



Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics

VYTARAS BRAZAUSKAS

*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee,
Wisconsin 53201, USA*
E-mail: vytaras@uwm.edu

ROBERT SERFLING

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA
E-mail: serfling@utdallas.edu

[Received January 10, 2000; Revised and Accepted March 15, 2001]

Abstract. 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. New estimators of “generalized quantile” type are introduced and compared with several well-established estimators, for the purpose of identifying which estimators provide favorable trade-offs between efficiency and robustness. Specifically, we examine asymptotic relative efficiency with respect to the (efficient but nonrobust) maximum likelihood estimator, and breakdown point. The new estimators, in particular the generalized median types, are found to dominate well-established and popular estimators corresponding to methods of trimming, least squares, and quantiles. Further, we establish that the least squares estimator is actually deficient with respect to both criteria and should become disfavored. The generalized median estimators manifest a general principle: “smoothing” followed by “medianing” produces a favorable trade-off between efficiency and robustness.

Key words. robust estimation, tail index, Pareto model, exponential model, generalized L-statistics

AMS 1991 Subject Classification. Primary—62F35;
Secondary—62G30.

1. Introduction and preliminaries

We treat robust estimation of the tail index of a two-parameter Pareto distribution, or, equivalently, of the scale of a two-parameter exponential distribution. New estimators of generalized quantile type are introduced and compared with several well-established types, for the purpose of identifying those estimators providing favorable trade-offs between efficiency and robustness.

Specifically, for the Pareto distribution $P(\sigma, \alpha)$ having cdf

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

where $\alpha > 0$ and $\sigma > 0$, we emphasize estimation of the shape parameter α characterizing the tail of the distribution, leaving as an unknown “nuisance parameter” the scale parameter σ . Equivalently, we consider the exponential distribution $E(\mu, \theta)$ having cdf

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

for $\theta > 0$ and $-\infty < \mu < \infty$. Since X has distribution F given by (1.1) if $X \stackrel{d}{=} \sigma e^{U/\alpha}$, where “ $\stackrel{d}{=}$ ” denotes “equal in distribution” and U is “standard exponential” $E(0, 1)$, equivalently $Z = \log X$ has cdf (1.2) and satisfies

$$Z \stackrel{d}{=} \mu + \theta U, \quad (1.3)$$

with $\mu = \log \sigma$ and $\theta = \alpha^{-1}$. Thus estimation of scale and shape in the model $P(\sigma, \alpha)$ is equivalent to estimation of location and scale in $E(\mu, \theta)$. Corresponding to (1.3), a useful linearization of (1.2) results by putting for z the quantile function $G^{-1}(p) = \{z : G(z) \geq p\}$, $0 < p < 1$, yielding

$$G^{-1}(p) = \mu + \theta(-\log(1-p)), 0 < p < 1. \quad (1.4)$$

The model $P(\sigma, \alpha)$ and variants are tractable and effective in many applications involving parametric modeling with high probability in the upper tail, including economics, finance, actuarial science, teletraffic, hydrology, reliability, structural engineering, and, recently, combinatorial search. See Arnold (1983), Johnson, Kotz, and Balakrishnan (1994, Chapter 20), Resnick (1997), Adler, Feldman and Taquq (1997), and Gomes, Selman and Crato (1997). Likewise, the model $E(\mu, \theta)$ has a wide domain of importance. For G a “lifetime” distribution, θ represents the mean life measured from the “minimum lifetime” μ as a starting point. In fatigue failure problems, μ represents a “sensitivity limit”. See Johnson, Kotz and Balakrishnan (1994, Chapter 19), for broad discussion.

Further, the models $P(\sigma, \alpha)$ and $E(\mu, \theta)$ arise in semiparametric contexts. For Y having cdf H with regularly varying tail, the conditional distribution of Y given $Y > t$ converges to a cdf of form (1.1) as $t \rightarrow \infty$, so that the upper ordered observations of a sample from H may be treated approximately as a Pareto sample. Similarly, for Y having cdf in the domain of attraction of an extreme value distribution and satisfying certain von Mises conditions, a suitable transformation of the data not depending upon parameters produces a sample whose upper ordered values behave as a sample from an exponential model. For details, see Galambos (1978), Bingham, Goldie and Teugels

(1987), Reiss (1989), Falk and Marohn (1993), and Falk (1994). In these cases, the model (1.1) applies with the parameter σ known. See Brazauskas and Serfling (2000) for a treatment under this assumption, with special emphasis on applications in actuarial science.

The present treatment will be carried out in terms of the model $E(\mu, \theta)$, focusing on estimation of θ and treating μ as an unknown nuisance parameter. Estimators under consideration will be evaluated on the basis of two competing criteria, efficiency and robustness, more specifically, asymptotic relative efficiency (ARE) with respect to the maximum likelihood estimator (MLE), and breakdown point (BP). The MLE being efficient but not robust, we seek to identify competitors having $BP > 0$ along with high ARE. New generalized quantile estimators for this problem are introduced and evaluated in Section 2, and in Section 3 they are compared with several well-established and popular estimators corresponding to methods of trimming, least squares, and quantiles. Excluded from consideration are the popular, but nonrobust, method of moments estimators. Nor need we consider the robust M -estimators that have been developed for this problem, as they do not improve upon the more directly formulated trimmed mean estimators that we shall consider (see Kimber, 1983b).

Our findings are as follows. For the relatively popular least squares type estimator we obtain a clarified perspective: it is both nonrobust and nonefficient and may reasonably be discarded from practice. The quantile and trimmed mean type estimators offer favorable trade-offs between efficiency and robustness, with the most efficient trimmed types outperforming the most efficient quantile types. (This corroborates studies by Kimber, 1983a,b; Gather, 1986, and Willemain et al., 1992, in which trimmed means fared well among competitors for efficient and robust estimation of θ in $E(0, \theta)$.) The new generalized quantile type estimators, however, in particular the generalized median types, dominate all the other competitors. Overall, the trimmed mean and generalized median statistics stand apart from the others as offering the most competitive trade-offs between efficiency and robustness, and these we recommend for practical use.

The above findings are based on large-sample efficiency and robustness criteria, ARE and BP, for $n \rightarrow \infty$. It turns out that the favorable performance of generalized median estimators holds also for small sample sizes, as shown by Brazauskas and Serfling (2001) in a simulation study with emphasis on sample sizes $n = 10$ and 25. Exact relative efficiency and exact breakdown points are used as criteria, and premium-protection plots based on outlier contamination models are developed to exhibit favorable efficiency—robustness trade-offs for the above collection of competing estimators.

The overall superiority of the generalized median estimators appears to be explained by a simple principle: “Smoothing” followed by “medianing” produces estimators possessing relatively high robustness at relatively small sacrifice of efficiency. (This also serves as an interpretation of the excellent performance of the Hodges-Lehmann location estimator in the classical nonparametric location problem.)

In the remainder of this introduction, we discuss the impact of errors in Pareto tail index estimation and formulate precisely our efficiency and robustness criteria.

1.1. Impact of errors

A small relative error in estimation of α in $P(\sigma, \alpha)$ can produce a large relative error in estimated quantiles or tail probabilities based on α . Even small improvements in estimation methods thus can yield significant improvements in applications.

