J. Statist. Comput. Simul., 2001, Vol. 70, pp. 1–19
Reprints available directly from the publisher
Photocopying permitted by license only

SMALL SAMPLE PERFORMANCE OF ROBUST ESTIMATORS OF TAIL PARAMETERS FOR PARETO AND EXPONENTIAL MODELS

VYTARAS BRAZAUSKAS[a,*] and ROBERT SERFLING[b,†]

[a]*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA;*
[b]*Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA*

(Received 21 May 2000; In final form 31 October 2000)

Robust estimation of tail index parameters is treated for (equivalent) two-parameter Pareto and exponential models. These distributions arise as parametric models in actuarial science, economics, telecommunications, and reliability, for example, as well as in semiparametric modeling of upper observations in samples from distributions which are regularly varying or in the domain of attraction of extreme value distributions. In a recent previous paper, new estimators of "generalized median" (GM) type were introduced and shown to provide more favorable trade-offs between efficiency and robustness than several well-established estimators, including those corresponding to methods of maximum likelihood, trimming, and quantiles. Here we establish — *via* simulation — that the superiority of the GM type estimators remains valid even for small sample sizes $n = 10$ and 25. To bridge between "small" and "large" sample sizes, we also include the cases $n = 50$ and 100. Further, we arrive at guidelines for selection of a particular GM estimator in practice, depending upon the sample size, upon whether protection is desired against upper outliers only, or against both upper and lower outliers, and upon whether the level of possible contamination by outliers is high or low. Comparisons of estimators are made on the basis of relative efficiency with respect to the maximum likelihood estimator, breakdown points, and premium-protection plots.

*Corresponding author.

†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.

Keywords and Phrases: Robust estimation; Tail index parameters; Pareto model; Exponential model; Generalized median statistics; Breakdown point; Premium/protection plot; Small-sample performance

AMS 1991 Subject Classifications: Primary: 62F35; Secondary: 62F11

1. INTRODUCTION AND PRELIMINARIES

Diverse parametric and semiparametric applications call for estimation of the tail index α of the two-parameter *Pareto* distribution $P(\sigma, \alpha)$ having cdf

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

where $\alpha > 0$ and $\sigma > 0$, or equivalently, *via* logarithmic transformation, of the scale parameter $\theta = \alpha^{-1}$ of the two-parameter *exponential* distribution $E(\mu, \theta)$ having cdf

$$G(z) = 1 - e^{-(z-\mu)/\theta}, \quad z \geq \mu, \tag{2}$$

for $\theta > 0$ and $-\infty < \mu < \infty$. For estimation of α, when (1) is indeed the true model for the data, the corresponding maximum likelihood estimator (MLE) based on (1) possesses optimal asymptotic efficiency. The performance of this estimator severely degrades, however, if the true model for the observed data departs somewhat from (1). Consequently, one desires to replace the MLE by an estimator which is *robust*, *i.e.*, which performs relatively well under departures from (1), at the cost of a sacrifice of some efficiency when (1) is actually valid.

Various types of competing estimators have been formulated and studied by simulation for fixed sample sizes. In the context of (1), Quandt (1966) compared the MLE with moments estimators, a least squares estimator, and certain "quantile" type estimators, on the basis of 100 samples each for sample sizes $n = 25$, 50, 100, 300, 500, 1000 and 2000; the MLE and quantile type estimators were judged to perform best. Koutrouvelis (1981) extended the list of quantile type estimators and compared them with the MLE and moments estimators, on the basis of 3200 samples for sample size $n = 25$, 800 samples for $n = 50$, and 160 samples for $n = 500$; one of the extended quantile type estimators was found most favorable overall. In the equivalent context of (2) with μ known, on the basis of 500 samples each of sizes $n = 5$, 10 and 20, Willemain *et al.* (1992) compared the MLE (the mean) with a collection of robust estimators, including a rather computational "transform" type estimator which they introduced and a trimmed mean type estimator introduced by Kimber (1983a, b); for protection against mild contamination from outliers the trimmed type estimator performed best overall, while for protection against heavy contamination the transform type performed best.

In a recent *large-sample* study, Brazauskas and Serfling (1999) introduced in the context of (1) new robust estimators of "*generalized median*" (GM) type and compared them with the maximum likelihood, quantile type, trimmed mean type, and other estimators. Using as *efficiency* criterion the *asymptotic relative efficiency* (ARE) with respect to the MLE and as *robustness criterion* the *breakdown point* (BP) (these are defined below), the GM type was seen to dominate all competitors, with the trimmed mean type second best. In Brazauskas and Serfling (2000), similar conclusions were obtained in a study of estimators of α for the *one-parameter* Pareto model given by (1) with σ known.

In the present paper we establish — *via* simulation — that the superiority of the GM type estimators remains valid even for small sample sizes $n=10$ and 25. To bridge between "small" and "large" sample sizes, we also include the cases $n=50$ and 100. Further, we arrive at guidelines for selection of a particular GM estimator in practice, depending upon whether protection is desired against upper outliers only, or against both upper and lower outliers, and whether the level of possible contamination by outliers is high or low. Our treatment is carried out in terms of the two-parameter model $\mathrm{P}(\sigma, \alpha)$ with σ unknown, confining attention to those estimators ranked as more favorable in the large-sample studies: the GM, trimmed mean, and quantile types (all defined precisely in Section 2). Of course, findings and comparisons regarding estimators $\hat{\alpha}$ of α in $\mathrm{P}(\sigma, \alpha)$ convert to corresponding statements about estimators $\hat{\theta} = \hat{\alpha}^{-1}$ of θ in $\mathrm{E}(\mu, \theta)$.

