_k)^2 \right] \\
= \frac{1}{k} \sum_{b=1}^{k} \left( \mathbb{E}((T_{b,k} - \xi)^2) + \mathbb{E}((T_k - \xi)^2) - 2 \mathbb{E}((T_{b,k} - \xi)(T_k - \xi)) \right) \\
= \xi^2 \left( 1 + \frac{eE_1(1)}{k} \right) (1 + o(1)) + \xi^2 \frac{e^2}{k} I_3(1 + o(1)) \\
- 2\xi^2 \frac{e}{k} (I_1 + I_2)(1 + o(1)) + Q_0^2(n/k) \left[ \frac{1}{k} \sum_{b=1}^{k} c_{b,k,p} - c_{k,p}^2 \right] (1 + o_p(1)) \\
= \xi^2 \left( 1 + \frac{eE_1(1)}{k} \right) (1 + o(1)) + \xi^2 \frac{e^2}{k} I_3(1 + o(1)) \\
- 2\xi^2 \frac{e}{k} (I_1 + I_2)(1 + o(1)) + Q_0^2(n/k) f(p)(1 + o_p(1)) \\
= \frac{C}{k} \xi^2(1 + o(1)) + Q_0^2(n/k) f(p)(1 + o_p(1))
\]

with

\[
C = 1 + eE_1(1) + e^2 I_3 - 2e(I_1 + I_2) \approx 0.502727.
\]

\[\square\]
3.3. **Optimal \( k \) in the Hall class.** We now make a further assumption on the regularly varying class, in order to get an explicit form of \( Q_0 \). Concretely, we assume the Hall class \cite{Hall1982}, which satisfies the property
\[
U(x) = A x^\xi (1 + D x^p (1 + o(1))), \quad x \to \infty.
\]
An immediate consequence then is the explicit expression
\[
Q_0(x) = -p D x^p (1 + o(1)).
\]
Hence,
\[
E \left[ \frac{1}{k} \sum_{b=1}^{k} (T_{b,k} - \bar{T}_k)^2 \right] = \frac{C}{k} \xi^2 (1 + o(1)) + p^2 D^2 (n/k)^2 f(p) (1 + o_p(1)).
\]
Recall that the classical Hill estimator for this class has AMSE given by
\[
\frac{\xi^2}{k} + \left( \frac{Q_0(n/k)}{1 - p} \right)^2
\]
which is minimized for
\[
k_0^* \sim (Q^2_0(n))^{-1/(1-2p)} \left( \frac{\xi^2 (1 - p)^2}{-2p} \right)^{1/(1-2p)}
\]
\[
= \left( \frac{n^{-2p} \xi^2 (1 - p)^2}{-2p^3 D^2} \right)^{1/(1-2p)}.
\]
see e.g. \cite[p.125]{Beirlant2004}). In a similar way, the minimizer of (22) is simply
\[
k^* \sim (Q^2_0(n))^{-1/(1-2p)} \left( \frac{C \xi^2}{-2p^3 D^2 f(p)} \right)^{1/(1-2p)}
\]
\[
= k_0^* \left( \frac{C}{(1 - p)^2 f(p)} \right)^{1/(1-2p)}.
\]

3.4. **Interpretation of \( \bar{T}_k \) as a weighted Hill estimator.** Observe that, for fixed \( k \),
\[
\bar{T}_k = \frac{1}{k} \sum_{b=1}^{k} T_{b,k} = \frac{1}{k} \sum_{i=1}^{k} \theta_i \log(X_{n-i+1,n}/X_{n-k,n})
\]
with
\[
\theta_i := \frac{1}{b(1 + \sum_{j=b+1}^{k} j^{-1})},
\]
so that one can interpret the estimator \( \bar{T}_k \) as a modification of the classical Hill estimator that uses different weights for different order statistics. It is not hard to see that asymptotically the correction factors behave like
\[
\theta_i \sim \log \left( \frac{\log(i/k) - 1}{\log(1 - 1/k) - 1} \right), \quad k \to \infty.
\]
Figure 6(left) highlights the accuracy of this approximation for \( k = 100 \) across different values of \( i \), and also illustrates the fact that the largest data point receives a weight
of almost 2 in this case, whereas on from the 20th-largest observation the weight is lower than for the classical Hill estimator, and the weight diminishes for smaller data points. Note that, as \( k \) increases, the weight of the largest observation grows above any bound, but extremely slowly, namely
\[
\theta_1 = \log(\log(k) + 1) - 1/k + O(1/k^3).
\]

Figure 6 (right) illustrates that even for a value as large as \( k = 10000(!) \), \( \theta_1 \) is still below 2.4.