For estimation of the quantile q_ε corresponding to upper tail probability ε , it follows from (1.1) that $q_\varepsilon = \sigma\varepsilon^{-1/\alpha}$. Thus, for \hat{q}_ε defined by putting $\hat{\alpha}$ for α , we have $\hat{q}_\varepsilon/q_\varepsilon = \varepsilon^{1/\alpha-1/\hat{\alpha}}$. Similarly, for estimation of the tail probability ε above a specified threshold q , it follows from (1.1) that $\varepsilon = (\sigma/q)^\alpha$. Thus, for $\hat{\varepsilon}$ defined by putting $\hat{\alpha}$ for α , we have $\hat{\varepsilon}/\varepsilon = (\sigma/q)^{\hat{\alpha}-\alpha} = \varepsilon^{(\hat{\alpha}/\alpha)-1}$. For example, for $\varepsilon = 0.001$, underestimation of $\alpha = 1$ by only 5% thus produces overestimation of $q_{0.001}$ by 44% and underestimation of $\alpha = 1.5$ by 5% produces overestimation of $q_{0.001}$ by 27%. Also, underestimation of any value of α by 5% produces overestimation of $\varepsilon = 0.001$ by 41% and of $\varepsilon = 0.0001$ by 58%. Likewise, for $\varepsilon = 0.001$, overestimation of $\alpha = 1$ by 10% produces underestimation of $q_{0.001}$ by 47% and overestimation of $\alpha = 1.5$ by 10% produces underestimation of $q_{0.001}$ by 34%. Also, overestimation of any value of α by 10% produces underestimation of $\varepsilon = 0.001$ by 50% and of $\varepsilon = 0.0001$ by 60%.

Tail probabilities and quantiles in the range of $\varepsilon = 0.001$ or $\varepsilon = 0.0001$ are common in actuarial and extreme value applications. For example, the Dutch government's standard for sea dikes is that the sea level not exceed the dike level in a given year with probability at least 0.9999 (see Dekkers and de Haan, 1989, for discussion).

1.2. 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 Z_1, \dots, Z_n from $E(\mu, \theta)$, the MLE of θ (for μ unknown) is readily derived:

$$\hat{\theta}_{ML} = \bar{Z}_n - Z_{n1},$$

where \bar{Z}_n denotes the sample mean and Z_{n1} the minimum sample value. The exact distribution of $\hat{\theta}_{ML}$ is given (e.g. Lehmann, 1983, Problem 1.5.18) by the statement

$$\frac{2n\hat{\theta}_{ML}}{\theta} \text{ has cdf } \chi_{2(n-1)}^2, \quad (1.5)$$

where χ_ν^2 denotes the chi-square distribution with ν degrees of freedom. This yields easily the asymptotic distribution: $\hat{\theta}_{ML}$ is asymptotically normal with mean θ and variance θ^2/n , denoted $\text{AN}(\theta, \theta^2/n)$.

For a competing estimator, efficiency is characterized in terms of its 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. All competing estimators $\hat{\theta}$ considered here will be $AN(\theta, c\theta^2/n)$ for some constant $c > 0$, yielding $ARE(\hat{\theta}, \hat{\theta}_{ML}) = c^{-1}$.

By standard asymptotic theory for transformations (Serfling, 1980, Chapter 3), for $\hat{\theta}_n$ $AN(\theta, c\theta^2/n)$ it follows that $\hat{\alpha}_n = 1/\hat{\theta}_n$ is $AN(\alpha, c\alpha^2/n)$ with the same constant c . Thus comparison of estimators with respect to ARE is the same whichever parameterization we use, that based on $P(\sigma, \alpha)$ or that based on $E(\mu, \theta)$.

1.3. Robustness criterion: Breakdown point

A popular and effective criterion for robustness of an estimator is its (finite-sample) breakdown point (BP), loosely characterized 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. In the present context, protection against upper contamination is sometimes more important than against lower contamination, 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.

We seek estimators possessing favorable ARE while also having $UBP > 0$ and $LBP > 0$ as well. In particular, $\hat{\theta}_{ML}$ is readily seen to have $LBP = UBP = 0$ and thus is nonrobust and rejected as a contender for robust estimation of θ .

An estimator $\hat{\theta}$ and its reciprocal $\hat{\alpha} = 1/\hat{\theta}$ for estimation of α possess numerically identical breakdown points, and so comparison of estimators with respect to BP is the same whether the parameterization based on $P(\sigma, \alpha)$ or that based on $E(\mu, \theta)$ is used.

2. Generalized quantile statistics

We introduce and study two estimators of ‘‘generalized quantile’’ (GQ) type for estimation of θ in $E(\mu, \theta)$, based on a sample Z_1, \dots, Z_n having distribution (1.2).

2.1. General formulation of GQ statistics

For a ‘‘kernel’’ $h(z_1, \dots, z_k)$ invariant under permutations of its k arguments, denote by H_G the cdf of $h(Z_1, \dots, Z_k)$ induced by the cdf G of the Z_i 's. If h is designed to make the target

parameter θ the median of the cdf H_G , then a natural estimator is given by the ‘‘generalized median statistic’’ $\hat{\theta}_{GM} = \text{Median}\{h(Z_{i_1}, \dots, Z_{i_k})\}$. More generally, for any $0 < p < 1$, if

$$\theta = H_G^{-1}(p) \quad (= p\text{th quantile of } H_G) \quad (2.1)$$

is satisfied, then an estimator of θ is generated by taking the p th quantile of a suitable sample cdf for estimation of H_G . For this we define

$$\hat{H}_n(y) = \binom{n}{k}^{-1} \sum \mathbf{1}\{h(Z_{i_1}, \dots, Z_{i_k}) \leq y\}, y \in \mathbb{R},$$

where the sum is over all k -sets of distinct indices $\{i_1, \dots, i_k\}$ from $\{1, \dots, n\}$. Since \hat{H}_n estimates H_G , the corresponding ‘‘generalized quantile statistic’’

$$\hat{\theta}_{GQ,p} = \hat{H}_n^{-1}(p) \quad (2.2)$$

estimates $\theta = H_G^{-1}(p)$. The estimator $\hat{\theta}_{GQ,p}$ being a special case of ‘‘generalized L -statistic’’, as treated by Serfling (1984) and Choudhury and Serfling (1988), we have: $\hat{\theta}_{GQ,p}$ is asymptotically normal with mean θ and variance

$$\frac{k^2 \zeta}{h_G^2(\theta)n}, \quad (2.3)$$

where h_G is the density of H_G , $\zeta = \text{Var}(w_h(Z))$, and $w_h(z) = P\{h(z, Z_1, \dots, Z_{k-1}) \leq \theta\}$. For the kernels selected below, (2.3) will have the form $\gamma\theta^2/n$ for some choice of γ . For n so large that the $O(n^k)$ computational complexity of $\hat{\theta}_{GQ,p}$ becomes prohibitive, one simply estimates $\hat{\theta}_{GQ,p}$ using only the evaluations $h(Z_{i_1}, \dots, Z_{i_k})$ for a random sample of size N of the $\binom{n}{k}$ subsets $\{i_1, \dots, i_k\}$, where N is quite large, say 10^6 or 10^8 . This renders the computational burden negligible while nevertheless maintaining any desired computational accuracy.