1.1. Relative Efficiency Criterion

As is well-known (*e.g.*, Arnold, 1983), for model (1) the MLE's of α and σ are given by

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

where $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the ordered values of the sample. Actually, we will use here the *unbiased* (but asymptotically equivalent) version of $\hat{\alpha}_{\text{ML}}$, namely

$$\hat{\alpha}_{\text{MLU}} = \frac{n-2}{n}\hat{\alpha}_{\text{ML}}.$$

As a benchmark for *efficiency* considerations, we use for fixed sample size the *exact* relative efficiency (RE) taken with respect to the MLU and based on the *mean square error* (MSE), *i.e.*,

$$\text{RE}(\hat{\alpha}, \hat{\alpha}_{\text{MLU}}) = \frac{\text{MSE of MLU}}{\text{MSE of } \hat{\alpha}}.$$

By comparison, the *large sample* relative efficiency may be defined as ARE = (asymptotic MSE of MLU)/(asymptotic MSE of $\hat{\alpha}$), or, equivalently, since all estimators under consideration are asymptotically unbiased, as (asymptotic variance of MLU)/(asymptotic variance of $\hat{\alpha}$).

1.2. Breakdown Point Criterion

A popular and effective criterion for robustness of an estimator is its (finite-sample) *breakdown point* (BP), loosely defined as the largest proportion of sample observations which may be corrupted without corrupting the estimator beyond any usefulness. It provides an index valid over a broad and *nonspecific* range of possible sources of contaminating data. We define separate versions for lower and upper contamination:

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 an uninformative limit not depending on the parameter being estimated.

Although estimators having both LBP > 0 and UBP > 0 are desired, we give priority to UBP, which is more important in typical applications. In particular, $\hat{\alpha}_{\text{ML}}$ is readily seen to have LBP = UBP = 0 and thus is *nonrobust* and rejected as a contender for robust estimation of α.

For comparison of estimators in terms of efficiency and robustness jointly, we will examine their RE's and UBP's together, paralleling the use of ARE and UBP together in Brazauskas and Serfling (1999, 2000).

1.3. "Premium *Versus* Protection" Approach

We shall also examine efficiency–robustness trade-offs *via* the *premium–protection* (PP) approach of Anscombe (1960), here employing a specific form of contamination model,

$$F = (1 - \varepsilon)\mathrm{P}(\sigma, \alpha) + \varepsilon\mathrm{P}(\sigma^*, \alpha), \tag{3}$$

where ε represents the probability that a sample observation comes from the distribution $\mathrm{P}(\sigma^*, \alpha)$ instead of $\mathrm{P}(\sigma, \alpha)$. In particular, we consider two cases:

model (3) with *upper* outliers (the case $\sigma^* \gg \sigma$), and
model (3) with *lower* outliers (the case $\sigma^* \ll \sigma$).

For each estimator T under consideration, corresponding "premium" and "protection" values are defined:

Premium The relative change (increase) in MSE due to use of T instead of the MLU in the null case C_0 (no contamination), *i.e.*,

$$\mathrm{Premium}(T) = \frac{\mathrm{MSE}(T, C_0) - \mathrm{MSE}(\mathrm{MLU}, C_0)}{\mathrm{MSE}(\mathrm{MLU}, C_0)},$$

and

Protection The relative change (preferably decrease) in MSE due to use of T instead of the MLU in a nonnull case C (contamination), *i.e.*,

$$\mathrm{Protection}(T, C) = \frac{\mathrm{MSE}(\mathrm{MLU}, C) - \mathrm{MSE}(T, C)}{\mathrm{MSE}(\mathrm{MLU}, C)}.$$

Premium *versus* protection works like an insurance policy. Favorable estimators T pay a low premium in terms of loss of efficiency in the null case, in return for high protection (lower MSE than that of the MLU) in the nonnull cases when the MLU is inefficient. For ease of comparison, estimators are displayed as points on so-called *PP-plots*, with an ideal estimator being located in the upper left corner of the PP-plot, reflecting maximal protection for minimal premium.

1.4. The Study

Models with *upper* and *lower* outliers were generated according to (3) for the following choices of parameters and sample sizes:

- $\alpha = 1.50$ (the center of a typical range of α's arising in practice)
- model with *upper* outliers: $\sigma = 1$, $\sigma^* = 1000$, $\varepsilon = 0.10$ and 0.20
- model with *lower* outliers: $\sigma = 1000$, $\sigma^* = 1$, $\varepsilon = 0.01$ and 0.20
- sample sizes: $n = 10$, 25, 50, 100.

The case $\varepsilon = 0.20$ represents a very severe level of contamination which is survived by only the best estimators, and the PP-plots identify these. The choice $\varepsilon = 0.01$ for lower outliers serves to illustrate how even a small amount of contamination can corrupt estimators with LBP $= 0$.

For each combination of parameters and sample size, and for each estimator considered, 25,000 samples were generated and the MSE evaluated. (For $n > 100$, the RE values approximate the corresponding ARE values so closely that large-sample results suffice. This is seen in Table 3.1, which for the uncontaminated case ($\varepsilon = 0$) presents for each estimator the RE's for $n = 10$, 25, 50, 100, and 200, and the ARE.)

