Robust and Efficient Estimation of the Tail Index of a One-Parameter Pareto Distribution

Vytaras Brazauskas[1]
University of Texas at Dallas

and

Robert Serfling[2]
University of Texas at Dallas

March, 1999

[1]Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA.

[2]Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Supported by grants from the Casualty Actuarial Society and Society of Actuaries, with administrative support from the Actuarial Education and Research Fund, and by NSF Grant DMS-9705209.

Abstract

Estimation of the tail index parameter of a one-parameter Pareto model has wide application in actuarial science. Here we introduce a new estimator of "generalized median" type and compare it with several well-established estimators associated with the methods of maximum likelihood, moments, trimming, least squares, quantiles, and percentile matching. Estimators are evaluated on the basis of two competing criteria, efficiency and robustness. The maximum likelihood estimator being efficient but nonrobust, we desire alternative estimators which retain a high degree of efficiency while also being robust. The method of moments and least squares estimators are found to be relatively deficient with respect to both criteria and should become disfavored, while the trimmed mean and generalized median estimators tend to dominate the other competitors, with the generalized median type performing best overall. These findings provide a basis for revision and updating of prevailing viewpoints. Applications to robust estimation of upper quantiles, tail probabilities, and actuarial quantities such as stop-loss and excess-of-loss reinsurance premiums that arise concerning solvency of portfolios, are discussed. Robust parametric methods are compared with (nonrobust) empirical nonparametric methods.

AMS 1991 Subject Classification: Primary 62F35, Secondary 62P05.

Key words and phrases: Pareto model; tail index; robust estimation; asymptotic relative efficiency; breakdown point; reinsurance.

1 Introduction

A useful and tractable parametric model with relatively high probability in the upper tail is the Pareto distribution $\mathrm{P}(\sigma, \alpha)$ having cdf

$$F(x) = 1 - \left(\frac{\sigma}{x}\right)^{\alpha}, \quad x \geq \sigma, \tag{1.1}$$

defined for $\alpha > 0$ and $\sigma > 0$. We treat estimation of the shape parameter α characterizing the tail, with the scale parameter σ assumed known. That is, we consider (1.1) as a "one-parameter" Pareto model, following Bierlant, Teugels, and Vynckier (1996) and Klugman, Panjer, and Willmot (1998). In actuarial applications, this model is appropriate when losses or claims below a certain level are not relevant, as for example when a deductible applies. In such a case, σ represents the deductible, or sometimes a lesser value in order to incorporate inflation into the model while ignoring data irrelevant to the issues under study. Besides its intrinsic interest, the parameter α plays a key role in connection with determination of extreme quantiles, upper tail probabilities, mean excess functions, and excess-of-loss and stop-loss reinsurance premiums, for example. Small relative errors in estimation of α can produce large relative errors in estimation of such quantities. More broadly, in the context of *semiparametric* modeling where one assumes merely a "Pareto-type" distribution, an *approximate* model for the *upper* observations of a sample is given by (1.1). This is apropos to situations where the upper tail of the model is not parametrically related to the lower and central parts and thus is estimated separately using only the uppermost observations. These various applications are discussed in detail in Section 6. For general overviews of the role of $\mathrm{P}(\sigma, \alpha)$ and variants in actuarial science, econometrics, and other fields, see Arnold (1983) and Johnson, Kotz, and Balakrishnan (1994), Chapter 20. New application contexts continue to arise: for example, the cost distributions of combinatorial search algorithms have recently been shown to exhibit Pareto-type tail behavior (see Gomes, Selman, and Crato (1997)).

For evaluation and comparison of estimators we shall employ two criteria, *efficiency* and *robustness*. More precisely, we use the *asymptotic relative efficiency* (ARE) of the estimator taken with respect to the maximum likelihood estimator (MLE), and the *breakdown point* (BP) of the estimator. These concepts, which are well-established and widely used in the general statistical literature, are defined and discussed precisely in Section 2. As the MLE is seen to be efficient but *nonrobust* (BP = 0), we seek competitors which have BP > 0 combined with high ARE.

In Section 3 we introduce a new "*generalized median*" type of estimator and investigate its performance with respect to our two-fold criteria. In Section 4 we similarly assess several well-established approaches in current use, namely those estimators corresponding to the methods of *moments*, *trimming*, *least squares*, *quantiles*, and *percentile matching*. Comparisons of all the estimators are presented in Section 5, with the purpose of identifing not only the superior estimators, but also the inferior ones which might be reasonably discarded from practical use or at least used more judiciously. We arrive at the following general conclusions. The percentile matching and trimmed mean type estimators tend to dominate, simultaneously with respect to both efficiency and robustness, the method of moments and least squares type competitors to the MLE. In particular, for the least squares type estimator, which has enjoyed long popularity in the Pareto problem literature, we arrive at a clarified perspective: it is both *nonrobust* and *nonefficient* and reasonably may be discarded. Also, the method of moments estimator is *nonrobust* and, except for relatively large values of α, *nonefficient*. It too may be discarded. The percentile matching and trimmed mean estimators, however, are dominated by the generalized median types, although the trimmed mean types are closely competitive and should remain in contention. Overall, it turns out that for the problem of estimation of α in $\mathrm{P}(\sigma, \alpha)$, the typically used maximum likelihood, method of moments, least squares, and percentile matching estimators can be improved upon by the trimmed mean type estimators and the new generalized median type estimator.

The "*trimmed mean*" estimator that we consider for estimation of α in $\mathrm{P}(\sigma, \alpha)$ is not well-known for this purpose but, rather, is a simple transform of a well-established estimator for estimation of θ in the *exponential* distribution $\mathrm{E}(\mu, \theta)$ having cdf

$$G(z) = 1 - e^{-(z-\mu)/\theta}, \quad z \geq \mu, \tag{1.2}$$

for $\theta > 0$ and $-\infty < \mu < \infty$. Note that a random variable X has distribution F given by (1.1) if

$$X \stackrel{\mathrm{d}}{=} \sigma\, e^{U/\alpha}, \tag{1.3}$$

where "$\stackrel{\mathrm{d}}{=}$" denotes "equal in distribution" and U is "standard exponential", i.e., $\mathrm{E}(0, 1)$. Equivalently, $Z = \log X$ has cdf (1.2) and satisfies

$$Z \stackrel{\mathrm{d}}{=} \mu + \theta\, U, \tag{1.4}$$

with $\mu = \log \sigma$ and $\theta = \alpha^{-1}$. Thus the problem of estimation of the scale and shape parameters σ and α in the model $\mathrm{P}(\sigma, \alpha)$ is equivalent, through logarithmic transformation of the data, to that of estimation of the location and scale parameters μ and θ in the model $\mathrm{E}(\mu, \theta)$. In the latter problem, certain *trimmed mean* type estimators of θ have become well-established as presenting a favorable trade-off between efficiency and robustness (see, for example, Kimber (1983a,b), Gather (1986), and Willemain *et al.* (1992)). The present study includes, therefore, the estimators $\hat{\alpha}$ defined by taking the reciprocal of trimmed mean type estimators $\hat{\theta}$ of θ. It is of interest that estimators of α defined through exploitation of this equivalence appear not to have received routine consideration in the literature on the Pareto problem.