2.2. Estimator 1

We start with the kernel $h_0(z_1, \dots, z_k) = k^{-1} \sum_{j=1}^k z_j - \min\{z_1, \dots, z_k\}$, which evaluated at Z_{i_1}, \dots, Z_{i_k} gives the MLE of θ based on just those observations. In order to modify h_0 to become p th quantile unbiased for estimation of θ , i.e. to satisfy condition (2.1), we use the fact that $(2k)h_0(Z_1, \dots, Z_k)/\theta$ has cdf $\chi_{2(k-1)}^2$, which follows by (1.5). Denoting by $M_{\nu,p}$ the p th quantile of χ_{ν}^2 , it follows that (2.1) is satisfied by cdf $H_G^{(1,p)}$ based on the kernel

$$h^{(1,p)}(z_1, \dots, z_k) = \frac{2k}{M_{2(k-1),p}} h_0(z_1, \dots, z_k).$$

Table 1. ARE($\hat{\theta}_{GQ,p}^{(1)}, \hat{\theta}_{ML}$) and $\gamma_{k,p}^{(1)}$, for $k = 2 : 10$ and $p = 0.10, 0.25, 0.50, 0.75, 0.90$.

	k								
	2	3	4	5	6	7	8	9	10
ARE($\hat{\theta}_{GQ,0.10}^{(1)}, \hat{\theta}_{ML}$)	0.75	0.64	0.65	0.68	0.72	0.75	0.77	0.79	0.81
$\gamma_{k,0.10}^{(1)}$	1.335	1.553	1.536	1.460	1.390	1.335	1.293	1.260	1.233
ARE($\hat{\theta}_{GQ,0.25}^{(1)}, \hat{\theta}_{ML}$)	0.74	0.70	0.74	0.78	0.81	0.84	0.86	0.87	0.89
$\gamma_{k,0.25}^{(1)}$	1.343	1.436	1.360	1.286	1.232	1.194	1.166	1.145	1.128
ARE($\hat{\theta}_{GQ,0.50}^{(1)}, \hat{\theta}_{ML}$)	0.72	0.74	0.80	0.85	0.88	0.90	0.91	0.92	0.93
$\gamma_{k,0.50}^{(1)}$	1.388	1.349	1.247	1.182	1.142	1.115	1.096	1.083	1.072
ARE($\hat{\theta}_{GQ,0.75}^{(1)}, \hat{\theta}_{ML}$)	0.64	0.72	0.79	0.83	0.86	0.88	0.90	0.91	0.92
$\gamma_{k,0.75}^{(1)}$	1.561	1.395	1.268	1.199	1.158	1.131	1.112	1.098	1.087
ARE($\hat{\theta}_{GQ,0.90}^{(1)}, \hat{\theta}_{ML}$)	0.49	0.59	0.67	0.73	0.76	0.79	0.81	0.83	0.87
$\gamma_{k,0.90}^{(1)}$	2.037	1.686	1.486	1.378	1.311	1.266	1.233	1.207	1.151

The corresponding estimator $\hat{\theta}_{GQ,p}^{(1)}$ is AN($\theta, \gamma_{k,p}^{(1)} \theta^2/n$) with $\gamma_{k,p}^{(1)}$ determined via (2.3). Values of $\gamma_{k,p}^{(1)}$ and the corresponding ARE = $1/\gamma_{k,p}^{(1)}$ for selected k and p are provided in Table 1 above. Regarding breakdown behavior, it is established in Lemma B.1 that $\hat{\theta}_{GQ,p}^{(1)}$ has (asymptotic) LBP = UBP given by $b(p) = \min\{p^{1/k}, 1 - p^{1/k}\}$, for which values for selected k and p are provided in Table 2 below. The function $b(p)$ attains maximum value $1/2$ at $p_{\max} = (1/2)^k$, for which values are provided for $k = 2 : 10$ in Table 3.

Examination of Tables 1–3 indicates the following. For fixed p , ARE increases with k while BP decreases. For fixed $k \geq 3$, ARE is favorable for middle values of p but decreases as p tends to 0 or 1, whereas BP is favorable for low values of p . The price of the optimal BP of $1/2$ is thus a serious degradation of ARE. The anomalous case $k = 2$, in which both ARE and BP improve as $p \rightarrow 0$, is outperformed by the $k \geq 4$ cases. Overall, the choice $p = 0.50$ offers a favorable trade-off: very strong ARE values combined with relatively high BP values. (Among the values of p in Table 1 the ARE is best uniformly over $k \geq 3$ for $p = 0.50$, but with a more refined range of p one finds that $p = 0.50$ is only nearly optimal.)

On the basis of these considerations, and for simplicity, we recommend the case $p = 0.50$ as suitable in practice. The estimator $\hat{\theta}_{GM}^{(1)} = \hat{\theta}_{GQ,0.50}^{(1)}$ has ARE included in Table 1 and LBP = UBP given by $b(1/2) = 1 - (1/2)^{1/k}$, these values included in Table 2.

Table 2. $b(p)$ for $k = 2 : 10$ and $p = 0.10, 0.25, 0.50, 0.75, 0.90$.

p	k								
	2	3	4	5	6	7	8	9	10
0.10	0.316	0.464	0.438	0.369	0.319	0.280	0.250	0.226	0.206
0.25	0.500	0.370	0.293	0.242	0.206	0.180	0.159	0.143	0.129
0.50	0.293	0.206	0.159	0.129	0.109	0.094	0.083	0.074	0.067
0.75	0.134	0.091	0.069	0.056	0.047	0.040	0.035	0.032	0.028
0.90	0.051	0.035	0.026	0.021	0.017	0.015	0.013	0.012	0.011

Table 3. Values of p_{\max} , for $k = 2 : 10$.

k								
2	3	4	5	6	7	8	9	10
0.250	0.125	0.062	0.031	0.016	0.008	0.004	0.002	0.001

Writing M_ν for $M_{\nu,0.50}$, the values $M_{2(k-1)}$ and multiplicative correction factors $2k/M_{2(k-1)}$ needed to construct $\hat{\theta}_{GM}^{(1)}$ for $k = 2 : 10$ are included in Table 4.

2.3. Estimator 2

An alternative to $\hat{\theta}_{GQ,p}^{(1)}$ that possesses greater ARE at the cost of $LBP = 0$ is developed as follows. We start with the kernel $\tilde{h}_0(z_1, \dots, z_k; \mu) = k^{-1} \sum_{j=1}^k z_j - \mu$, which involves the unknown nuisance parameter μ as well as arguments z_1, \dots, z_k to be filled in with sample values. Using the well-known and easily proved result that $(2k)\tilde{h}_0(Z_1, \dots, Z_k; \mu)/\theta$ has cdf χ_{2k}^2 , we see that (2.1) is satisfied by the cdf $H_{G,\mu}^{(2,p)}$ based on the kernel

$$h^{(2,p)}(z_1, \dots, z_k; \mu) = \frac{2k}{M_{2k,p}} \tilde{h}_0(z_1, \dots, z_k; \mu).$$

The corresponding statistic $\tilde{\theta}_{GQ,p}^{(2)}$ cannot, however, serve as an actual estimator since it involves the unknown μ in its definition. We therefore put for μ its classical MLE Z_{n1} based on the full sample, thus producing the strictly data-based kernel

$$\tilde{h}^{(2,p)}(z_1, \dots, z_k; Z_{n1}) = \frac{2k}{M_{2k,p}} \tilde{h}_0(z_1, \dots, z_k; Z_{n1}).$$

Table 4. $M_{2(k-1)}$, M_{2k} , $2k/M_{2(k-1)}$, and $2k/M_{2k}$, for $k = 2 : 10$.

k	$M_{2(k-1)}$	$2k/M_{2(k-1)}$	$2k/M_{2k}$
2	1.3863*	2.88538	1.19165
3	3.3567	1.78747	1.12189
4	5.3481	1.49585	1.08931
5	7.3441	1.36163	1.07046
6	9.3418	1.28455	1.05187
7	11.3403	1.23453	1.04953
8	13.3393	1.19947	1.04313
9	15.3385	1.17352	1.03819
10	17.3379	1.15354	1.03427
11	19.3374	—	—

Note. * $2 \log 2 = 1.3863$.

