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

Vytaras Brazauskas* and Robert Serfling†

Abstract

Estimation of the tail index parameter of a single-parameter Pareto model has wide application in actuarial and other sciences. Here we examine various estimators from the standpoint of two competing criteria: efficiency and robustness against upper outliers. With the maximum likelihood estimator (MLE) being efficient but nonrobust, we desire alternative estimators that retain a relatively high degree of efficiency while also being adequately robust. A new generalized median type estimator is introduced and compared with the MLE and several well-established estimators associated with the methods of moments, trimming, least squares, quantiles, and percentile matching. 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. The generalized median type performs best overall. These findings provide a basis for revision and updating of prevailing viewpoints. Other topics discussed are 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. Robust parametric methods are compared with empirical nonparametric methods, which are typically nonrobust.

1. Introduction

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

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

defined for $\alpha > 0$ and $\sigma > 0$. Here we treat estimation of the shape parameter α that characterizes the tail, with the scale parameter σ assumed known. That is, we consider Equation (1.1) as a single-parameter Pareto model, following Beirlant, Teugels, and Vynckier (1996) and Klugman, Panjer, and Willmot (1998). In actuarial applications, Equation (1.1) with σ known is appropriate when losses or claims below a certain level are not relevant (for example, when a deductible applies). In such a case, σ can represent the deductible, or sometimes a lesser value in order to incorporate inflation into the model, while ignoring data irrelevant to the issues under study. The parameter α, on which we shall focus, 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. As small relative errors in estimation of α can produce large relative errors in estimation of such quantities, improvements in estimation of α can yield significant favorable impact in applications (see Section 6.1 for precise discussion).

More broadly, in the context of semiparametric modeling, where one assumes merely a "Pareto-type" distribution, an approximate model for the

*Vytaras Brazauskas is in the Department of Mathematical Sciences at the University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201.

†Robert Serfling is in the Department of Mathematical Sciences Department, 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.

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 $P(\sigma, \alpha)$ and variants in actuarial science, econometrics, and other fields (see Arnold 1983 and Johnson, Kotz, and Balakrishnan 1994, chap. 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 estimation of α, the MLE enjoys great popularity. Indeed, 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 should replace the MLE by an estimator that maintains satisfactory high performance over a specified range of departures from the ideal model, while not being much less efficient than the MLE when the ideal model is fully accurate.

In this paper we evaluate and compare a number of estimators for α, employing two criteria representing efficiency and robustness. More specifically, for each estimator under consideration, its asymptotic relative efficiency (ARE), taken with respect to the MLE, and its breakdown point (BP) are evaluated. These concepts, well-established and widely used in the general statistical literature, are defined and discussed precisely in Section 2. We see how the MLE is nonrobust in the present problem, leading us to seek competing estimators that are robust while retaining relatively high ARE.

In Section 3, a new generalized median type of estimator is introduced and its performance is investigated 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. A comparison of all the estimators is presented in Section 5, with the purpose of identifying not only the superior estimators, but also the inferior ones that should be used with caution, if not perhaps discarded from practical use. We arrive at the following general conclusions. The quantile, percentile matching, and trimmed mean type estimators all 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. Also, the method of moments estimator is nonrobust and, except for relatively large values of α, nonefficient.

These two types of estimators should be used with great caution. The quantile and percentile matching type estimators are more favorable but dominated by the trimmed mean type estimators, which, in turn, are dominated by the new generalized median type. Overall, it turns out that for the problem of efficient and robust estimation of α in $P(\sigma, \alpha)$, the typically used maximum likelihood, method of moments, least squares, quantile, and percentile matching estimators can be improved upon with respect to ARE and BP by the trimmed mean and, especially, the new generalized median type estimators.

The trimmed mean estimator that we consider here for estimation of α in $P(\sigma, \alpha)$ is not well-known for this purpose; rather, it is a simple transformation of an estimator well-known for estimation of θ in the exponential distribution $E(\mu, \theta)$ having cdf

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

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

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

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

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

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 $P(\sigma, \alpha)$ is equivalent, through logarithmic transformation of the data, to that of estimation of the location and scale parameters μ and θ in the model $E(\mu, \theta)$. 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 Kimber 1983a,b, Gather 1986, and Willemain et al. 1992). The present study includes 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, do not appear to have received routine consideration in the literature on the Pareto problem.

We can view the MLE as a special case of "optimization-based" estimators as described in Klugman, Panjer, and Willmot (1998, sect. 2.5). Among such estimators the MLE is the most efficient and enjoys other advantages as well. As such, the other members of this class become of interest primarily because they offer more flexibility in their mathematical treatment. Besides being dominated by the MLE with respect to efficiency, they typically are inadequate with respect to robustness. Therefore, among optimization-based estimators, we consider only the MLE.

Estimation for the "two-parameter" version of Equation 1.1 (i.e., with σ unknown), is included in our treatment of least squares, quantile, and percentile matching estimators in Sections 4.3–4.5 and is discussed briefly in Section 7. A full treatment of the two-parameter case is found in Brazauskas and Serfling (1999), along with more detailed statistical theory underlying the results of the present paper.

The use of the new generalized median type of estimator for other model distributions in actuarial and other sciences is also of interest; however, it is more complicated and will be explored in separate investigations.

2. Efficiency Versus Robustness, and Two Criteria

Here we introduce the precise efficiency and robustness criteria to be employed.

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 $P(\sigma, \alpha)$ as described by (1.1) with σ known, the MLE of α is readily derived as in Arnold (1983),

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

Its exact distribution theory is described by the statement that

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

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 by $\mathrm{AN}(\alpha, \alpha^2/n)$, meaning

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

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

For a competing estimator, efficiency is characterized in terms of 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, each estimator $\hat{\alpha}$ for α considered here is $\mathrm{AN}(\alpha, c\alpha^2/n)$ for some constant $c > 0$, from which it follows that

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

General treatment of ARE as a statistical tool are available in Lehmann (1983) and Serfling (1980).

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 that perform relatively strongly when an ample amount of data are provided, to be

distinguished from those that do not perform as strongly. The weaker estimators can then be eliminated from further consideration, while the stronger estimators can 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 BP, loosely characterized as the largest proportion of sample observations that can 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 depending only on the sample size n, then we typically use as our criterion its limit value as $n \to \infty$. The BP of an estimator measures the degree to which it remains uninfluenced by the presence of outlying observations, which possibly (but not with certainty) could 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 can differ in importance and impact, so we define separate versions:

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

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

In the present context of estimation of α in $P(\sigma, \alpha)$ with σ known, the most extreme form of lower corruption is to take observations to the lower limit σ, while the most extreme form of upper contamination is to take observations to ∞. (If σ is unknown, however, lower contamination can include taking observations to 0.) As contamination of the upper type is of greater concern in typical applications, we emphasize UBP in the present treatment.

More precisely, upper outliers are not necessarily infinite but instead can consist of very high layer data points, in cases where the model under consideration has been truncated above by some realistic upper bound. In the sense of small UBP, *robustness* means that the sensitivity of the estimator to upper data points is held within judicious limits. Such a viewpoint has application, for example, to commercial excess liability. Even when the fraction of contamination is very small and the model is truncated from above, a nonrobust estimator of a tail index parameter can still err by a significant percentage, in the presence of upper layer data points of dubious status. Thus, we regard characterization of an estimator's UBP on the basis of performance in the presence of contaminating data points at "infinity" to be apropos to the case of truncated models.

Let us now examine the MLE with respect to LBP and UBP. By the classical law of large numbers of probability theory, we have, with probability 1, that as $n \to \infty$

$$n^{-1} \sum_{i=1}^{n} \log X_i \to E(\log X) = \alpha^{-1} + \log \sigma,$$

whence follows the reassuring consistency property that, with probability 1, $\hat{\alpha}_{\mathrm{ML}} \to \alpha$ as $n \to \infty$. 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 $\hat{\alpha}_{\mathrm{ML}} \to 0$. That is, corruption of a single data value by upper contamination can render the MLE completely uninformative. Thus $\hat{\alpha}_{\mathrm{ML}}$ has UBP = 0 and hence is nonrobust against upper outliers. Lower contamination is seen to modify the MLE less severely, rendering it completely uninformative only when all observations are taken to the lower limit σ. Thus LBP = $(n - 1)/n \to 1$, which is optimal. On the basis of its extremely poor UBP, however, the MLE should be used with caution for estimation of α unless upper outliers are of no concern.

The notion of breakdown point has antecedents in Hodges (1967) and Hampel (1971), but it 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 do not come from a presupposed ideal model G but rather from a specified contamination model; that is, the data

are assumed to have distribution of form $F = (1 - \epsilon)G + \epsilon H$, where H belongs to some specified 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 BP as the criterion, however, is that it provides a robustness index valid over a broad and nonspecific range of possible sources of contaminating data and levels ϵ.

