Extreme Value Analysis for Partitioned Insurance Losses

by John B. Henry III and Ping-Hung Hsieh

ABSTRACT

The heavy-tailed nature of insurance claims requires that special attention be put into the analysis of the tail behavior of a loss distribution. It has been demonstrated that the distribution of large claims of several lines of insurance have Pareto-type tails. As a result, estimating the tail index, which is a measure of the heavy-tailedness of a distribution, has received a great deal of attention. Although numerous tail index estimators have been proposed in the literature, many of them require detailed knowledge of individual losses and are thus inappropriate for insurance data in partitioned form. In this study we bridge this gap by developing a tail index estimator suitable for partitioned loss data. This estimator is robust in the sense that no particular global density is assumed for the loss distribution. Instead we focus only on fitting the model in the tail of the distribution where it is believed that the Pareto-type form holds. Strengths and weaknesses of the proposed estimator are explored through simulation and an application of the estimator to real world partitioned insurance data is given.

KEYWORDS

Heavy-tailed distribution; slowly varying function; partitioned (grouped) data; (re)insurance losses; tail index estimation

1. Introduction

The heavy-tailed nature of insurance claims requires that special attention be put into the analysis of the tail of a loss distribution. Since a few large claims can significantly impact an insurance portfolio, statistical methods that deal with extreme losses have become necessary for actuaries. For example, in order to price certain reinsurance treaties, it is often necessary for actuaries to model losses in excess of some high threshold value, i.e., to model the largest k upper order statistics. Beirlant and Teugels (1992), McNeil (1997), Embrechts, Resnick, and Samorodnitsky (1999), Beirlant, Matthys, and Dierckx (2001), Cebrián, Denuit, and Lambert (2003), and Matthys et al. (2004) provide additional examples where statistical methods were developed to deal with extreme insurance losses.

Extreme value theory has become one of the main theories in developing statistical models for extreme insurance losses. The theory states that the tail of a typical loss distribution $F_X(x)$ can be approximated by a Pareto-type function. That is, $1 - F_X(x) = \ell(x)x^{-\alpha}$, $x > D$ where $\ell : \mathbb{R}^+ \to \mathbb{R}^+$ is a Lebesgue measurable function slowly varying at infinity, i.e., $\lim_{x\to\infty} \ell(tx)/\ell(x) = 1$ for all $t > 0$. The parameter α is known in the literature as the Pareto tail index that measures the heavy-tailedness of the loss distribution. See, for example, Finkelstein, Tucker, and Veeh (2006). Many distributions commonly seen in modeling insurance losses have Pareto-type tails. They include the Pareto, generalized Pareto, Burr, Fréchet, half T, F, inverse gamma, and log gamma distributions. Following the theory, an actuary may assume that the tail of the loss distribution, where extreme losses occur, can be approximated by a Pareto-type function without making specific assumption on the global density. With an estimate of the Pareto index parameter, the actuary can then estimate quantities of interest that are related to extreme losses, e.g., expected loss above a high retention limit. The approximation of a Pareto-type function has been demonstrated to be reasonable for many lines of insurance. Numerous tail index estimators have also been proposed in the literature, including earlier contributions by Hill (1975) and Pickands (1975) in which the Hill estimator has become somewhat of a benchmark to which later proposed estimators are compared. A survey of existing estimators, including their advantages and disadvantages, can be found in Brazauskas and Serfling (2000), Hsieh (2002), and Beirlant et al. (2004).

Insurance loss data reported in partitioned form are common in practice. The frequencies of losses occurred in certain loss intervals for numerous lines of insurance can often be found in companies' reports or in government publications. Individual loss data are typically proprietary to the company and may not be available to its competitors in the industry. Despite the number of tail-index estimators proposed in the literature, many, if not all, of them require the use of individual loss data, and thus are inappropriate for tail-index estimation under the constraint of partitioned data. This paper intends to expand the horizon of tail-index estimation by applying extreme value theory to partitioned loss data. The main objective is to propose a robust tail-index estimator for partitioned loss data. The estimator is robust in the sense that no global density is assumed and the Pareto function is used to approximate the tail of a large class of distributions commonly used in modeling insurance loss data. This approach is advantageous because fitting a global density to losses can lead to errors when making tail inference in the event that the true loss distribution does not have the assumed density. Instead, we rely on the extreme value theory and focus only on fitting the tail of the distribution without assuming a specific global density. In addition, we will demonstrate the loss of efficiency by using the partitioned data versus individual data through simulation.

The remainder of the paper is arranged as follows. Several tail-index estimators are reviewed in Section 2. Except for the Hill and Pickands estimators, both of which have historical values, and the former serving as a benchmark in our simulation, the rest of the review is intended to be a supplement to the excellent reviews of Brazauskas and Serfling (2000), Hsieh (2002), and Beirlant et al. (2004). The derivation of the proposed estimator and an examination of its theoretical properties are worked out in Section 3. In Section 4, a simulation study is conducted to assess the performance of the proposed estimator. Two questions guide the design of the simulation: first, what is the efficiency lost by using data in partitioned form, and, second, what is the penalty of model misspecification? The simulation results are discussed in Section 5. Insurance applications are given in Section 6 using actual grouped insurance losses, followed by concluding remarks in Section 7.

2. Literature review

In this section we consider tail-index estimators for a loss random variable (r.v.) X taking values on the positive real line $\mathbb{R}^+$ with nondegenerate distribution function F_X. We assume that the loss distribution has a Pareto-like tail in the sense that

$$\mathrm{P}(X > x) = \ell(x)x^{-\alpha}, \quad \text{as} \quad x \to \infty, \quad (2.1)$$

where $\alpha > 0$. In this case the probability that a loss exceeds a level x can be closely approximated by $Cx^{-\alpha}$ when x is larger than some threshold D. We will denote the tail probability function by $\bar{F}_X(x) := 1 - F_X(x)$. Let $\{X_k : 1 \le k \le n\}$ be a sequence of independent copies of X and denote the descending order statistics by $X^{(1)} \ge X^{(2)} \ge \cdots \ge X^{(n)}$.

In the following subsections, we discuss several estimators for the tail index α. Some noteworthy estimators that are not discussed below are the method of moments, probability-weighted moments, elemental percentile, Bayes estimator with conjugate priors, and hybrid estimators. A description of these can be found, for example, in Hsieh (2002) and the references therein.

2.1. The Hill and Pickands estimators

Hill (1975) proposed the tail-index estimator

$$\hat{\alpha}_H = \frac{k+1}{\sum_{i=1}^{k} i \log\left(\frac{X^{(i)}}{X^{(i+1)}}\right)} \quad (2.2)$$

based on a maximum likelihood argument where $k \in \{1,2,\ldots,n-1\}$. The Hill estimator is closely related to the mean excess function $e(u) = \mathrm{E}\{X - u \mid X > u\}$. In particular, the *empirical* mean excess function is given by $e_n(u) = [\mathrm{card}\Lambda_n(u)]^{-1} \sum_{j\in\Lambda_n(u)}(X_j - u)$, where $\mathrm{card}\Lambda_n(u)$ denotes the number of elements in the set $\{j : X_j - u > 0, j = 1,\ldots,n\}$. Then, letting $e_n^*(u)$ denote the empirical mean excess function of the log transformed variables, we have $e_n^*(\log X^{(k+1)}) = (1/k)\sum_{i=1}^{k}(\log X^{(i)} - \log X^{(k+1)})$. As a result, we see that $\hat{\alpha}_H = ((k+1)/k)e_n^*(\log X^{(k+1)})^{-1}$. That is, the Hill estimator is asymptotically equal to the reciprocal of the empirical mean excess function of $\log X$ evaluated at the threshold $\log X^{(k+1)}$.

An important feature of the Hill estimator to keep in mind is the variance-bias tradeoff that occurs when choosing the number of upper order statistics to use. Choosing too many of the largest order statistics can lead to a biased estimator, while too few increases the variability of the estimator. See Embrechts, Klüppelberg, and Mikosch (1997) for a further variance-bias tradeoff discussion and Hall (1990), Dekkers and de Haan (1993), Dupuis (1999), and Hsieh (1999) for methods for determining the number of upper-order statistics or threshold to use. Properties of the Hill estimator can be found in Embrechts, Klüppelberg, and Mikosch (1997) and the references therein.

Pickands (1975) proposed an estimator that matches the 0.5 and 0.75 quantiles of the generalized Pareto distribution (GPD) with quantile

estimates. More specifically, for a GPD r.v. X with distribution function

$$G(x;\xi,\sigma) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi} 1_{(0,\infty)}(x),$$

it is easy to show that

$$\frac{G^{-1}(0.75) - G^{-1}(0.5)}{G^{-1}(0.5)} = 2^{\xi}.$$

Then denoting 0.5 and 0.75 quantile estimates by $\hat{q}_1$ and $\hat{q}_2$, respectively, we have

$$\hat{\xi} = \frac{\log\left(\frac{\hat{q}_2 - \hat{q}_1}{\hat{q}_1}\right)}{\log 2}.$$

Pickands proposed, for n independent copies of X, using $\hat{q}_1 = X^{(m)} - X^{(4m)}$ and $\hat{q}_2 = X^{(2m)} - X^{(4m)}$ where $n \gg m \geq 1$. Then noting that the tail index for a GPD r.v. is given by $\alpha = 1/\xi$, the resulting tail index estimate is, for $X^{(k)} \geq D$,

$$\hat{\alpha}_P = \frac{\log 2}{\log\left(\frac{X^{(m)} - X^{(2m)}}{X^{(2m)} - X^{(4m)}}\right)}, \tag{2.3}$$

where $k \geq 4m \geq 4$.

For consistency and asymptotic results, see Dekkers and de Haan (1989). While the simplicity of the Pickands estimator is an attractive feature, it makes use of only three upper-order statistics and can have a large asymptotic variance. Generalized versions of the Pickands estimator can be found, for example, in Segers (2005). See Section 2.2.2.

2.2. Some recent tail-index estimators

2.2.1. Censored data estimator

In the case of moderate right censoring, Beirlant and Guillou (2001) proposed an estimator based on the slope of the Pareto quantile plot, excluding the censored data. This can be useful in situations when there has been a policy limit or when a reinsurer has covered losses in the portfolio exceeding some well-defined retention level. Letting N_c denote the number of censored losses, the estimator is

$$\hat{\alpha}_{N_c}(k) = \frac{k - N_c}{\sum_{i=N_c+1}^{k} \log \frac{X^{(i)}}{X^{(k+1)}} + N_c \log \frac{X^{(N_c+1)}}{X^{(k+1)}}}, \tag{2.4}$$

where $k \in \{N_c + 1, \ldots, n-1\}$. This estimator is equivalent to the Hill estimator (except for the change from $k+1$ to k, which is asymptotically negligible) in the case of no censoring (i.e., $N_c = 0$). It is argued by Beirlant and Guillou (2001) that typically no more than 5% of observations should be censored for an effective use of this method.

2.2.2. Location invariant estimators

It is pointed out by Fraga Alves (2001) that, for modeling large claims in an insurance portfolio, it is desirable for an estimator of α to have the same distribution for the excesses taken over any possible fixed deductible. For this reason, location invariance is clearly a desirable property for an estimator of α. Fraga Alves (2001) introduced a Hill-type estimator that is made location invariant by a random shift. The location-invariant estimator is

$$\hat{\alpha}_{k_0,k} = \frac{k_0}{\sum_{i=1}^{k_0} \log \frac{X^{(i)} - X^{(k+1)}}{X^{(k_0+1)} - X^{(k+1)}}}, \tag{2.5}$$

where k_0 is a secondary value chosen with $k_0 < k$. An algorithm is included in Fraga Alves (2001) to estimate the optimal k_0, and to make a bias correction adjustment to $\hat{\alpha}_{k_0,k}$.