Table 5. ARE($\hat{\theta}_{GM}^{(2)}, \hat{\theta}_{ML}$) and $\gamma_{k,0.50}^{(2)}$, for $k = 2 : 10$.

	k								
	2	3	4	5	6	7	8	9	10
ARE($\hat{\theta}_{GM}^{(2)}, \hat{\theta}_{ML}$)	0.78	0.88	0.92	0.94	0.96	0.97	0.97	0.98	0.98
$\gamma_{k,0.50}^{(2)}$	1.280	1.141	1.088	1.061	1.044	1.035	1.028	1.023	1.019

The presence of Z_{n1} in each kernel evaluation makes the corresponding estimator $\hat{\theta}_{GQ,p}^{(2)}$ nonrobust against lower outliers, i.e. LBP = 0. By Lemma B.2, however, $\hat{\theta}_{GQ,p}^{(2)}$ is robust against upper outliers, having (asymptotic) UBP given by $c(p) = 1 - p^{1/k}$.

Since $\hat{\theta}_{GQ,p}^{(2)}$ involves a kernel with an estimated parameter, the asymptotic normality result given in (2.3) does not apply directly. But $\hat{\theta}_{GQ,p}^{(2)}$ can be approximated by $\tilde{\theta}_{GQ,p}^{(2)}$ sufficiently well to share the asymptotic normality of the latter. Noting that $|\hat{\theta}_{GQ,p}^{(2)} - \tilde{\theta}_{GQ,p}^{(2)}| \leq |Z_{n1} - \mu|$ and using the easily established result

$$Z_{n1} - \mu = O_p(n^{-1}), n \rightarrow \infty, \tag{2.4}$$

we obtain $n^{1/2}(\hat{\theta}_{GQ,p}^{(2)} - \tilde{\theta}_{GQ,p}^{(2)}) \xrightarrow{p} 0$. By Slutsky's theorem follows that $\hat{\theta}_{GQ,p}^{(2)}$, like $\tilde{\theta}_{GQ,p}^{(2)}$, is AN($\theta, \gamma_{k,p}^{(2)}\theta^2/n$), with $\gamma_{k,p}^{(2)}$ determined via (2.3).

From considerations similar to those for Estimator 1, we find that $p = 0.50$ again yields a favorable trade-off between efficiency and robustness and we henceforth focus on this case, using the kernel $h^{(2)}(z_1, \dots, z_k; \mu) = \frac{2k}{M_{2k}} \tilde{h}_0(z_1, \dots, z_k; \mu)$ with Z_{n1} for μ . For kernel sizes $k = 2 : 10$, values of M_{2k} and multiplicative correction factors $2k/M_{2k}$ needed to construct the desired estimator $\hat{\theta}_{GM}^{(2)}$ appear in Table 4 above. Since the BP functions $b(\cdot)$ and $c(\cdot)$ agree for $p = 0.50$, i.e. $c(0.50) = b(0.50)$, $\hat{\theta}_{GM}^{(1)}$ and $\hat{\theta}_{GM}^{(2)}$ have the same UBP, given by the $p = 0.50$ row of Table 2 above. Values of $\gamma_{k,0.50}^{(2)}$ and the ARE = $1/\gamma_{k,0.50}^{(2)}$ of $\hat{\theta}_{GM}^{(2)}$ for $k = 2 : 10$ are given in Table 5.

The above results show that $\hat{\theta}_{GM}^{(2)}$ significantly outperforms $\hat{\theta}_{GM}^{(1)}$ with respect to asymptotic efficiency, and simulation studies confirm this comparison for small n as well. We recommend Estimator 1 if protection against both upper and lower outliers is desired, but Estimator 2 if protection against lower outliers is not of concern.

2.3.1. A further modification. Unlike Estimator 1, which is exactly p th quantile unbiased for θ , Estimator 2 is only asymptotically p th quantile unbiased. For small sample size applications, it can be made exactly p th quantile unbiased by replacing $M_{2k,p}$ in the definition of its kernel by the p th quantile of the cdf of $(2k)h_0(Z_1, \dots, Z_k; Z_{n1})/\theta$. By (1.3) and Corollary A.1, the relevant cdf is the mixture

$$\left(1 - \frac{k}{n}\right)\chi_{2k}^2 + \frac{k}{n}\chi_2^2(k-1).$$

3. Comparisons and conclusions

We evaluate three popular types of estimator for θ in $E(\mu, \theta)$ with respect to our joint criteria of efficiency and robustness and then compare these and the GQ estimators, in order to arrive at an overall perspective.

3.1. Trimmed mean estimators

A trimmed mean for specified α and β satisfying $0 \leq \alpha < 1$ and $0 \leq \beta < 1 - \alpha$ is formed by discarding the proportion α lowermost observations and proportion β uppermost observations and then averaging the remaining ones in some sense. In particular, by analogy with the case of μ known treated by Kimber (1983a,b), but replacing μ by Z_{n1} (as in Section 2.3), we introduce the estimator

$$\hat{\theta}_T = \sum_{i=2}^n c_{ni}(Z_{ni} - Z_{n1}), \tag{3.1}$$

with $c_{ni} = 0$ for $2 \leq i \leq [(n-1)\alpha] + 1$ and $n - [(n-1)\beta] + 1 \leq i \leq n$, and $c_{ni} = 1/d(\alpha, \beta, n-1)$ for $[(n-1)\alpha] + 2 \leq i \leq n - [(n-1)\beta]$, where $[\cdot]$ denotes ‘‘greatest integer part’’ and $d(\alpha, \beta, n) = \sum_{j=[n\alpha]+1}^{n-n\beta} \sum_{i=1}^j (n-i+1)^{-1}$. (This choice of c_{ni} ’s makes $\hat{\theta}_T$ mean-unbiased.) Due to the presence of Z_{n1} , robustness is gained only against upper outliers, i.e. LBP = 0 and UBP = $[(n-1)\beta]/n \rightarrow \beta, n \rightarrow \infty$. To obtain the ARE of $\hat{\theta}_T$, we use the fact that $\hat{\theta}_T$ has the same distribution as

$$\hat{\theta}_T^{(0)} = \sum_{i=2}^n c_{ni}(Z_{n-1,i} - \mu), \tag{3.2}$$

which is $AN(\theta, D_{\alpha\beta}\theta^2/n)$ with $D_{\alpha\beta}$ computable by standard results for L-statistics (e.g. Serfling (1980), Section 8.2.4, or Lehmann (1983), Section 5.4). It follows that existing ARE results for $\hat{\theta}_T^{(0)}$ in comparison with the MLE of θ when μ is known apply unchanged for $\hat{\theta}_T$ in comparison with $\hat{\theta}_{ML}$, with $ARE = D_{\alpha\beta}^{-1}$. In particular, therefore, for $\alpha = 0$ or $\alpha = \beta$, and β taking values 0.05, 0.10, 0.15, 0.20, and 0.25, the values of ARE and corresponding values of $D_{0\beta}$ and $D_{\beta\beta}$ may be taken from Table 1 of Kimber (1983a) and Table II of Kimber (1983b). For each of these five choices of β the ARE’s for the two cases

Table 6. $ARE(\hat{\theta}_T, \hat{\theta}_{ML})$ and $D_{\alpha\beta}$ for $\alpha = 0$ or $\alpha = \beta$, and selected β .