On the basis of these MSE's, the GM, trimmed mean, and quantile type estimators are compared in Section 3 from the standpoints of efficiency and robustness considered separately (Sections 3.1 and 3.2, respectively), as well as from the standpoint of efficiency–robustness trade-offs (Sections 3.3 and 3.4). Conclusions and recommendations are presented in Section 3.5.

2. THE ESTIMATORS

The MLE and MLU were given in Section 1. Here we briefly introduce the other estimators in the study. Further discussion is available in Brazauskas (1999) and Brazauskas and Serfling (1999, 2000).

2.1. Generalized Median Estimators

Generalized median (GM) estimators are defined by taking the median of the evaluations $h(X_{i_1}, \ldots, X_{i_k})$ of a given kernel $h(x_1, \ldots, x_k)$ over all subsets of observations taken k at a time, that is, corresponding to all $\binom{n}{k}$ k-sets $\{i_1, \ldots, i_k\}$ of distinct indices from $\{1, \ldots, n\}$. See Serfling (1984, 2000) for general discussion. In Brazauskas and Serfling (1999), such estimators were considered for the parameter α in the case of σ unknown:

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

with two particular choices of kernel $h(x_1, \ldots, x_k)$:

$$h^{(1)}(x_1, \ldots, x_k) = \frac{1}{C_k} \frac{1}{k^{-1} \sum_{j=1}^{k} \log x_j - \log \min\{x_1, \ldots, x_k\}}$$

and

$$h^{(2)}(x_1, \ldots, x_k; X_{(1)}) = \frac{1}{C_{n,k}} \frac{1}{k^{-1} \sum_{j=1}^{k} \log x_j - \log X_{(1)}},$$

where C_k and $C_{n,k}$ are multiplicative median-unbiasing factors, *i.e.*, chosen so that in each case the distribution of $h^{(j)}(X_{i_1}, \ldots, X_{i_k}), j = 1, 2$, has *median* α. (Note that the kernel $h^{(2)}$ depends both on a k-set of the data and the minimum order statistic $X_{(1)}$ of the whole sample. Except when $k = n$, this differs from $h^{(1)}$, which uses the minimum of the particular k-set forming the arguments of the kernel.) Let us denote the corresponding GM estimators by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ and $\hat{\alpha}_{\mathrm{GM}}^{(2)}$, respectively. Values of C_k and $C_{n,k}$ are provided in the following tables. (For $n = 50, 100$, and 200, $C_{n,k}$ is given by a very accurate approximation, $C_{n,k} \approx k/[k(1 - 1/n) - 1/3]$.)

Although one may consider GM statistics for other choices of median-unbiased kernel, our particular choices have special appeal, as

TABLE 2.1 Values of C_k, for $k = 2:10$

				k				
2	3	4	5	6	7	8	9	10
2.89	1.79	1.50	1.36	1.28	1.23	1.20	1.17	1.15

TABLE 2.2 Values of $C_{n,k}$ for $k=2:10$ and $n=10, 25, 50, 100, 200$

n	k = 2	3	4	5	6	7	8	9	10
10	1.36	1.27	1.23	1.20	1.19	1.17	1.17	1.16	1.15
25	1.25	1.18	1.14	1.12	1.11	1.10	1.09	1.08	1.08
50	1.23	1.15	1.12	1.09	1.08	1.07	1.07	1.06	1.06
100	1.21	1.14	1.10	1.08	1.07	1.06	1.05	1.05	1.05
200	1.21	1.13	1.10	1.08	1.06	1.05	1.05	1.04	1.04

follows. Note that each evaluation of the kernel $h^{(1)}$ is essentially the MLE based on that particular subsample, thus endowing this kernel with the efficiency of the MLE in extracting information about α from a given subsample. The modification $h^{(2)}$ is similarly motivated.

For estimation of σ when $\hat{\alpha}_{\mathrm{GM}}$ is used, we shall use simply $\hat{\sigma}_{\mathrm{ML}} = X_{(1)}$.

2.2. Trimmed Mean Estimators

For specified β_1 and β_2 satisfying $0 \le \beta_1,\ \beta_2 < 1/2$, a trimmed mean is formed by discarding the proportion β_1 lowermost observations and the proportion β_2 uppermost observations and averaging the remaining ones in some sense. In particular, for α we introduce the trimmed mean estimator

$$\hat{\alpha}_{\mathrm{T}} = \left(\sum_{i=1}^{n} c_{ni}(\log X_{(i)} - \log X_{(1)}) \right)^{-1},$$

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

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

These estimators correspond to the trimmed mean estimators introduced and studied by Kimber (1983a, b) for the equivalent problem of estimation of $\theta = \alpha^{-1}$ in the model $\mathrm{E}(\mu, \theta)$ with μ known. Note that various choices of weights $d(\beta_1, \beta_2, n)$ are possible, *e.g.*, a choice that minimizes MSE of $\hat{\theta}_{\mathrm{T}}$. Here, however, we follow the existing literature where the above d's are a choice making $\hat{\theta}_{\mathrm{T}} = \hat{\alpha}_{\mathrm{T}}^{-1}$ mean-unbiased for θ.

For estimation of σ when $\hat{\alpha}_{\mathrm{T}}$ is used, we use $\hat{\sigma}_{\mathrm{ML}} = X_{(1)}$.