We may view the maximum likelihood estimator (MLE) as a special case of *minimum distance* or "optimization-based"estimators as described, for example, in Klugman, Panjer, and Willmot (1998), §2.5. Among these the MLE is the most efficient and enjoys other advantages as well, so that the other members of this class become considered only to the extent that they offer more flexibility in their mathematical treatment. Besides being dominated by the MLE with respect to efficiency, they typically are inadequate also with respect to robustness. Therefore, among such estimators we consider here only the MLE.

Estimation for the "two-parameter" version of (1.1), i.e., with σ *unknown*, is briefly discussed in Section 7. This case is treated fully in Brazauskas and Serfling (1999), where also is found more detailed statistical theory underlying the results of the present paper.

2 Efficiency versus Robustness, and Two Criteria

For typical parametric models, the MLE proves highly *efficient*: for large sample size n, it attains in its approximating normal distribution the *minimum* possible variance among a large class of competing estimators. Typically, however, it is *nonrobust*: in the presence of departures of the actual data from the assumed parametric model, for example if the sample includes unrepresentative "outliers", the performance of the MLE degrades severely. One thus seeks to replace the MLE by a competitor which trades off some degree of efficiency in return for a favorable degree of robustness. That is, we desire an estimator which maintains satisfactorily high performance over a specified range of departures from the "ideal" model, while also being not too less efficient than the MLE when the "ideal" model is indeed fully accurate. Let us now introduce some precise notions of efficiency and robustness criteria.

2.1 Efficiency Criterion: Asymptotic Relative Efficiency

In terms of its optimum asymptotic variance, the MLE provides a quantitative benchmark for efficiency considerations. In particular, for a sample $X_1, \ldots, X_n$ from the model $\mathrm{P}(\sigma, \alpha)$ as described by (1.1), the MLE of α is readily derived (e.g., Arnold (1983)),

$$\hat{\alpha}_{\mathrm{ML}} = \frac{1}{n^{-1} \sum_{i=1}^{n} \log X_i \; - \; \log \sigma},$$

and its exact distribution theory is described by the statement that

$$\frac{2n\,\alpha}{\hat{\alpha}_{\mathrm{ML}}} \quad \text{has cdf} \quad \chi^2_{2n}, \tag{2.1}$$

where χ^2_ν denotes the chi-square distribution with ν degrees of freedom. This yields easily the asymptotic distribution: $\hat{\alpha}_{\mathrm{ML}}$ *is asymptotically normal with mean* α *and variance* α^2/n, denoted $\mathrm{AN}(\alpha, \alpha^2/n)$, by which is meant that

$$\frac{n^{1/2}(\hat{\alpha}_{\mathrm{ML}} - \alpha)}{\alpha} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1),$$

where "$\xrightarrow{\mathrm{d}}$" denotes "converges in distribution" and $\mathrm{N}(0,1)$ denotes the standard normal distribution.

For a competing estimator, efficiency is characterized in terms of *asymptotic relative efficiency* (ARE) with respect to the MLE, defined as the limiting ratio of respective sample sizes at which the two estimators perform equivalently with respect to the variance criterion.

In particular, the estimators $\hat{\alpha}$ for α that we shall consider will in every case be $\mathrm{AN}(\alpha, c\,\alpha^2/n)$ for some constant $c > 0$, from which it follows that

$$\mathrm{ARE}(\hat{\alpha}, \hat{\alpha}_{\mathrm{ML}}) = \frac{1}{c}.$$

Note that the ARE provides a *large-sample* index of comparison whose numerical value is not expected to apply precisely for any fixed small or moderate sample size n. Such an index permits estimators which perform relatively strongly when an ample amount of data is provided to be distinguished from those which do not. The weaker estimators may then be eliminated from further consideration, while the stronger estimators may be further compared using additional criteria of choice and through simulation studies for selected fixed sample sizes.

2.2 Robustness Criterion: Breakdown Point

A popular and effective criterion for robustness of an estimator is its *breakdown point* (BP), loosely characterized as the largest proportion of sample observations which themselves may be corrupted without the estimator itself becoming corrupted. When the BP is well-defined as a quantity not depending on the particular sample values but only on the sample size n, then we typically take as our criterion its limit value as $n \to \infty$. The BP measures the degree to which the estimator remains uninfluenced by the presence of outlying observations which could possibly be due to contamination of the dataset rather than being properly representative of the target parametric model. Depending on the context, protection against upper and lower contamination may differ in importance, so we define separate versions:

> *Lower (Upper) Breakdown Point*: the largest proportion of *lower* (*upper*) sample observations which may be taken to a lower (an upper) limit without taking the estimator to a limit not depending on the parameter being estimated.

Clearly, estimators which have *nonzero* breakdown points while possessing relatively high efficiency are desired.

In the present context of estimation of α in $\mathrm{P}(\sigma, \alpha)$ with σ known, the most extreme form of *lower* corruption is to take observations to the lower limit σ, and this does not greatly impact estimation of α. On the other hand, the most extreme form of *upper* contamination is to take observations to ∞, which does indeed greatly impact the estimation of α. Consequently, we shall be concerned here primarily with UBP. (In the case σ *unknown*, however, lower contamination can include taking observations to 0, which does seriously impact estimation of α, so that LBP becomes considered. See Section 7.)

Let us examine the MLE with respect to BP. By the classical law of large numbers of probability theory, as $n \to \infty$ we have

$$n^{-1} \sum_{i=1}^{n} \log X_i \xrightarrow{\text{a.s.}} E(\log X) = \alpha^{-1} + \log \sigma,$$

whence follows the reassuring property that $\hat{\alpha}_{\rm ML} \xrightarrow{\text{a.s.}} \alpha$. On the other hand, for any fixed n, if even a single X_i is taken to ∞, then $n^{-1}\sum_{i=1}^n \log X_i \to \infty$ and consequently $\hat{\alpha}_{\rm ML} \to 0$. That is, corruption of a single data value by upper contamination can render the estimator completely uninformative. Thus $\hat{\alpha}_{\rm ML}$ has UBP $= 0$ and hence is *nonrobust* against upper outliers. On this basis, we reject the MLE as a contender for estimation of α.

The notion of breakdown point has antecedents in Hodges (1967) and Hampel (1971) but became widely popularized beginning with Donoho and Huber (1983). A related method for studying robustness of an estimator is to evaluate its performance when the data comes not from a presupposed "ideal" model G but rather from a specified "contamination model". I.e., the data is assumed to have distribution of form $F = (1-\varepsilon)G + \varepsilon H$, where H belongs to some given class of possible contaminating distributions, and ε represents the probability that a sample observation comes from the distribution H instead of G. The special advantage of using BP as criterion, however, is that it provides a robustness index valid over a *broad* and *nonspecific* range of possible sources of contaminating data.

3 A Generalized Median Estimator

Consider a sample $X_1, \ldots, X_n$ from the model $\mathrm{P}(\sigma, \alpha)$ as described by (1.1). We introduce and evaluate a new estimator which in the comparisons of Section 5 will be seen to improve upon established estimators with respect to both our efficiency and robustness criteria applied simultaneously.