Generalized Pickands estimators described in Segers (2005) are also location invariant and are linear combinations of log-spacings of order statistics. In particular, let Λ denote the collection of all signed Borel measures λ on $(0,1]$ such that

$$\lambda((0,1]) = 0, \qquad \int \log(1/t)|\lambda|(dt) < \infty,$$

$$\text{and} \qquad \int \log(1/t)\lambda(dt) = 1.$$

Then for $\lambda \in \Lambda$ and $0 < c < 1$, the generalized Pickands estimators are given by

$$\hat{\alpha}_k(c,\lambda) = \left(\sum_{i=1}^{k} \left[\lambda\left(\frac{i}{k}\right) - \lambda\left(\frac{i-1}{k}\right) \right] \times \log(X^{(1+\lfloor cj \rfloor)} - X^{(i+1)}) \right)^{-1}. \tag{2.6}$$

See Segers (2005) for examples using different measures λ and theoretical properties of the generalized Pickands estimators. See also Drees (1998) for a general theory of location and scale invariant tail-index estimators that can be written as Hadamard differentiable continuous functionals of the empirical tail quantile function.

2.2.3. Generalized median estimator

Brazauskas and Serfling (2000) proposed a class of generalized median (GM) estimators with the goal of retaining a relatively high degree of efficiency while also being adequately robust. The GM estimator is found by considering, for $X^{(k)} \geq D$ and $r \in \{2,\ldots,k\}$, the median of a kernel h evaluated over all $\binom{k}{r}$ subsets of $X^{(1)},\ldots,X^{(k)}$. The GM estimator is then given by

$$\hat{\alpha}_r = \text{med}\{h(X^{(i_1)}),\ldots,h(X^{(i_r)})\}, \tag{2.7}$$

where $\{i_1,\ldots,i_r\}$ corresponds to a set of distinct indices from $\{1,\ldots,k\}$. Examples of kernels h, properties of the GM estimators, and comparison between the GM estimators and several other estimators can be found in Brazauskas and Serfling (2000).

2.2.4. Probability integral transform statistic estimator

Finkelstein, Tucker, and Veeh (2006) describe a probability integral transform statistic (PITS) estimator for the tail-index parameter of a Pareto distribution. They develop the PITS estimator through an easily understandable and sound probabilistic argument. The PITS estimator is shown to be comparable to the best robust estimators. Consider first a random sample of Pareto random variables $X_1,\ldots,X_n$, each with common distribution function $F(x) = 1 - (D/x)^{\alpha}$ for $x \geq D$ where $D > 0$ is known and $\alpha > 0$. Then defining

$$G_{n,t}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D}{X_i}\right)^{\beta t},$$

where $t > 0$, observe that

$$G_{n,t}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\bar{F}(X_i)^t \overset{\text{d}}{=} = \frac{1}{n}\sum_{i=1}^{n}U_i^t,$$

where $U_1,\ldots,U_n$ are i.i.d Uniform $(0,1)$ random variables. Applying the Strong Law of Large Numbers yields

$$G_{n,t}(\alpha) \xrightarrow{\text{p}} E(U_1^t) = (t+1)^{-1}.$$

Using the idea of method of moment estimation, the PITS estimator is the solution of the equation $G_{n,t}(\beta) = (t+1)^{-1}$. The tuning parameter $t > 0$ is used to adjust between robustness and efficiency. See Finkelstein, Tucker, and Veeh (2006) for details. In the case D is unknown, one can consider

$$G_{n,t,k}(\beta) := \frac{1}{n}\sum_{i=1}^{k}\left(\frac{X^{(k+1)}}{X^{(i)}}\right)^{\beta t},$$

for $k \in \{1,2,\ldots,n-1\}$ and use the same approach to arrive at a PITS estimator for the tail-index α.

3. Tail-index estimator for partitioned data

Let $\{X_k : 1 \leq k \leq n\}$ be a sequence of independent copies of a loss random variable X satisfying (2.1). Suppose that losses are grouped into classes $\{I_i = (a_i, a_{i-1}]\}_{i=1,\ldots,g}$, where $\infty = a_0 > a_1 > \cdots > a_g > 0$. Assuming the loss distribution has the Pareto-type form above a threshold D, we take $0 < D \leq a_k$ without loss of generality for some $k \in \{2,3,\ldots,g\}$. We let $N_1,\ldots,N_g$ denote the frequencies with which $(X_1,\ldots,X_n)$ take values in $\{I_i = (a_i, a_{i-1}]\}_{i=1,\ldots,g}$. That is, $N_i = \text{card}\{j : a_{i+1} < X_j \leq a_i, \quad 1 \leq j \leq n\}, \quad i = 1,\ldots,g.$

The likelihood function is then defined as

$$L_1 = \frac{n!}{\prod_{i=1}^{r} n_i!} \prod_{i=1}^{g} \left(\int_{a_i}^{a_{i-1}} f_X(x) d\mu(x) \right)^{n_i},$$

where f_X is the density of X with respect to Lebesgue measure μ. Hence

$$L_1 \propto \prod_{i=1}^{g} (\bar{F}_X(a_i) - \bar{F}_X(a_{i-1}))^{n_i}.$$

Then setting $\bar{F}_X(x)$ equal to $\ell(x)x^{-\alpha}$ for $x \geq a_k \geq D$, we consider the conditional likelihood function $L(\alpha \mid n_1, \ldots, n_k)$ proportional to

$$L_k(\alpha) = \prod_{i=1}^{k} \left(\frac{\bar{F}_X(a_i) - \bar{F}_X(a_{i-1})}{\bar{F}_X(a_k)} \right)^{n_i} \approx \prod_{i=1}^{k} \left(\frac{a_i^{-\alpha} - a_{i-1}^{-\alpha}}{a_k^{-\alpha}} \right)^{n_i}, \quad (3.1)$$

where $a_0^{-\alpha}$ is set to 0. The proposed tail-index estimator is given by

$$G_k := \arg\max L_k(\alpha) \quad (3.2)$$

where $k \in \{2, 3, \ldots, g\}$. That is, G_k equals the value of α that maximizes the likelihood function L_k defined in Eq. (3.1). The lemma below shows that G_k exists and is a unique maximum likelihood estimator for α. As a result, one is able to obtain maximum likelihood estimates for tail probabilities and mean excess loss by using the invariance property of maximum likelihood estimators. These formulas are given in Section 6.

LEMMA Existence and uniqueness of the proposed estimator
G_k in Eq. (3.2) exists and is unique.

PROOF Define $b_i := \log(a_i/a_k)$ for $i = 1, \ldots, k$ and $u_i := a_i/a_{i-1}$ for $i = 2, \ldots, k$. Using Eq. (3.1), consider the log-likelihood function

$$\log L_k(\alpha) = \alpha n_1 \log(a_k/a_1) + \sum_{i=2}^{k} n_i \log \left(\frac{a_i^{-\alpha} - a_{i-1}^{-\alpha}}{a_k^{-\alpha}} \right).$$

Then it is easy to show using calculus that

$$\frac{\partial \log L_k(\alpha)}{\partial \alpha} = -n_1 b_1 - \sum_{i=2}^{k} n_i \left(\frac{b_i}{1 - u_i^{\alpha}} + \frac{b_{i-1}}{1 - u_i^{-\alpha}} \right).$$

Noting that $u_i < 1$ for each i and $b_i > 0$ for $i \geq 2$, we have

$$\frac{\partial \log L_k(\alpha)}{\partial \alpha} \longrightarrow \begin{cases} -\sum_{i=1}^{k} n_i b_i < 0, & \alpha \uparrow +\infty, \\ +\infty, & \alpha \downarrow 0. \end{cases}$$

The result follows by noting that $b_i > b_{i-1}$ implies

$$\frac{\partial^2 \log L_k(\alpha)}{\partial \alpha^2} = \frac{(b_i - b_{i-1}) \log u_i}{(2 \sinh(\alpha \log(u_i)/2))^2} < 0.$$

■

4. Performance assessment

In this section, we conduct a simulation to study the performance of the proposed tail estimator G_k. The two key questions guiding the design of the simulation are, first, what is the efficiency lost due to the use of partitioned data, and, second, how robust is the proposed estimator with respect to model misspecification?

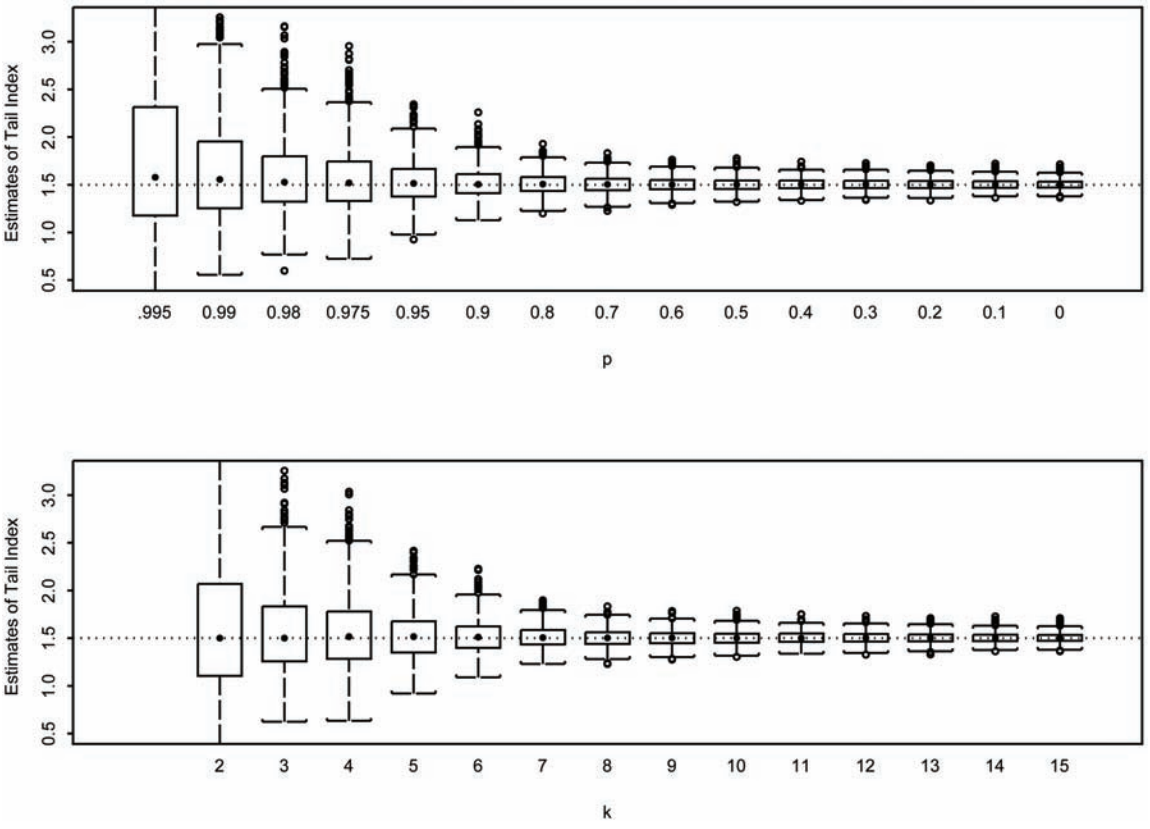
Specifically, m samples of size n are generated from a distribution $F(x)$ with the mean $\mu < \infty$, standard deviation σ and $x \geq 0$. The domain of $F(x)$, $\mathbb{R}^+$, is partitioned into g nonoverlapping intervals, $I_1, \ldots, I_g$. That is, $I_i \cap I_j = \emptyset$ for $1 \leq i \neq j \leq g$ and $\mathbb{R}^+ = \cup_{i=1}^{g} I_i$. The individual observations in each sample are then grouped with respect to the partition, and frequencies n_i in each interval, $i = 1, \ldots, g$, are recorded. In this paper, we report the simulation results obtained from using $m = 1000$ (samples), $n = 1000$ (observations), $g = 15$ (intervals), and the partition $I_i = (F^{-1}(p_i), F^{-1}(p_{i-1}))$, where $\{p_j\}_0^{15} = \{1.00, 0.995, 0.99, 0.98, 0.975, 0.95, 0.90, (0.80, 0.70, \ldots)0.00\}$ for $i = 1, 2, \ldots, g$, and $F^{-1}(p) = \inf\{x : F(x) \geq p\}$. We consider four distributions commonly used in modeling insurance losses. They include the Pareto with a parameter α, generalized Pareto with parameters γ and σ, Burr with parameters λ, θ, and τ, and the half T distribution with degrees of freedom ϕ. The parameterizations of these distributions are given in Table 1.