2.3. Quantile Type Estimators

Quantile estimators based on $k \geq 2$ quantile levels $0 < p_1 < \cdots < p_k < 1$ are defined as follows:

$$\hat{\alpha}_{\mathrm{Q}} = \left(\sum_{i=1}^{k} b_i \log X_{(\lceil np_i \rceil)} \right)^{-1},$$
$$\hat{\sigma}_{\mathrm{Q}} = \exp\{ \log X_{(\lceil np_1 \rceil)} - u_1 / \hat{\alpha}_{\mathrm{Q}} \},$$

with

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

and

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

where $u_i = -\log(1 - p_i)$, $1 \leq i \leq k$, and $\lceil x \rceil$ denotes the least integer $\geq x$. Such estimators were introduced and studied for the Pareto problem by Quandt (1966) for $k = 2$ and by Koutrouvelis (1981) for general k, and for the equivalent exponential problem by Sarhan, Greenberg, and Ogawa (1963) with μ known and by Saleh and Ali (1966) with μ unknown.

As argued in Saleh and Ali (1966), the optimal choice of p_1 is

$$p_1^{\circ} = \frac{1}{n + .5}. \tag{4}$$

As discussed in Brazauskas and Serfling (1999), a reduction to the case of μ known can be carried out, which permits the optimal choices of $p_2, \ldots, p_k$ then to be found from Sarhan, Greenberg and Ogawa (1963). (These values minimize the generalized variance, *i.e.*, the determinant of the asymptotic covariance matrix, of the estimators of σ and α, subject to (4).) Denoting the corresponding estimator of α by $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},k}$, we have in particular:

- For $k=2$, *i.e.*, for $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$, the optimal p_i's are $p_1 = p_1^{\circ}$ and $p_2 = .80$.
- For $k=5$, *i.e.*, for $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},5}$, the optimal p_i's are $p_1 = p_1^{\circ}$, $p_2 = .45$, $p_3 = .74$, $p_4 = .91$, and $p_5 = .98$.

We also consider a nonoptimal case which is robust also with respect to lower outliers:

- For $k=5$, take $p_1 = .13$, $p_2 = .32$, $p_3 = .50$, $p_4 = .69$, and $p_5 = .87$. We denote this estimator as $\hat{\alpha}_{\mathrm{Q}}^{*}$.

3. COMPARISONS AND CONCLUSIONS

The above-mentioned estimators are compared in Section 3.1 with respect to *efficiency* (*i.e.*, RE under the "null" model), in Section 3.2 with respect to *robustness* (*via* UBP), and in Sections 3.3 and 3.4 with respect to *efficiency–robustness trade-offs* (*i.e.*, by looking at RE and UBP jointly, and *via* PP-plots). In Section 3.5 we present conclusions and recommendations.

3.1. Efficiency Comparisons

Comparison of estimators on the basis of efficiency alone is carried out for the case of no contamination ($\varepsilon = 0$). Table 3.1 provides the RE's for selected sample sizes $n = 10, 25, 50, 100$ and 200, along with the ARE's (for $n = \infty$). In Section 3.3 the RE's for $n = 10, 25, 50$ and ∞ are examined again, in conjunction with the corresponding UBP values.

Remark 1 (for Tab. 3.1) The approximate sampling error of the RE's is ± 0.0089. Only 500,000 randomly chosen kernel evaluations were used in cases where $\binom{n}{k}$ exceeded this number, contributing at most an additional ± 0.0001 error to the RE. These cases are indicated with an asterisk (∗). Thus the overall sampling error of any RE is approximately ± 0.009.

Remark 2 (for Tab. 3.1) The above sampling error suffices to explain some small deviations from monotonicity across rows that occur in some rows of Table 3.1 for large sample sizes ($n = 100$ and 200). Also, a rather larger deviation from monotonicity across rows occurs for the GM estimators for $k = 5$, due to the fact that for the very small sample size $n = 10$ each relevant kernel evaluation is based on a very informative full half of the sample. (For the case $k = 10$, where there is only one kernel evaluation, we obtain just the MLE itself, except modified by a factor that reduces its MSE from that of the MLU.) Finally, non-monotonicity across rows for the optimal Q estimators results from the smallness of the sample sizes $n = 10$, 25 and 50, because the "optimal" quantile levels defining the estimators are based on large n asymptotics.

TABLE 3.1 Values of RE for selected n

	RE					
	n					
Estimator	10	25	50	100	200	∞
$\hat{\alpha}_Q^{opt,2}$	.69	.67	.67	.61	.65	.65
$\hat{\alpha}_Q^{opt,5}$	.84	.75	.97	.92	.93	.93
$\hat{\alpha}_Q^{*}$ $(k = 5)$	.32	.56	.67	.67	.69	.72
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .25$	.46	.57	.63	.64	.66	.67
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .20$	.46	.62	.67	.70	.71	.72
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .15$	.60	.72	.74	.76	.77	.78
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .10$	.60	.78	.80	.82	.84	.85
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .05$	.75	.84	.89	.89	.91	.92
$\hat{\alpha}_{GM}^{(1)}$, $k = 2$	.64	.69	.70	.71	.71	.72
$\hat{\alpha}_{GM}^{(1)}$, $k = 3$	.74	.73	.74	.74	.73*	.74
$\hat{\alpha}_{GM}^{(1)}$, $k = 4$	.78	.79	.80	.77*	.79*	.80
$\hat{\alpha}_{GM}^{(1)}$, $k = 5$	.92	.84	.79*	.81*	.83*	.85
$\hat{\alpha}_{GM}^{(1)}$, $k = 10$	.85	.80*	.87*	.89*	.91*	.93
$\hat{\alpha}_{GM}^{(2)}$, $k = 2$	.65	.72	.76	.78	.78	.78
$\hat{\alpha}_{GM}^{(2)}$, $k = 3$	.82	.83	.86	.87	.88*	.88
$\hat{\alpha}_{GM}^{(2)}$, $k = 4$	.83	.89	.91	.90*	.92*	.92
$\hat{\alpha}_{GM}^{(2)}$, $k = 5$	.97	.92	.91*	.92*	.94*	.94
$\hat{\alpha}_{GM}^{(2)}$, $k = 10$	.85	.91*	.95*	.96*	.97*	.98