3. A Generalized Median Estimator

Consider a sample $X_1, \ldots, X_n$ from the model $P(\sigma, \alpha)$ as described by Equation (1.1). We introduce and evaluate a new estimator that, 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 a given choice of integer $k \geq 1$, 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. In order to modify h_0 to make any evaluation with sample values median unbiased for estimation of α, we use the fact that

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

which follows from Equation (2.1). Denoting the median of χ^2_ν by M_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; that is, the cdf H_F of $h(X_1, \ldots, X_k)$ satisfies

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

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, that is, 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 (GM) statistic

$$\hat{\alpha}_{\text{GM}} = \text{Median}\{h(X_{i_1}, \ldots, X_{i_k})\} \qquad (3.2)$$

for estimation of α. (For $k = 1$, this is just the median of the transformed observations $(\log X_i - \log \sigma)^{-1}$, $1 \leq i \leq n$, adjusted to be median-unbiased for estimation of α.)

Values of M_{2k} and the multiplicative correction factors $M_{2k}/(2k)$ needed in constructing $\hat{\alpha}_{\text{GM}}$ are provided for kernel sizes $k = 1 : 10$ in Table 1.

Although one can consider GM statistics for other choices of kernel that are median-unbiased, for α, the present choice has special appeal. Note that each evaluation $h_0(X_{i_1}, \ldots, X_{i_k})$ is the MLE of α based on just the observations $X_{i_1}, \ldots, X_{i_k}$ (for $k = n$, the MLE is based on the full sample). This endows h_0 and its modification h with the efficiency of the MLE in extracting relevant information about α from any given subsample.

3.1 Asymptotic Normality

The estimator $\hat{\alpha}_{\text{GM}}$ 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). Therefore, $\hat{\alpha}_{\text{GM}}$ 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, $\zeta = \text{Var}(w_h(X))$, 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. Table 2 provides these AREs and the corresponding values of γ_k for $k = 1 : 10$.

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

	k									
	1	2	3	4	5	6	7	8	9	10
M_{2k}	1.386	3.357	5.348	7.344	9.342	11.340	13.339	15.339	17.338	19.337
$M_{2k}/(2k)$	0.693	0.839	0.891	0.918	0.934	0.945	0.953	0.959	0.963	0.967

Table 2
ARE($\hat{\alpha}_{GM}$, $\hat{\alpha}_{ML}$) and γ_k, for $k = 1 : 10$

	k									
	1	2	3	4	5	6	7	8	9	10
ARE($\hat{\alpha}_{GM}$, $\hat{\alpha}_{ML}$)	0.64	0.78	0.88	0.92	0.94	0.96	0.97	0.97	0.98	0.98
γ_k	1.563	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 of the statistic $\hat{\alpha}_{GM}$, it can be shown (see the Appendix) that

$$\text{UBP} = n^{-1} \max_{1 \le m \le n} \left\{ m : \frac{\binom{n-m}{k}}{\binom{n}{k}} \ge \frac{1}{2} \right\} \rightarrow 1 - \left(\frac{1}{2}\right)^{1/k}, \quad n \rightarrow \infty, \quad (3.3)$$

$$\text{LBP} = n^{-1} \max_{1 \le m \le n} \left\{ m : \frac{\binom{m}{k}}{\binom{n}{k}} \le \frac{1}{2} \right\} \rightarrow \left(\frac{1}{2}\right)^{1/k}, \quad n \rightarrow \infty. \quad (3.4)$$

Values of the limits in Equations (3.3) and (3.4) for $k = 1 : 10$ are given in Table 3.

From examination of Tables 2 and 3, we see that ARE and LBP increase with k while UBP decreases. From a practical standpoint, the case $k = 1$ can be ignored. While it is included for completeness, the low ARE of 0.64 is not competitive, and UBP as high as 0.50 is not especially needed. Serious interest begins with the case $k = 2$, with marginally competitive ARE of 0.78 and very strong UBP of 0.293. While UBP decreases with k to 0.067 at $k = 10$, this value is not too small for practical application and the corresponding ARE of 0.98 is excellent. Thus the cases $k = 2 : 10$ represent a practical range of trade-offs between robustness and efficiency.

3.3 Computational Considerations

The computational burden of computing $\hat{\alpha}_{GM}$ grows with n as $O(n^k)$, which, for large n, could become prohibitive. For example, for $k = 5$, the number of kernel evaluations needed for $n = 25$ is only $\binom{25}{5} = 53{,}130$, whereas for $n = 500$, the number is 255,244,687,600. When the number of evaluations needed exceeds $N = 10^6$ or 10^8, one can simply estimate the estimator $\hat{\alpha}_{GM}$ by using only the evaluations $h(X_{i_1}, \ldots, X_{i_k})$ for a random sample size N of the $\binom{n}{k}$ possible k-sets $\{i_1, \ldots, i_k\}$. Such an approach renders the computational burden negligible but maintains any desired degree of numerical accuracy. For example, for $k = 5$ and $n = 500$, the computation of $\hat{\alpha}_{GM}$ on a Pentium II 400 MHz laptop computer with $N = 10^7$ evaluations requires only 50 seconds, whereas the exact calculation using all evaluations would require approximately 354 hours. For $k = 10$ and $n = 500$, the computation with $N = 10^7$ evaluations requires 90 seconds. As n increases for fixed k, these computation times do not increase because N is held fixed.

4. REVIEW OF ESTABLISHED ESTIMATORS

Continuing the setting of Section 3, we focus on estimation of the parameter α, based on a sample

Table 3
Asymptotic LBP and UBP of $\hat{\alpha}_{GM}$, for $k = 1 : 10$

	k									
	1	2	3	4	5	6	7	8	9	10
LBP	0.500	0.707	0.794	0.841	0.871	0.891	0.906	0.917	0.926	0.933
UBP	0.500	0.293	0.206	0.159	0.129	0.109	0.094	0.083	0.074	0.067

$X_1, \ldots, X_n$ having cdf F corresponding to the $P(\sigma, \alpha)$ model. Denote the ordered sample values by

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

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

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

and $\bar{Z}_n = n^{-1} \sum_{i=1}^n Z_i$. 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 $P(\sigma, \alpha)$, we utilize the formula $EX = \sigma\alpha/(\alpha - 1)$, which is valid provided that $\alpha > 1$. The corresponding method of moments estimator is then obtained by solving the equation $\bar{X}_n = \sigma\hat{\alpha}/(\hat{\alpha} - 1)$, yielding

$$\hat{\alpha}_{\text{MM}} = \frac{\bar{X}_n}{\bar{X}_n - \sigma}$$

By the law of large numbers, as $n \to \infty$ we have $\bar{X}_n \to EX$ and thus $\hat{\alpha}_{\text{MM}} \to \alpha$. However, for any fixed n, if even a single X_i is taken to ∞, then $\bar{X}_n \to \infty$ and, consequently, $\hat{\alpha}_{\text{MM}} \to 1$. That is, the estimator can be rendered uninformative by upper corruption of even a single observation. Thus, $\hat{\alpha}_{\text{MM}}$ has UBP = 0 and 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, we also examine the ARE. Using standard central limit theory and

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

which is valid provided $\alpha > 2$, we have $\hat{\alpha}_{\text{MM}}$ is $\text{AN}(\alpha, \alpha(\alpha - 1)^2/(\alpha - 2)n)$. It follows that the ARE of $\hat{\alpha}_{\text{MM}}$, with respect to $\hat{\alpha}_{\text{ML}}$, is $\alpha(\alpha - 2)/(\alpha - 1)^2$, which approaches 1 as $\alpha \to \infty$; however, for typical values of α, it is poor. For example, for $2 \le \alpha \le 2.5$, we have $0 \le \text{ARE} \le 0.56$, and for $2.5 \le \alpha \le 3$, we have $0.56 \le \text{ARE} \le 0.75$.

In conclusion, the method of moments estimator defined for $\alpha > 1$ exhibits neither satisfactory robustness nor satisfactory efficiency. While method of moments estimators can also be defined when $\alpha \le 1$, these fail to satisfy consistency (convergence to α).

4.2 Trimmed Mean Estimators

For the problem of robust estimation of θ in the model $E(\mu, \theta)$, 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, for estimation of θ in $E(\mu, \theta)$, Kimber (1983a,b) defines

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

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 $\hat{\theta}_{\text{T}}$ mean unbiased.) Robustness is gained against lower outliers if $[n\beta_1] > 1$ and against upper outliers if $[n\beta_2] > 1$. Indeed, the trimmed mean estimator $\hat{\theta}_{\text{T}}$ is completely unaffected by taking the proportion β_1 lowermost observations to the lower limit μ or by taking the proportion β_2 uppermost observations to $+\infty$, so that $\text{LBP} = [n\beta_1]/n \to \beta_1$ and $\text{UBP} = [n\beta_2]/n \to \beta_2$, $n \to \infty$. These BPs apply also to the corresponding estimator of α given by $\hat{\alpha}_{\text{T}} = \hat{\theta}_{\text{T}}^{-1}$. (In the case of no trimming, that is, $\beta_1 = \beta_2 = 0$, note that $d(0, 0, n) = n$ and thus $\hat{\alpha}_{\text{T}}$ reduces to the MLE discussed previously.)