Table 1. Tail index parameters and mean excess functions for selected distributions

Distribution	$\bar{F}_X(x) = 1 - F_X(x)$	Parameters	$e(u)^a$	Tail Index
Pareto	$\left(\frac{D}{x}\right)^{\alpha} 1_{(D,\infty)}(x)$	$D, \alpha > 0$	$\frac{u}{\alpha - 1}$, for $\alpha > 1$	α
GPD	$\left(1 + \frac{\gamma}{\sigma} x\right)^{-1/\gamma} 1_{(0,\infty)}(x)$	$\gamma, \sigma > 0$	$\frac{\sigma + \gamma u}{1 - \gamma}$, for $\gamma^{-1} > 1$	γ^{-1}
Burr	$\left(\frac{\lambda}{\lambda + x^{\tau}}\right)^{\alpha} 1_{(0,\infty)}(x)$	$\lambda, \tau, \alpha > 0$	$\frac{u}{\alpha\tau - 1}(1 + o(1))$, for $\alpha\tau > 1$	$\alpha\tau$
Half-T	$\frac{2\Gamma\left(\frac{\phi+1}{2}\right)}{\sqrt{\phi\pi}(\phi/2)} \int_x^{\infty} \left(1 + \frac{y^2}{\phi}\right)^{-(\phi+1)/2} dy\, 1_{(0,\infty)}(x)$	$\phi > 0$	$\frac{u}{\phi - 1}(1 + o(1))$, for $\phi > 1$	ϕ

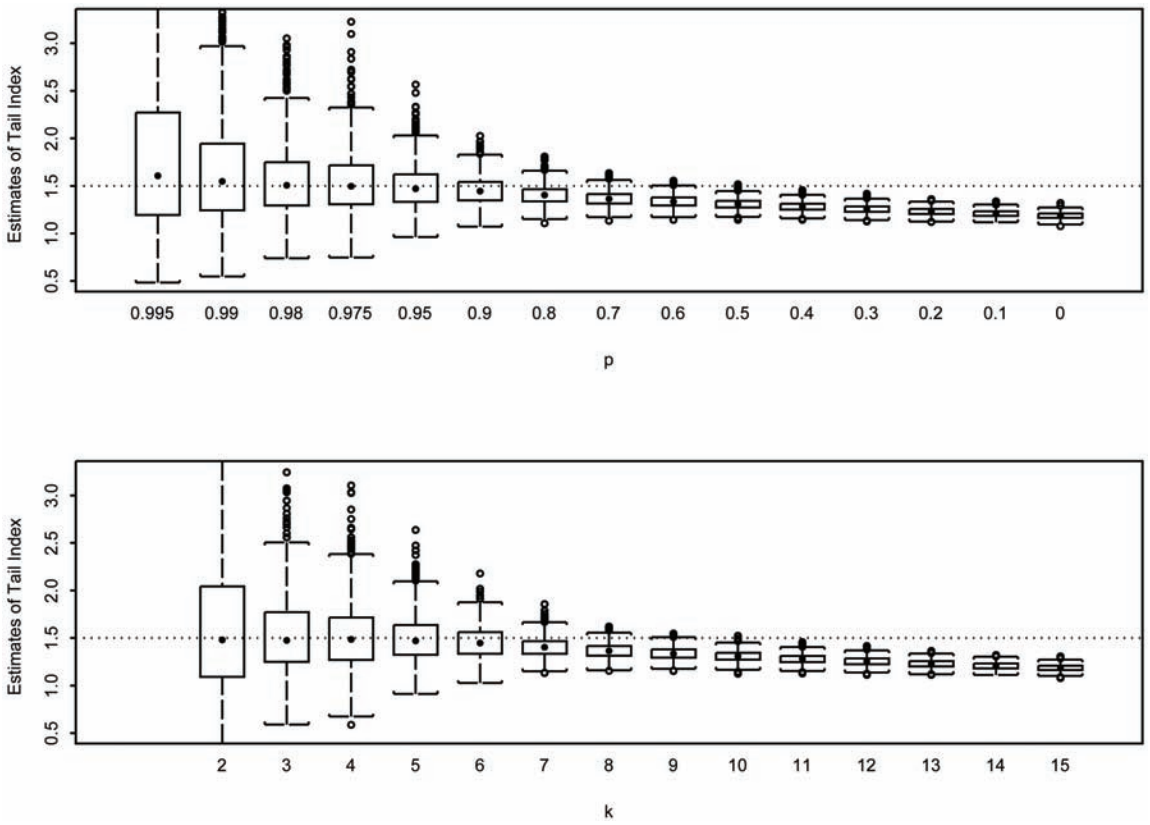
[a]The asymptotic relations are to be understood for $u \to \infty$.

Figure 1. Performance of Hill (top) and G_k (bottom) estimators for underlying Pareto model with true tail index $\alpha = 1.5$ ($D = 1$, $\alpha = 1.5$). Hill estimates use all order statistics above $F^{-1}(p)$ where F is the distribution function of the underlying distribution. Tail index estimates using grouped data are found using Eq. (3.2) for the given number of upper interval counts k. Sample size = number of replications = 1000.



With simulated data in two different formats, the exact values as well as values in partitioned form, we compare the performance of the proposed estimator G_k using frequencies in the intervals I_i where $\inf I_i \geq D$ to that of the Hill estimator using all $x_i \geq D$, as well as to that of the maximum likelihood estimator using all frequencies n_i or all x_i. In Figures 1–4, we report the loss in efficiency due to the use of partitioned data. The Hill estimates for α in the jth box-plot, from left to right, are calculated using the largest $n_1 + \cdots + n_j$ order statistics. The es-

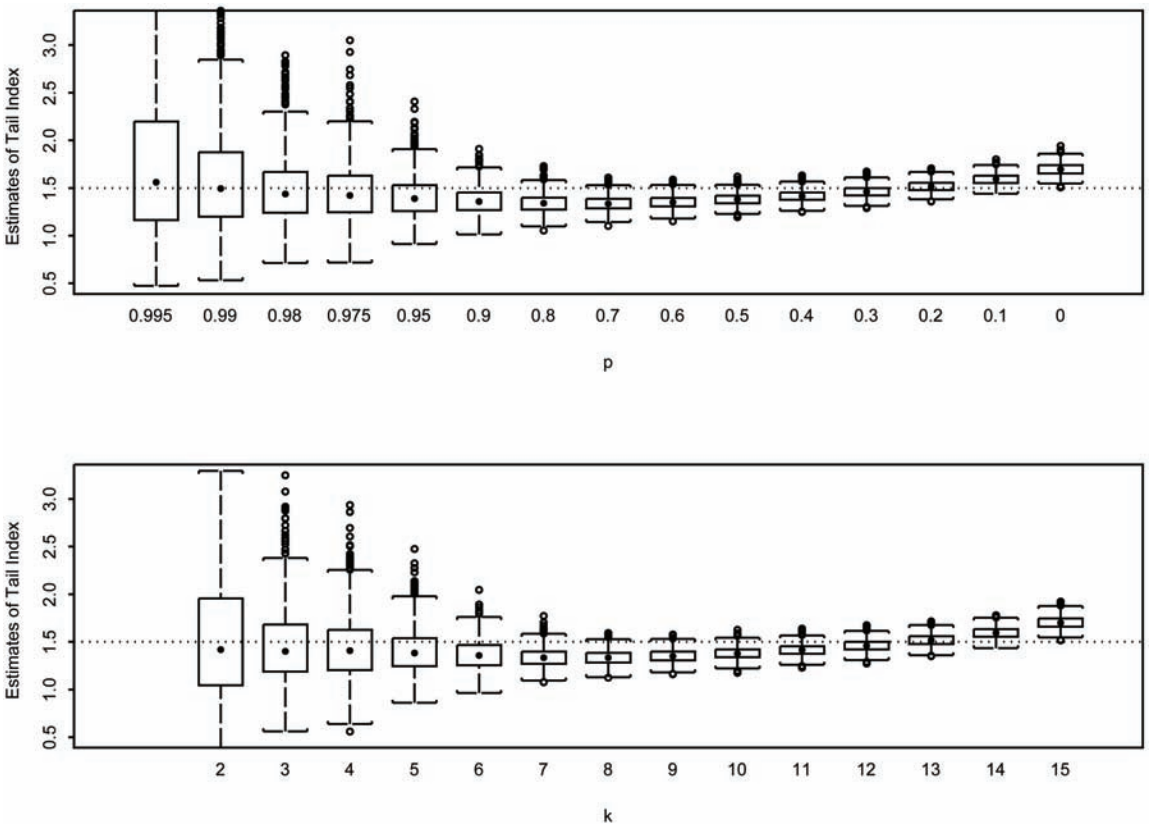
Figure 2. Performance of Hill (top) and G_k (bottom) estimators for underlying generalized Pareto model with true tail index $\alpha = 1.5$ ($\gamma = 1/1.5$, $\sigma = 1$). Hill estimates use all order statistics above $F^{-1}(p)$ where F is the distribution function of the underlying distribution. Tail index estimates using grouped data are found using Eq. (3.2) for the given number of upper interval counts k. Sample size = number of replications = 1000.



timates from the proposed estimator in the jth box-plot, from left to right, are calculated using Eq. (3.2) with $k = j + 1$, for $j = 1,2,\ldots,14$. We notice that in Figures 1–4 the proposed estimator behaves similar to the Hill estimator. In addition, we take the tail estimates that comprise each box-plot to calculate the root mean squared error (RMSE). That is, for the jth box-plot, $\text{RMSE}_j = m^{-1}\sum_{i=1}^{m}(\hat{\alpha}_{ji} - \alpha)^2$, where $m = 1000$, the true tail index $\alpha = 1.50$, and $\hat{\alpha}_{ji}$ represents the ith tail-index estimate in the jth box-plot. The dashed line in each panel represents the true tail-index parameter value. To quantify the loss of efficiency, we further define efficiency (EFF) as the ratio of RMSE_j obtained from the proposed estimator to RMSE_j obtained from the Hill estimator. The results are reported in Table 2.

To examine the robustness of the proposed estimator against model misspecification, we compare the proposed estimator using frequencies in the top 6 and 7 intervals, which correspond to the 90th and 80th percentiles of the true underlying distribution, to four maximum likelihood (ML) estimators using all 15 frequencies $N_1,\ldots,N_{15}$. These four ML estimators differ in the assumed underlying distributions. They include Pareto (ML_Pareto), generalized Pareto (ML_GPD), Burr (ML_Burr), and half T (ML_T). Following our simulation design, it allows one of the four ML estimates to be the target estimate since this particular estimate is obtained by assuming the correct underlying distribution and by using the entire sample (all 15 frequencies) in estimation. The performance of the Hill estimator

Figure 3. Performance of Hill (top) and G_k (bottom) estimators for underlying Burr model with true tail index $\alpha = 1.5$ ($\lambda = 1.2$, $\theta = 4/2$, $\tau = 3/4$). Hill estimates use all order statistics above $F^{-1}(p)$ where F is the distribution function of the underlying distribution. Tail index estimates using grouped data are found using Eq. (3.2) for the given number of upper interval counts k. Sample size = number of replications = 1000.