	β				
	0.05	0.10	0.15	0.20	0.25
$ARE(\hat{\theta}_T, \hat{\theta}_{ML})$	0.92	0.85	0.78	0.72	0.67
$D_{0\beta}, D_{\beta\beta}$	1.09	1.18	1.28	1.39	1.49

$\alpha = 0$ and $\alpha = \beta$ happen to agree within two decimal places. Table 6 provides ARE and D values for comparison of $\hat{\theta}_T$ with $\hat{\theta}_{ML}$.

3.2. Estimators based on the linearized model

Certain estimators popular in the present problem are motivated by the linear model (1.4). With the usual sample cdf $\hat{G}_n(z) = n^{-1} \sum_{i=1}^n \mathbf{1}\{Z_{ni} \leq z\}$, $z \in \mathbb{R}$, and corresponding sample quantile function

$$\hat{G}_n^{-1}(p) = Z_{n, [np]}, 0 < p < 1, \quad (3.3)$$

where $Z_{n1} \leq Z_{n2} \leq \dots \leq Z_{nn}$ denote the ordered sample values and $[x]$ denotes the least integer $\geq x$, and defining $\varepsilon = \hat{G}_n^{-1}(p) - G^{-1}(p)$, we obtain an exact reexpression of (1.4) as

$$\hat{G}_n^{-1}(p) = \mu + \theta u + \varepsilon, 0 < p < 1, \quad (3.4)$$

where $u = -\log(1-p)$, $0 < p < 1$. Choosing n values of p such that $\hat{G}_n^{-1}(p)$ generates the set of order statistics $\{Z_{ni}, 1 \leq i \leq n\}$ as per relation (3.3), i.e. for choices $p_{n1}^*, \dots, p_{nn}^*$ satisfying

$$p_{n1}^* \in \left(0, \frac{1}{n}\right], p_{n2}^* \in \left(\frac{1}{n}, \frac{2}{n}\right], \dots, p_{n, n-1}^* \in \left(\frac{n-2}{n}, \frac{n-1}{n}\right], p_{nn}^* \in \left(\frac{n-1}{n}, 1\right), \quad (3.5)$$

we obtain from (3.4) a set of n equations for the two unknowns μ and θ :

$$Z_{ni} = \mu + \theta u_{ni}^* + \varepsilon_{ni}^*, \quad 1 \leq i \leq n, \quad (3.6)$$

where $u_{ni}^* = -\log(1-p_{ni}^*)$ and $\varepsilon_{ni}^* = Z_{ni} - G^{-1}(p_{ni}^*)$, $1 \leq i \leq n$. These equations may be interpreted from the standpoint of the usual linear regression model, yielding estimates of μ and θ by fitting a straight line through the scatterplot of points $\{(Z_{ni}, u_{ni}^*), 1 \leq i \leq n\}$, or through a strategically selected subset, applying either ordinary or weighted least squares. Two well-established scenarios are discussed below.

3.2.1. Least squares estimators. Using the full set of equations (3.6), with $p_{ni}^* = p_{ni}$, where $p_{ni} = i/n$, $1 \leq i \leq n-1$, and $p_{nn} = n/(n+1)$, ordinary least squares regression yields a so-called ‘‘least squares’’ estimator of θ ,

$$\hat{\theta}_{LS} = \frac{n^{-1} \sum_{i=1}^n c_{ni} Z_{ni} - \bar{c}_n \bar{Z}_n}{n^{-1} \sum_{i=1}^n c_{ni}^2 - (\bar{c}_n)^2},$$

where $c_{ni} = -\log(1 - p_{ni})$ and $\bar{c}_n = n^{-1} \sum_{i=1}^n c_{ni}$. Clearly, $LBP = UBP = 0$ for $\hat{\theta}_{LS}$ and hence it is nonrobust.

In the context of the Pareto model, some discussions in the literature suggest that the LS estimator is competitive in efficiency with the MLE. We clarify, therefore, the performance of $\hat{\theta}_{LS}$ by establishing (in Appendix C) the following result on its asymptotic normality, which indicates that it is actually poor in efficiency, having $ARE = 0.50$.

Theorem 3.1: *The estimator $\hat{\theta}_{LS}$ is $AN(\theta, 2\theta^2/n)$.*

3.2.2. Estimators based on k selected quantiles. Here we use a selected subset of the equations (3.6). For given integer k , let values $0 < p_1 < \dots < p_k < 1$ be chosen and suppose that $n > k$ is sufficiently large that the p_i 's fall in k distinct members of the subintervals in (3.5). In this case the k equations from (3.6) corresponding to $\hat{G}_n^{-1}(p_i)$ in (3.4), $1 \leq i \leq k$, are given by

$$Z_{n, \lceil np_i \rceil} = \mu + \theta u_i + \varepsilon_{ni}, \quad 1 \leq i \leq k, \tag{3.7}$$

where $u_i = -\log(1 - p_i)$ and $\varepsilon_{ni} = Z_{n, \lceil np_i \rceil} - G^{-1}(p_i)$, $1 \leq i \leq k$. Thus estimates of μ and θ result from fitting a straight line to the scatterplot of points $\{(Z_{n, \lceil np_i \rceil}, u_i), 1 \leq i \leq k\}$. In particular (see Saleh and Ali, 1966, for details), weighted least squares regression yields

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

where $b_1 = -L^{-1}(u_2 - u_1)/(e^{u_2} - e^{u_1})$, $b_i = L^{-1}[(u_i - u_{i-1})/(e^{u_i} - e^{u_{i-1}}) - (u_{i+1} - u_i)/(e^{u_{i+1}} - e^{u_i})]$, $2 \leq i \leq k - 1$, $b_k = L^{-1}(u_k - u_{k-1})/(e^{u_k} - e^{u_{k-1}})$, and $L = \sum_{i=2}^k (u_i - u_{i-1})^2 / (e^{u_i} - e^{u_{i-1}})$. In the equivalent context of estimation of α in Pareto(σ, α), this is the ‘‘quantile’’ approach 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.

We see that $\hat{\theta}_Q$ has $LBP = p_1$ and $UBP = 1 - p_k$ and is robust if p_1 and p_k are bounded away from 0 and 1. Further, $\hat{\theta}_Q$ is $AN(\theta, L^{-1}\theta^2/n)$ and thus has $ARE = L$. The asymptotic normality holds not only for fixed p_1, \dots, p_k but also for p_1 tending to 0 at rate $O(n^{-1})$ and/or for p_2, \dots, p_k having limits in $(0, 1)$.