TABLE 3.2 Values of UBP for selected n

	UBP				
	n				
Estimator	10	25	50	100	∞
$\hat{\alpha}_Q^{\text{opt},2}$	.10	.16	.18	.20	.20
$\hat{\alpha}_Q^{\text{opt},5}$	.00	.00	.00	.01	.02
$\hat{\alpha}_Q^* \ (k=5)$	.10	.12	.12	.13	.13
$\hat{\alpha}_T,\ \beta_1=\beta_2=.25$	.20	.24	.24	.25	.25
$\hat{\alpha}_T,\ \beta_1=\beta_2=.20$	.20	.20	.20	.20	.20
$\hat{\alpha}_T,\ \beta_1=\beta_2=.15$	.10	.12	.14	.15	.15
$\hat{\alpha}_T,\ \beta_1=\beta_2=.10$	.10	.08	.10	.10	.10
$\hat{\alpha}_T,\ \beta_1=\beta_2=.05$	.00	.04	.04	.05	.05
$\hat{\alpha}_{GM}^{(i)},\ k=2$	.20	.28	.28	.29	.29
$\hat{\alpha}_{GM}^{(i)},\ k=3$	.10	.16	.20	.20	.21
$\hat{\alpha}_{GM}^{(i)},\ k=4$	.10	.12	.14	.15	.16
$\hat{\alpha}_{GM}^{(i)},\ k=5$	.10	.08	.12	.12	.13
$\hat{\alpha}_{GM}^{(i)},\ k=10$	.00	.04	.06	.06	.07

TABLE 3.3 Values of RE and UBP for selected n, and asymptotic LBP

	n								
	10		25		50		∞		
Estimator	*RE*	*UBP*	*RE*	*UBP*	*RE*	*UBP*	*ARE*	*LBP*	*UBP*
$\hat{\alpha}_Q^{\text{opt},2}$	.69	.10	.67	.16	.67	.18	.65	0	.20
$\hat{\alpha}_Q^{\text{opt},5}$	.84	.00	.75	.00	.97	.00	.93	0	.02
$\hat{\alpha}_Q^* \ (k=5)$	.32	.10	.56	.12	.67	.12	.72	.13	.13
$\hat{\alpha}_T,\ \beta_1=\beta_2=.25$	.46	.20	.57	.24	.63	.24	.67	0	.25
$\hat{\alpha}_T,\ \beta_1=\beta_2=.20$	.46	.20	.62	.20	.67	.20	.72	0	.20
$\hat{\alpha}_T,\ \beta_1=\beta_2=.15$	.60	.10	.72	.12	.74	.14	.78	0	.15
$\hat{\alpha}_T,\ \beta_1=\beta_2=.10$	.60	.10	.78	.08	.80	.10	.85	0	.10
$\hat{\alpha}_T,\ \beta_1=\beta_2=.05$	.75	.00	.84	.04	.89	.04	.92	0	.05
$\hat{\alpha}_{GM}^{(1)},\ k=2$	.64	.20	.69	.28	.70	.28	.72	.29	.29
$\hat{\alpha}_{GM}^{(1)},\ k=3$	.74	.10	.73	.16	.74	.20	.74	.21	.21
$\hat{\alpha}_{GM}^{(1)},\ k=4$	.78	.10	.79	.12	.80	.14	.80	.16	.16
$\hat{\alpha}_{GM}^{(1)},\ k=5$	.92	.10	.84	.08	.79*	.12	.85	.13	.13
$\hat{\alpha}_{GM}^{(1)},\ k=10$	.85	.00	.80*	.04	.87*	.06	.93	.07	.07
$\hat{\alpha}_{GM}^{(2)},\ k=2$	.65	.20	.72	.28	.76	.28	.78	0	.29
$\hat{\alpha}_{GM}^{(2)},\ k=3$	.82	.10	.83	.16	.86	.20	.88	0	.21
$\hat{\alpha}_{GM}^{(2)},\ k=4$	.83	.10	.89	.12	.91	.14	.92	0	.16
$\hat{\alpha}_{GM}^{(2)},\ k=5$	.97	.10	.92	.08	.91*	.12	.94	0	.13
$\hat{\alpha}_{GM}^{(2)},\ k=10$	.85	.00	.91*	.04	.95*	.06	.98	0	.07

3.2. Robustness Comparisons

For the estimators under study, values of UBP (finite sample and asymptotic) established in Brazauskas and Serfling (1999) are provided in Table 3.2 for sample sizes $n = 10, 25, 50, 100$ and ∞. In Section 3.3 we examine the asymptotic UBP's again, along with asymptotic LBP, in conjunction with the corresponding RE values.