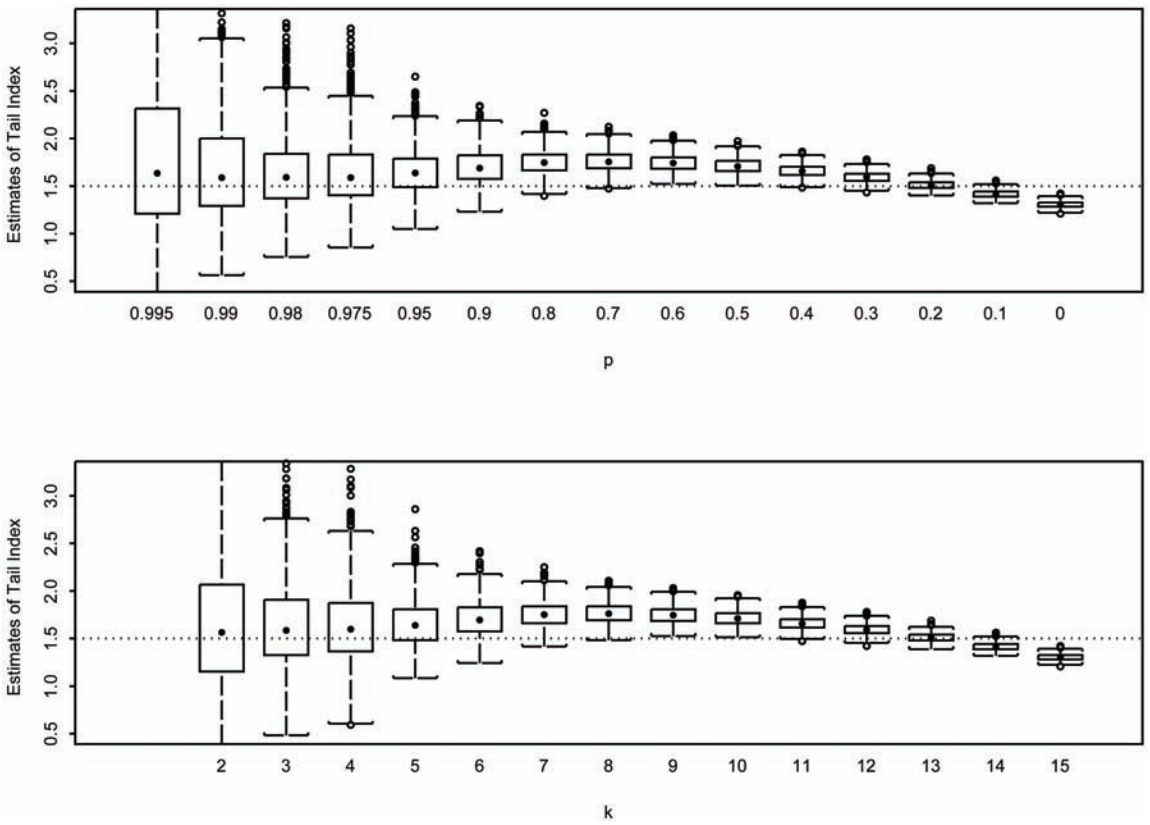
using observations above the 90th and 80th percentiles of the true distribution is also compared to those of the four similarly defined ML estimators that use the entire sample in estimation. With the tail estimates, we then calculate the expected loss exceeding the 95th percentile of the true distribution, $e(q_{.95}) = \mathrm{E}\{X - q_{.95} \mid X > q_{.95}\}$. The resulting expected losses are reported in Figures 5–8. In addition, we quantify these figures by calculating RMSE and EFF (see Table 3). Note that EFF in this table is defined as the ratio of RMSE of an estimator to that of the ML estimator that assumes the correct underlying distribution. Hence, if the true underlying distribution is Pareto, then EFF = 1 for ML_Pareto.

The simulation results for sample sizes 100, 250, and 500 are reported in the Appendix.

5. Discussion of simulation results

The simulation conducted in the previous section illustrates the loss of efficiency in using partitioned data. There is no doubt that efficiency is lost with the use of partitioned data simply because fewer data points are used in maximizing the likelihood function. This is evident from those box-plots in the far left in Figures 1–4 and from the EFF measures in the first few columns in Table 2 when only observations exceeding the 95th percentile are used in estimation. For example, as shown in Table 2, when the underlying distribution is Pareto, the RMSE for the Hill estimator using observations exceeding the 99th percentile and the RMSE for the proposed estimator using the frequencies from the top two intervals are 0.75 and 4.47, respectively, giving

Figure 4. Performance of Hill (top) and G_k (bottom) estimators for underlying half T model with true tail index $\alpha = 1.5$ ($\phi = 1.5$). Hill estimates use all order statistics above $F^{-1}(p)$ where F is the distribution function of the underlying distribution. Tail index estimates using grouped data are found using Eq. (3.2) for the given number of upper interval counts k. Sample size = number of replications = 1000.



EFF = 5.99. This implies that parameter estimation error, measured in RMSE, can be 5.99 times higher with the use of partitioned data than with the use of individual data. However, the amount of error quickly diminishes. With only the top three frequencies (N_1, N_2, and N_3) in use, the EFF is below 1.20 for all four distributions. Using the top five frequencies or more, the EFF never exceeds 1.10 and quickly approaches 1.01. The parameter estimation error between the use of partitioned data and of individual data becomes negligible.

The tables in Appendix A show results where sample sizes 500, 250, and 100 were used in the simulation. For $n = 500$, the EFF never exceeds 1.10 and quickly approaches 1.01 for all four distributions when the top five frequencies (i.e., upper 5%) are in use. This is similar to the findings with $n = 1000$. For $n = 250$, the top 6 (i.e., upper 10%), and for $n = 100$, the top seven frequencies (i.e., upper 20%) must be included for the EFF to go below 1.10. Our simulation seems to suggest that, for sample size between 100 and 1000, the loss of efficiency due to grouped data is minimal if 20% or more of the observations are included in estimating the tail index.

Figures 1–4 also reveal a typical problem in tail-index estimation. When taking only few data points in estimation, the resulting estimates exhibit large variance, whereas if taking more data points than necessary, the bias of the estimates seems evident. This variance-bias tradeoff suggests the development of a threshold selection process to determine a threshold above which

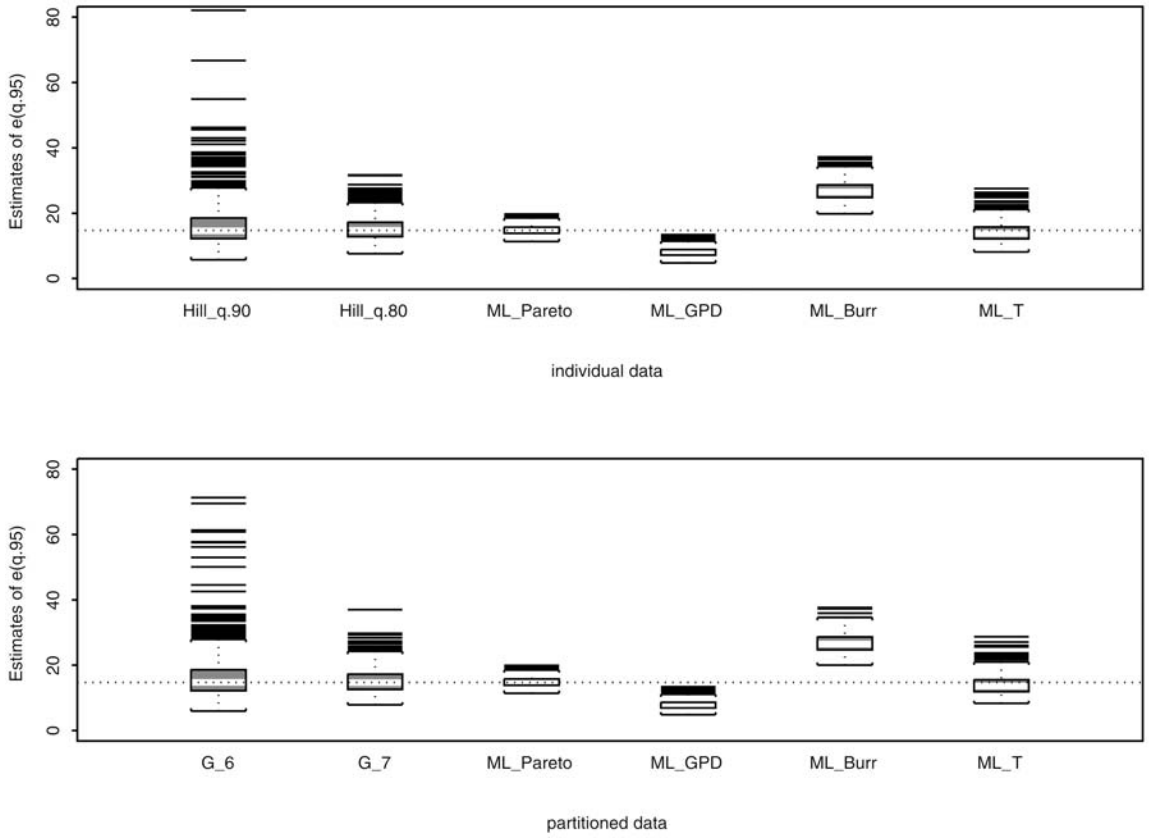
Table 2. Loss of efficiency with the use of partitioned data, $n = 1000$

Threshold D used in the Hill estimator	$q_{.99}$	$q_{.98}$	$q_{.975}$	$q_{.95}$	$q_{.90}$	$q_{.80}$	$q_{.70}$	$q_{.60}$	$q_{.50}$	$q_{.40}$	$q_{.30}$	$q_{.20}$	$q_{.10}$	$q_{.00}$
No. of top intervals k used in G_k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
							True distribution: Pareto							
Cutoff D	**21.54**	**13.57**	**11.7**	**7.37**	**4.64**	**2.92**	**2.23**	**1.84**	**1.59**	**1.41**	**1.27**	**1.16**	**1.07**	**1**
Hill	0.75	0.41	0.34	0.23	0.15	0.11	0.09	0.08	0.07	0.06	0.06	0.05	0.05	0.05
G_k	4.47	0.48	0.39	0.24	0.16	0.11	0.09	0.08	0.07	0.06	0.06	0.05	0.05	0.05
Efficiency	5.99	1.19	1.14	1.07	1.03	1.03	1.03	1.02	1.02	1.02	1.01	1.01	1.01	1.01
							True distribution: generalized Pareto							
Cutoff D	**31.82**	**19.86**	**17.04**	**10.55**	**6.46**	**3.89**	**2.85**	**2.26**	**1.88**	**1.61**	**1.4**	**1.24**	**1.11**	**1**
Hill	0.66	0.38	0.33	0.21	0.15	0.14	0.15	0.18	0.20	0.23	0.25	0.27	0.29	0.32
G_k	3.25	0.44	0.35	0.23	0.16	0.14	0.15	0.18	0.20	0.23	0.25	0.27	0.30	0.32
Efficiency	4.95	1.15	1.08	1.07	1.05	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
							True distribution: Burr							
Cutoff D	**24.87**	**15.12**	**12.86**	**7.7**	**4.57**	**2.69**	**1.99**	**1.62**	**1.39**	**1.25**	**1.14**	**1.07**	**1.03**	**1**
Hill	0.79	0.44	0.39	0.30	0.27	0.27	0.27	0.26	0.22	0.17	0.11	0.05	0.09	0.20
G_k	4.43	0.52	0.44	0.32	0.28	0.27	0.28	0.26	0.22	0.17	0.11	0.05	0.09	0.20
Efficiency	5.61	1.17	1.12	1.06	1.03	1.01	1.01	1.01	1.00	1.00	1.00	1.00	1.02	1.01
							True distribution: Half T							
Cutoff D	**18.82**	**12.2**	**10.64**	**7.02**	**4.71**	**3.2**	**2.55**	**2.15**	**1.87**	**1.65**	**1.47**	**1.3**	**1.15**	**1**
Hill	0.66	0.38	0.32	0.22	0.19	0.19	0.18	0.16	0.13	0.10	0.07	0.06	0.11	0.21
G_k	3.42	0.45	0.37	0.24	0.20	0.19	0.18	0.16	0.14	0.10	0.07	0.06	0.11	0.21
Efficiency	5.19	1.18	1.13	1.08	1.03	1.02	1.01	1.01	1.01	1.01	1.00	1.02	1.02	1.02

the assumed Pareto-type functional form holds. In other words, we should not include any data points that are below the threshold in estimation to avoid bias because the assumed functional form is no longer valid. In addition to the diagnostic plot approach described in the next section, we may also consider an analytic approach to selecting the threshold for a given sample. We may start with the frequencies N_1 and N_2 in the first two intervals I_1 and I_2, and sequentially include frequencies in the adjacent intervals by testing whether the assumed functional form holds. We could perhaps make use of the fact that, conditional on $\sum_{i=1}^{k} N_i = \sum_{i=1}^{k} n_i$, $N_j \sim$ Binomial$(\sum_{i=1}^{k} n_i, p_{jk}(\alpha))$, where $p_{jk}(\alpha) = (a_j^{-\alpha} - a_{j-1}^{-\alpha})/a_k^{-\alpha}$. See, for example, Hsieh (1999) and Dupuis (1999).