For given choice of integer $k \geq 2$, we introduce a "kernel"

$$h_0(x_1, \ldots, x_k) = \left(k^{-1} \sum_{j=1}^{k} \log x_j \; - \; \log \sigma \right)^{-1}$$

whose arguments are to be filled in with sample values. Note that a particular evaluation $h_0(X_{i_1}, \ldots, X_{i_k})$ is merely the MLE of α based on just the observations $X_{i_1}, \ldots, X_{i_k}$ (for $k = n$, the MLE based on the full sample). In order to modify h_0 to make it *median unbiased* for estimation of α, we use the fact that

$$(2k\alpha)\, h_0^{-1}(X_1, \ldots, X_k) \quad \text{has cdf} \quad \chi^2_{2k},$$

which follows from (2.1). Denoting by M_ν the median of χ^2_ν, it follows that the kernel

$$h(x_1, \ldots, x_k) = \frac{M_{2k}}{2k}\, h_0(x_1, \ldots, x_k)$$

is median unbiased for α, i.e., the cdf H_F of $h(X_1, \ldots, X_k)$ satisfies

$$\alpha = \text{median of } H_F = H_F^{-1}(\tfrac{1}{2}), \tag{3.1}$$

where $H^{-1}(p), 0 < p < 1$, denotes the quantile function of a cdf H. Then a natural estimator of α is generated by taking the median of the evaluations $h(X_{i_1}, \ldots, X_{i_k})$ of the kernel h over all subsets of observations taken k at a time, i.e., corresponding to all $\binom{n}{k}$ k-sets $\{i_1, \ldots, i_k\}$ of *distinct* indices from $\{1, \ldots, n\}$. This yields the "generalized median statistic"

$$\hat{\alpha}_{\mathrm{GM}} = \mathrm{Median}\{h(Z_{i_1}, \ldots, Z_{i_k})\} \tag{3.2}$$

for estimation of α. For kernel sizes $k = 2 : 10$, values of M_{2k} and the multiplicative correction factors $M_{2k}/(2k)$ needed in constructing $\hat{\alpha}_{\mathrm{GM}}$ are provided in the following table.

Table 3.1. M_{2k} and $M_{2k}/(2k)$, for $k = 2 : 10$.

	k								
	2	3	4	5	6	7	8	9	10
M_{2k}	3.3567	5.3481	7.3441	9.3418	11.3403	13.3393	15.3385	17.3379	19.3374
$M_{2k}/(2k)$	.839	.891	.918	.934	.945	.953	.959	.963	.967

3.1 Asymptotic Normality

The estimator $\hat{\alpha}_{\mathrm{GM}}$ is a special case of generalized L-statistic ("GL-statistic"), for which asymptotic normality has been established under broad conditions by Serfling (1984) and Choudhury and Serfling (1988). We have, therefore: $\hat{\alpha}_{\mathrm{GM}}$ *is asymptotically normal with mean* α *and variance* $k^2\zeta/h_F^2(\alpha)n$, where h_F denotes the density of H_F, $\zeta = \mathrm{Var}(w_h(X))$, and $w_h(x) = P\{h(x, X_1, \ldots, X_{k-1}) \leq \alpha\}$. It turns out that this variance is of form $\gamma_k \alpha^2/n$ and thus $1/\gamma_k$ represents the ARE. The following table provides these ARE's and the corresponding values of γ_k for $k = 2 : 10$.

Table 3.2. ARE($\hat{\alpha}_{\mathrm{GM}}$, $\hat{\alpha}_{\mathrm{ML}}$) and γ_k, for $k = 2 : 10$.

	k								
	2	3	4	5	6	7	8	9	10
ARE($\hat{\alpha}_{\mathrm{GM}}$, $\hat{\alpha}_{\mathrm{ML}}$)	.78	.88	.92	.94	.96	.97	.97	.98	.98
γ_k	1.280	1.141	1.088	1.061	1.044	1.035	1.028	1.023	1.019

3.2 Breakdown Points

Regarding breakdown behavior for the statistic $\hat{\alpha}_{\mathrm{GM}}$, it can be shown that

$$\mathrm{UBP} \;=\; n^{-1} \max_{1 \leq m \leq n} \left\{ m : \frac{\binom{n-m}{k}}{\binom{n}{k}} \geq \tfrac{1}{2} \right\} \longrightarrow 1 - (\tfrac{1}{2})^{1/k}, \quad n \to \infty, \tag{3.3}$$

$$\mathrm{LBP} \;=\; n^{-1} \max_{1 \leq m \leq n} \left\{ m : \frac{\binom{m}{k}}{\binom{n}{k}} \leq \tfrac{1}{2} \right\} \longrightarrow (\tfrac{1}{2})^{1/k}, \quad n \to \infty. \tag{3.4}$$

Values of the limits (3.3) and (3.4) for $k = 2 : 10$ are given in the following table.

Table 3.3. Asymptotic LBP and UBP of $\hat{\alpha}_{\mathrm{GM}}$, for $k = 2:10$.

	k								
	2	3	4	5	6	7	8	9	10
LBP	.707	.794	.841	.871	.891	.906	.917	.926	.933
UBP	.293	.206	.159	.129	.109	.094	.083	.074	.067

From examination of Tables 3.2 and 3.3, we see that ARE increases with k while BP decreases.

3.3 Computational Considerations

The computational burden of computing $\hat{\alpha}_{\mathrm{GM}}$ grows with n as $O(n^k)$, which for large n could become prohibitive. In such a case, however, one can simply *estimate* the estimator $\hat{\alpha}_{\mathrm{GM}}$ by using only the evaluations $h(X_{i_1}, \ldots, X_{i_k})$ for a random sample of size N of the $\binom{n}{k}$ k-sets $\{i_1, \ldots, i_k\}$, where $N = 10^6$ or 10^8, say. Such an approach renders the computational burden negligible while nevertheless maintaining any desired degree of numerical accuracy.

4 Review of Established Estimators

Continuing the setting of Section 3, we focus on estimation of the parameter α, based on a sample $X_1, \ldots, X_n$ having cdf F corresponding to the $\mathrm{P}(\sigma, \alpha)$ model. Denote the ordered sample values by

$$X_{n1} \leq X_{n2} \leq \cdots \leq X_{nn}$$

and the sample mean by $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$. For convenience we also use the notation

$$Z_i = \log X_i \text{ and } Z_{ni} = \log X_{ni}, \text{ for } 1 \leq i \leq n,$$

and $\overline{Z}_n = n^{-1} \sum_{i=1}^n Z_i$. Here we review the properties of the methods of moments, trimmed mean, least squares, percentile matching, and quantile based estimators.