4. A RATIO STATISTIC

Once a \( k^* \) has been selected, it is important to be able to statistically assess whether the remaining upper tail differs significantly from the one of a pure Pareto. In order to recognize whether a Pareto tail has been achieved or not, we have seen that flatness of the lower-trimmed Hill estimator is desirable. Inspired by the T-statistic introduced in [Bhattacharya et al. (2017)], we introduce the ratio statistics

\[
R_{b,k} = \frac{T_{b+1,k}}{T_{b,k}}, \quad b = 1, \ldots, k-1,
\]

quantities which we expect to be close to one. Although these statistics do not have the property of being i.i.d. and hence test sizes have to be calibrated using Monte Carlo simulation, an advantage which carries over to the present setting is that they do not depend on \( \xi \). Indeed,

\[
R_{b,k} = \frac{d}{\omega(b, k)} \left( 1 + \frac{\log(\Gamma_{b+1}/\Gamma_{k+1})}{\sum_{i=1}^{b} \log(\Gamma_i/\Gamma_{k+1})} \right),
\]

by the order statistics property of the Poisson process, where \( \Gamma_m = \sum_{i=1}^{m} E_i \), and \( E_i, \ i = 1, 2, \ldots, \) is an i.i.d. sequence of independent unit-rate exponential random variables. This invariance with respect to the \( \xi \) parameter permits to assess the goodness of selection of a threshold \( k^* \) as follows:

1. Simulate the \( R_{b,k^*} \) statistics \( N_{MC} \) times, and call them

\[
R_{b,k^*}^{(m)}, \quad m = 1, \ldots, N_{MC}, \ b = 2, \ldots, k^* - 1.
\]
(2) For fixed $\alpha \in (0,1)$, find the empirical $\alpha/2$ and $1-\alpha/2$ quantiles corresponding to each of the $b = 2, \ldots, k^* - 1$ samples,

$$R_{b,k^*}^m, \quad m = 1, \ldots, N_{MC},$$

and call them $(q_1, q_2)_2, \ldots, (q_1, q_2)_{k^* - 1}$.

(3) Count the proportion of the trajectories $R_{b,k^*}^m, b = 2, \ldots, k^* - 1$, which fall outside of their confidence interval $(q_1, q_2)_b$ for some $2 \leq b \leq k^* - 1$. Call this proportion $\alpha_r$.

(4) If $\alpha_r$ is, up to some tolerance level, too large (too small), go to step (2) and decrease (increase) $\alpha$ to a value within its two last values.

(5) Plot the $R_{b,k^*}, b = 2, \ldots, k^* - 1$, from the data, together with the last set of quantiles $(q_1, q_2)_1, \ldots, (q_1, q_2)_{k^*}$. It is also a good idea, for visualization, to plot the standardized version

$$R_{b,k^*}^m - q_{1,b} \over q_{2,b} - q_{1,b}, \quad b = 2, \ldots, k^* - 1,$$

which for a pure Pareto tail is expected by construction to lie between 0 and 1 in $100(1-\alpha)$% of the cases.

**Example 4.1.** For the Burr sample of Figure 4, we compare taking $k^* = 326$ and $k^* = 600$ in the plots of Figure 7. The first number, $k^* = 326$ is precisely the one that minimizes the expected empirical variance, according to the parameters of the Burr sample and to formula (24), with $p$ chosen to be $-1$. The number of Monte Carlo simulations was in each case $N_{MC} = 10000$, and the significance level is $\alpha = 0.05$. Observe how the fit is good for $k = 326$, but is outside the bands for $k = 600$.