Remark (for Tab. 3.2) For each k, $\hat{\alpha}_{\mathrm{GM}}^{(i)}$, $i = 1, 2$, have the same UBP values. A couple of non-monotonicities across the rows (for $\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = .10$ and $\hat{\alpha}_{\mathrm{GM}}^{(i)}$, $k = 5$, $i = 1, 2$) are due to the discrete nature of the UBP functions, which are not monotone in the range of small n (≤ 25).

3.3. Efficiency – Robustness Trade-offs: RE *Versus* BP

In Table 3.3 we examine RE and UBP together, enabling us to select estimators which provide favorable trade-offs between RE and UBP. For completeness, we also include the asymptotic LBP, for cases when this might be important.

Remark (for Tab. 3.3) RE's marked with (∗) are based on 500,000 randomly chosen kernel evaluations when $\binom{n}{k}$ exceeds 500,000. The approximate sampling error of RE's is ± 0.009.

For sample size $n = 10$, the following conclusions emerge:

- Quantile type estimators dominate the trimmed types. For example, $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .10$ with RE $= .60$ and UBP $= .10$ is dominated by $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$ with RE $= .69$ and UBP $= .10$. (We do not consider cases with UBP $= 0$, which are nonrobust.)
- The generalized median type estimators, however, improve upon both the quantile and trimmed type estimators. For example, $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$ with RE $= .69$ and UBP $= .10$ is dominated by $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 5$ with RE $= .97$ and UBP $= .10$, as well as by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 5$ with RE $= .92$ and UBP $= .10$.

For sample size $n = 25$, the following conclusions emerge:

- Quantile and trimmed type estimators offer comparable trade-offs between RE and UBP.
- The generalized median type estimators, however, again improve upon both the quantile and trimmed type estimators. For example,

$\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$ with RE = .67 and UBP = .16 is dominated by $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 3$ with RE = .83 and UBP = .16, as well as by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 3$ with RE = .73 and UBP = .16. And $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .10$ with RE = .78 and UBP = .08 is dominated by $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 5$ with RE = .92 and UBP = .08, as well as by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 5$ with RE = .84 and UBP = .08.

For sample size $n = 50$, the following conclusions emerge:

- Trimmed type estimators show some improvement over quantile types. For example, $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$ with RE = .67 and UBP = .18 is slightly improved by $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .20$ with RE = .67 and UBP = .20.
- Again, however, the generalized median type estimators improve upon both the quantile and trimmed type estimators. For example, $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},2}$ with RE = .67 and UBP = .18 is dominated by $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 3$ RE = .86 and UBP = .20, as well as by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 3$ with RE = .74 and UBP = .20. And $\hat{\alpha}_{\mathrm{T}}$ for $\beta_1 = \beta_2 = .15$ with RE = .74 and UBP = .14 is dominated by $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 4$ with RE = .91 and UBP = .14, as well as by $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 4$ with RE = .80 and UBP = .14.

As the sample size $n \to \infty$, the trimmed means continue to outperform the quantile types but in turn are dominated quite definitively by the generalized median types. (This asymptotic finding has already been given by Brazauskas and Serfling (1999).)

3.3.1. Overall Perspective

The generalized median type estimators are superior to the quantile and trimmed types. If lower outliers are not of concern, then for protection up to 10% contamination by upper outliers a suitable choice is $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 4$ or 5, and for protection up to 20% contamination a suitable choice is $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 2$, which trades off some additional RE in exchange for higher UBP. If, however, lower outliers are also of concern, then for protection up to 10% contamination by either upper or lower outliers a suitable choice is $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ with k ranging from 4 (for small n) to 10 (for larger n), and for protection up to 20% contamination a suitable choice is $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ with $k = 2$ (for small n) or 3 (for larger n).

In sum, a good overall choice is $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ with $k = 3$, which can be improved to the more efficient $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ with $k = 3$ if lower outliers are not of concern, and in either case may be tightened to $k = 2$ if n is very small (≈ 10). (Note that compared to the estimators $\hat{\alpha}_{\mathrm{GM}}^{(2)}$, the estimators $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ trade off additional RE in exchange for nonzero LBP, except the case $n = 10$ and $k = 10$ where both estimators coincide.)

3.4. Efficiency – Robustness Trade-offs: PP-plots

Another way to explore efficiency – robustness trade-offs for the estimators under consideration is provided by the PP-plots discussed in Section 1. We exhibit a collection of PP-plots with the following features:

- In each PP-plot, the estimators offering the best trade-offs between efficiency and robustness are those not dominated by other points, *i.e.*, those points with no other points to their "northwest". These non-dominated points constitute an "efficient – robust (ER) frontier" of options and are connected by straight lines. Among the points on the ER frontier, one selects according to one's utility function for (premium, protection).
- For each choice of ε, a PP-plot is given for each of the sample sizes $n = 10, 25, 50$, and 100, and the four plots are displayed in a single captioned figure.
- For upper outliers, we consider $\varepsilon = .10$ (Fig. 3.1) and $\varepsilon = .20$ (Fig. 3.2). For lower outliers, we consider $\varepsilon = .01$ (Fig. 3.3) and $\varepsilon = .20$ (Fig. 3.4).
- In these figures, the following notation for the estimators is used:

  $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt,k}}$ is denoted as Qk,
  $\hat{\alpha}_{\mathrm{Q}}^{*}$ is denoted as Q,
  $\hat{\alpha}_{\mathrm{GM}}^{(1)}$, $k = m$, is denoted as Am,
  $\hat{\alpha}_{\mathrm{GM}}^{(2)}$, $k = m$, is denoted as Bm,
  $\hat{\alpha}_{\mathrm{T}}$, $\beta_1 = \beta_2 = p/100$, is denoted as T$p$.