To consider ARE, we utilize the fact that $\hat{\alpha}_T$ is $\text{AN}(\alpha, D_{\beta_1,\beta_2}\alpha^2/n)$, with D_{β_1,β_2} computable following general methods for L-statistics in Serfling

(1980), Section 8.2.4, or Lehmann (1983), Section 5.4. Thus,

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

In particular, for $\beta_1 = \beta_2$ or $\beta_1 = 0$, and β_2 taking values 0.05, 0.10, 0.15, 0.20, and 0.25, values of $\mathrm{ARE}(\hat{\alpha}_T, \hat{\alpha}_{ML})$, and thus through Equation (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 AREs for the two cases $\beta_1 = 0$ and $\beta_1 = \beta_2$ agree (coincidentally) within two decimal places. This leads to ARE and corresponding D values as listed in Table 4.

4.3 Estimators Based on a Regression Approach

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

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

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

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

and the corresponding sample quantile function

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

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

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

where $u = -\log(1-p)$, $0 < p < 1$. Taking n choices of p, such that $\hat{F}_n^{-1}(p)$ generates the set of order statistics $\{X_{ni}, 1 \leq i \leq n\}$ via Equation (4.4), that is, for choices $p^*_{n1}, \ldots, p^*_{nn}$ satisfying

$$p^*_{n1} \in \left(0, \frac{1}{n}\right], \; p^*_{n2} \in \left(\frac{1}{n}, \frac{2}{n}\right], \; \ldots,$$
$$p^*_{n,n-1} \in \left(\frac{n-2}{n}, \frac{n-1}{n}\right], \; p^*_{nn} \in \left(\frac{n-1}{n}, 1\right), \tag{4.6}$$

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

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

where $u^*_{ni} = -\log(1 - p^*_{ni})$ and $\epsilon^*_{ni} = Z_{ni} - \log F^{-1}(p^*_{ni})$, $1 \leq i \leq n$.

The equations in (4.7) can be utilized via interpretation 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

$$(Z_{ni}, u^*_{ni}), \; 1 \leq i \leq n,$$

and these, in turn, yield estimates of σ and α. We may select either the full set of all n points or a strategic subset. In particular, in Section 4.4, we

Table 4

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

	β_2				
	0.05	**0.10**	**0.15**	**0.20**	**0.25**
$\mathrm{ARE}(\hat{\alpha}_T, \hat{\alpha}_{ML})$	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

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 Section 4.5, we take just a specified number of k points and apply weighted least squares to obtain "quantile" estimators, 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 Section 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 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

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

$$\hat{\mu}_{\mathrm{LS}} = \bar{Z}_n - \hat{\theta}_{\mathrm{LS}}\bar{c}_n;$$

and hence

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

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

Discussions of versions of this least squares estimator by Quandt (1966, p. 60), and Arnold (1983, p. 202) suggest 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 $\hat{\alpha}_{\mathrm{LS}}$ is AN(α, $2\alpha^2/n$), clarifying that $\hat{\alpha}_{\mathrm{LS}}$ is actually poor in efficiency, having ARE = 0.50. In sum, $\hat{\alpha}_{\mathrm{LS}}$ is both nonrobust and nonefficient.

4.5 Estimators Based on k Selected Quantiles

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

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

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

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

using weighted least squares based on the asymptotic covariance matrix of the ϵ_{ni}'s. This yields $\hat{\alpha}_{\mathrm{Q}} = \hat{\theta}_{\mathrm{Q}}^{-1}$ (see Koutrouvelis 1981 and Saleh and Ali 1966 for details) for estimation of α, where

$$\hat{\theta}_{\mathrm{Q}} = \sum_{i=1}^k b_i Z_{n,\lceil np_i\rceil}, \quad (4.10)$$

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

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

and for σ the estimator $\exp\hat{\mu}_{\mathrm{Q}} = \exp\{Z_{n,\lceil np_i\rceil} - \hat{\theta}_{\mathrm{Q}}u_1\}$. The estimators $\hat{\theta}_{\mathrm{Q}}$ and $\hat{\mu}_{\mathrm{Q}}$ can be characterized as the asymptotically best linear unbiased estimates of θ and μ based on the selected k quantiles.

In the context of estimation of the $P(\sigma, \alpha)$ model with σ unknown, 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 a discussion. In the equivalent context of estimation of the $E(\mu, \theta)$ model with μ unknown, this approach was treated by Harter (1961) and Sarhan, Greenberg, and Ogawa (1963) for $k = 2$ and by Saleh and Ali (1966) for arbitrary $k \geq 2$. Also, Sarhan, Greenberg, and Ogawa (1963) treated $E(\mu, \theta)$ with μ known for arbitrary $k \geq 1$. As will be discussed below in Section 4.5.4, these latter results yield a quantile approach to the $P(\sigma, \alpha)$ model with σ known.

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, which is analogous to the method of moments. In the present situation, this corresponds to the case $k = 2$, for which the weighted least squares solution of the equations in (4.8) is given by setting the ϵ_{ni}'s equal to 0 and solving the resulting linear system of two equations in two unknowns (see Section 4.5.2). Choosing k larger than the number of parameters, however, produces higher efficiency, as we shall see in the following section.

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

Furthermore, $\hat{\alpha}_Q$ is AN(α, $L^{-1}\alpha^2/n$). Thus $\hat{\alpha}_Q$ has ARE = L.

4.5.1 Optimal Choice of Quantiles

It is of interest, of course, to choose the quantile levels $p_1, \ldots, p_k$ optimally, in the sense of minimizing the generalized variance of the asymptotic joint distribution of the estimators $\hat{\theta}_Q$ and $\hat{\mu}_Q$. Following Saleh and Ali (1966), an optimal choice of p_1 is found to be

$$p_1^o = \frac{1}{n + 0.5} \tag{4.11}$$

and optimal choices of the remaining p_i's are obtained by minimizing the generalized variance with respect to $p_2, \ldots, p_k$, subject to Equation (4.11). Because Equation (4.11) results in $Z_{n,\lceil np_1 \rceil} = Z_{n1}$ and the sum of the b_i's is 0, the optimal estimator can be expressed as

$$\hat{\alpha}_Q^{\text{opt},k} = \left(\sum_{i=2}^{k} b_i (Z_{n,\lceil np_i \rceil} - Z_{n1}) \right)^{-1}. \tag{4.12}$$

As discussed in Brazauskas and Serfling (1999), it follows that the optimal choices of $p_2, \ldots, p_k$ are given by $p_i = \lambda_{k-1,i-1}$, $2 \leq i \leq k$, where $\lambda_{m,1}, \ldots, \lambda_{m,m}$ are the asymptotically optimal quantiles for estimation of θ in the one-parameter exponential model $E(0, \theta)$ by a linear function of m order statistics. These values are derived by Sarhan, Greenberg, and Ogawa (1963) and listed for $m = 1 : 15$ in their Table 3. For the cases $m = 1$ and 2, fixed sample analogues of the $\lambda_{m,i}$'s for $n = 1 : 100$ are provided by Harter (1961).

4.5.2 The Case $k = 2$

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

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

and (with p_1 and p_2 chosen optimally)

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

Furthermore, we have

$$\text{Var}(\hat{\alpha}_Q) = \frac{e^{u_2} - e^{u_1}}{(u_2 - u_1)^2} \frac{\alpha^2}{n} + O(n^{-2}), \tag{4.13}$$

as $n \to \infty$ with p_1 fixed or tending to 0 at rate $O(n^{-1})$ and p_2 fixed. Also, $\text{Var}(\hat{\alpha}_Q^{\text{opt},2})$ satisfies (4.13) with u_1 replaced by 0.

4.5.3 Examples

In particular,

- For $k = 2$, the optimal quantile levels are $p_1 = p_1^o$ and $p_2 = 0.80$. For the corresponding estimator, $\hat{\alpha}_Q^{\text{opt},2}$, we have LBP = 0, UBP = 0.20, and ARE = 0.648.
- 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 (in connection with a different version of Pareto distribution with two unknown parameters), is $p_1 =$

0.40 and $p_2 = 0.70$. For the corresponding estimator, we have LBP = 0.40, UBP = 0.30, and ARE = 0.288.
- For $k = 5$, the optimal p_i's are $p_1 = p_1^o$, $p_2 = 0.45$, $p_3 = 0.74$, $p_4 = 0.91$, and $p_5 = 0.98$, and for $\hat{\alpha}_Q^{\text{opt},5}$ we have LBP = 0, UBP = 0.02, and ARE = 0.926.

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

- For the $k = 2$ estimator based on $p_1 = 0.10$ and $p_2 = 0.90$, we have LBP = UBP = 0.10 and ARE = 0.543.
- For the $k = 4$ estimator based on $p_1 = p_1^o$, $p_2 = 0.25$, $p_3 = 0.50$, and $p_4 = 0.75$, we have LBP = 0, UBP = 0.25, and ARE = 0.735.
- For the $k = 5$ estimator based on $p_1 = 0.13$, $p_2 = 0.32$, $p_3 = 0.50$, $p_4 = 0.69$, and $p_5 = 0.87$, we have LBP = UBP = 0.13 and ARE = 0.73. In later references, we designate this estimator by $\hat{\alpha}_Q^*$.