4.1 The Method of Moments Estimator

In the classical "method of moments" approach to estimation, estimators are produced by solving equations formed by equating low order sample and population moments of suitably chosen random variables. In particular, for the model $\mathrm{P}(\sigma, \alpha)$ we utilize the formula

$$EX = \frac{\sigma\,\alpha}{\alpha - 1},$$

which is valid provided that $\alpha > 1$. The corresponding method of moments estimator is then obtained by solving the equation

$$\overline{X}_n = \frac{\sigma\,\hat{\alpha}}{\hat{\alpha} - 1},$$

yielding

$$\hat{\alpha}_{\mathrm{MM}} = \frac{\overline{X}_n}{\overline{X}_n - \sigma}$$

By the law of large numbers, as $n \to \infty$ we have $\overline{X}_n \to EX$ and thus $\hat{\alpha}_{\mathrm{MM}} \to \alpha$. However, for any fixed n, if even a single X_i is taken to ∞, then $\overline{X}_n \to \infty$ and consequently $\hat{\alpha}_{\mathrm{MM}} \to 1$. That is, the estimator can be rendered uninformative by upper corruption of even a single observation. Thus $\hat{\alpha}_{\mathrm{MM}}$ has UBP = 0 and hence is *nonrobust* against upper outliers. This suffices for rejection of this estimator, even though its behavior against lower outliers is more stable.

For the sake of a more complete comparison with the MLE, however, we also examine the ARE. Using standard central limit theory and the fact that

$$\mathrm{Var}(X) = \frac{\sigma^2\,\alpha}{(\alpha-1)^2\,(\alpha-2)},$$

which is valid provided that $\alpha > 2$, we have that $\hat{\alpha}_{\mathrm{MM}}$ is $\mathrm{AN}(\alpha, \alpha(\alpha-1)^2/(\alpha-2)n)$. It follows that the ARE of $\hat{\alpha}_{\mathrm{MM}}$ with respect to $\hat{\alpha}_{\mathrm{ML}}$ is $\alpha(\alpha-2)/(\alpha-1)^2$, which approaches 1 as $\alpha \to \infty$ but for typical values of α is poor. For example, for $2 \le \alpha \le 2.5$ we have $0 \le \mathrm{ARE} \le .56$, and for $2.5 \le \alpha \le 3$, we have $.56 \le \mathrm{ARE} \le .75$.

In conclusion, the method of moments estimator is defined only for $\alpha > 1$ and when defined exhibits neither satisfactory robustness nor satisfactory efficiency.

4.2 Trimmed Mean Estimators

For the problem of robust estimation of θ in the model $\mathrm{E}(\mu, \theta)$, trimmed mean estimators have been introduced and investigated by Kimber (1983a,b), Gather (1986), and Willemain *et al.* (1992), among others. Here we consider the corresponding estimators of α defined by $\hat{\alpha} = \hat{\theta}^{-1}$.

For specified β_1 and β_2 satisfying $0 \le \beta_1 < 1$ and $0 \le \beta_2 < 1 - \beta_1$, a trimmed mean is formed by discarding the proportion β_1 lowermost observations and proportion β_2 uppermost observations and then averaging the remaining ones in some sense. In particular, Kimber (1983a,b) defines

$$\hat{\theta}_{\mathrm{T}} = \sum_{i=1}^{n} c_{ni}(Z_{ni} - \mu), \tag{4.1}$$

with $c_{ni} = 0$ for $1 \le i \le [n\beta_1]$, $= 0$ for $n - [n\beta_2] + 1 \le i \le n$, and $= 1/d(\beta_1, \beta_2, n)$ for $[n\beta_1] + 1 \le i \le n - [n\beta_2]$, where $[\cdot]$ denotes "greatest integer part" and

$$d(\beta_1, \beta_2, n) = \sum_{j=[n\beta_1]+1}^{n-[n\beta_2]} \sum_{i=1}^{j} (n - i + 1)^{-1}.$$

(This choice of c_{ni}'s makes $\hat{\theta}_{\mathrm{T}}$ *mean unbiased*.) Robustness against lower outliers is gained if $[n\beta_1] > 1$ and against upper outliers if $[n\beta_2] > 1$. Indeed, the trimmed mean estimator is completely unaffected by taking the proportion β_1 lowermost observations to a lower limit ($\log \mu$ if σ known, 0 otherwise), or by taking the proportion β_2 uppermost observations to $+\infty$, so that LBP $= [n\beta_1]/n \to \beta_1$ and UBP $= [n\beta_2]/n \to \beta_2$, $n \to \infty$. These BP's apply also to the corresponding estimator of α given by $\hat{\alpha}_{\mathrm{T}} = \hat{\theta}_{\mathrm{T}}^{-1}$. (In the case of no trimming, i.e., $\beta_1 = \beta_2 = 0$, note that $d(0,0,n) = n$ and thus $\hat{\alpha}_{\mathrm{T}}$ reduces to the MLE considered above.)

To consider ARE, we utilize the fact that $\hat{\alpha}_{\mathrm{T}}$ is AN$(\alpha, D_{\beta_1,\beta_2}\alpha^2/n)$, with D_{β_1,β_2} computable following general methods for L-statistics in Serfling (1980), §8.2.4, or Lehmann (1983), §5.4. Thus

$$\mathrm{ARE}(\hat{\alpha}_{\mathrm{T}}, \hat{\alpha}_{\mathrm{ML}}) = \frac{1}{D_{\beta_1,\beta_2}}. \tag{4.2}$$

In particular, for $\beta_1 = \beta_2$ or $\beta_1 = 0$, and β_2 taking values .05, .10, .15, .20, and .25, values of ARE$(\hat{\alpha}_{\mathrm{T}}, \hat{\alpha}_{\mathrm{ML}})$, and thus via (4.2) corresponding values of D_{0,β_2} and D_{β_1,β_2} are found from Table 1 of Kimber (1983a) and Table II of Kimber (1983b). For each of these five choices of β_2, the ARE's for the two cases $\beta_1 = 0$ and $\beta_1 = \beta_2$ turn out to agree within two decimal places. This leads to ARE and corresponding D values as listed in the following table.

Table 4.1. ARE$(\hat{\alpha}_{\mathrm{T}}, \hat{\alpha}_{\mathrm{ML}})$ and D_{β_1,β_2} for $\beta_1 = 0$ or $\beta_1 = \beta_2$, and selected β_2.

	β_2				
	.05	.10	.15	.20	.25
ARE$(\hat{\alpha}_{\mathrm{T}}, \hat{\alpha}_{\mathrm{ML}})$	0.92	0.85	0.78	0.72	0.67
D_{0,β_2}, D_{β_2,β_2}	1.09	1.18	1.28	1.39	1.49

4.3 Estimators Based on a Regression Approach

The "least squares", "quantile based", and "percentile matching" estimators to be considered in §§4.4–4.5 below may be viewed as special cases of a regression approach based on *linearization* of the model (1.1) in terms of its parameters. We describe this approach here, and, for compatibility with existing literature, consider estimation of α with σ treated as an unknown nuisance parameter. Representing the model (1.1) by its *quantile function*

$$F^{-1}(p) = \inf\{x : F(x) \geq p\} = \sigma(1-p)^{-1/\alpha}, \quad 0 < p < 1,$$

and then taking logarithms, we arrive at

$$\log F^{-1}(p) = \log \sigma + \alpha^{-1}\left(-\log(1-p)\right), \quad 0 < p < 1. \tag{4.3}$$