If the underlying distribution is known, then the ML estimator is a common choice for parameter estimation. The ML estimate and the quantities derived from the estimate, e.g., the mean excess value $e(u)$, possess desirable statistical properties. However, the true underlying distribution is typically unknown in practice, and the penalty of model misspecification and possibly subsequent misinformed decisions may not be negligible. Our simulation results shown in Figures 5–8 and in Table 3 illustrate the robustness of our proposed estimator and the penalty of model misspecification. It is clear from Table 1 that a reliable estimate of the tail index is crucial for estimating the mean excess function $e(u)$. The estimation error of $e(u)$ can be substantial without a reliable tail index estimator. For example, as reported in Table 3, when the true distribution is Pareto, the estimation error of $e(u)$, measured as RMSE, for the four ML estimators using individual data and partitioned data ranges from 1.15 to 12.08, and from 1.16 to 12.09, respectively. ML_Pareto, not surprisingly, has the lowest RMSE because it assumes the correct un-

Figure 5. Estimation of mean excess value $e(q_{.95})$. ML estimates are calculated under the assumption of the specified distributions. The true distribution F is Pareto with tail index $\alpha = 1.5$. The top plot uses all data, and the bottom plot uses grouped data. The $\text{Hill}\_q_{.90}$ and $\text{Hill}\_q_{.80}$ use all order statistics larger than $q_{.90} = F^{-1}(.90)$ and $q_{.80} = F^{-1}(.80)$. The G_6 and G_7 use the counts from top 6 and 7 intervals. Sample size = number of replications = 1000.



derlying distribution and utilizes the entire sample. However, if the distribution is mistakenly assumed, then the RMSE can be 2, 6, or even 10 times higher than that of ML_Pareto. In contrast, the RMSEs of the proposed estimator and the Hill estimator, despite using only a fraction of the data, stay relatively close to the best RMSE across all four assumed distributions, providing the robustness against model misspecification. The same conclusion can be drawn even with a sample size $n = 100$; see the tables in Appendix B.

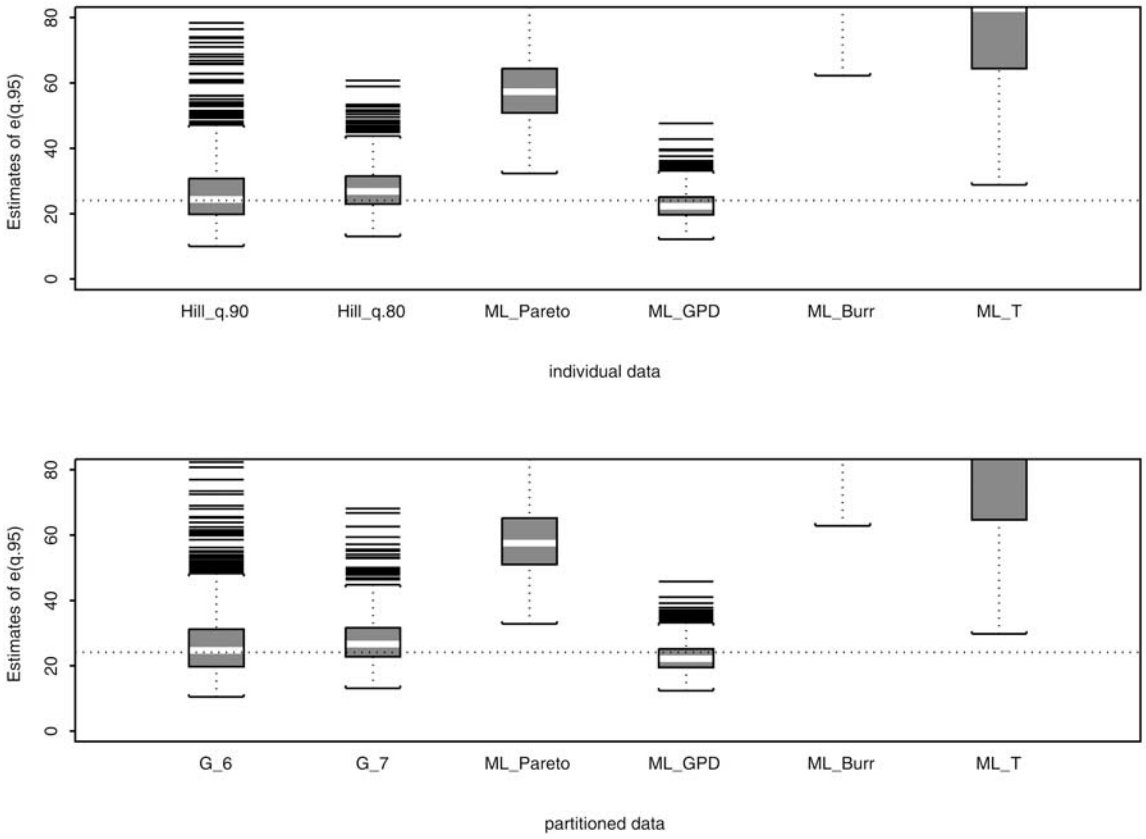
Table 3 also highlights a problem often encountered in practice: the ML algorithm may not converge properly, leading to abnormal estimates. This is evident from the ML_Burr column where the ML algorithm did not converge in several iterations, resulting in insensible estimates, and thus, large RMSE.

Finally, the Hill and G_k estimators largely underestimate $e(q_{.95})$ when the true underlying distribution is half T (Figure 8). This is the result of the variance-bias tradeoff previously discussed. By using frequencies in the top 6 or 7 intervals, we have taken data from the area of distribution that the Pareto tail approximation does not hold. Once again, a threshold selection method is necessary to identify the optimal number k of frequencies to be used in G_k.

6. Applications to insurance

In this section we apply the proposed tail index estimator to actual insurance data available only in a partitioned form. The observed losses,

Figure 6. Estimation of mean excess value $e(q_{.95})$. ML estimates are calculated under the assumption of the specified distributions. The true distribution F is generalized Pareto with tail index $\alpha = 1.5$ ($\gamma = 1/1.5$, $\sigma = 1$). The top plot uses all data, and the bottom plot uses grouped data. The $\text{Hill\_}q_{.90}$ and $\text{Hill\_}q_{.80}$ use all order statistics larger than $q_{.90} = F^{-1}(.90)$ and $q_{.80} = F^{-1}(.80)$. The G_6 and G_7 use the counts from top 6 and 7 intervals. Sample size = number of replications = 1000.



summarized in Table 4, are taken from Hogg and Klugman (1984) and consist of Homeowners 02 policies in California during accident year 1977 supplied by the Insurance Services Office (ISO). Losses were developed to 27 months and include only policies with a $100 deductible.

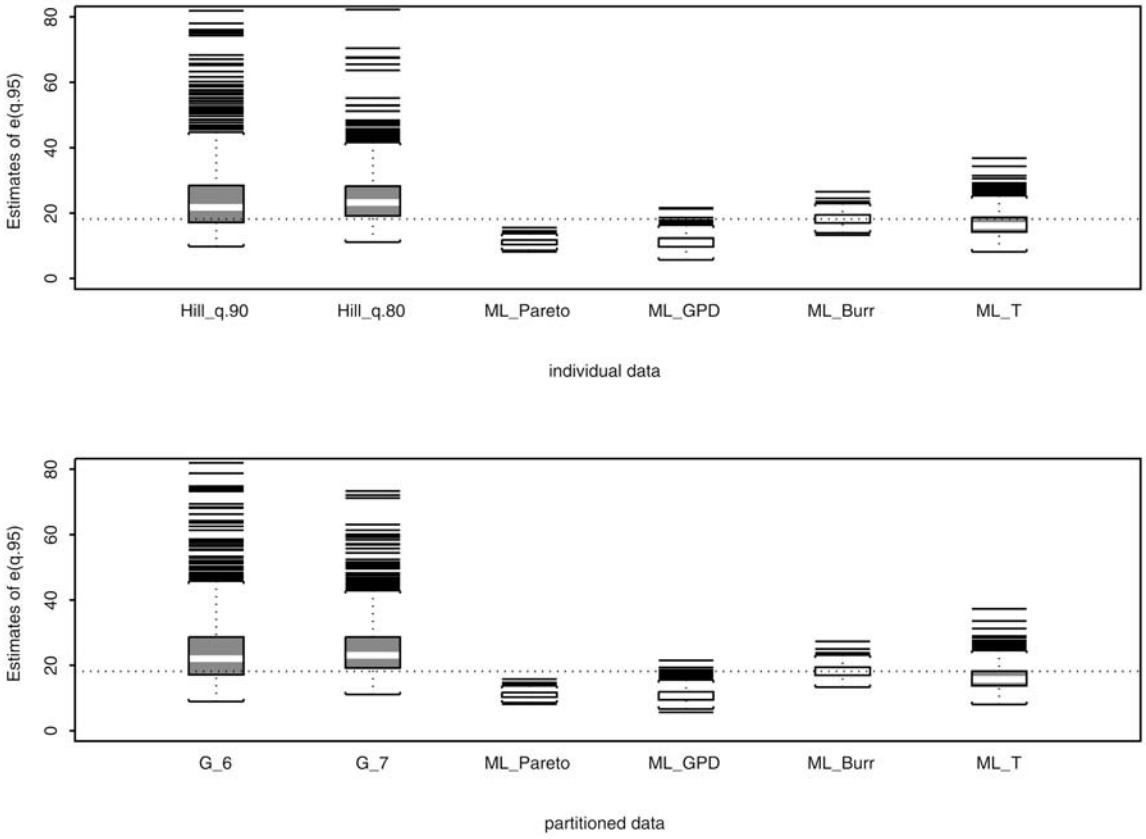
To determine the threshold above which to fit the Pareto tail and estimate the tail index, we look for a range in which the α estimates are stable. We use a plot similar to the Hill plot (see, for example, Embrechts, Klüppelberg, and Mikosch (1997) and Drees, DeHaan, and Resnick (2000)), but modify it to be applicable for partitioned losses. Under our general framework, we consider the plot

$$\{(k, G_k) : k = 2, \ldots, g\}, \tag{6.1}$$

where k is the number of top groups used to find G_k, and look for a range of k values where the plot is approximately level. This plot is given in Figure 9 for the above insurance example. Notice that the plot is roughly linear for thresholds between 500 and 1100 (see also Table 4, $5 \leq j \leq 8$). We use $a_k := 500$ ($k = 8$) as the threshold and obtain $G_k = 0.7905$. This tail index suggests no finite mean for the loss distribution.