4.5.4 The Single-Parameter Model

As noted in Section 4.5.1, asymptotically optimal quantiles for estimation of θ in $E(0, \theta)$ have been derived in Sarhan, Greenberg, and Ogawa (1963) and are closely related to those for estimation of θ in $E(\mu, \theta)$ with μ unknown. Because the model $P(\sigma, \alpha)$, with σ known, can equivalently be treated via the model $E(\mu, \theta)$ with μ known, which, in turn, without loss of generality, can be reduced to $E(0, \theta)$, it follows that in $P(\sigma, \alpha)$, with σ known, the asymptotically optimal linear estimator of α based on m quantiles is given by

$$\tilde{\alpha}_Q^{\text{opt},m} = \left(\sum_{i=1}^{m} w_i Z_{n,\lceil n\lambda_{m,i} \rceil} \right)^{-1}, \qquad (4.14)$$

with w_i's given by the b_i's for $k = m + 1$ via $w_i = b_{i+1}$, and the $\lambda_{m,i}$'s by the p_i's for $k = m + 1$ via $\lambda_{m,i} = p_{i+1}$, for $1 \le i \le m$. It follows also that $\hat{\alpha}_Q^{\text{opt},k}$ and $\tilde{\alpha}_Q^{\text{opt},k-1}$ have the same ARE and the same UBP, but the LBPs = 0 and p_2, respectively. Thus, from the examples considered thus far, we can state:

- For $m = 1$, the optimal quantile level is $\lambda_{1,1} = 0.80$. For the corresponding estimator, $\tilde{\alpha}_Q^{\text{opt},1}$, we have LBP = 0.80, UBP = 0.20, and ARE = 0.648.
- For $m = 4$, the optimal λ's are $\lambda_{4,1} = 0.45$, $\lambda_{4,2} = 0.74$, $\lambda_{4,3} = 0.91$, and $\lambda_{4,4} = 0.98$, and for $\tilde{\alpha}_Q^{\text{opt},4}$, we have LBP = 0.45, UBP = 0.02, and ARE = 0.926.

More generally, the estimator $\hat{\alpha}_Q = \hat{\theta}_Q^{-1}$, with $\hat{\theta}_Q$ given in Equation (4.10) is equal in distribution to that given by

$$\tilde{\alpha}_Q = \left(\sum_{i=1}^{k-1} b_{i+1} Z_{n-1,\lceil (n-1)p_{i+1} \rceil} \right)^{-1}. \qquad (4.15)$$

For example, in this single-parameter case, the percentile matching approach with the single quantile level p yields the estimator

$$\tilde{\alpha}_Q = (b_2 Z_{n-1,\lceil (n-1)p \rceil})^{-1}$$

with LBP = p, UBP = $1 - p$, and through Equation (4.13), ARE = $(1 - p)(\log(1 - p))^2/p$.

In particular, for $p = 0.7$ (as selected in an example of percentile matching in Klugman, Panjer, and Willmot 1998, p. 47), for the one-parameter exponential distribution, we have LBP = 0.70, UBP = 0.30, and ARE = 0.621. (Compare the optimal quantile level $\lambda_{1,1} = 0.80$ with ARE = 0.648, as previously shown.)

5. COMPARISONS AND CONCLUSIONS

In Table 5, important cases of the estimators of α

Table 5
ARE and UBP for Selected Estimators of α

Estimator	ARE	UBP
MLE	1	0
$\hat{\alpha}_{MM}$ $(2 < \alpha \le 2.5)$	≤0.56	0
$\hat{\alpha}_{LS}$	0.50	0
$\hat{\alpha}_Q^{\text{opt},2}$	0.65	0.20
$\hat{\alpha}_Q^*$ (k = 5)	0.73	0.13
$\hat{\alpha}_Q^{\text{opt},5}$	0.93	0.02
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .25$	0.67	0.25
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .20$	0.72	0.20
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .15$	0.78	0.15
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .10$	0.85	0.10
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .05$	0.92	0.05
$\hat{\alpha}_T$, $\beta_1 = \beta_2 = .04$	0.93	0.04
$\hat{\alpha}_{GM}$, $k = 2$	0.78	0.29
$\hat{\alpha}_{GM}$, $k = 3$	0.88	0.21
$\hat{\alpha}_{GM}$, $k = 4$	0.92	0.16
$\hat{\alpha}_{GM}$, $k = 5$	0.94	0.13
$\hat{\alpha}_{GM}$, $k = 10$	0.98	0.07

considered in Sections 3 and 4 are compared from the standpoint of efficiency versus robustness, with the latter represented by UBP. The following conclusions are quite evident:

- The method of moments and least squares estimators are neither efficient nor robust and, thus, are not competitive.
- The quantile type estimators are improved upon by the trimmed type estimators. For example, $\hat{\alpha}_Q^{opt}$ for $k = 2$, with ARE = 0.65 and UBP = 0.20, is dominated by $\hat{\alpha}_T$ for $\beta_1 = \beta_2 = 0.20$, with ARE = 0.72 and UBP = 0.20. Also, $\hat{\alpha}_Q^{opt}$ for $k = 5$, with ARE = 0.93 and UBP = 0.02, is dominated by $\hat{\alpha}_T$ for $\beta_1 = \beta_2 = 0.04$, with ARE = 0.93 and UBP = 0.04. Finally, the quantile estimator $\hat{\alpha}_Q^*$, with ARE = 0.73 and UBP = 0.13, is improved upon by $\hat{\alpha}_T$ for $\beta_1 = \beta_2 = 0.15$, with ARE = 0.78 and UBP = 0.15.
- In turn, the trimmed type estimators and the quantile estimators are improved upon by the generalized median type estimators. For example, $\hat{\alpha}_T$ for $\beta_1 = \beta_2 = 0.20$, with ARE = 0.72 and UBP = 0.20, is dominated by $\hat{\alpha}_{GM}$ for $k = 3$, with ARE = 0.88 and UBP = 0.21. Likewise, $\hat{\alpha}_T$ for $\beta_1 = \beta_2 = 0.05$, with ARE = 0.92 and UBP = 0.05, is dominated by $\hat{\alpha}_{GM}$ for $k = 10$, with ARE = 0.98 and UBP = 0.07. Finally, $\hat{\alpha}_T$ for $\beta_1 = \beta_2 = 0.15$, with ARE = 0.78 and UBP = 0.15, is improved upon by $\hat{\alpha}_{GM}$ for $k = 4$, with ARE = 0.92 and UBP = 0.16.

Interpretive Conclusion

The superiority of the GM estimators can be explained by the following general principle:

> Smoothing of the data, accomplished 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 with respect to UBP and competitors should be considered. Among competitors, the new generalized median approach dominates the others simultaneously with respect to ARE and UBP and should become incorporated into practical use. Closely competitive is the trimmed mean approach. The less competitive quantile and percentile matching approaches should be used perhaps more cautiously, and the method of moments and least squares approaches should be used with extreme caution.

On the other hand, so as to avoid overstating the case for the GM estimators, we note that ARE and UBP are not the only criteria to consider. As one referee stated:

> I believe that this alternative, the generalized median estimator, should be among the estimators that an actuary considers when estimating the parameter of a single-parameter Pareto model. . . . For a given problem, I believe the prudent actuary should look at several estimators (including the maximum likelihood estimator), keeping in mind the strengths and weaknesses of each of them.

6. APPLICATIONS

Here we consider various applications utilizing estimation of α in the model $P(\sigma, \alpha)$. 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. Furthermore, 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 Equation (1.1) that

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

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

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

Consequently, for $\epsilon = 0.001$, underestimation of $\alpha = 1$ by only 5% produces overestimation of $q_{.001}$ by 44%, and underestimation of $\alpha = 1.5$ by 5% produces overestimation of $q_{.001}$ by 27%. Likewise, overestimation of $\alpha = 1$ by 10% produces underestimation of $q_{.001}$ by 47%, and overestimation of $\alpha = 1.5$ by 10% produces underestimation of $q_{.001}$ by 34% and underestimation of $q_{.0001}$ by 43%.

Tail probabilities in the range of $\epsilon = 0.001$ or even $\epsilon = 0.0001$ are common in actuarial and extreme value applications. For example, high quantiles are used in the percentile principle of reinsurance premium calculation as seen in 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 a probability less than 0.0001 (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 Equations (1.1) or (6.1) that

$$\epsilon = \left(\frac{\sigma}{q}\right)^{\alpha}. \tag{6.3}$$

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

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

Consequently, underestimation of any value of α by 5% produces overestimation of the tail probabilities $\epsilon = 0.001$ by 41% and $\epsilon = 0.0001$ by 58%. Likewise, overestimation of any value of α by 10% produces underestimation of $\epsilon = 0.001$ by 50% and $\epsilon = 0.0001$ 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 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. 102–16), and Beirlant, Teugels, and Vynckier (1996, pp. 122–6), for a detailed discussion.

A broad and effective assumption for semiparametric modeling of the tail of an individual claim distribution or of an aggregate claims distribution 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

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

that is, the conditional distribution of an observation, given that it exceeds a threshold d, becomes for large d approximately a single-parameter Pareto distribution, $P(1, \alpha)$. 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 can apply methods designed for $P(1, \alpha)$ samples to the upper-ordered values of a sample from H. In particular, for determination of reinsurance premiums as discussed thus far, methods for estimation of quantiles and tail probabilities of $P(1, \alpha)$ are relevant, along with the considerations in Sections 6.1–6.2. By way of illustration, in an analysis of $n = 429$ Norwegian fire insurance claims for 1981, Beirlant, Teugels, and Vynckier (1996) arrive at $k(n) = 94$ as the suitable number of upper-ordered values to employ in the estimation of α and of the upper 0.001 quantile of the claim distribution.