As argued in Saleh and Ali (1966) (see also Koutrouvelis, 1981), the optimal choice of p_1 is

$$p_1^\circ = \frac{1}{n + 0.5} \tag{3.8}$$

and optimal choices of p_2, \dots, p_k are then obtained by minimization of the generalized variance of the joint estimators of μ and θ , subject to (3.8). Alternatively, however, the

optimal p_2, \dots, p_k can be found directly by reduction to the case of μ known; i.e. making use of (3.8) and noting that the sum of the b_i 's is 0, we may express the optimal estimator as

$$\hat{\theta}_Q^{opt,k} = \sum_{i=2}^k b_i(Z_{n, \lceil np_i \rceil} - Z_{n,1}),$$

for which it follows by (2.4) that the optimal choices of p_2, \dots, p_k are 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 model $E(0, \theta)$ by a linear function of $k - 1$ order statistics. In particular,

- For $k = 2$ the optimal p_i 's are $p_1 = p_1^\circ$ and $p_2 = 0.80$, and for $\hat{\theta}_Q^{opt,2}$ we have LBP = 0, UBP = 0.20, and ARE = 0.649.
- For $k = 5$ the optimal p_i 's are $p_1 = p_1^\circ, p_2 = 0.45, p_3 = 0.74, p_4 = 0.91$, and $p_5 = 0.98$, and for $\hat{\theta}_Q^{opt,5}$ we have LBP = 0, UBP = 0.02, and ARE = 0.926.

These examples indicate that the increase in ARE of $\hat{\theta}_Q$ by choosing optimal levels and taking k larger is accompanied by severe reduction in BP. If one desires to maintain relatively high BP, then nonoptimal quantile levels must be selected. For example:

- With $k = 2, p_1 = 0.10$, and $p_2 = 0.90$, for $\hat{\theta}_Q$ we have LBP = UBP = 0.10 and ARE = 0.543.
- With $k = 4, p_1 = p_1^\circ, p_2 = 0.25, p_3 = 0.50$, and $p_4 = 0.75$, for $\hat{\theta}_Q$ we have LBP = 0, UBP = 0.25, and ARE = 0.735.
- With $k = 5, p_1 = 0.13, p_2 = 0.32, p_3 = 0.50, p_4 = 0.69$, and $p_5 = 0.87$, we have LBP = UBP = 0.13 and ARE = 0.73. For later reference, we designate this estimator by $\hat{\theta}_Q^*$.

3.3. Comparisons and conclusions

In Table 7 below we examine together the *ML, LS, Q, T*, and *GM* estimators with respect to our joint criteria of efficiency and robustness.

The following conclusions emerge:

- The ‘‘least squares’’ estimator is neither robust nor efficient and thus is not competitive.
- Optimal ‘‘quantile’’ type estimators are improved upon by the ‘‘trimmed’’ types. For example, $\hat{\theta}_Q^{opt}$ for $k = 2$ with ARE = 0.65, LBP = 0, and UBP = 0.20 is dominated by $\hat{\theta}_T$ for $\alpha = \beta = 0.20$ with ARE = 0.72, LBP = 0, and UBP = 0.20. Also, $\hat{\theta}_Q^{opt}$ for $k = 5$ with ARE = 0.93, LBP = 0, and UBP = 0.02 is dominated by $\hat{\theta}_T$ for $\alpha = \beta = 0.04$ with ARE = 0.93, LBP = 0, and UBP = 0.04.

Table 7. ARE and BP for selected estimators of θ .

Estimator	ARE	LBP	UBP
Mle	1	0	0
$\hat{\theta}_{LS}$	0.50	0	0
$\hat{\theta}_Q^{opt,2}$	0.65	0	0.20
$\hat{\theta}_Q^{opt,5}$	0.93	0	0.02
$\hat{\theta}_Q^*$ ($k = 5$)	0.73	0.13	0.13
$\hat{\theta}_T, \alpha = \beta = 0.25$	0.67	0	0.25
$\hat{\theta}_T, \alpha = \beta = 0.20$	0.72	0	0.20
$\hat{\theta}_T, \alpha = \beta = 0.15$	0.78	0	0.15
$\hat{\theta}_T, \alpha = \beta = 0.10$	0.85	0	0.10
$\hat{\theta}_T, \alpha = \beta = 0.05$	0.92	0	0.05
$\hat{\theta}_T, \alpha = \beta = 0.04$	0.93	0	0.04
$\hat{\theta}_{GM}^{(1)}, k = 2$	0.72	0.29	0.29
$\hat{\theta}_{GM}^{(1)}, k = 3$	0.74	0.21	0.21
$\hat{\theta}_{GM}^{(1)}, k = 4$	0.80	0.16	0.16
$\hat{\theta}_{GM}^{(1)}, k = 5$	0.85	0.13	0.13
$\hat{\theta}_{GM}^{(1)}, k = 10$	0.93	0.07	0.07
$\hat{\theta}_{GM}^{(2)}, k = 2$	0.78	0	0.29
$\hat{\theta}_{GM}^{(2)}, k = 3$	0.88	0	0.21
$\hat{\theta}_{GM}^{(2)}, k = 4$	0.92	0	0.16
$\hat{\theta}_{GM}^{(2)}, k = 5$	0.94	0	0.13
$\hat{\theta}_{GM}^{(2)}, k = 10$	0.98	0	0.07

- In turn, the “trimmed” types (and thus also the optimal quantile types) are improved upon by the “generalized median” types. For example, $\hat{\theta}_T$ for $\alpha = \beta = 0.20$ with ARE = 0.72, LBP = 0, and UBP = 0.20 is dominated by $\hat{\theta}_{GM}^{(1)}$ for $k = 2$ with ARE = 0.72 and LBP = UBP = 0.29, by $\hat{\theta}_{GM}^{(1)}$ for $k = 3$ with ARE = 0.74 and LBP = UBP = 0.21, and by $\hat{\theta}_{GM}^{(2)}$ for $k = 3$ with ARE = 0.88, LBP = 0, and UBP = 0.21. Likewise, $\hat{\theta}_T$ for $\alpha = \beta = 0.05$ with ARE = 0.92, LBP = 0, and UBP = 0.05 is dominated by $\hat{\theta}_{GM}^{(1)}$ for $k = 10$ with ARE = 0.93 and LBP = UBP = 0.07, and by $\hat{\theta}_{GM}^{(2)}$ for $k = 10$ with ARE = 0.98 and LBP = 0, UBP = 0.07.
- Further, the nonoptimal quantile estimators are improved upon by the “generalized median” types. For example, $\hat{\theta}_Q^*$ with ARE = 0.73 and LBP = UBP = 0.13 is improved upon by $\hat{\theta}_{GM}^{(1)}$ for $k = 5$ with ARE = 0.85, and LBP = UBP = 0.13, as well as by $\hat{\theta}_{GM}^{(1)}$ for $k = 3$ with ARE = 0.74, and LBP = UBP = 0.21.

Interpretive conclusion:

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

Practical Recommendations:

The maximum likelihood estimator 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 approach should be used more cautiously if not dropped, and the standard least squares approach should be dropped from practical use.

Appendix A: A result for standard exponential

For a sample U_1, \dots, U_n from $E(0, 1)$, denote the ordered sample values by $U_{n1} \leq U_{n2} \leq \dots \leq U_{nn}$. A simple but productive distributional result is the following.