A sample analogue of (4.3) based on $X_1, \ldots, X_n$ is then obtained by introducing the usual sample cdf for estimation of F, i.e.,

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\}, \quad -\infty < x < \infty,$$

and the corresponding sample quantile function

$$\widehat{F}_n^{-1}(p) = X_{n,\lceil np \rceil}, \quad 0 < p < 1, \tag{4.4}$$

where $\lceil x \rceil$ denotes the least integer $\geq x$. Now defining $\varepsilon = \log \widehat{F}_n^{-1}(p) - \log F^{-1}(p)$ and substituting $\widehat{F}_n^{-1}(p)$ for $F^{-1}(p)$ into (4.3), we obtain an exact reexpression of model (1.1) as

$$\log \widehat{F}_n^{-1}(p) = \log \sigma + \alpha^{-1}\, u + \varepsilon, \quad 0 < p < 1, \tag{4.5}$$

where $u = -\log(1-p)$, $0 < p < 1$. Taking n choices of p, such that $\widehat{F}_n^{-1}(p)$ generates the set of order statistics $\{X_{ni},\ 1 \leq i \leq n\}$ as per relation (4.4), i.e., for choices $p_{n1}^*, \ldots, p_{nn}^*$ satisfying

$$p_{n1}^* \in (0, \tfrac{1}{n}], \quad p_{n2}^* \in (\tfrac{1}{n}, \tfrac{2}{n}], \quad \ldots, \quad p_{n,n-1}^* \in (\tfrac{n-2}{n}, \tfrac{n-1}{n}], \quad p_{nn}^* \in (\tfrac{n-1}{n}, 1), \tag{4.6}$$

we obtain from (4.5) a set of n equations for the two unknowns σ and α (equivalently, $\log \sigma$ and α^{-1}):

$$Z_{ni} = \log \sigma + \alpha^{-1}\, u_{ni}^* + \varepsilon_{ni}^*, \quad 1 \leq i \leq n, \tag{4.7}$$

where $u_{ni}^* = -\log(1 - p_{ni}^*)$ and $\varepsilon_{ni}^* = Z_{ni} - \log F^{-1}(p_{ni}^*)$, $1 \leq i \leq n$. The equations (4.7) may be utilized via interpretion from the standpoint of the usual linear regression model. Thus estimates of $\mu = \log \sigma$ and $\theta = \alpha^{-1}$ result by fitting a straight line through the scatterplot of points

$$(\, Z_{ni},\, u_{ni}^* \,), \quad 1 \leq i \leq n,$$

and these in turn yield estimates of σ (if unknown) and α. We may select either the full set of all n points or a strategic subset. In particular, in §4.4 we take the full set of points and apply ordinary least squares to obtain a version of the "least squares" estimator considered by Quandt (1966) and Arnold (1983). In §4.5 we take just k points, as considered by Quandt (1966) for $k = 2$ and Sarhan, Greenberg, and Ogawa (1963), Saleh and Ali (1966), and Koutrouvelis (1981) for arbitrary $k \geq 2$.

4.4 Least Squares Estimators

Applying the regression-based approach of §4.3 using the full set of equations (4.7), with the p_{ni}^*'s given by $p_{ni}^* = p_{ni}$, where

$$p_{ni} = \frac{i}{n},\ 1 \leq i \leq n-1, \text{ and } p_{nn} = \frac{n}{n+1},$$

the usual (ordinary) least squares regression estimators yield so-called "least squares" estimators of σ and α. Defining $c_{ni} = -\log(1 - p_{ni})$ and $\bar{c}_n = n^{-1} \sum_{i=1}^n c_{ni}$, we thus arrive at

$$\begin{aligned}
\hat{\theta}_{\text{LS}} &= \frac{n^{-1} \sum_{i=1}^n c_{ni} Z_{ni} - \bar{c}_n \overline{Z}_n}{n^{-1} \sum_{i=1}^n c_{ni}^2 - (\bar{c}_n)^2}, \\
\hat{\mu}_{\text{LS}} &= \overline{Z}_n - \hat{\theta}_{\text{LS}}\, \bar{c}_n.
\end{aligned}$$

and hence

$$\hat{\alpha}_{\text{LS}} = \hat{\theta}_{\text{LS}}^{-1}, \quad \hat{\sigma}_{\text{LS}} = e^{\hat{\mu}_{\text{LS}}}.$$

Focusing on the estimator $\hat{\alpha}_{\text{LS}}$, it is readily seen that its UBP is 0 and hence this estimator is *nonrobust.* (Further, for the case σ unknown, so that $X_i \to 0$ is a possible form of corruption, we have LBP $= 0$.)

In discussions of versions of this least squares estimator by Quandt (1966), p. 60, and Arnold (1983), p. 202, it is suggested that the corresponding estimators are consistent as well as competitive in efficiency with the maximum likelihood estimators. A rigorous treatment of efficiency now available in Brazauskas and Serfling (1999) establishes, however, that $\hat{\alpha}_{\text{LS}}$ is AN$(\alpha, 2\alpha^2/n)$, clarifing that $\hat{\alpha}_{\text{LS}}$ is actually *poor in efficiency*, having ARE = .50. In sum, $\hat{\alpha}_{\text{LS}}$ is both *nonrobust* and *nonefficient.*

4.5 Estimators Based on k Selected Quantiles

Here we consider the regression-based approach of §4.3 using *weighted least squares* with a selected subset of the equations (4.7). Let integer k be given and fixed, select values $0 < p_1 < \cdots < p_k < 1$, and suppose that $n > k$ is sufficiently large that the p_i's fall in k distinct members of the subintervals in (4.6). In this case the k equations from (4.7) corresponding to $\widehat{F}_n^{-1}(p_i)$ in (4.5), $1 \leq i \leq k$, are given by

$$Z_{n,\lceil np_i \rceil} = \log \sigma + \alpha^{-1} u_i + \varepsilon_{ni}, \quad 1 \leq i \leq k, \tag{4.8}$$

where $u_i = -\log(1 - p_i)$ and $\varepsilon_{ni} = Z_{n,\lceil np_i \rceil} - \log F^{-1}(p_i)$, $1 \leq i \leq k$. Thus estimates of $\mu = \log \sigma$ and $\theta = \alpha^{-1}$ result from fitting a straight line to the scatterplot of points

$$(\, Z_{n,\lceil np_i \rceil},\, u_i \,), \quad 1 \leq i \leq k. \tag{4.9}$$

When the number k of quantiles is chosen to equal the number of unknown parameters of the model, the method corresponds to what is called *percentile matching* by Klugman, Panjer, and Willmot (1998). That is, equations for the k parameters are produced by equating k sample and model quantiles, analogous to the method of moments.