3.4.1. Model with Upper Outliers

DISCUSSION OF FIGURES 3.1 AND 3.2

(i) *The Case* $\varepsilon = .10$ For $n = 10$, all estimators are poor (as might be expected), with $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 2, 4$ and 10 forming the ER frontier. The worst performance is exhibited by $\hat{\alpha}_{\mathrm{Q}}^{*}$. For $n = 25$, 50 and 100, the heavier trimmed estimators and $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 2:5$ and 10, form the ER frontier, substantially beating all competitors.

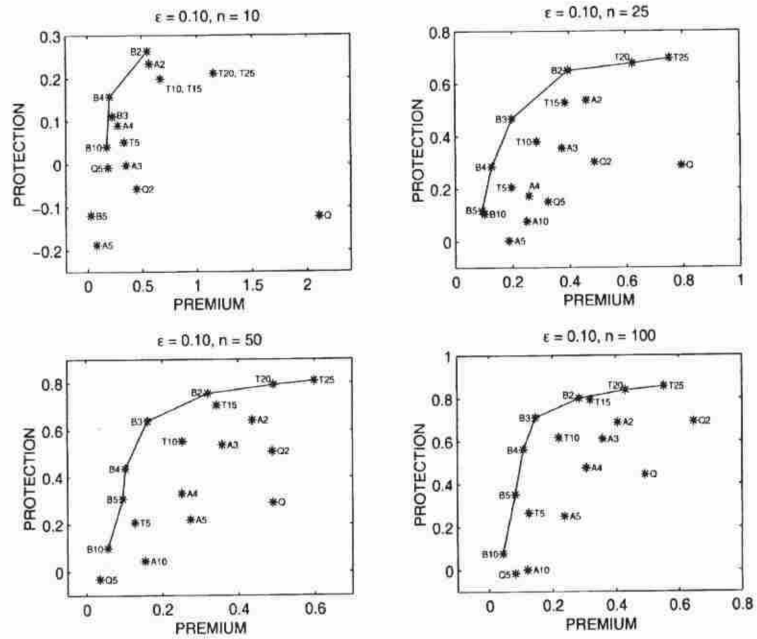

FIGURE 3.1 Model with upper outliers and $\varepsilon = .10$.

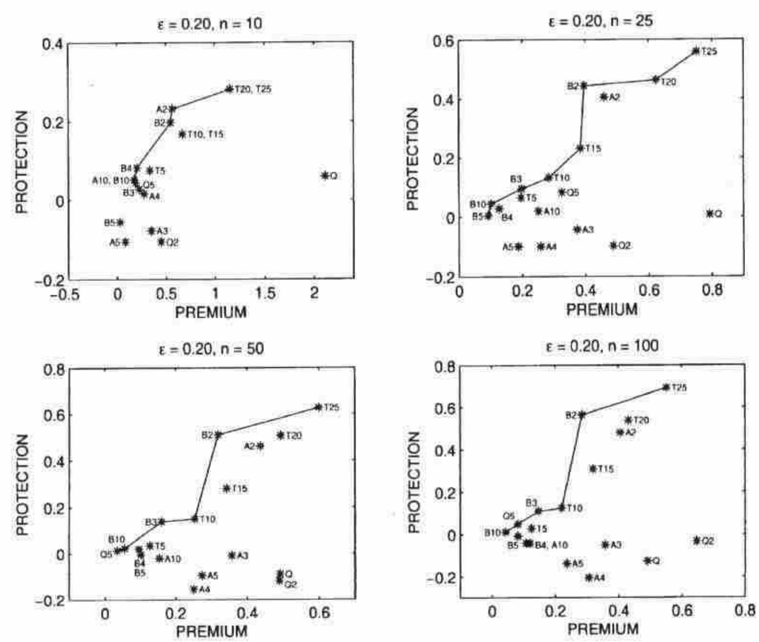

FIGURE 3.2 Model with upper outliers and $\varepsilon = .20$.

(ii) *The Case* $\varepsilon = .20$ For $n = 10$, again all estimators are poor, with the heavier trimmed and $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ (for $k = 2$) and $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ (for $k = 2$ and 4) forming the ER frontier. For $n = 25$, 50, and 100, the estimator $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ for $k = 2$ jumps significantly towards "northwest", offering a very strong performance among estimators on the ER frontier.

3.4.2. Model with Lower Outliers

Discussion of Figures 3.3 and 3.4

(i) *The case* $\varepsilon = .01$ For $n = 10$, all estimators are poor (as might be expected), with only $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 3$ and 4 even exhibiting positive protection. For $n = 25$, the ER frontier consists of just $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 5$. For $n = 50$ and 100, the most attractive estimator is $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 10$, although $\hat{\alpha}_{\mathrm{Q}}^{\mathrm{opt},5}$ is also on the ER frontier, and $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 5$ is competitive.

(ii) *The case* $\varepsilon = .20$ For all sample sizes, the estimators $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 2, 3$ and 4 are the most attractive.