Next, we consider some important quantities in modeling large insurance claims, such as extreme tail probabilities, extreme quantiles, and mean excess loss, given that losses are available only in partitioned form. Under the setup described in Section 3, $\bar{F}(x) = \mathrm{P}(X > x)$ can be approximated

Figure 7. Estimation of mean excess value $e(q_{.95})$. ML estimates are calculated under the assumption of the specified distributions. The true distribution F is Burr with tail index $\alpha = 1.5$ ($\lambda = 1.2$, $\theta = 4/2$, $\tau = 3/4$). The top plot uses all data, and the bottom plot uses grouped data. The Hill_$q_{.90}$ and Hill_$q_{.80}$ use all order statistics larger than $q_{.90} = F^{-1}(.9)$ and $q_{.80} = F^{-1}(.8)$. The G_6 and G_7 use the counts from top 6 and 7 intervals. Sample size = number of replications = 1000.



by

$$\hat{\bar{F}}(x) = \begin{cases} \bar{F}_n(a_k)(x/a_k)^{-G_k} & \text{if} \quad x > a_k \\ \bar{F}_n(x) & \text{if} \quad x \le a_k, \end{cases} \tag{6.2}$$

where F_n is the empirical d.f. for the losses $X_1, \ldots, X_n$. In Figure 10 this approximation is illustrated for the above Fire loss data with $x > a_k = 500$. Notice how closely the fitted tail probabilities are to the empirical tail probabilities.

Similarly, one can also approximate the conditional tail probability $\mathrm{P}(X > x \mid X > a_k)$ by $(x/a_k)^{-G_k}$. An extreme quantile of the loss distribution, q_p, is defined by the relationship $\bar{F}(q_p) = 1 - p$ where p is close to 1 (say, $F_n(a_k) < p < 1$). Setting $\hat{\bar{F}}(x)$ equal to $1 - p$ and solving for q_p in Eq. (6.2) yields the following estimate for the extreme quantile q_p:

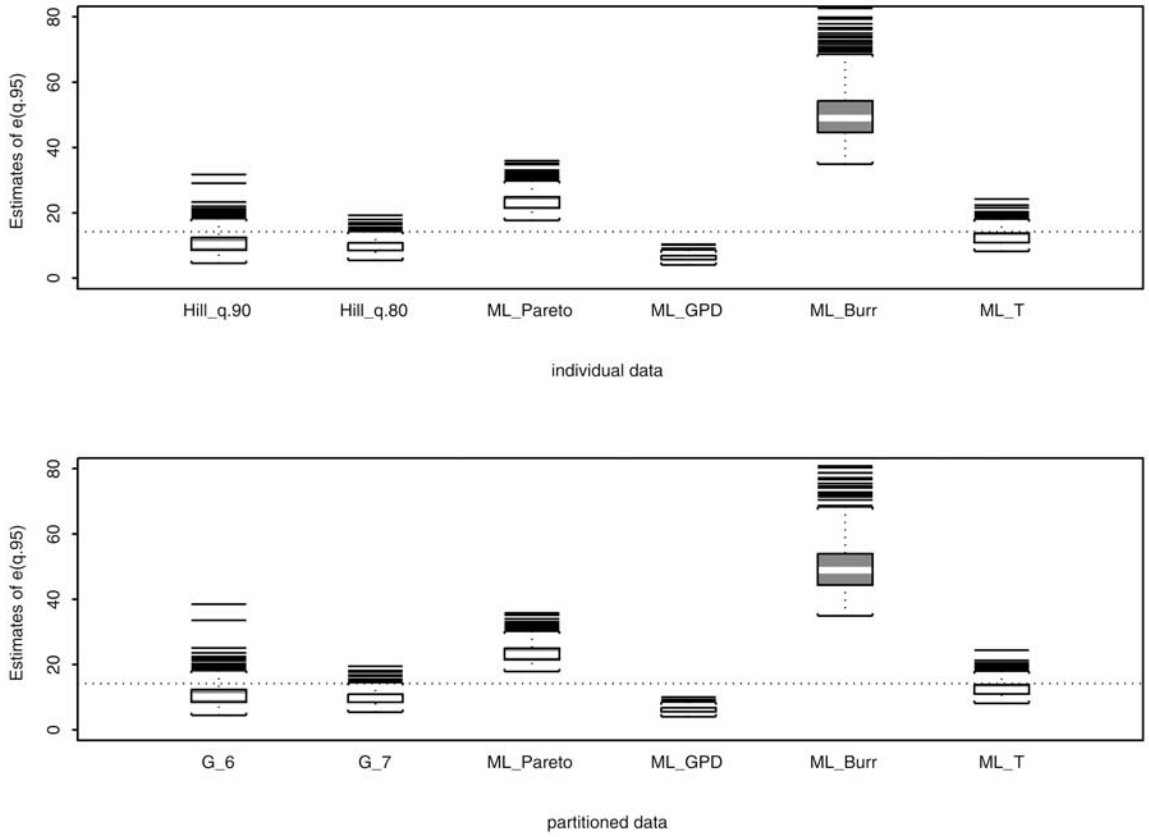
$$\hat{q}_p = a_k \left(\frac{1-p}{\bar{F}_n(a_k)} \right)^{-1/G_k}. \tag{6.3}$$

As an example, we estimate the .99 quantile to be $\hat{q}_{.99} = \$57,315$ using the above Fire loss data. The mean excess loss above a high threshold is important in premium determination and is given by $e(u) = \mathrm{E}\{X - u \mid X > u\}$. For $u > a_k$, the mean excess loss can be approximated by

$$\hat{e}(u) = \frac{u}{G_k - 1}, \tag{6.4}$$

for $G_k > 1$. In this example, however, $\hat{e}(u)$ is not available because $G_k \le 1$.

Figure 8. Estimation of mean excess value $e(q_{.95})$. ML estimates are calculated under the assumption of the specified distributions. The true distribution F is half T with tail index $\alpha = 1.5$ ($\phi = 1.5$). The top plot uses all data, and the bottom plot uses grouped data. The $\text{Hill}\_q_{.90}$ and $\text{Hill}\_q_{.80}$ use all order statistics larger than $q_{.90} = F^{-1}(.9)$ and $q_{.80} = F^{-1}(.8)$. The G_6 and G_7 use the counts from top 6 and 7 intervals. Sample size = number of replications = 1000.



7. Summary and conclusion

It has been shown that losses for many lines of insurance possess Pareto-type tails. For this reason, tail index estimation, which is a measure of the heavy-tailedness of a distribution, is an important problem for actuaries. Most estimators, however, cannot be used when loss data are available only in a partitioned form. The proposed estimator possesses the attractive features of (1) being applicable when loss data are available only in a partitioned form, and (2) being robust with respect to a large class of distributions commonly used in modeling insurance losses. We also showed that tail index estimates can be misleading if one misspecifies the distribution when trying to fit a global density. We have demonstrated that the proposed estimator compares favorably to the Hill estimator that uses individual data, and provided an example showing its effectiveness using actual insurance loss data.

Acknowledgments

The authors thank the financial support provided by the Committee on Knowledge Extension Research of the Society of Actuaries.

References

Beirlant, J., and J. L. Teugels, "Modeling Large Claims in Non-life Insurance," *Insurance: Mathematics and Economics* 11, 1992, pp. 17–29.

Beirlant, J., and A. Guillou, "Pareto Index Estimation Under Moderate Right Censoring, *Scandinavian Actuarial Journal* 2, 2001, pp. 111–125.

Table 3. Robustness of the proposed estimator against the underlying distribution, n = 1000

				RMSE			
	True $F(x)$	Hill_$q_{.90}$	Hill_$q_{.80}$	ML_Pareto	ML_GPD	ML_Burr	ML_T
Individual data	Pareto	3.72	2.70	1.15	6.79	12.08	2.36
	GPD	6.55	5.81	34.34	3.83	120.84	71.69
	Burr	8.00	7.25	7.15	7.33	1.47	3.30
	Half T	4.33	4.68	9.07	7.93	35.47	2.53
	True $F(x)$	G_6	G_7	ML_Pareto	ML_GPD	ML_Burr	ML_T
Partitioned data	Pareto	3.85	2.79	1.16	6.98	12.09	2.45
	GPD	7.04	5.88	34.87	3.93	122.17	74.32
	Burr	8.49	7.57	7.22	7.61	1.49	3.48
	Half T	4.41	4.68	9.21	8.04	35.32	2.55
				Efficiency			
	True $F(x)$	Hill_$q_{.90}$	Hill_$q_{.80}$	ML_Pareto	ML_GPD	ML_Burr	ML_T
Individual data	Pareto	3.24	2.35	1.00	5.92	10.52	2.05
	GPD	1.71	1.52	8.97	1.00	31.56	18.72
	Burr	5.44	4.94	4.87	4.99	1.00	2.25
	Half T	1.71	1.85	3.59	3.14	14.02	1.00
	True $F(x)$	G_6	G_7	ML_Pareto	ML_GPD	ML_Burr	ML_T
Partitioned data	Pareto	3.31	2.40	1.00	6.01	10.40	2.11
	GPD	1.79	1.50	8.88	1.00	31.11	18.92
	Burr	5.71	5.08	4.85	5.11	1.00	2.34
	Half T	1.73	1.84	3.62	3.16	13.87	1.00

Table 4. Homeowners physical damage

			Fire		
j	a_j	a_{j-1}	$1-F_n(a_j)^a$	$\bar{x}_j{}^b$	$\hat{\alpha}_j{}^c$
1	50100	∞	1.21	78278	NA
2	25100	50100	3.03	35486	1.3286
3	10100	25100	5.83	16419	0.8779
4	5100	10100	9.00	7135	0.759
5	1100	5100	30.85	2256	0.7902
6	850	1100	37.99	974	0.7938
7	600	850	49.65	715	0.7873
8	500	600	57.55	555	0.7905
9	400	500	66.68	452	0.7684
10	350	400	71.91	378	0.7478
11	300	350	77.70	328	0.7203
12	250	300	83.69	278	0.6812
13	211	250	88.64	233	0.6435
14	200	211	89.90	207	0.6303
15	175	200	93.46	191	0.6026
16	156	175	95.61	167	0.5753
17	150	156	96.11	154	0.5653
18	125	150	98.92	141	0.5258
19	100	125	100.00	117	0.4743
			n = 7534		

[a] Proportion of losses observed greater than a_j (given as a percentage).
[b] Average of losses between a_j and a_{j-1}.
[c] Estimator given in Eq. (3.2) using $k = j$.

Beirlant, J., G. Matthys, and G. Dierckx, "Heavy Tailed Distributions and Rating," *ASTIN Bulletin* 31, 2001, pp. 37–58.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels, *Statistics of Extremes: Theory and Applications*, Hoboken, NJ: Wiley, 2004.

Brazauskas, V., and R. Serfling, "Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution," *North American Actuarial Journal* 4 (4), 2000, pp. 12–27.

Cebrián, A., M. Denuit, and P. Lambert, "Generalized Pareto Fit to the Society of Actuaries' Large Claims Database," *North American Actuarial Journal* 7 (3), 2003, pp. 18–36.

Dekkers, A. L. M., and L. De Haan, "On the Estimation of the Extreme-Value Index and Large Quantile Estimation," *The Annals of Statistics* 17, 1989, pp. 1795–1832.

Dekkers, A. L. M., and L. De Haan, "Optimal Choice of Sample Fraction in Extreme Value Estimation," *Journal of Multivariate Analysis* 47, 1993, pp. 173–195.

Drees, H., "On Smooth Statistical Tail Functionals," *Scandinavian Journal of Statistics* 25, 1998, pp. 187–210.

Drees, H., L. de Haan, and S. Resnick, "How to Make a Hill Plot," *Annals of Statistics* 28, 2000, pp. 254–274.

Dupuis, D. J., "Exceedances Over High Thresholds: A Guide to Threshold Selection," *Extremes* 1, 1999, pp. 251–261.

Embrechts, P., C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, New York: Springer, 1997.