Lemma A.1: For $n > 1$ and $1 \leq k \leq n$, the joint cdf of $U_1 - U_{n1}, \dots, U_k - U_{nk}$ is

$$\begin{aligned}
 &F_{U_1 - U_{n1}, \dots, U_k - U_{nk}}(t_1, \dots, t_k) \\
 &= \frac{1}{n} \left[\sum_{j=1}^k \prod_{1 \leq i \leq k; i \neq j} (1 - e^{-t_i}) + (n - k) \prod_{i=1}^k (1 - e^{-t_i}) \right], \tag{A.1}
 \end{aligned}$$

for $t_1 \geq 0, \dots, t_k \geq 0$.

Proof: From

$$\begin{aligned}
 &P\{U_1 - U_{n1} \leq t_1, \dots, U_n - U_{n1} \leq t_n\} \\
 &= \sum_{j=1}^n P\{U_1 - U_{n1} \leq t_1, \dots, U_n - U_{n1} \leq t_n, U_{n1} = U_j\} \\
 &= \sum_{j=1}^n P\{0 \leq U_i - U_j \leq t_i; 1 \leq i \leq n, i \neq j\} \\
 &= \sum_{j=1}^n \int_{u_j=0}^{\infty} \left(\prod_{1 \leq i \leq n; i \neq j} \int_{u_i=u_j}^{u_j+t_i} \right) e^{-u_1 - \dots - u_n} du_1 \dots du_n,
 \end{aligned}$$

we have (A.1) with $k = n$, from which we obtain (A.1) for $1 \leq k \leq n - 1$ by integrating out the variables t_{k+1}, \dots, t_n . □

In terms of the cdf $G_0(u) = 1 - e^{-u}$, $u > 0$, of $E(0, 1)$, we may reexpress (A.1) in the form

$$\begin{aligned} & F_{U_1 - U_{n_1}, \dots, U_k - U_{n_k}}(t_1, \dots, t_k) \\ &= \frac{1}{n} G_0(t_2) \dots G_0(t_k) + \frac{1}{n} G_0(t_1) G_0(t_3) \dots G_0(t_k) \\ &+ \dots + \frac{1}{n} G_0(t_1) \dots G_0(t_{k-1}) + \frac{n-k}{n} G_0(t_1) \dots G_0(t_k), \end{aligned} \quad (\text{A.2})$$

for $t_1 \geq 0, \dots, t_k \geq 0$. The representation in (A.2) is a mixture of products, each involving either $k - 1$ or k factors G_0 .

Corollary A.1:

$$2 \sum_{i=1}^k (U_i - U_{n_i}) \text{ has cdf } \left(1 - \frac{k}{n}\right) \chi_{2k}^2 + \frac{k}{n} \chi_{2(k-1)}^2. \quad (\text{A.3})$$

Proof: By (A.2) the random variable in (A.3) is a sum of independent standard exponentials with the number of summands equal to k with probability $(n - k)/n$ and $k - 1$ with probability k/n . Since a standard exponential variate U may be expressed as $1/2$ times a χ_{2m}^2 random variable, we have for any integer $m \geq 1$ that $2 \sum_{i=1}^m U_i$ has cdf χ_{2m}^2 . Thus (A.3) follows. \square

Appendix B: Breakdown points of GQ statistics

For a GQ statistic $\hat{\theta}_{GQ,p} = \hat{H}_n^{-1}(p)$, corresponding to some kernel $h(z_1, \dots, z_k)$ and choice of $0 < p < 1$, we define breakdown to occur if either

- B1** The fraction of evaluations $h(Z_{i_1}, \dots, Z_{i_k})$ taken spuriously to a lower limit t_0 exceeds p (resulting in the estimator taking the value t_0), or
- B2** The fraction of evaluations $h(Z_{i_1}, \dots, Z_{i_k})$ taken spuriously to an upper limit t_1 exceeds $1 - p$ (resulting in the estimator taking the value t_1).

The LBP (UBP) of $\hat{\theta}_{GQ,p}$ is then given by the largest fraction of sample values Z_1, \dots, Z_n which may be taken to a lower (an upper) limit L without either of B1 or B2 occurring.

In particular, for the estimator $\hat{\theta}_{GQ,p}^{(1)}$ considered in Section 2.2 we obtain

Lemma B.1: *The estimator $\hat{\theta}_{GQ,p}^{(1)}$ has LBP = UBP given by*

$$n^{-1} \max_{1 \leq m \leq n} \left\{ m : \frac{\binom{m}{k}}{\binom{n}{k}} \leq p \leq \frac{\binom{n-m}{k}}{\binom{n}{k}} \right\} \rightarrow \min\{p^{1/k}, 1 - p^{1/k}\}, \quad n \rightarrow \infty. \quad (\text{B.1})$$

Proof: Writing the relevant kernel as $h_0(z_1, \dots, z_k) = k^{-1} \sum_{j=1}^k [z_j - \min\{z_1, \dots, z_k\}] \geq 0$, we note that $h_0(z_1, \dots, z_k) \rightarrow +\infty$ if and only if at least one summand $\rightarrow +\infty$. Now, if $j < k$ arguments $\rightarrow +\infty$ while the others remain fixed, then clearly $h_0(z_1, \dots, z_k) \rightarrow +\infty$. Also, if all k arguments $\rightarrow +\infty$ and

$$\max\{z_1, \dots, z_k\} - \min\{z_1, \dots, z_k\} \rightarrow +\infty, \quad (\text{B.2})$$

then again $h_0(z_1, \dots, z_k) \rightarrow +\infty$. Consequently, in order to avoid B2 through upper contamination, the number m of upper contaminating observations must satisfy

$$\frac{\binom{n}{k} - \binom{n-m}{k}}{\binom{n}{k}} \leq 1 - p. \quad (\text{B.3})$$

Further, if all k arguments $\rightarrow +\infty$ and

$$\max\{z_1, \dots, z_k\} - \min\{z_1, \dots, z_k\} \rightarrow 0, \quad (\text{B.4})$$

then $h_0(z_1, \dots, z_k) \rightarrow 0$. Consequently, in order to avoid B1 through upper contamination, the number m must also satisfy

$$\frac{\binom{m}{k}}{\binom{n}{k}} \leq p. \quad (\text{B.5})$$

Finally, if all k arguments $\rightarrow +\infty$ but neither (B.2) nor (B.4) holds, then $h_0(z_1, \dots, z_k)$ neither $\rightarrow +\infty$ nor $\rightarrow 0$. It follows that the UBP is given by the quantity in (B.1), and a similar argument covers the LBP. Simple combinatorics yield the convergence in (B.1). \square

For the GQ estimator $\hat{\theta}_{GQ,p}^{(2)}$ considered in Section 2.3, the relevant kernel may be expressed as $\tilde{h}_0(z_1, \dots, z_k; Z_{n1}) = k^{-1} \sum_{j=1}^k (z_j - Z_{n1}) \geq 0$. It is readily seen that if a single observation $\rightarrow -\infty$, thus causing $Z_{n1} \rightarrow -\infty$, then all evaluations $\tilde{h}_0(Z_{i_1}, \dots, Z_{i_k}) \rightarrow +\infty$, resulting in B2. Consequently, LBP = 0 for the estimator $\hat{\theta}_{GQ,p}^{(2)}$. Also, upper contamination does not cause B1 but can cause B2 unless the number m of upper contaminating observations satisfies (B.3). We thus arrive at