**Remark 4.1.** This approach can only be considered as a selection procedure itself if the corresponding sequential testing is adjusted to have the correct size. In other words, if the above algorithm is used multiple times to choose $k$, the rejection probability will exceed the desired $\alpha$ level. An alternative is to take sequential values of $k$ into the algorithm, which makes the routine highly computationally intensive. Hence, we presently recommend it solely as a goodness of selection evaluation.

### 5. Simulations

We perform a simulation study based on three different and common distributions which belong to the Hall class (21). We consider simulating $N_{sim} = 1000$ times from the following three distributions, with four sub-cases for each distribution, for varying sample size and parameters:

- The Burr distribution, with tail given by

$$F(x) = \left( \frac{\eta}{\eta + x^\tau} \right)^\lambda, \quad x > 0, \quad \eta, \tau, \lambda > 0,$$

which implies by Taylor expansion that

$$\xi = \frac{1}{\lambda \tau}, \quad A = \eta^{1/\tau}, \quad D = -\frac{1}{\tau}, \quad p = -\frac{1}{\lambda}.$$

We consider for $n = 100, 500$ the two sets of parameters $\eta = 1, \lambda = 2, \tau = 1/2$; and $\eta = 3/2, \lambda = 1/2, \tau = 2$.  

Figure 7. Standardized R-statistic for the Burr sample of Figure 4 (all parameters set to 1), for two choices of threshold: $k = 326, 600$, respectively. $N_{MC} = 10000$ and $\alpha = 0.05$.

- The Fréchet distribution with tail
  $\overline{F}(x) = 1 - \exp(-x^{-\alpha}), \quad \alpha > 0,$
  which implies
  $\xi = \frac{1}{\alpha}, \quad A = 1, \quad D = -\frac{1}{2\alpha}, \quad p = -1.$
  We consider for $n = 100, 500$ the two parameters $\alpha = 1, 1/2$.

- The Generalized Pareto Distribution (GPD) distribution, with tail given by
  $\overline{F}(x) = \left(1 + \frac{\gamma x}{\sigma}\right)^{-1/\gamma}, \quad \gamma, \sigma > 0,$
  which implies
  $\xi = \gamma, \quad A = \frac{\sigma}{\gamma}, \quad D = -1, \quad p = -\gamma.$
  We consider for $n = 100, 500$ the two sets of parameters $\gamma = 1/2, \sigma = 2$; and $\gamma = 5/2, \sigma = 1$.

For each sample we evaluate the Hill estimator
$$H_k = T_{k,k}$$
and the average LTH estimator
$$T_k = \frac{1}{k} \sum_{b=1}^{k} T_{b,k}$$
at three particular choices of $k$:

(i) The AMSE minimizer $[23]$ for the Hill estimator, i.e.
$$k_{opt} \sim \left(\frac{n^{-2p}\xi^2(1-p)^2}{-2p^3D^2}\right)^{1/(1-2p)}.$$
Observe that this optimal threshold depends not only on the second order index \( p \) but also on \( D \) and \( \xi \) itself. For benchmark purposes, we will assume a perfect estimation of \( \xi, D, p \) by imputing the true values of these quantities. Note that in Beirlant et al. (2002) this estimator was identified to almost always outperform other alternatives like Danielsson et al. (2001), Drees and Kaufmann (1998) and Guillou and Hall (2001) when imputing estimates for the characteristics of the tail behavior.

(ii) An estimator of \( k_0^* \) from (23) obtained as follows. Motivated by (22), we compute \( k^* \) as the minimizer of the empirical variance (the search beginning at 1/5 of the sample size, to avoid degeneracies) of the trimmed Hill estimator, as a function of \( b \), and using (24) to set

\[
    k_0 := k^* \left( \frac{(1 - p)^2 f(p)}{C} \right)^{-1/(1-2p)}.
\]

Observe that while we still have to input \( p \), here prior knowledge of \( \xi, D \) is no longer needed. We choose \( p = -1 \) as the canonical choice.