Figure 9. Tail index estimation for fire loss data. The estimates for α using Eq. (3.2) are stable in the range $5 \leq k \leq 8$. This suggests to choose the cutoff $a_8 = 500$ as the threshold and to use the observed counts in top 8 intervals in Eq. (3.2).

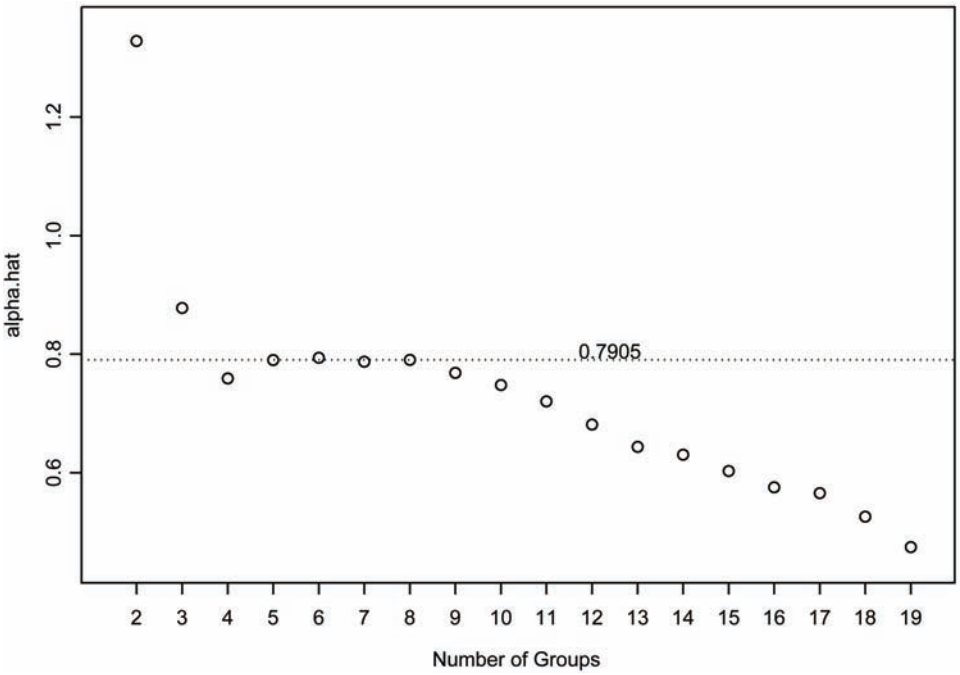
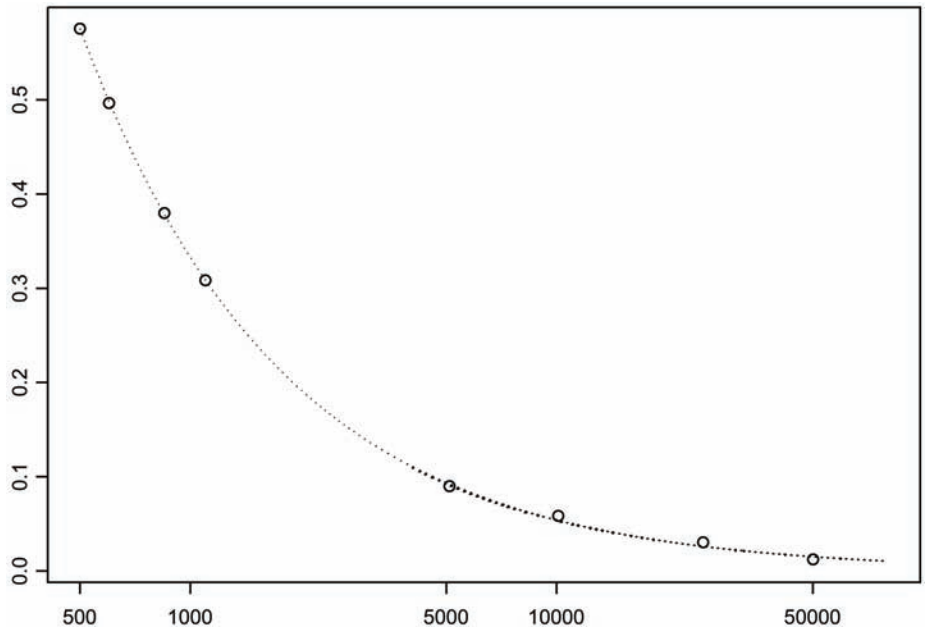


Figure 10. Comparison of empirical and fitted tail probabilities for fire loss data. $\bar{F}_n(x)$ is given by open circles and $\hat{\bar{F}}(x)$ by the dashed line where $\hat{\alpha} = 0.7905$ and $a_k = 500$. Note that the x axis is on log scale.



Embrechts, P., S. I. Resnick, and G. Samorodnitsky, "Extreme Value Theory as a Risk Management Tool," *North American Actuarial Journal* 3 (2), 1999, pp. 30–41.

Finkelstein, M., H. G. Tucker, and J. A. Veeh, "Pareto Tail Index Estimation Revisited," *North American Actuarial Journal* 10 (1), 2006, pp. 1–10.

Fraga Alves, M. I., "A Location Invariant Hill-Type Estimator," *Extremes* 4, 2001, pp. 199–217.

Hall, P., "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems," *Journal of Multivariate Analysis* 32, 1990, pp. 177–203.

Hill, B. M., "A Simple General Approach to Inference About the Tail of a Distribution," *Annals of Statistics* 3, 1975, pp. 1163–1174.

Hogg, R. V., and S. A. Klugman, *Loss Distributions*, New York: Wiley, 1984.

Hsieh, P.-H., "Robustness of Tail Index Estimation," *Journal of Computational and Graphical Statistics* 8, 1999, pp. 318–332.

Hsieh, P.-H., "An Exploratory First Step in Teletraffic Data Modeling: Evaluation of Long-Run Performance of Parameter Estimators," *Computational Statistics and Data Analysis* 40, 2002, pp. 263–283.

Matthys, G., E. Delafosse, A. Guillou, and Beirlant, J., "Estimating Catastrophic Quantile Levels for Heavy-Tailed Distributions," *Insurance: Mathematics and Economics* 34, 2004, pp. 517–537.

McNeil, A. J., "Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory," *ASTIN Bulletin* 27, 1997, pp. 117–138.

Pickands, J. III, "Statistical Inference Using Extreme Order Statistics," *Annals of Statistics* 3, 1975, pp. 119–131.

Segers, J., "Generalized Pickands Estimators for the Extreme Value Index," *Journal of Statistical Planning and Inference* 128, 2005, pp. 381–396.

Appendix A

Table 5. Loss of efficiency with the use of partitioned data, $n = 500$

Threshold D used in the Hill estimator	$q_{.99}$	$q_{.98}$	$q_{.975}$	$q_{.95}$	$q_{.90}$	$q_{.80}$	$q_{.70}$	$q_{.60}$	$q_{.50}$	$q_{.40}$	$q_{.30}$	$q_{.20}$	$q_{.10}$	$q_{.00}$
No. of top intervals k used in G_k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
							True distribution: Pareto							
Cutoff D	**21.54**	**13.57**	**11.7**	**7.37**	**4.64**	**2.92**	**2.23**	**1.84**	**1.59**	**1.41**	**1.27**	**1.16**	**1.07**	**1**
Hill	8.50	0.81	0.62	0.35	0.23	0.15	0.12	0.11	0.09	0.09	0.08	0.07	0.07	0.07
G_k	11.97	3.74	0.66	0.37	0.24	0.16	0.13	0.11	0.10	0.09	0.08	0.08	0.07	0.07
Efficiency	1.41	4.62	1.07	1.05	1.04	1.03	1.03	1.02	1.02	1.01	1.01	1.01	1.01	1.01
							True distribution: generalized Pareto							
Cutoff D	**31.82**	**19.86**	**17.04**	**10.55**	**6.46**	**3.89**	**2.85**	**2.26**	**1.88**	**1.61**	**1.4**	**1.24**	**1.11**	**1**
Hill	1.87	0.87	0.56	0.33	0.22	0.17	0.17	0.18	0.21	0.23	0.25	0.27	0.29	0.31
G_k	11.02	2.85	0.60	0.36	0.23	0.18	0.17	0.19	0.21	0.23	0.25	0.27	0.30	0.32
Efficiency	5.90	3.27	1.07	1.10	1.07	1.03	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
							True distribution: Burr							
Cutoff D	**24.87**	**15.12**	**12.86**	**7.7**	**4.57**	**2.69**	**1.99**	**1.62**	**1.39**	**1.25**	**1.14**	**1.07**	**1.03**	**1**
Hill	3.51	0.61	0.49	0.31	0.24	0.21	0.19	0.17	0.15	0.12	0.09	0.09	0.13	0.22
G_k	11.36	2.42	0.57	0.33	0.25	0.22	0.20	0.18	0.15	0.12	0.09	0.09	0.13	0.22
Efficiency	3.24	3.99	1.17	1.08	1.05	1.03	1.02	1.01	1.01	1.01	1.01	1.02	1.03	1.02
							True distribution: Half T							
Cutoff D	**18.82**	**12.2**	**10.64**	**7.02**	**4.71**	**3.2**	**2.55**	**2.15**	**1.87**	**1.65**	**1.47**	**1.3**	**1.15**	**1**
Hill	1.76	1.02	0.68	0.42	0.34	0.32	0.31	0.28	0.24	0.19	0.13	0.07	0.10	0.20
G_k	10.18	4.25	0.73	0.44	0.35	0.32	0.31	0.29	0.24	0.19	0.12	0.07	0.10	0.20
Efficiency	5.78	4.17	1.07	1.05	1.03	1.01	1.01	1.01	1.01	1.00	1.00	1.00	1.01	1.01

Table 6. Loss of efficiency with the use of partitioned data, $n = 250$

Threshold D used in the Hill estimator	$q_{.99}$	$q_{.98}$	$q_{.975}$	$q_{.95}$	$q_{.90}$	$q_{.80}$	$q_{.70}$	$q_{.60}$	$q_{.50}$	$q_{.40}$	$q_{.30}$	$q_{.20}$	$q_{.10}$	$q_{.00}$
No. of top intervals k used in G_k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
							True distribution: Pareto							
Cutoff D	**21.54**	**13.57**	**11.7**	**7.37**	**4.64**	**2.92**	**2.23**	**1.84**	**1.59**	**1.41**	**1.27**	**1.16**	**1.07**	**1**
Hill	13.75	2.87	1.53	0.63	0.35	0.23	0.18	0.15	0.14	0.13	0.12	0.11	0.10	0.10
G_k	18.42	11.82	8.83	1.50	0.37	0.23	0.19	0.16	0.14	0.13	0.12	0.11	0.10	0.10
Efficiency	1.34	4.13	5.78	2.38	1.05	1.03	1.02	1.02	1.01	1.02	1.01	1.01	1.01	1.01
							True distribution: generalized Pareto							
Cutoff D	**31.82**	**19.86**	**17.04**	**10.55**	**6.46**	**3.89**	**2.85**	**2.26**	**1.88**	**1.61**	**1.4**	**1.24**	**1.11**	**1**
Hill	259.25	3.04	1.41	0.71	0.32	0.21	0.19	0.20	0.21	0.23	0.25	0.27	0.30	0.32
G_k	17.83	10.84	9.03	2.16	0.34	0.22	0.19	0.20	0.22	0.24	0.25	0.27	0.30	0.32
Efficiency	0.07	3.57	6.40	3.03	1.08	1.04	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
							True distribution: Burr							
Cutoff D	**24.87**	**15.12**	**12.86**	**7.7**	**4.57**	**2.69**	**1.99**	**1.62**	**1.39**	**1.25**	**1.14**	**1.07**	**1.03**	**1**
Hill	20.05	10.45	1.17	0.47	0.30	0.23	0.21	0.19	0.17	0.14	0.12	0.12	0.16	0.24
G_k	17.48	10.80	6.88	1.59	0.32	0.24	0.22	0.19	0.17	0.15	0.12	0.12	0.16	0.25
Efficiency	0.87	1.03	5.86	3.38	1.07	1.04	1.02	1.02	1.02	1.01	1.01	1.01	1.02	1.02
							True distribution: Half T							
Cutoff D	**18.82**	**12.2**	**10.64**	**7.02**	**4.71**	**3.2**	**2.55**	**2.15**	**1.87**	**1.65**	**1.47**	**1.3**	**1.15**	**1**
Hill	15.17	29.00	1.40	0.63	0.46	0.37	0.33	0.30	0.26	0.21	0.15	0.10	0.12	0.21
G_k	18.07	11.26	8.05	1.54	0.48	0.37	0.34	0.30	0.26	0.21	0.15	0.10	0.12	0.21
Efficiency	1.19	0.39	5.77	2.44	1.05	1.02	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Table 7. Loss of efficiency with the use of partitioned data, $n = 100$