In particular (see Koutrouvelis (1981) and Saleh and Ali (1966) for details), we obtain using (4.9) the estimator $\hat{\alpha}_{\text{Q}} = \hat{\theta}_{\text{Q}}^{-1}$, where

$$\hat{\theta}_{\text{Q}} = \sum_{i=1}^{k} b_i \, Z_{n,\lceil np_i \rceil},$$

with

$$\begin{aligned}
b_1 &= -\frac{1}{L}\,\frac{u_2 - u_1}{e^{u_2} - e^{u_1}}, \\
b_i &= \frac{1}{L}\left[\frac{u_i - u_{i-1}}{e^{u_i} - e^{u_{i-1}}} - \frac{u_{i+1} - u_i}{e^{u_{i+1}} - e^{u_i}}\right], \quad 2 \leq i \leq k-1, \\
b_k &= \frac{1}{L}\,\frac{u_k - u_{k-1}}{e^{u_k} - e^{u_{k-1}}},
\end{aligned}$$

and

$$L = \sum_{i=2}^{k} \frac{(u_i - u_{i-1})^2}{e^{u_i} - e^{u_{i-1}}}.$$

In the context of estimation of the $\mathrm{P}(\sigma, \alpha)$ model, this quantile approach was introduced for $k = 2$ by Quandt (1966) and considered for arbitrary $k \geq 2$ by Koutrouvelis (1981). See also Arnold (1983), p. 201, for discussion. In the equivalent context of estimation of the $\mathrm{E}(\mu, \theta)$ model, this approach was treated by Saleh and Ali (1966) and Sarhan, Greenberg, and Ogawa (1963).

We note that $\hat{\alpha}_{\mathrm{Q}}$ is *robust* if p_1 and p_k are bounded away from 0 and 1. Namely, $\hat{\alpha}_{\mathrm{Q}}$ is completely unaffected by taking the proportion p_1 of lower X_i's to a lower limit, or by taking the proportion p_2 of upper X_i's to an upper limit, so that LBP $= p_1$ and UBP $= 1 - p_k$.

Further, $\hat{\alpha}_{\mathrm{Q}}$ is $\mathrm{AN}(\alpha, L^{-1}\alpha^2/n)$. Thus $\hat{\alpha}_{\mathrm{Q}}$ has ARE $= L$.

It is of interest, of course, to choose the quantile levels $p_1, \ldots, p_k$ *optimally*, Following Saleh and Ali (1966), the optimal choice of p_1 is found to be

$$p_1^{\circ} = \frac{1}{n + .5} \tag{4.10}$$

and optimal choices of the remaining p_i's are obtained by minimizing the generalized variance of the joint estimators of α and σ with respect to $p_2, \ldots, p_k$, subject to (4.10). Since (4.10) results in $Z_{n,\lceil np_1 \rceil} = Z_{n1}$ and the sum of the b_i's is 0, the optimal estimator may be expressed as

$$\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},k} = \left(\sum_{i=2}^{k} b_i (Z_{n,\lceil np_i \rceil} - Z_{n1}) \right)^{-1}.$$

It follows, as discussed in Brazauskas and Serfling (1999), that the optimal choices of $p_2, \ldots, p_k$ are in fact just those derived by Sarhan, Greenberg, and Ogawa (1963) and listed for $k = 2(1)16$ in their Table 3, for asymptotically optimal estimation of θ in the *one-parameter* exponential model $\mathrm{E}(0, \theta)$ by a linear function of $k - 1$ order statistics.

4.5.1 The Case $k = 2$

Formulas simplify nicely for the case $k = 2$. We have

$$\hat{\alpha}_{\mathrm{Q}} = \left(\frac{Z_{n,\lceil np_2 \rceil} - Z_{n,\lceil np_1 \rceil}}{u_2 - u_1} \right)^{-1}$$

and

$$\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2} = \left(\frac{Z_{n,\lceil np_2 \rceil} - Z_{n1}}{u_2} \right)^{-1}.$$

Further, we have

$$\mathrm{Var}(\hat{\alpha}_{\mathrm{Q}}) = \frac{e^{u_2} - e^{u_1}}{(u_2 - u_1)^2} \frac{\alpha^2}{n} + O(n^{-2}), \tag{4.11}$$

as $n \to \infty$ with p_1 fixed or tending to 0 at rate $\mathrm{O}(n^{-1})$ and p_2 fixed. Also, $\mathrm{Var}(\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2})$ satisfies (4.11) with u_1 replaced by 0.

4.5.2 Examples

In particular,

- For $k = 2$ the optimal quantile levels are $p_1 = p_1^\circ$ and $p_2 = .80$. For the corresponding estimator $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$ we have LBP = 0, UBP = .20, and ARE = .649.

- Another choice of p_1 and p_2 for $k = 2$, used in an example of percentile matching in Klugman, Panjer, and Willmot (1998), p. 47, is $p_1 = .40$ and $p_2 = .70$. For the corresponding estimator we have LBP = .40, UBP = .30, and ARE = .625.

- For $k = 5$ the optimal p_i's are $p_1 = p_1^\circ$, $p_2 = .45$, $p_3 = .74$, $p_4 = .91$, and $p_5 = .98$, and for $\hat{\theta}_{\mathrm{Q}}^{\mathrm{opt},5}$ we have LBP = 0, UBP = .02, and ARE = .926.

These examples indicate that efficiency can be increased by choosing the quantile levels optimally and taking k larger, but at the expense of severe reduction in breakdown point. If, rather, one desires relatively high breakdown points, then nonoptimal levels must be selected. For example:

- For the $k = 2$ estimator based on $p_1 = .10$ and $p_2 = .90$, we have LBP = UBP = .10 and ARE = .543.

- For the $k = 4$ estimator based on $p_1 = p_1^\circ$, $p_2 = 0.25$, $p_3 = 0.50$, and $p_4 = 0.75$, we have LBP =0, UBP = .25, and ARE = .735.

- For the $k = 5$ estimator based on $p_1 = .13$, $p_2 = .32$, $p_3 = .50$, $p_4 = .69$, and $p_5 = .87$, we have LBP = UBP = .13 and ARE = .73. For later reference, we designate this estimator by $\hat{\alpha}_{\mathrm{Q}}^*$.

5 Comparisons and Conclusions

In Table 5.1 below, important cases of the estimators of α considered in Sections 3 and 4 are compared from the standpoint of *efficiency versus robustness.*

Table 5.1. ARE and UBP for selected estimators of α.

Estimator	ARE	UBP
MLE	1	0
$\hat{\alpha}_{\mathrm{MM}}$ $(2 < \alpha \leq 2.5)$	$\leq .56$	0
$\hat{\alpha}_{\mathrm{LS}}$	.50	0
$\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$	.65	.20
$\hat{\alpha}_{\mathrm{Q}}^{*}$ $(k=5)$	.73	.13
$\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},5}$	.93	.02
$\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = .25$	.67	.25
$\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = .20$	.72	.20
$\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = .15$	.78	.15
$\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = .10$	.85	.10
$\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = .05$	.92	.05
$\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = .04$	.93	.04
$\hat{\alpha}_{\mathrm{GM}}$, $k=2$	.78	.29
$\hat{\alpha}_{\mathrm{GM}}$, $k=3$	.88	.21
$\hat{\alpha}_{\mathrm{GM}}$, $k=4$	.92	.16
$\hat{\alpha}_{\mathrm{GM}}$, $k=5$	.94	.13
$\hat{\alpha}_{\mathrm{GM}}$, $k=10$	.98	.07

The following conclusions are quite evident:

- The "method of moments" and "least squares" estimators are neither robust nor efficient and thus are not competitive.