In this semiparametric context, the MLE of α based on treating the $k(n)$ upper-order statistics as a sample from $P(1, \alpha)$ is known as the Hill estimator (introduced by Hill 1975). It is efficient and provides the benchmark against which competing estimators are compared, but it is nonrobust, being seriously influenced by any extreme outliers that are not representative of the model being estimated. See Beirlant, Teugels, and Vynckier (1996, chap. 2), for a 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 that 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

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

that is, the conditional expectation of the excess above threshold d for an observation that exceeds d. For $P(\sigma, \alpha)$ with $\alpha > 1$ we have

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

independently of σ, 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 through Equation (6.6) this may be carried out by robust estimation of α in $P(\sigma, \alpha)$. In this regard, we note that for $\alpha = 1.5$ and large d, underestimation of α by 10% produces overestimation of $e(d)$ by 43%, and overestimation of α by 10% produces underestimation of $e(d)$ by 23%. For such robust estimation through Equation (6.6), estimators $\hat{\alpha}$ (for example, $\hat{\alpha}_{GM}$ or $\hat{\alpha}_T$), restricted to just the excess values, can 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, such as 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 $T(\hat{H}_n)$, where $\hat{H}_n$ denotes a sample analogue estimator of H, as introduced in Section 4.3 for estimation of F given by Equation (1.1). While this has the advantage of not depending directly upon parametric assumptions, which can 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 $P(\sigma, \alpha)$ through Equation (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. Furthermore, 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, sec. 2.2 and 2.6).

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

7. ESTIMATION OF α IN $P(\sigma, \alpha)$ WITH σ UNKNOWN

In Section 4 we considered certain estimators of α in $P(\sigma, \alpha)$ with σ unknown. In this case, the MLE of α is the estimator produced by substituting the minimum sample observation X_{n1} for σ in $\hat{\alpha}_{ML}$. Like $\hat{\alpha}_{ML}$, this modified MLE satisfies UBP = 0 and is AN(α, α^2/n); it also satisfies LBP = 0. Similarly modified versions of the trimmed, generalized median, and least squares estimators are treated in Brazauskas and Serfling (1999), where it is found that, in this case, the method of moments and least squares estimators are least competitive, while the quantile and trimmed-type estimators are competitive but dominated overall by the generalized median estimators.

ACKNOWLEDGMENT

We are very appreciative of valuable insights and comments provided by Dr. G. L. Thompson, two referees, and the Associate Editor, leading to many improvements in the paper. Also, support of the second author 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

ARNOLD, B.C. 1983. *Pareto Distributions.* Fairland, Md.: International Cooperative Publishing House.

BEIRLANT, J., TEUGELS, J.F., AND VYNCKIER, P. 1996. *Practical Analysis of Extreme Values.* Leuven, Belgium: Leuven University Press.

BRAZAUSKAS, V., AND SERFLING, R. 1999. "Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics." Submitted for publication.

CHOUDHURY, J., AND SERFLING, R. 1988. "Generalized Order Statistics, Bahadur Representations, and Sequential Nonparametric Fixed-Width Confidence Intervals." *Journal of Statistical Planning and Inference* 19:269–82.

DAYKIN, C.D., PENTIKAINEN, T., AND PESONEN, M. 1994. *Practical Risk Theory for Actuaries.* London: Chapman & Hall.

DEKKERS, A.L.M., AND DE HAAN, L. 1989. "On the Estimation of the Extreme-Value Index and Large Quantile Estimation." *Annals of Statistics* 17:1795–832.

DONOHO, D.L., AND HUBER, P.J. 1983. "The Notion of Breakdown Point." In *A Festschrift for Erich L. Lehmann,* ed. by P.J. Bickel, K.A. Doksum, and J.L. Hodges, Jr. Belmont, Calif.: Wadsworth, pp. 157–84.

GATHER, U. 1986. "Robust Estimation of the Mean of the Exponential Distribution in Outlier Situation." *Communications in Statistics, Part A—Theory and Methods* 15:2323–45.

GERBER, H.U. 1979. *An Introduction to Mathematical Risk Theory.* S.S. Huebner Foundation for Insurance Education, University of Pennsylvania.

GOMES, C.P., SELMAN, B., AND CRATO, N. 1997. "Heavy-Tailed Distributions in Combinatorial Search." In *Principles and Practice of Constraint Programming CP-97,* ed. G. Smolka. Lecture Notes in Computer Science, New York: Springer, 1330:121–35.

HAMPEL, F.R. 1971. "A General Qualitative Definition of Robustness." *Annals of Mathematical Statistics* 42:1887–96.

HARTER, H.L. 1961. "Estimating the Parameters of Negative Exponential Populations from One or Two Order Statistics." *Annals of Mathematical Statistics* 32:1078–90.

HILL, B.M. 1975. "A Simple General Approach to Inference About the Tail of a Distribution." *Annals of Statistics* 3:1163–74.

HODGES, J.L., JR. 1967. "Efficiency in Normal Samples and Tolerance of Extreme Values for Some Estimates of Location." *Proceedings of Fifth Berkely Symposium on Mathematical Statistics and Probability,* Volume 1, 163–8.

JOHNSON, N.L., KOTZ, S., AND BALAKRISHNAN, N. 1994. *Continuous Univariate Distributions,* Volume 1, 2nd edition. New York: Wiley.

KIMBER, A.C. 1983a. "Trimming in Gamma Samples." *Applied Statistics* 32:7–14.

KIMBER, A.C. 1983b. "Comparison of Some Robust Estimators of Scale in Gamma Samples with Known Shape." *Journal of Statistical Computation and Simulation* 18:273–86.

KLUGMAN, S.A., PANJER, H.H., AND WILLMOT, G.E. 1998. *Loss Models: From Data to Decisions.* New York: Wiley.

KOUTROUVELIS, I.A. 1981. "Large-Sample Quantile Estimation in Pareto Laws." *Communications in Statistics, Part A—Theory and Methods* 10:189–201.

LEHMANN, E.L. 1983. *Theory of Point Estimation.* New York: Wiley, (transferred to Wadsworth, 1991).

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." *Annals of Mathematical Statistics* 37:143–51.

SARHAN, A.E., GREENBERG, B.G., AND OGAWA, J. 1963. "Simplified Estimates for the Exponential Distribution." *Annals of Mathematical Statistics* 34:102–16.

SERFLING, R.J. 1980. *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

SERFLING, R. 1984. "Generalized *L*-, *M*- and *R*-statistics." *Annals of Statistics* 12:76–86.

WILLEMAIN, T.R., ET AL. 1992. "Robust Estimation Methods for Exponential Data: A Monte-Carlo Comparison." *Communications in Statistics, Part B—Simulation and Computation* 21:1043–75.

APPENDIX

BREAKDOWN POINTS OF GENERALIZED MEDIAN STATISTICS

Consider estimation of a parameter η by a GM statistic $\hat{\eta}_{GM}$ corresponding to some kernel $h(x_1, \ldots, x_k)$. The LBP of $\hat{\eta}_{GM}$ is based on its performance under the influence of lower contaminating observations in the sample and the UBP on upper contamination. In each case, taken separately, we consider breakdown to occur if either

B1 The fraction of evaluations $h(X_{i_1}, \ldots, X_{i_k})$ taken spuriously to a lower limit t_0 exceeds 0.5 (which results in the GM estimator taking the value t_0), or

B2 The fraction of evaluations $h(X_{i_1}, \ldots, X_{i_k})$ taken spuriously to an upper limit t_1 exceeds 0.5 (which results in the GM estimator taking the value t_1).

For simplicity, less extreme forms of breakdown that can occur due to the influence of contaminating values will not be considered. Thus, for present purposes, the LBP of $\hat{\eta}_{GM}$ is defined as the largest proportion of the sample values $X_1, \ldots, X_n$, which can be taken to a lower limit L_0 without either B1 or B2 occurring, and the UBP as the largest proportion, which can be taken to an upper limit L_1 without either B1 or B2 occurring.

In particular, we now treat the GM estimator of α based on the kernel h_0 of Section 3, which, for convenience, we write as $h_0 = g^{-1}$, where

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

$$= k^{-1} \sum_{j=1}^{k} (\log x_j - \log \sigma) \geq 0.$$

Note that contamination causing the GM estimator based on h_0 to satisfy B1 or B2 causes the GM estimator based on g to satisfy B2 or B1, respectively, and conversely. It follows that the GM estimators based on h_0 and g have the same LBPs and the same UBPs.

For the kernel g, upper contamination does not cause B1 but can cause B2 unless the number m of upper contaminating observations satisfies

$$\frac{\binom{n}{k} - \binom{n-m}{k}}{\binom{n}{k}} \leq 0.5.$$

Likewise, lower contaminating observations do not take a kernel evaluation to $+\infty$ but can take it to 0 if all k arguments are taken to $\log \sigma$; that is, lower contamination does not cause B2 but can cause B1 unless the number m of lower contaminating observations satisfies

$$\frac{\binom{m}{k}}{\binom{n}{k}} \leq 0.5.$$

This leads to the UBP and LBP given by Equations (3.3) and (3.4), respectively.