Threshold D used in the Hill estimator	$q_{.99}$	$q_{.98}$	$q_{.975}$	$q_{.95}$	$q_{.90}$	$q_{.80}$	$q_{.70}$	$q_{.60}$	$q_{.50}$	$q_{.40}$	$q_{.30}$	$q_{.20}$	$q_{.10}$	$q_{.00}$
No. of top intervals k used in G_k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
							True distribution: Pareto							
Cutoff D	**21.54**	**13.57**	**11.7**	**7.37**	**4.64**	**2.92**	**2.23**	**1.84**	**1.59**	**1.41**	**1.27**	**1.16**	**1.07**	**1**
Hill	23.57	12.26	316.23	5.30	0.72	0.38	0.30	0.25	0.22	0.20	0.18	0.17	0.16	0.15
G_k	23.72	19.80	26.88	9.91	3.33	0.40	0.30	0.26	0.23	0.20	0.18	0.17	0.16	0.15
Efficiency	1.01	1.61	0.08	1.87	4.63	1.03	1.02	1.02	1.02	1.02	1.01	1.01	1.01	1.01
							True distribution: generalized Pareto							
Cutoff D	**31.82**	**19.86**	**17.04**	**10.55**	**6.46**	**3.89**	**2.85**	**2.26**	**1.88**	**1.61**	**1.4**	**1.24**	**1.11**	**1**
Hill	290.38	15.53	10.81	3.05	0.79	0.35	0.27	0.25	0.25	0.25	0.27	0.28	0.30	0.32
G_k	22.92	19.10	24.32	10.50	2.71	0.37	0.28	0.25	0.25	0.26	0.27	0.28	0.30	0.32
Efficiency	0.08	1.23	2.25	3.45	3.44	1.07	1.02	1.02	1.01	1.01	1.01	1.01	1.01	1.01
							True distribution: Burr							
Cutoff D	**24.87**	**15.12**	**12.86**	**7.7**	**4.57**	**2.69**	**1.99**	**1.62**	**1.39**	**1.25**	**1.14**	**1.07**	**1.03**	**1**
Hill	81.77	13.49	9.19	2.98	0.58	0.35	0.29	0.26	0.23	0.21	0.19	0.19	0.22	0.29
G_k	22.17	19.22	27.52	9.24	2.33	0.37	0.30	0.26	0.23	0.21	0.20	0.19	0.22	0.30
Efficiency	0.27	1.42	2.99	3.10	3.98	1.05	1.03	1.02	1.01	1.01	1.01	1.01	1.02	1.02
							True distribution: Half T							
Cutoff D	**18.82**	**12.2**	**10.64**	**7.02**	**4.71**	**3.2**	**2.55**	**2.15**	**1.87**	**1.65**	**1.47**	**1.3**	**1.15**	**1**
Hill	88.80	38.79	42.35	4.35	0.91	0.59	0.48	0.40	0.34	0.27	0.21	0.15	0.14	0.21
G_k	25.75	21.83	32.76	14.40	3.70	0.61	0.49	0.41	0.35	0.27	0.21	0.15	0.15	0.22
Efficiency	0.29	0.56	0.77	3.31	4.09	1.03	1.02	1.02	1.02	1.01	1.01	1.01	1.01	1.01

Appendix B

Table 8. Robustness of the proposed estimator against the underlying distribution, $n = 500$

				RMSE			
	True $F(x)$	Hill_$q_{.90}$	Hill_$q_{.80}$	ML_Pareto	ML_GPD	ML_Burr	ML_T
Individual data	Pareto	5.01	3.52	1.62	6.92	12.31	3.09
	GPD	10.20	8.26	34.90	5.49	126.22	77.25
	Burr	11.05	9.99	7.15	7.42	2.06	4.43
	Half T	5.07	4.95	9.43	7.99	36.51	3.09
	True $F(x)$	G_6	G_7	ML_Pareto	ML_GPD	ML_Burr	ML_T
Partitioned data	Pareto	5.33	3.60	1.63	7.06	12.34	3.20
	GPD	11.13	8.78	35.57	5.52	128.47	78.05
	Burr	11.65	10.58	7.23	7.67	2.07	4.59
	Half T	5.19	4.99	9.58	8.09	36.39	3.10
				Efficiency			
	True $F(x)$	Hill_$q_{.90}$	Hill_$q_{.80}$	ML_Pareto	ML_GPD	ML_Burr	ML_T
Individual data	Pareto	3.10	2.18	1.00	4.27	7.61	1.91
	GPD	1.86	1.50	6.35	1.00	22.97	14.06
	Burr	5.36	4.85	3.47	3.60	1.00	2.15
	Half T	1.64	1.60	3.05	2.59	11.82	1.00
	True $F(x)$	G_6	G_7	ML_Pareto	ML_GPD	ML_Burr	ML_T
Partitioned data	Pareto	3.27	2.20	1.00	4.33	7.56	1.96
	GPD	2.02	1.59	6.45	1.00	23.29	14.15
	Burr	5.63	5.11	3.49	3.71	1.00	2.22
	Half T	1.67	1.61	3.09	2.61	11.73	1.00

Table 9. Robustness of the proposed estimator against the underlying distribution, $n = 250$

				RMSE			
	True $F(x)$	**Hill**$_{q.90}$	**Hill**$_{q.80}$	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Individual data	**Pareto**	7.43	5.27	2.35	7.04	12.76	4.40
	GPD	13.39	10.67	38.21	7.08	137.87	79.16
	Burr	13.97	11.94	7.19	8.00	2.90	6.01
	Half T	5.98	5.14	9.98	7.89	38.17	3.92
	True $F(x)$	G_6	G_7	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Partitioned data	**Pareto**	8.02	5.49	2.37	7.16	12.80	4.48
	GPD	14.26	10.99	38.79	7.17	139.55	80.78
	Burr	14.12	12.85	7.25	8.19	2.92	6.14
	Half T	6.18	5.24	10.11	8.00	37.97	3.95
				Efficiency			
	True $F(x)$	**Hill**$_{q.90}$	**Hill**$_{q.80}$	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Individual data	**Pareto**	3.17	2.25	1.00	3.00	5.44	1.87
	GPD	1.89	1.51	5.40	1.00	19.47	11.18
	Burr	4.81	4.12	2.48	2.76	1.00	2.07
	Half T	1.53	1.31	2.54	2.01	9.73	1.00
	True $F(x)$	G_6	G_7	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Partitioned data	**Pareto**	3.38	2.31	1.00	3.02	5.40	1.89
	GPD	1.99	1.53	5.41	1.00	19.46	11.26
	Burr	4.84	4.40	2.49	2.81	1.00	2.11
	Half T	1.56	1.33	2.56	2.02	9.61	1.00

Table 10. Robustness of the proposed estimator against the underlying distribution, $n = 100$

				RMSE			
	True $F(x)$	**Hill$_{.90}$**	**Hill$_{.80}$**	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Individual data	**Pareto**	11.02	7.98	3.59	7.38	13.96	6.72
	GPD	16.94	17.26	50.89	11.52	152.79	69.69
	Burr	14.85	16.20	7.25	9.21	4.66	9.78
	Half T	8.10	6.73	11.06	8.25	41.89	5.63
	True $F(x)$	G_6	G_7	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Partitioned data	**Pareto**	11.47	8.32	3.61	7.48	14.00	6.77
	GPD	17.38	15.89	51.00	12.04	149.07	69.94
	Burr	14.82	16.36	7.33	9.31	4.69	9.89
	Half T	8.31	6.85	11.32	8.29	42.15	5.71
				Efficiency			
	True $F(x)$	**Hill$_{.90}$**	**Hill$_{.80}$**	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Individual data	**Pareto**	3.07	2.22	1.00	2.06	3.89	1.87
	GPD	1.47	1.50	4.42	1.00	13.26	6.05
	Burr	3.19	3.48	1.56	1.98	1.00	2.10
	Half T	1.44	1.19	1.96	1.46	7.44	1.00
	True $F(x)$	G_6	G_7	**ML_Pareto**	**ML_GPD**	**ML_Burr**	**ML_T**
Partitioned data	**Pareto**	3.17	2.30	1.00	2.07	3.87	1.87
	GPD	1.44	1.32	4.24	1.00	12.38	5.81
	Burr	3.16	3.49	1.56	1.99	1.00	2.11
	Half T	1.46	1.20	1.98	1.45	7.38	1.00

Appendix C

Table 11. Bias in estimating $e(q_{.95})$

	Individual Data								Grouped Data			
	Hill		ML Estimator				Proposed		ML Estimator			
N	$q_{.90}$	$q_{.80}$	Pareto	GPD	Burr	halfT	G_6	G_7	Pareto	GPD	Burr	halfT
						True $F(x)$: Pareto						
100	1.11	1.33	0.21	−6.67	12.44	0.06	1.30	1.38	0.21	−6.77	12.44	−0.11
250	0.95	0.55	0.12	−6.80	12.13	−0.60	1.18	0.63	0.12	−6.95	12.14	−0.75
500	0.24	0.08	0.09	−6.82	12.02	−0.77	0.44	0.11	0.10	−6.97	12.05	−0.91
1000	0.56	0.25	0.05	−6.74	11.93	−0.76	0.60	0.23	0.06	−6.93	11.94	−0.96
						True $F(x)$: GPD						
100	−0.19	4.85	40.77	−0.77	123.27	49.34	−0.26	4.14	40.73	−0.58	120.32	48.43
250	1.99	3.71	34.97	−1.73	122.99	63.86	2.21	3.72	35.45	−1.71	123.80	64.50
500	2.19	3.65	33.40	−1.71	118.62	66.24	2.72	3.83	34.00	−1.65	120.31	67.10
1000	1.20	3.15	33.68	−1.67	117.75	65.97	1.46	3.07	34.16	−1.67	118.79	67.80
						True $F(x)$: Burr						
100	2.94	7.16	−6.80	−6.49	0.61	0.18	2.21	6.74	−6.89	−6.75	0.61	−0.26
250	5.47	6.08	−7.05	−7.34	0.16	−1.66	5.68	6.66	−7.12	−7.53	0.18	−1.99
500	5.73	6.63	−7.09	−7.15	0.06	−1.58	5.95	6.89	−7.17	−7.43	0.05	−2.00
1000	4.77	5.56	−7.12	−7.21	0.01	−1.78	5.05	5.76	−7.19	−7.50	0.01	−2.21
						True $F(x)$: Half T						
100	−3.43	−4.27	9.67	−8.01	38.50	−1.39	−3.53	−4.16	9.84	−8.05	38.45	−1.36
250	−3.00	−4.21	9.33	−7.80	36.73	−1.42	−2.90	−4.27	9.46	−7.92	36.50	−1.42
500	−3.70	−4.64	9.12	−7.95	35.86	−1.84	−3.68	−4.66	9.26	−8.05	35.70	−1.81
1000	−3.72	−4.53	8.92	−7.91	35.15	−1.85	−3.73	−4.53	9.05	−8.01	34.98	−1.82