(iii) As in (ii), but using the true parameter of \( p \), in order to quantify how the removal of a potential misspecification of \( p \) by the canonical choice \( p = -1 \) affects the estimators (this complements Beirlant et al. (2002), where it was concluded from simulation studies for various estimators that this potential misspecification does not seem to be of major importance).

We then plot the bias, variance and MSE of each resulting estimator as a function of \( k \).

The results are given in Figures 8, 9 for the Burr case; Figures 10, 11 for the Fréchet case; and Figures 12, 13 for the GPD. We observe that the behaviour is very similar for the three families (which is not uncommon in this context, cf. (Beirlant et al., 2002, p.178)). For the Hill estimator, the threshold \( k_0 \) selected via the method of minimizing the empirical variance of the trimmed Hill follows very closely the theoretical optimal AMSE minimizer \( k_{opt} \) (although the true values of \( D \) and \( \xi \) are then not used!), and in the case where the true \( p \) is used, the finite-sample MSE of the former can in some cases equal or even improve on the latter. The same behaviour is observed within the three \( \overline{T} \)-estimators. When comparing Hill against \( \overline{T} \)-estimators, the latter improve the bias and MSE for nearly all \( k \), and in most cases also the variance (except for very heavy tails (\( \xi \geq 1 \) and small values of \( k \)). Remarkably, the estimator \( \overline{T}_{k_0, p=-1} \), where the canonical \( p = -1 \) is used, is highly competitive against the theoretically optimal \( H_{k_{opt}} \), especially so for \( \xi \leq 1 \), where an improvement can be observed. This is not a contradiction, since the optimality of \( H_{k_{opt}} \) refers to choices for \( k \) within the class of \( H_k \), whereas the \( \overline{T}_k \) estimators span a different class (visible in the weighting interpretation of Section 3.4), and when \( k \) is optimized w.r.t. AMSE in that class, even better performance can be feasible, which, however, is not the subject of the present paper.

6. INSURANCE DATA

Let us now consider a real-life insurance data set consisting of 837 motor third party liability (MTPL) insurance claims from the period 1995-2010 that was studied
intensively in Albrecher et al. (2017) (where it is referred to as "Company A"). These data are right-censored, and were also analyzed recently combining survival analysis techniques and expert information in Bladt et al. (2019). Here, we focus only on the ultimates, see Figure 14, which are the actual final claim sizes for the settled claims and an expert prediction of the total payment until closure for all claims that are still open.

Figure 8. Burr distribution, parameters $\eta = 1$, $\lambda = 2$, $\tau = 1/2$. Top: Violin plots for $n = 100, 500$ of the estimators $H_{k_0}$, $H_{\tilde{k}_0}$, $p=-1$, $H_{\tilde{k}_0}^*, p=-1/\lambda$, $T_{\tilde{k}_0}$, $T_{\tilde{k}_0}^*, p=-1$, $T_{\tilde{k}_0}^*, p=-1/\lambda$. Bottom: diagnostics of $T_k$ (blue) and $H_k$ (red) as a function of $k$. 