Discussions on this paper can be submitted until April 1, 2001. The author reserves the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.

used to calculate the required premium by selecting "Tools," "Goal Seek," and then inputting L3; 100,000; and D2 in the "Set cell," "To value," and "By changing cell" boxes, respectively.

As shown in Table 2, a premium of $567.71 per year is required (see cell D2).

UL Policy

The spreadsheet application has been used as a computer lab exercise in a risk management and insurance course and a personal finance course. Once an original sample policy is created and the funding premium calculated, the intent of the computer lab spreadsheet application is to enable students to determine the premium on their individually generated policies. A conceptual pedagogic benefit is the premium determination process demonstrates the inter-relationships between mortality charges, expenses, fees, and interest rate assumptions.

Table 3 illustrates the results of the above funding premium solution in a policy year format. The spreadsheet values are written to be self-explanatory. However, Mort. Chg may require one clarification. It is equal to current mortality charge, or 60% of the 1980 CSO rate of $2.11 for a male age 35. If Guar/Curr COI (E4) were set equal to 1 for guaranteed mortality charges, the Mort. Chg in Table 3 would be equal to the 1980 CSO rate of $2.11.

In conclusion, there are pedagogic benefits of modeling life insurance funding in a spreadsheet application, especially when class time is limited. This spreadsheet application also provides an additional intuitive understanding of the sophisticated mathematical and statistical models common to actuarial science.

REFERENCES

A.M. BEST'S FLITCRAFT COMPEND. 1993. Oldwick, NJ: A. M. Best Company.

BLACK, K., AND H. D. SKIPPER. 1994. *Life Insurance.* Englewood Cliffs, NJ: Prentice Hall.

BOWERS, N. L., H. U. GERBER, J. C. HICKMAN, D. A. JONES, AND C. J. NESBITT. 1997. *Actuarial Mathematics.* Schaumburg, IL: Society of Actuaries.

CARR, D. L., AND D. A. FRENCH. 1989. "Interest Scenario Pricing," I-440C Study Note, Education and Examination Committee, Society of Actuaries, (440–34–89).

CHERRY, C. D. 2000. "Calculating Funding Premiums for Universal Life Insurance," *North American Actuarial Journal* 4(2): 20–7.

CHUNG, Y., AND H. D. SKIPPER. 1987. "The Effect of Interest Rates on Surrender Values," *Journal of Risk and Insurance* 54(4): 341–7.

GOLD, A. D. 1994. "10-Year History of Universal Life Policy Survey," *Best's Review, Life/Health Insurance Edition* 95(8): 39–41.

JORDAN, C. W., JR. 1991. *Society of Actuaries' Textbook on Life Contingencies,* Schaumburg, IL: Society of Actuaries.

LANGE, D. R., T. LUCY-BOULER, AND R. M. JOHNSON. 1995. "A Cash Value Sensitivity Analysis Approach To Evaluating Universal Life Insurance Policies," *Academy of Financial Services—Paper Abstracts,* 25.

"Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution," Vytaras Brazauskas and Robert Serfling, October 2000.

MARTIN BILODEAU*

M-estimator and constrained M-estimator (CM) of the tail index of a single-parameter Pareto distribution are investigated and compared to the generalized median (GM) estimator of Brazauskas and Serfling (B&S). Criteria used for comparisons are the asymptotic relative efficiency, upper breakdown point, and gross error sensitivity.

The Pareto distribution $P(d, \alpha)$ has cdf $F(x) = 1 - (d/x)^{\alpha}$, $x \geq d$, defined for parameters $\alpha > 0$ and $d > 0$. The parameter d representing a deductible is assumed known. The transformation $Z = \log(X)$ yields the scale model $Z = \mu + \theta U$, where U follows the standard exponential distribution with pdf $g(u) = \exp(-u)$, $u \geq 0$. The reparametrization is given by $\mu = \log(d)$ and $\theta = 1/\alpha$. The problem is, thus, to estimate the scale parameter θ in this scale model. Subtraction of known μ to Z allows one to assume, without loss of generality, that $\mu = 0$. Based on a sample $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$ the problem is, thus, to estimate θ in the scale model $Z = \theta U$. This problem is treated not only for the standard exponential distribution, but for variables U distributed on the interval $[0, \infty)$ satisfying weak assumptions.

* Martin Bilodeau, Ph.D., A.S.A., is in the Département de mathématiques et de statistique and Centre de recherches mathématiques of Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Canada H3C 3J7, email: bilodeau@dms.umontreal.ca.

M and CM Estimators

An adaptation of the M-estimator of Huber (1981) and of the constrained M-estimator of Kent and Tyler (1996) is presented. Properties of these estimators can be found in Bilodeau (2001).

Define the function

$$\rho_0(u) = \begin{cases} 3u - 3u^2 + u^3, & 0 \le u \le 1 \\ 1, & u > 1. \end{cases} \quad \text{(D1)}$$

An M-estimate of scale $v(\mathcal{Z})$ is defined as a solution v of the equation

$$\text{ave}[\rho(Z_i/v)] = \epsilon$$

for some given ϵ in the interval $(0, 1)$. The notation "ave" refers to the average taken over values indexed by $i = 1, \ldots, n$. The function $\rho(\cdot)$ is defined as $\rho(u) = \rho_0(u/\lambda_0)$, where λ_0 is the unique solution λ of the equation

$$E[\rho_0(U/\lambda)] = \epsilon.$$

A CM-estimator of scale $v(\mathcal{Z})$ is defined as a solution to the problem:

$$\min_{v>0} \text{ave}[\rho(Z_i/v)] + \ln(v)$$

subject to the constraint

$$\text{ave}[\rho(Z_i/v)] \le \epsilon\rho(\infty).$$

The function $\rho(\cdot)$ is defined, for a given constant $C > 0$, as $\rho(u) = C\rho_0(u/\lambda_0)$. The constant λ_0 is the unique solution of the optimization problem

$$\min_{\lambda \ge \lambda_L} E[C\rho_0(U/\lambda)] + \ln(\lambda),$$

where λ_L is the unique solution λ to $E[\rho_0(U/\lambda)] = \epsilon$.

Global and Local Measures of Robustness

The breakdown point of an estimator is a criteria for assessing the *global* robustness under contamination of the sample. As in B&S, upper breakdown point (UBP) is used for contamination in the upper tail, whereas lower breakdown point (LBP) is used for contamination in the lower tail. Estimators with a high UBP (LBP) remain bounded away from infinity (zero) when a large proportion of the sample is sent to infinity (zero). Let $v(\mathcal{Z})$ be an estimate calculated from the original sample $\mathcal{Z}$. The finite sample UBP is defined as $\min\{m : \sup_{\mathcal{Z}(m)} v(\mathcal{Z}(m)) = \infty\}/n$, where $\mathcal{Z}(m)$ is obtained from $\mathcal{Z}$ by sending at most m of the points $Z_1, \ldots, Z_n$ to infinity. It represents the smallest fraction of the sample that, when sent to infinity, can send the estimate beyond any given bound. Similarly, finite sample LBP is defined as the smallest fraction of the sample that, when sent to zero, can send the estimate to zero.

Let $\lfloor x \rfloor$ be the greatest integer smaller than or equal to x. Similarly, $\lceil x \rceil$ will denote the smallest integer greater than or equal to x. M-estimator and CM-estimator have the same finite sample breakdown points given by UBP $= \lceil n\epsilon \rceil/n$ and LBP $= \lfloor n(1 - \epsilon) + 1 \rfloor/n$. Asymptotically, as $n \to \infty$, UBP $\to \epsilon$ and LBP $\to 1 - \epsilon$.

Whereas breakdown point is a measure of global robustness, gross error sensitivity (GES) is a measure of *local* robustness. GES is defined with the influence function (IF). Each estimator has a corresponding functional which defines uniquely the parameter being estimated. For a scale model, the parameter θ being estimated is a solution of an estimating equation of the form

$$E[\psi(Z/v)] = 0.$$

The influence function describes the effect of an infinitesimal contamination at point z on the estimate. The influence function of the functional $v(\cdot)$ at the cdf F_Z is defined pointwise by

$$\text{IF}(z; v(F_Z)) = \lim_{h \downarrow 0+} \frac{v((1-h)F_Z + h\delta_z) - v(F_Z)}{h},$$

if this limit exists. The quantity δ_z is the cdf of the constant z defined as

$$\delta_z(t) = \begin{cases} 0, & t < z \\ 1, & t \ge z. \end{cases}$$

If we replace F_Z by the empirical cdf F_{n-1}, which puts a mass of $1/(n-1)$ at each point $Z_1, \ldots, Z_{n-1}$, and h by $1/n$, then the IF is approximately n times the increase in the estimate when one additional observation z is added to a large sample of size $n - 1$. A bounded influence function is, thus, a desirable measure of local robustness. Moreover, it is preferable that the absolute value of the IF be as small as possible, which leads to the GES measure defined by

$$\text{GES} = \sup_{z \ge 0} |\text{IF}(z; v(F_Z))|.$$

The scale model $Z = \theta U$ is invariant with respect to the group of transformations $Z_i \mapsto aZ_i$,

$a > 0$. It induces the parameter transformation $\theta \mapsto a\theta$. Let $a\mathcal{Z} = \{aZ_1, \ldots, aZ_n\}$. An estimator $\hat{\theta}(\mathcal{Z})$ is said *equivariant* if it undergoes the same transformation, that is, $\hat{\theta}(a\mathcal{Z}) = a\hat{\theta}(\mathcal{Z})$. All estimators studied by B&S including the generalized median estimator, as well as the M-estimator and CM-estimator studied here, are equivariant. As a consequence, all these estimators have a GES of the form GES $= \theta G$ for a certain constant G independent of θ. Hence, G can be taken as a measure of GES.