- The "quantile" type estimators are improved upon by the "trimmed" type estimators. For example, $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt}}$ for $k=2$, with ARE $= .65$ and UBP $= .20$, is dominated by $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .20$, with ARE $= .72$ and UBP $= .20$. Also, $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt}}$ for $k=5$, with ARE $= .93$ and UBP $= .02$, is dominated by $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .04$, with ARE $= .93$ and UBP $= .04$. Finally, the quantile estimator $\hat{\alpha}_{\mathrm{Q}}^{*}$, with ARE $= .73$ and UBP $= .13$, is improved upon by $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .15$, with ARE $= .78$ and UBP $= .15$.

- In turn, the "trimmed" type estimators (and thus also the quantile estimators) are improved upon by the "generalized median" type estimators. For example, $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .20$, with ARE $= .72$ and UBP $= .20$, is dominated by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k=3$, with ARE $= .88$ and UBP $= .21$. Likewise, $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .05$, with ARE $= .92$ and UBP $= .05$, is dominated by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k=10$, with ARE $= .98$ and UBP $= .07$. Finally, $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .15$, with ARE $= .78$ and UBP $= .15$, is improved upon by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k=4$, with ARE $= .92$ and UBP $= .16$.

Interpretative Conclusion. The superiority of the generalized median estimators may be explained by the following general principle:

> *Smoothing* of the data by evaluating a function of a few observations at a time over all corresponding subsets of the data, followed by *medianing* applied to these function evaluations, yields a very favorable combination of *efficiency* and *robustness.*

Practical Recommendations:

> The MLE is efficient but not robust and should be replaced by a competitor. The new generalized median approach dominates the other competitors and should become incorporated into practice. The trimmed mean approach remains competitive, however, and should remain in practical use. The less competitive quantile and percentile matching approaches should be used more cautiously if not dropped, and the standard least squares approach should be dropped from practical use.

6 Applications

Here we consider various applications utilizing estimation of α in the model $\mathrm{P}(\sigma, \alpha)$. We find that a small relative error in estimation of α can produce a large relative error in estimated quantities based on α. Thus even small improvements in methods of estimation of the tail index α can yield substantial impact in applications. Further, for robust estimation of quantities based on α, robust estimation of α itself is crucial.

6.1 Estimation of an Upper Quantile

For estimation of the quantile q_ε corresponding to upper tail probability ε, it follows from (1.1) that

$$q_\varepsilon = \sigma \varepsilon^{-1/\alpha}. \tag{6.1}$$

Thus for the estimator $\hat{q}_\varepsilon$ defined by putting $\hat{\alpha}$ for α in (6.1) we have

$$\frac{\hat{q}_\varepsilon}{q_\varepsilon} = \varepsilon^{1/\alpha - 1/\hat{\alpha}}. \tag{6.2}$$

Consequently, for $\alpha = 1$ and $\varepsilon = .001$, *overestimation* of α by 10% produces *underestimation* of $q_{.001}$ by 47%. Likewise, for $\alpha = 1.5$ and $\varepsilon = .001$, overestimation of α by 10% produces underestimation of $q_{.001}$ by 34%. For $\alpha = 1.5$ and $\varepsilon = .0001$, overestimation of α by 10% produces underestimation of $q_{.0001}$ by 43%.

Tail probabilities in the range of $\varepsilon = .001$ or even $\varepsilon = .0001$ are common in actuarial and extreme value applications. For example, high quantiles are used in the percentile principle of reinsurance premium calculation (see Gerber (1979)). Also, after the 1953 flood disaster, the Dutch government set a standard for sea dikes that the sea level should not exceed the dike level in any given year except with probability less than .0001 (see Dekkers and de Haan (1989) for discussion).

6.2 Estimation of an Upper Tail Probability

Reversing the previous illustration, for estimation of the tail probability ε above a specified threshold q, it follows from (1.1) or (6.1) that

$$\varepsilon = (\frac{\sigma}{q})^{\alpha}. \tag{6.3}$$

Thus for the estimator $\hat{\varepsilon}$ defined by putting $\hat{\alpha}$ for α in (6.3) we have

$$\frac{\hat{\varepsilon}}{\varepsilon} = \left(\frac{\sigma}{q}\right)^{\hat{\alpha}-\alpha} = \varepsilon^{(\hat{\alpha}/\alpha)-1}. \tag{6.4}$$

Consequently, for estimation of the tail probability $\varepsilon = .001$, *overestimation* of α by 10% causes *underestimation* of ε by 50%. Likewise, for estimation of the tail probability $\varepsilon = .0001$, overestimation of α by 10% causes underestimation of ε by 60%.

6.3 Solvency of Portfolios

For analysis of the solvency of a portfolio, some methods (Ramlau-Hansen (1988)) involve upper quantiles of the *total claims* distribution as well as of the individual claim distribution. In particular, for determination of the premium for stop-loss reinsurance, the tail of the distribution of the *reinsured amount* of the total claims is directly related to the tail of the total claims distribution. Also, for determination of the premium for *excess-of-loss reinsurance*, one method requires rewriting the deductible as a quantile of the individual claim distribution and estimating the corresponding tail probability. See Daykin, Pentikäinen, and Pesonen (1994), pp. 102–116, and Beirlant, Teugels, and Vynckier (1996), pp. 122–126, for detailed discussion.

For *semiparametric* modeling of the tail of an individual claim distribution or of an aggregate claims distribution, a broad and effective assumption is a "*Pareto type*" distribution: a distribution H for which the survival function $1 - H(d)$ tends to 0 at a polynomial rate $d^{-\alpha}$ as $d \to \infty$, for some index α. In such a case we have

$$\lim_{d\to\infty} \frac{1 - H(dx)}{1 - H(d)} = x^{-\alpha}, \tag{6.5}$$

i.e., the conditional distribution of an observation, given that it exceeds a threshold d, becomes for large d approximately a one-parameter Pareto distribution, $\mathrm{P}(1, \alpha)$. See Beirlant, Teugels, and Vynckier (1996), pp. 29 and 51, and Brazauskas and Serfling (1999) for discussion. It follows that for estimation of α one may apply to the upper ordered values of a sample from H methods designed for $\mathrm{P}(1,\alpha)$ samples. In particular, for determination of reinsurance premiums as discussed above, methods for estimation of quantiles and tail probabilities of $\mathrm{P}(1,\alpha)$ are relevant, along with the considerations in §§6.1–6.2 above. By way of illustration, in an analysis of $n = 429$ Norwegian fire insurance claims for the year 1981, Beirlant, Teugels, and Vynckier (1996) arrive at $k(n) = 94$ as the suitable number of upper ordered values to employ in estimation of α and of the upper .001 quantile of the claim distribution.