Figure 9. Burr distribution, parameters $\eta = 3/2, \lambda = 1/2, \tau = 2$.
Top: Violin plots for $n = 100, 500$ of the estimators $H_{k_{0}^{*}}, H_{k_{0}^{*}, p = 1}, H_{k_{0}^{*}, p = -1}$.
Bottom: diagnostics of $\overline{T}_{k}$ (blue) and $H_{k}$ (red) as a function of $k$. 
Figure 10. Fréchet distribution, parameter $\alpha = 1$. Top: Violin plots for $n = 100, 500$ of the estimators $H_k, H_{k_{\hat{p}}}^{*}, p = -1; \bar{T}_k, \bar{T}_{k_{\hat{p}}}, p = -1$. Bottom: diagnostics of $\bar{T}_k$ (blue) and $H_k$ (red) as a function of $k$. 
Figure 11. Fréchet distribution, parameter $\alpha = 1/2$. Top: Violin plots for $n = 100, 500$ of the estimators $H_{k0}$, $H_{\hat{k}0}$, $T_{k0}$, $T_{\hat{k}0}$, $p = -1$. Bottom: diagnostics of $T_k$ (blue) and $H_k$ (red) as a function of $k$. 
Figure 12. GPD distribution, parameters $\gamma = 1/2$, $\sigma = 2$. Top: Violin plots for $n = 100, 500$ of the estimators $H_{k,0}^*$, $H_{k,0}^{\gamma}$, $H_{k,0}^{\gamma}$, $\overline{H}_{k,0}^{\gamma}$, $\overline{T}_{k,0}^{\gamma}$, $\overline{T}_{k,0}^{\gamma}$, $\overline{T}_{k,0}^{\gamma}$. Bottom: diagnostics of $\overline{T}_k$ (blue) and $H_k$ (red) as a function of $k$. 
Figure 13. GPD distribution, parameters $\gamma = 5/2$, $\sigma = 1$. Top: Violin plots for $n = 100, 500$ of the estimators $H_{k_0}$, $H_{k_0^*}$, $T_{k_0^*}$, $T_{k_0^*, p=-1}$, $H_{k_0^*, p=\gamma}$, $T_{k_0^*, p=-\gamma}$. Bottom: diagnostics of $T_k$ (blue) and $H_k$ (red) as a function of $k$. 
In Figure 14 we depict the lower-trimmed Hill plots and the usual Hill plot, together with the empirical variance. As in the simulation studies in Section 5 in order to avoid degeneracies, we only look at candidates for the minimizer to the right of $n/5$, which corresponds to 167 in this case. The minimum empirical variance is then obtained for $k^* = 222$. Using the canonical choice $p = -1$, we have that $k_0 = 222/2.62421 \approx 85$. Note that for the same choice of $p = -1$, and using the prior eyeballed estimate $\xi \approx 0.5$, and consequently $D = -0.5$, we get by (23) the suggestion $k_0^* \approx 112$ (which might be considered the classical choice of the threshold in this case).

The corresponding estimates of $\xi$ are given by $H_{k_0} = 0.508$, $H_{k_0^*} = 0.560$, $T_{k_0} = 0.480$, $T_{k_0^*} = 0.525$.

The simulation studies of Section 5 may suggest the third of the above numbers to be the most reliable estimate here. The ratio statistic test in Figure 16 suggests that for both thresholds the sample is Pareto in the tail (with only a slight issue for the two largest observations).

In (Albrecher et al., 2017, p.99), a splicing point was suggested for this data set at around $k = 20$, based on expert opinion. A semi-automated option using our method for detecting this splicing point would be to replace the left limit $k = 167$ by a very small number (in this case $k = 4$ is chosen after visual inspection of the erratic nature of the empirical variance for the first three), and then to apply our method, which leads to the detection of the minimum variance at $k = 38$ (which is clearly visible in Figure 15). Under the assumption $p = -1$ this then leads to $k \approx 14$ as a suggested splicing point. Note that in the nature of the present data set, the ultimates for the highest claims have intrinsic uncertainty (as they are just estimates of the final closed claim size), and a more systematic way to approach this particular situation would be to combine the trimming of the Hill estimator from below and above, which is not the focus of the present paper.

7. Conclusion

In this paper, we showed that trimming the Hill estimator from the left can lead to favorable properties in connection with the expected empirical variance of the tail index estimators in extreme value statistics. For the Hall class, we established asymptotic results on the behavior of this expected empirical variance, which allows to develop a guideline for the choice of the optimal threshold in the tail index estimation problem. It turns out that there is an intrinsic link between this optimal threshold...
Figure 15. MTPL insurance ultimates. Top left: Lower-trimmed Hill estimator (LTH) estimator for varying lower trimming $b$, for $k$ uniformly spaced from 1 to 837 in units of 20. Top right: Hill plot. Bottom: empirical variance of the LTH as a function of $k$. The dotted line is the left limit for candidates, and the solid line is the resulting minimum.