The maximum likelihood estimates (MLE) with its smallest asymptotic variance provides a quantitative measure of efficiency. The MLE, $\hat{\theta}_{ML} = \text{ave}(Z_i)$, has the asymptotic distribution $n^{1/2}(\hat{\theta}_{ML} - \theta) \xrightarrow{d} N(0, \theta^2)$. All the asymptotically normal and equivariant estimators have an asymptotic distribution of the form $n^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, c\theta^2)$ for some constant $c > 0$. Among such estimators, it has become customary to use

$$\text{ARE}(\hat{\theta}, \hat{\theta}_{ML}) = \frac{1}{c}$$

as a measure of relative efficiency. The asymptotic relative efficiency (ARE) is a relative measure of the asymptotic variance of an estimator when all the observations follow the scale model.

Comparison of Estimators

B&S investigated the performance of some robust estimators: the generalized median, the trimmed mean, and estimators based on selected quantiles. The generalized median was found to outperform the other robust estimators with respect to UBP and ARE. Such comparisons have a serious shortcoming. Namely, they do not measure the effect of infinitesimal perturbations. For example, a large enough tuning constant C of the CM-estimator will bring the ARE close to 1, even for an UBP = 0.5. However, such an estimator could suffer a large increase when one spurious observation is added to the data. In fact, when the CM-estimator is tuned to have a very high ARE, then the GES reaches an unacceptable level.

Table 1 contains a CM-estimator ($\epsilon = 0.5$ and $C = 20$) with an ARE = 0.99 and GES = 8.51. Compared to an M-estimator ($\epsilon = 0.5$), with the same breakdown point, the ARE is 30% higher but the GES is 360% higher. Sometimes, for example, when $\epsilon = .29$ or .21, a suitable choice of the tuning constant C will improve both values. However, the improvements are slight. In other cases, for example $\epsilon = .4$, a CM-estimator cannot improve ARE and GES simultaneously. Note that, for small values of C, the CM-estimator ($\epsilon = .29$ and $C = 4$) corresponds to an M-estimator with the same ϵ.

Table 1
UBP, ARE, and GES Index *G* of Different Estimators

Estimator	UBP	ARE	*G*
GM, $k = 1$	0.50	0.64	1.44
GM, $k = 2$	0.29	0.78	1.90
GM, $k = 3$	0.21	0.88	2.27
GM, $k = 4$	0.16	0.92	2.60
GM, $k = 5$	0.13	0.94	2.88
GM, $k = 10$	0.07	0.98	4.02
M, $\epsilon = 0.50$	0.50	0.76	1.85
M, $\epsilon = 0.29$	0.29	0.93	3.21
M, $\epsilon = 0.21$	0.21	0.97	4.49
M, $\epsilon = 0.16$	0.16	0.98	5.96
M, $\epsilon = 0.13$	0.13	0.99	7.39
M, $\epsilon = 0.07$	0.07	0.997	13.97
CM, $\epsilon = 0.50$, $C = 20$	0.50	0.99	8.51
CM, $\epsilon = 0.29$, $C = 20$	0.29	0.99	8.51
CM, $\epsilon = 0.29$, $C = 7.8$	0.29	0.94	3.17
CM, $\epsilon = 0.29$, $C = 4.0$	0.29	0.93	3.21
CM, $\epsilon = 0.21$, $C = 10.5$	0.21	0.97	4.33

Note: The $\rho_0(\cdot)$ function is given by (1) and U is a standard exponential variable.

The M-estimator and CM-estimator are now compared to the generalized median. The order of the kernel of the generalized median is denoted by k as in B&S. At every level of UBP, the M-estimator is more efficient than the generalized median. However, the generalized median is more B-robust (Hampel et al. 1986, p. 87) in the sense of having lower GES. This should not come as a surprise as it is well known (Hampel et al. 1986, p. 133) that the ordinary median ($k = 1$) is the optimal B-robust estimator of the location of a symmetric distribution. Strictly speaking, however, this result does not apply here since the χ^2_{2k} density is not symmetric. Values of G for generalized median were obtained from Serfling (1984, Equation 2.12),

$$G_{GM} = \frac{k}{2M_{2k}g_{2k}(M_{2k})},$$

where $g_{2k}(x)$ is the χ^2_{2k} density and M_{2k} is the median of χ^2_{2k}.

The computational burden of an M-estimator does not depend on UBP. It grows with n as $O(n)$ compared to $O(n^k)$ for the generalized median. Exact evaluation of an M-estimator for $\epsilon = .07$ and $n = 500$ requires only five seconds with the S-PLUS 2000 software on a Pentium III 500 MHz computer.

In conclusion, M, CM, and GM estimators are globally robust; they can all reach a high breakdown point. For a given UBP; M and, to a larger extent, CM estimators are more efficient than the GM estimator. However, the GM estimator is more locally robust.

REFERENCES

BILODEAU, M. 2001. "Robust Estimation of the Tail Index of a Pareto Distribution." Technical Report, Université de Montréal.

HAMPEL, F. R., E. M. RONCHETTI, P. J. ROUSSEEUW, AND W. A. STAHEL. 1986. *Robust Statistics: The Approach Based on Influence Functions.* New York: John Wiley & Sons.

HUBER, P. J. 1981. *Robust Statistics.* New York: John Wiley & Sons.

KENT, J. T., AND D. E. TYLER. 1996. "Constrained M-Estimation for Multivariate Location and Scatter," *Annals of Statistics* 24: 1346–70.

SERFLING, R. J. 1984. "Generalized L-, M- and R-statistics," *Annals of Statistics* 12: 76–86.

Authors' Reply

We heartily thank Dr. Bilodeau for joining our effort to emphasize robustness and improve upon conventional estimators in the problem of estimating the tail index α of the Pareto distribution $F(x) = 1 - (\sigma/x)^{\alpha}$, $x \geq \sigma$, with σ known. In our paper we introduced a new estimator of the generalized median (GM) type and found it to dominate conventional estimators by offering better trade-offs between asymptotic relative efficiency (ARE) as an efficiency criterion and upper breakdown point (UBP) as a robustness criterion.

Dr. Bilodeau's discussion very nicely augments our paper by considering further estimators (M and CM) and suggesting the use of a further robustness criterion, the gross error sensitivity (GES), which is the maximum magnitude of the influence curve. For extensive treatment and discussion of influence curves and the robustness measures based on them, see Hampel (1974), Serfling (1980), and Hampel et al. (1986). In particular, we note that the average of the evaluations of the influence curve at the sample values may be interpreted as a first-order approximation to the estimation error. Then the GES represents the maximum possible contribution any observation can make in this average. On the other hand, the UBP represents the largest fraction of upper observations which may be corrupted without destroying the stability of the estimator. It is desirable to have high UBP and low GES combined with high ARE.

In Dr. Bilodeau's Table 1, the GM, M, and CM estimators are compared using ARE, UBP, and G = GES/α together. It also is of interest to compare with trimmed mean (TM) estimators, which in our paper were found to be the ones most closely competitive with the GM estimators. Also, just as we introduced *upper* breakdown point in order to confine attention to upper outliers, we follow Kimber (1983a,b) and use the notion of *upper* gross error sensitivity (GES^+) given by the maximum of the influence curve over *upper* values of x.

As before, our TM estimator for α is $\hat{\theta}^{-1}$, where $\hat{\theta}$ is the TM estimator of θ in the equivalent problem of estimation of θ in the exponential model $E(\mu, \theta)$ with cdf $\tilde{F}(x) = 1 - e^{-(x-\mu)/\theta}$, $x \geq \mu$. It is readily seen that $G^+ = \text{GES}^+/\alpha$ for the Pareto problem is identical with GES^+/θ as computed for the model $E(\mu, \theta)$. Without loss of generality we set $\mu = 0$ and compute G^+ under the model $E(0, \theta)$. Denoting by β_1 the lower proportion of observations trimmed and by β_2 the upper proportion trimmed, the relevant TM functional for θ is defined by

$$T(\tilde{F}) = d(\beta_1, \beta_2)^{-1} \int_{\beta_1}^{1-\beta_2} \tilde{F}^{-1}(p)\, dp,$$

where $d(\beta_1, \beta_2) = 1 - \beta_1 - \beta_2 + (1 - \beta_1)\log(1 - \beta_1) + \beta_2 \log \beta_2$, a factor making the sample analogue $\hat{\theta}$ of $T(\tilde{F})$ a consistent estimator of θ. A calculation of the influence curve (as in Kimber 1983b) then yields the following formula for GES^+/θ:

$$G^+ = d(\beta_1, \beta_2)^{-1}[\log(1 - \beta_1) - \log(1 - \beta_2) - (\beta_2 - \beta_1)].$$

For β_2 taking values 0.05, 0.10, 0.15, 0.20, and 0.25, we consider two cases, (i) $\beta_1 = \beta_2$ and (ii) $\beta_1 = 0$. For these choices of β_2, the G^+ values in cases (i) and (ii) differ but the UBPs are the same and (as noted in our paper) the AREs agree within two decimal places. In Table A, we provide UBP, ARE, and G^+ values for these TM estimators, paralleling Dr. Bilodeau's Table 1 which covers the GM, M, and

CM estimators. Also, the selection of GM estimators in his Table 1 can be slightly improved, and, in Table B, we provide an updated version of the part of his table concerning the GM estimators.