Lemma B.3: For the statistic $\hat{\theta}_{GQ,p}^{(2)}$, LBP = 0 and

$$UBP = n^{-1} \max_{1 \leq m \leq n} \left\{ m : \frac{\binom{n-m}{k}}{\binom{n}{k}} \geq p \right\} \rightarrow 1 - p^{1/k}, n \rightarrow \infty. \quad (\text{B.6})$$

Appendix C: Proof of Theorem 3.1

Proof: From interpretation of the terms \bar{c}_n and $n^{-1} \sum_{i=1}^n c_{ni}^2$ as Riemann approximations to $-\int_0^1 \log(1-x)dx = 1$, and $\int_0^1 \log^2(1-x)dx = 2$, respectively, it is seen that the denominator of $\hat{\theta}_{LS}$ converges to 1 as $n \rightarrow \infty$. Denoting the numerator of $\hat{\theta}_{LS}$ by T_n , it thus suffices to show that T_n is AN($\theta, 2\theta^2/n$). Using (1.3) and $\bar{U} = O_p(1)$ along with $\bar{c}_n = 1 + O(\frac{\log n}{n})$, $n \rightarrow \infty$, which follows readily by a refinement of Stirling's formula due to Robbins (1965), we obtain

$$\begin{aligned} T_n &= n^{-1} \sum_{i=1}^n c_{ni}(Z_{ni} - \bar{Z}_n) \stackrel{d}{=} \theta n^{-1} \sum_{i=1}^n c_{ni}(U_{ni} - \bar{U}) \\ &= -\theta n^{-1} \sum_{i=1}^n J_n(i/n)U_{ni} + O_p\left(\frac{\log n}{n}\right), \end{aligned} \quad (\text{C.1})$$

where $J_n(t) = 1 - c_{ni}$ for $(i-1)/n < t \leq i/n$, $1 \leq i \leq n$. Since $J_n(t) \rightarrow J(t) = 1 + \log(1-t)$, $n \rightarrow \infty$, and $E|U|^r < \infty$ for $r > 0$, the conditions of Theorem 7.5.1 of Sen (1981) for L-statistics of form (C.1) are satisfied and we obtain that $T_n \stackrel{d}{=} \theta + \theta n^{-1} \sum_{i=1}^n A(U_i) + o_p(n^{-1/2})$, where $A(u) = -\int_{-\infty}^{\infty} [\mathbf{1}\{y \geq u\} - G_0(y)]J(G_0(y))dy$ with G_0 the standard exponential cdf defined in Appendix A. It follows that $A(u) = -u + u^2/2$, $EA(U) = 0$, and $EA^2(U) = 2$, completing the proof. \square

Acknowledgment

We thank Dr. G. L. Thompson for very insightful and helpful remarks on earlier drafts. Support by grants from the Casualty Actuarial Society and Society of Actuaries, with administrative support the Actuarial Education and Research Fund, and by NSF Grant DMS-9705209, and useful comments by members of the Project Oversight Group of the Actuarial Education and Research Fund, are gratefully acknowledged.

References

- Adler, R.J., Feldman, R., and Taqqu, M., (eds.) *A User's Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy Tailed Distributions*, Birkhäuser, Boston, 1997.
- Arnold, B.C., *Pareto Distributions*, International Cooperative Publishing House, Fairland, Maryland, 1983.
- Bingham, N.H., Goldie, C.M., and Teugels, J.L., *Regular Variation*, Cambridge University Press, Cambridge, 1987.
- Brazauskas, V. and Serfling, R., "Robust and efficient estimation of the tail index of a single-parameter Pareto distribution," *North American Actuarial J.* 4, 12–27, (2000).
- Brazauskas, V. and Serfling, R., "Small sample performance of robust estimators of tail parameters for Pareto and exponential models," *J. Statist. Comput. Simul. in press.* (2001).
- Choudhury, J. and Serfling, R., "Generalized order statistics, Bahadur representations and sequential nonparametric fixed-width confidence intervals," *J. Statist. Plann. and Inf.* 19, 269–282, (1988).
- Dekkers, A.L.M. and de Haan, L., "On the estimation of the extreme-value index and large quantile estimation," *Ann. Statist.* 17, 1795–1832, (1989).
- Falk, M., "Extreme quantile estimation in δ -neighborhoods of generalized Pareto distributions," *Statist. Prob. Letters* 20, 9–21, (1994).
- Falk, M. and Marohn, F., "Von Mises conditions revisited," *Ann. Statist.* 21, 1310–1328, (1993).
- Galambos, J., *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York, 1978.
- Gather, U., "Robust estimation of the mean of the exponential distribution in outlier situation," *Commun. Statist.—Theor. Meth.* 15, 2323–2345, (1986).
- Gomes, C.P., Selman, B., and Crato, N., "Heavy-tailed distributions in combinatorial search," in G. Smolka, (ed.), *Principles and Practice of Constraint Programming CP-97, Lecture Notes in Computer Science* 1330 121–135. Springer, New York, 1997.
- Johnson, N.L., Kotz, S., and Balakrishnan, N., *Continuous Univariate Distributions*, Vol 1, 2nd ed, Wiley, New York, 1994.
- Kimber, A.C., "Trimming in gamma samples," *Appl. Statist.* 32, 7–14, (1983a).
- Kimber, A.C., "Comparison of some robust estimators of scale in gamma samples with known shape," *J. Statist. Comput. Simul.* 18, 273–286, (1983b).
- Koutrouvelis, I.A., "Large-sample quantile estimation in Pareto laws," *Commun. Statist.—Theor. Meth.* A10, 189–201, (1981).
- Lehmann, E.L., *Theory of Point Estimation*, Wiley, New York (transferred to Wadsworth, 1991) 1983.
- Quandt, R.E., "Old and new methods of estimation and the Pareto distribution," *Metrika* 10, 55–82, (1966).
- Reiss, R.-D., *Approximate Distributions of Order Statistics*, Springer, New York, 1989.
- Resnick, S.I., "Heavy tail modeling and teletraffic data (with discussion)," *Ann. Statist.* 25, 1805–1869, (1997).
- Robbins, H., "A remark on Stirling's formula," *Amer. Math. Monthly* 62, 26–29, (1965).
- Saleh, A.K., Md. E., and Ali, M.M., "Asymptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution," *Ann. Math. Statist.* 37, 143–151, (1966).
- Sarhan, A.E., Greenberg, B.G., and Ogawa, J., "Simplified estimates for the exponential distribution," *Ann. Math. Statist.* 34, 102–116, (1963).
- Sen, P.K., *Sequential Nonparametrics*, Wiley, New York, 1981.
- Serfling, R.J., *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
- Serfling, R., "Generalized L -, M - and R -Statistics," *Ann. Statist.* 12, 76–86, (1984).
- Willemain, T.R., Allahverdi, A., DeSautels, P., Eldredge, J., Gur, O., Miller, M., Panos, G., Srinivasan, A., Surtihadi, J., and Topal, E., "Robust estimation methods for exponential data: A monte-carlo comparison," *Commun. Statist.—Simula* 21, 1043–1075, (1992).