and the classical optimal threshold for the Hill estimator. Since in the trimming context the identification of the optimal threshold is much more insensitive on the tail characteristics (it only depends on the $p$-parameter in the Hall class, not on $D$ nor on the tail index $\xi$), this link allows to circumvent the classical problem in threshold selection for the Hill estimator. As a by-product, by suitable averaging we develop a novel tail index estimator which assigns a non-uniform weight to each observation in a natural way, relies on fewer assumptions on the tail characteristics, is simple to implement and outperforms the classical Hill estimator in most cases. The latter is illustrated in extensive simulation studies. In addition, the technique is applied to a real-life insurance data set that was previously studied by other techniques. We conclude by noting that the approach taken in this paper is in principle also applicable for the potential improvement of tail index estimators other than the Hill estimator. Further possible directions of future research include the combination of left trimming with right trimming in situations with possible outliers, as well as the consideration of possibly censored data.
Figure 16. Standardized R-statistic for the MTPL ultimates, for the two threshold choices $k = 85$ (left) and $k = 112$ (right). $N_{MC} = 100,000$ and $\alpha = 0.05$.

Acknowledgement. H.A. acknowledges financial support from the Swiss National Science Foundation Project 200021_168993.

References

Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). Reinsurance: Actuarial and Statistical Aspects. John Wiley & Sons, Chichester.

Beirlant, J., Boniphace, E., and Dierckx, G. (2011). Generalized sum plots. REVSTAT-Statistical Journal, 9(2):181–198.

Beirlant, J., Dierckx, G., Guillou, A., and Stãricã, C. (2002). On exponential representations of log-spacings of extreme order statistics. Extremes, 5(2):157–180.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). Statistics of extremes: theory and applications. John Wiley & Sons, Chichester.

Beirlant, J., Vynckier, P., and Teugels, J. L. (1996). Tail index estimation, paretov quantile plots regression diagnostics. Journal of the American Statistical Association, 91(436):1659–1667.

Bhattacharya, S., Kallitsis, M., and Stoev, S. (2017). Trimming the Hill estimator: robustness, optimality and adaptivity. arXiv preprint arXiv:1705.03088.

Bladt, M., Albrecher, H., and Beirlant, J. (2019). Combined tail estimation using censored data and expert information. Preprint, University of Lausanne, submitted for publication.

Csörgö, S., Deheuvels, P., and Mason, D. (1985). Kernel estimates of the tail index of a distribution. Ann. Statist., 13(3):1050–1077.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate analysis, 76(2):226–248.

de Haan, L. and Ferreira, A. (2007). Extreme value theory: an introduction. Springer Science & Business Media.

De Sousa, B. and Michailidis, G. (2004). A diagnostic plot for estimating the tail index of a distribution. Journal of Computational and Graphical Statistics, 13(4):974–995.
Draisma, G., de Haan, L., Peng, L., and Pereira, T. T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes*, 2(4):367–404.

Drees, H., de Haan, L., and Resnick, S. (2000). How to make a Hill plot. *The Annals of Statistics*, 28(1):254–274.

Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75(2):149–172.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, second edition.

Gomes, M. I., de Haan, L., and Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):31–52.

Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review*, 83(2):263–292.

Gomes, M. I. and Oliveira, O. (2001). The bootstrap methodology in statistics of extremes choice of the optimal sample fraction. *Extremes*, 4(4):331–358.

Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (var) estimator. *Journal of the American Statistical Association*, 102(477):280–292.

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):293–305.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B*, 44(1):37–42.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis*, 32(2):177–203.

Hall, P., Welsh, A., et al. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1):331–341.

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

(M. Bladt) Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne, CH-1015 Lausanne, Switzerland

E-mail address: martin.bladt@unil.ch

(H. Albrecher) Department of Actuarial Science, Faculty of Business and Economics and Swiss Finance Institute, University of Lausanne, CH-1015 Lausanne, Switzerland

E-mail address: hansjoerg.albrecher@unil.ch

(J. Beirlant) Department of Mathematics, KU Leuven, Celestijnenlaan 200B, B-3001 Leuven, Belgium

E-mail address: jan.beirlant@kuleuven.be