For the choices of GM, M, and CM estimators considered in Table B and in Dr. Bilodeau's Table 1, GES = GES^+. Thus, from consideration of all three tables together, we conclude:

- The GM and TM estimators offer the best trade-offs between ARE and GES^+, while the M and CM estimators offer the best trade-offs between ARE and UBP.
- For any given level of trade-off between ARE and GES^+, the GM estimators dominate the TM estimators by virtue of better UBP.

We hope that these perspectives will assist practitioners in comparing GM, M, CM, and TM estimators.

We note for the TM estimators that, as might be expected, increased trimming yields improved G^+. It is of interest, however, that this is not so for the GES measure. For example, for case (ii) above, for $\beta_2 = 0.05, 0.10, 0.15$, and 0.20, we have GES = G^+, and these values decrease, whereas for $\beta_2 = 0.21$, we have GES = $1.68 \neq 1.71 = G^+$, and the GES values (now equal to G^-) begin increasing with β_2. This is because, for $\beta_2 \geq 0.21$, the associated influence curve in case (ii) has maximum magnitude for x at the origin $\mu = 0$ rather than for $x \to \infty$. Thus, in the case of a distribution with support on an interval of form $[A, \infty)$ for finite A, a TM estimator with upper trimming but no lower trimming ultimately becomes more influenced by inliers than by outliers, as the trimming level increases. In such cases, if only upper outliers are of concern, the GES measure can be misleading and GES^+ should be used instead.

In addition to comparing estimators via numerical efficiency and robustness measures, it is also of value to consider what the estimators actually do with the data. Within a multiplicative factor, the GM estimator is the median of all the maximum likelihood estimators based on subsamples of a given size k. The TM estimators simply average the observations with zero weight to outlying observations and constant weight to inner observations. The M and CM estimators are produced by solving specified equations or optimization problems based on the function $\rho_0(\cdot)$ given by Dr. Bilodeau's Equation (D1). Thus, operationally, the GM, M, CM, and TM estimators have very different intuitive interpretations.

Finally, we briefly mention further work that has been carried out in connection with our paper under discussion. In Brazauskas and Serfling (2001) we perform a small sample simulation study of the same collection of estimators, using as criteria UBP and exact instead of asymptotic relative efficiency. In addition, we employ a specific contamination model for outliers and evaluate efficiency-robustness trade-offs corresponding to the absence or presence of contamination. The results again favor the GM estimators over conventional ones.

In Brazauskas and Serfling (2000) we treat the related problem of estimation of α with σ unknown. Again the GM estimators dominate the conventional ones with respect to ARE and UBP. Furthermore, it can be shown that the ARE and GES^+ values in this case agree exactly with corresponding values for the case of σ known, so that the trade-offs between ARE and GES^+ are identical in the two cases.

Table A
UBP, ARE, and G^+ for TM Estimators, for Selected β_2

β_2	UBP	ARE	G^+, Case (i)	G^+, Case (ii)
0.25	0.25	0.67	1.62	1.58
0.20	0.20	0.72	1.72	1.69
0.15	0.15	0.78	1.87	1.85
0.10	0.10	0.85	2.10	2.09
0.05	0.05	0.92	2.56	2.56

Table B
UBP, ARE, and G^+ for Selected GM Estimators

Estimator	UBP	ARE	G^+
GM, $k = 1$	0.50	0.64	1.44
GM, $k = 2$	0.29	0.78	1.90
GM, $k = 3$	0.21	0.88	2.27
GM, $k = 4$	0.16	0.92	2.60
GM, $k = 5$	0.13	0.94	2.88
GM, $k = 7$	0.09	0.97	3.38
GM, $k = 9$	0.07	0.98	3.82

REFERENCES

BRAZAUSKAS, V., AND R. SERFLING. 2000. "Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics," *Extremes* 3: 231–49.

———. 2001. "Small Sample Performance of Robust Estimators of Tail Parameters for Pareto and Exponential Models,"

Journal of Statistical Computation and Simulation 70(August), in press.

HAMPEL, F. R. 1974. "The Influence Curve and Its Role in Robust Estimation," *Journal of American Statistical Association* 69: 383–97.

HAMPEL, F. R., E. M. RONCHETTI, P. J. ROUSSEEUW, AND W. A. STAHEL. 1986. *Robust Statistics: The Approach Based on Influence Functions.* New York: John Wiley and Sons.

KIMBER, A. C. 1983a. Trimming in Gamma Samples. *Applied Statistics* 32: 7–14.

KIMBER, A. C. 1983b. "Comparison of Some Robust Estimators of Scale in Gamma Samples with Known Shape," *Journal of Statistical Computation and Simulation* 18: 273–86.

SERFLING, R. J. 1980. *Approximation Theorems of Mathematical Statistics.* New York: John Wiley and Sons.

"Valuing Equity-Indexed Annuities," Serena Tiong, October 2000

YA-CHUN HUANG*

Equity-indexed annuities (EIAs) have been the fastest growing annuity products since 1995. This product line is successful because it provides to the policyholder the upside market potential without a downside market risk. The most popular product designs, such as the point-to-point, the cliquet, and the lookback have been presented in Dr. Tiong's informative paper. She successfully applied the method of Esscher transforms to evaluate these EIA products and also discussed several important issues for pricing.

As will become clear, although the feature of the lookback design is attractive to the clients, the fact that its higher cost ends up with a lower participation rate is of great concern for insurers. Whereas the focus of this discussion is mainly on the valuation of a guarantee embedded in a lookback EIA contract, the response to changes in various economic factors is also examined through several sensitivity tests.

A Continuous Lookback (High-Water-Mark) Design

In practice, the high-water-mark design credits interest as a portion of the percentage growth in the underlying index value from the initial index value to the highest value the index has attained over the measurement period. The measurement can be daily, monthly, or annually. Here, we consider a lookback design in a continuous setting. At the maturity time T, a customer is promised the maximum of a fixed value K or the highest return during the term of the policy, subject to a participation rate α. The participation rate α determines a portion of return credited to the customer's account.

To derive a closed-form solution for the value of this lookback contract, further assume that the risk-free interest rate is constant, no death and no surrender occur before time T, and the market is frictionless. The problem is formulated as follows: Assume that we are given a filtered probability space $(\Omega, \mathcal{F}, F, P)$. The filtration F specifies how the information is revealed in time and is a family $\{\mathcal{F}_t\}$ of increasing σ-algebras on $(\Omega, \mathcal{F})$, $\mathcal{F}_t \subset \mathcal{F}$. We denote $\mathcal{F}_t$ the σ-algebra generated by the Brownian motion up to time t. Let $S(t)$ denote the time-t value of an index, which pays dividends $\delta S(t)dt$ between t and $t + dt$. It is assumed that the index follows a geometric Brownian motion, $S(t) = S(0)e^{X(t)}$, where $\{X(t)\}$ is a Brownian motion with instantaneous drift μ, and instantaneous variance σ^2.

Suppose that a customer invested in an equity-indexed deferred annuity for a single premium f at the inception of the contract. In practice, the insurer promises the total value to a customer's account at time T as

$$\max\left\{f + f\alpha\,\frac{\max_{s\in[0,T]} S(s) - S(0)}{S(0)}, K\right\}.$$

Using $S(s) = S(0)e^{X(s)}$ and $\max_{s\in[0,T]} S(s) = S(0)e^{\max_{s\in[0,T]} X(s)}$, the formula becomes

$$\max\{f(1 + \alpha(e^{\max_{s\in[0,T]} X(s)} - 1)), K\}.$$

In a continuous setting, use the following approximation instead:

$$(e^{\max_{s\in[0,T]} X(s)})^{\alpha} \simeq (1 + \alpha(e^{\max_{s\in[0,T]} X(s)} - 1)).$$

The guaranteed payoff to the customer at time T is now determined by the formula:

$$\max\{fe^{\alpha \max_{s\in[0,T]} X(s)}, K\}. \qquad \text{(D1)}$$

* Ya-Chun Huang is a Ph.D. Student, Program in Applied Mathematical and Computational Sciences, The University of Iowa, Iowa City, IA 52242, yhuang@math.uiowa.edu.