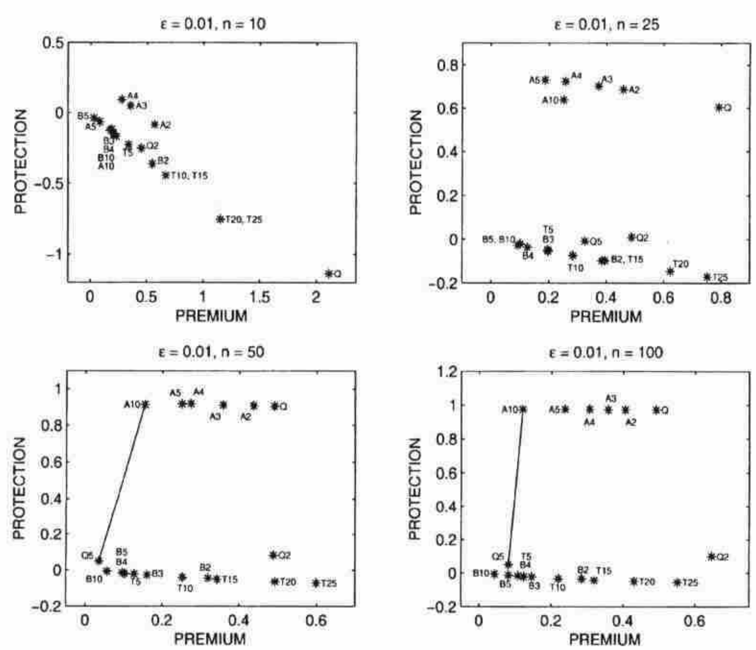

FIGURE 3.3 Model with lower outliers and $\varepsilon = .01$.

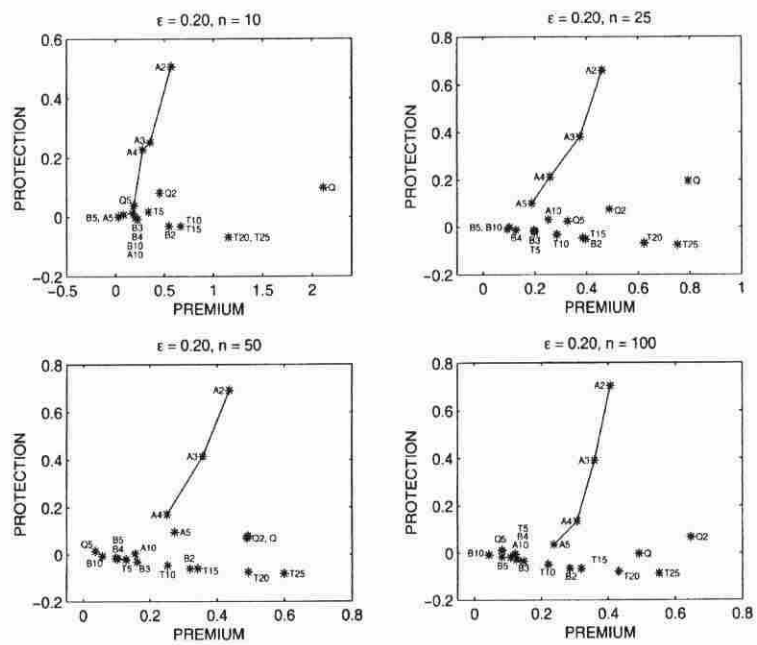

FIGURE 3.4 Model with lower outliers and $\varepsilon = .20$.

3.5. Conclusions and Recommendations

Although based on different approaches, the findings of Sections 3.3 and 3.4 tend to be in agreement. From a very conservative standpoint, the estimator $\hat{\alpha}_{\mathrm{GM}}^{(1)}$ for $k = 2$ provides for any sample size a relatively favorable efficiency–robustness trade-off with broad protection against both upper and lower outliers and allowing rather high levels of contamination. Its RE values range from .64 to .72 as n increases from 10 to ∞, and its BP values are .20 or above. We wish to do better, however, whenever possible. Thus, if lower outliers are not of concern, in the cases $n \geq 25$ we may replace this estimator by $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ with $k = 2$, yielding improved RE values from .72 to .78 instead of .69 to .72. Further, if, for example, protection against a contamination level of only 10% suffices, then we may advance to $\hat{\alpha}_{\mathrm{GM}}^{(2)}$ with $k = 5$ for $n = 10$ and $k = 4$ for $n \geq 25$, yielding RE values of .97 and .89 to .92, respectively, and corresponding BP values of .10 and $\geq .12$, respectively.

Acknowledgments

The authors are very appreciative of valuable suggestions provided by an anonymous referee. These have led to many improvements in the paper. Also, support of the second author by grants from the Casualty Actuarial Society and the Society of Actuaries, with administrative support from the Actuarial Education and Research Fund, and by NSF Grant DMS-9705209, is gratefully acknowledged.

References

Anscombe, F. J. (1960) Rejection of outliers. *Technometrics*, **2**, 123–147.

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

Brazauskas, V. (1999) *Robust and Nonparametric Statistical Methods for Pareto Distribution Tail Index Estimation, with Actuarial Science Applications. Ph.D. Dissertation*, University of Texas, Dallas.

Brazauskas, V. and Serfling, R. (1999) *Robust Estimation of Tail Parameters for Two-parameter Pareto and Exponential Models via Generalized Quantile Statistics*. Preprint.

Brazauskas, V. and Serfling, R. (2000) Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial J.*, **4**(4), 12–27.

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

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

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

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

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.

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

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

Serfling, R. (2000) "Robust and nonparametric estimation *via* generalized L-statistics: theory, applications, and perspectives", In: *Advances in Methodological and Applied Aspects of Probability and Statistics*, Balakrishnan, N. (Ed.), pp. 197–217. Gordon & Breach.

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