In this semiparametric context, the MLE of α based on the treating the $k(n)$ upper order statistics as a sample from $\mathrm{P}(1,\alpha)$ is known as the *Hill estimator* (introduced by Hill (1975)). It is efficient and thus provides the benchmark against which competing estimators are compared, but it is nonrobust, being seriously influenced by any extreme outliers which are not representative of the model being estimated. See Beirlant, Teugels, and Vynckier (1996), Chapter 2, for review of various "excess values" estimators including the Hill estimator and weighted least squares versions. Adaptation of our generalized median estimator considered in Section 4 should yield a further competitor to the Hill estimator which competes well with respect to efficiency while also achieving a high degree of robustness.

A further quantity sometimes used as a principle for setting reinsurance premiums is the *mean excess function*

$$e(d) = E(X - d | X > d),$$

i.e., the conditional expectation of the excess above threshold d for an observation that exceeds d. For $\mathrm{P}(1,\alpha)$ we have

$$e(d) = \frac{d}{\alpha - 1},$$

and for *Pareto type* distributions we have

$$\lim_{d \to \infty} \frac{e(d)}{d} = \frac{1}{\alpha - 1}. \tag{6.6}$$

Thus robust estimation of $e(d)$ becomes of interest, and via (6.6) this may be carried out by robust estimation of α in $\mathrm{P}(1,\alpha)$. In this regard, we note that for $\alpha = 1.5$ and large d, overestimation of α by 10% produces underestimation of $e(d)$ by 26%. For such robust estimation via (6.6), the estimators $\hat{\alpha}_{\mathrm{T}}$ and $\hat{\alpha}_{\mathrm{GM}}$ restricted to just the excess values may be used. We also note an empirical "*trimmed mean excess values*" estimator for $e(d)$ proposed by Beirlant, Teugels, and Vynckier (1996), p. 45, for robust estimation of $e(d)$.

6.4 Empirical versus Parametric Methods

Many important features of an underlying loss or claim distribution H can be represented as functionals of H, for example the *mean*, *variance*, *standard deviation*, *coefficient of variation*, *skewness*, *kurtosis*, and *k-th factorial moments*. To this list we can add such functionals as the *mean excess function*, the *loss elimination ratio*, and various types of *reinsurance premiums*.

One can consider *nonparametric* estimation, of which "*empirical estimation*" represents a standard approach. Simply, one estimates a functional $T(H)$ by $T(\widehat{H}_n)$, where $\widehat{H}_n$ denotes a sample analogue estimator of H, as introduced in §4.3 for estimation of F given by (1.1). While this has the advantage of not depending directly upon parametric assumptions, which may be of questionable validity, such estimators often give up too much efficiency in return for too little robustness.

In the case of *parametric* modeling of H, for example as $\mathrm{P}(\sigma, \alpha)$ via (1.1), one represents such functionals as explicit functions of the parameters σ and α and obtains estimates by substitution of $\hat{\alpha}$ for α. Here disasters due to lack of complete validity of parametric assumptions are avoided by employing *robust* methods. Further, parametric approaches are in keeping with the principle of parsimony in modeling while also permitting inferences to be made beyond the range of the actual observed data. For more detailed discussion and illustration, see Klugman, Panjer, and Willmot (1998), §§2.2 and 2.6.

In the present paper we have emphasized *robust parametric* estimation over the *empirical nonparametric* approach.

7 Estimation of α in $\mathrm{P}(\sigma, \alpha)$ with σ Unknown

With σ *unknown*, the MLE of α is the estimator produced by substituting for σ in $\hat{\alpha}_{\mathrm{ML}}$ the minimum sample observation X_{n1}. Like $\hat{\alpha}_{\mathrm{ML}}$, the modified estimator $\tilde{\alpha}_{\mathrm{ML}}$ satisfies LBP $=$ UBP $= 0$ and is $\mathrm{AN}(\alpha, \alpha^2/n)$. Similarly modified generalized median, method of moments, trimmed mean, least squares, and quantile based estimators are treated in Brazauskas and Serfling (1999). It is found that the method of moments and least squares estimators are non-competitive, while the quantile and trimmed type estimators are competitive but dominated overall by the generalized median estimators.

Acknowledgment

We thank Dr. G. L. Thompson for helpful comments on an earlier draft. Also, support by grants from the Casualty Actuarial Society and Society of Actuaries, with administrative support from the Actuarial Education and Research Fund, and by NSF Grant DMS-9705209, is gratefully acknowledged.

References

[1] Arnold, B. C. (1983). *Pareto Distributions.* International Cooperative Publishing House, Fairland, Maryland.

[2] Bierlant, J., Teugels, J. F., and Vynckier, P. (1996). *Practical Analysis of Extreme Values*. Leuven University Press, Leuven, Belgium.

[3] Brazauskas, V., and Serfling, R. (1999). Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. Preprint.

[4] Choudhury, J., and Serfling, R. (1988). Generalized order statistics, Bahadur representations, and sequential nonparametric fixed-width confidence intervals. *J. Statist. Plann. and Inf.* **19** 269–282.

[5] Daykin, C. D., Pentikäinen, T., and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*. Chapman & Hall, London.

[6] Dekkers, A. L. M., and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17** 1795–1832.

[7] Donoho, D. L., and Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum, and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

[8] Gather, U. (1986). Robust estimation of the mean of the exponential distribution in outlier situation. *Commun. Statist.—Theor. Meth.* **15** 2323–2345.

[9] Gerber, H. U. (1979). *An Introduction to Mathematical Risk Theory*. S. S. Huebner Foundation for Insurance Education, University of Pennsylvania.

[10] Gomes, C. P., Selman, B., and Crato, N. (1997). Heavy-tailed distributions in combinatorial search. In Smolka, G. (ed.), *Principles and Practice of Constraint Programming CP-97*, Lecture Notes in Computer Science **1330** 121–135. Springer, New York.

[11] Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

[12] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **13** 331–341.

[13] Hodges, J. L., Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berk. Symp. on Math. Statist. and Prob.*, Vol. 1, 163–168.

[14] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1, 2nd edition. Wiley, New York.

[15] Kimber, A. C. (1983a). Trimming in gamma samples. *Appl. Statist.* **32** 7–14.

[16] Kimber, A. C. (1983b). Comparison of some robust estimators of scale in gamma samples with known shape. *J. Statist. Comput. Simul.* **18** 273–286.

[17] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models: From Data to Decisions.* Wiley, New York.

[18] Koutrouvelis, I. A. (1981). Large-sample quantile estimation in Pareto laws. *Commun. Statist. – Theor. Meth.* **A10** 189–201.

[19] Lehmann, E. L. (1983). *Theory of Point Estimation.* Wiley, New York (transferred to Wadsworth, 1991).

[20] Quandt, R. E. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika* **10** 55–82.

[21] Saleh, A. K. Md. E., and Ali, M. M. (1966). Asymptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution. *Ann. Math. Statist.* **37** 143–151.

[22] Sarhan, A. E., Greenberg, B. G., and Ogawa, J. (1963). Simplified estimates for the exponential distribution. *Ann. Math. Statist.* **34** 102–116.

[23] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[24] Serfling, R. (1984) Generalized L-, M- and R-statistics. *Ann. Statist.* **12** 76–86.

[25] Willemain, T. R., *et al.* (1992). Robust estimation methods for exponential data: a Monte-Carlo comparison. *Commun. Statist.—Simula.* **21** 1043–1075.