

Favorable Estimators for Fitting Pareto Models: A Study  
Using Goodness-of-Fit Measures with Actual Data

Vytaras Brazauskas <sup>1</sup>

*University of Wisconsin-Milwaukee*

and

Robert Serfling <sup>2</sup>

*University of Texas at Dallas*

To appear in the *ASTIN Bulletin*, 2004

<sup>1</sup>Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA. E-mail: [vytaras@uwm.edu](mailto:vytaras@uwm.edu) Supported by a grant from the Actuarial Education and Research Fund.

<sup>2</sup>Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. E-mail: [serfling@utdallas.edu](mailto:serfling@utdallas.edu) 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-0103698.

## **Abstract**

Several recent papers treated robust and efficient estimation of tail index parameters for (equivalent) Pareto and truncated exponential models, for large and small samples. New robust estimators of “generalized median” (GM) and “trimmed mean” (T) type were introduced and shown to provide more favorable trade-offs between efficiency and robustness than several well-established estimators, including those corresponding to methods of maximum likelihood, quantiles, and percentile matching. Here we investigate performance of the above mentioned estimators on real data and establish – via the use of goodness-of-fit measures – that favorable theoretical properties of the GM and T type estimators translate into an excellent practical performance. Further, we arrive at guidelines for Pareto model diagnostics, testing, and selection of particular robust estimators in practice. Model fits provided by the estimators are ranked and compared on the basis of Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling statistics.

# 1 Introduction and Preliminaries

A single-parameter Pareto distribution plays a very significant role in actuarial modeling because of its conceptual simplicity and ease of applicability in practice. The cdf of the Pareto  $P(\sigma, \alpha)$  model is given by

$$F(x) = 1 - (\sigma/x)^\alpha, \quad x > \sigma, \quad (1)$$

where  $\alpha > 0$  is the shape parameter that characterizes the tail of the distribution and  $\sigma > 0$  is the scale parameter. When  $\sigma$  is assumed known, the  $P(\sigma, \alpha)$  model is called a single-parameter Pareto model. The assumption of  $\sigma$  *known* is quite typical in the actuarial literature because, as for example Philbrick (1985) states, “although there may be situations where this value must be estimated, in virtually all insurance applications this value will be selected in advance.” (See also discussion by Rytgaard (1990).)

In several recent papers Brazauskas and Serfling (2000a,b; 2001) treated robust and efficient estimation of the tail index  $\alpha$  for various setups: for *large-* and *small-samples* and for *one-* and *two-parameter* models (corresponding to  $\sigma$  *known* or *unknown*). Developments presented there utilized a well-known equivalence relation between model (1) and the truncated exponential distribution  $E(\mu, \theta)$  having cdf

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

for  $\theta > 0$  and  $-\infty < \mu < \infty$ . Specifically, if random variable  $X$  has cdf  $F$  given by (1) then variable  $Z = \log X$  has cdf  $G$  given by (2), with  $\mu = \log \sigma$  and  $\theta = \alpha^{-1}$ .

In large-sample studies, for example, new robust estimators of “generalized median” (GM) type were introduced and “trimmed mean” (T) type estimators were adapted from the  $E(\mu, \theta)$  model literature. These estimators were then compared with the maximum likelihood, quantile type, percentile matching, and other estimators. Using as *efficiency* criterion the *asymptotic relative efficiency* (ARE) with respect to the *maximum likelihood estimator* (MLE) and as *robustness criterion* the *breakdown point* (BP) (this is defined in Section 1.2), the GM type was seen to dominate all competitors, with the T type second best. From a practical point of view, the ARE is equivalent to the *accuracy* of the estimator and can be interpreted in terms of the length of the confidence interval (see Section 1.3 for precise discussion).

In the present paper we investigate performance of the above mentioned estimators on real data and establish – via the use of goodness-of-fit measures – that favorable theoretical properties of the GM and T type estimators translate into an excellent practical performance. The goodness-of-fit measures, defined in Section 1.1, are used here for two purposes: (i) to (formally) test the appropriateness of the estimated Pareto model for a particular data set when  $\alpha$  is estimated by the MLE (this is defined in Section 1.3), and (ii) to evaluate and compare Pareto fits when various estimators (not only the MLE) of  $\alpha$  are employed.

In the actuarial literature the issue of goodness-of-fit is addressed through a combination of informal methods and formal statistical tests. Most informal techniques are based on the difference (absolute or relative) between the fitted and empirical values of relevant

quantities, such as the *number of claims* or *expected value of claims* for different claim layers. Additionally, for the Pareto model in particular, comparisons of  $\hat{\alpha}$  with a typical value of  $\alpha$  for the same insurance line of the entire industry are also used in the literature. (See Patrick (1980), Philbrick (1985), and Reichle and Yonkunas (1985).) Regarding formal approaches, tests based on *Kolmogorov-Smirnov (KS)* and  $\chi^2$  statistics seem to have a leading role (see, e.g., Philbrick and Jurschak (1981)). More extensive discussion on model validation principles is available in Klugman, Panjer, and Willmot (1998), Section 2.9.

As is well-known in the statistical literature (e.g., D'Agostino and Stephens (1986), p. 110), the  $\chi^2$  test is less powerful than tests based on the empirical cdf. Therefore, here we use three widely popular goodness-of-fit measures which are based on the empirical cdf – the above-mentioned KS statistic, the *Cramér-von Mises (CvM)* statistic, and the *Anderson-Darling (AD)* statistic. All these statistics (though emphasizing different aspects of discrepancy) measure the distance in some sense between the fitted model cdf  $\hat{F}$  and the empirical cdf  $F_n$ . Thus estimators that lead to smaller values of these statistics are preferable.

The paper is organized as follows. First, in Section 2, we define precisely several estimators for the parameter  $\alpha$  in (1). Next, in Section 3, the data sets are introduced, a method for data de-grouping is described, and preliminary data visualization and diagnostic tools are applied to the sets. Finally, comparisons and conclusions are presented in Section 4. Also, in Section 4.2, we arrive at guidelines for Pareto model diagnostics, testing, and selection of particular robust estimators in practice.

In the remainder of this introduction, we formulate precisely our performance criteria.

## 1.1 Goodness-of-Fit Measures

Let us consider a sample  $X_1, \dots, X_n$  and denote the ordered sample values by  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  and the empirical cdf by

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

Also, for an estimator  $\hat{\alpha}$ , let  $\hat{F}(X_{(j)})$  denote the probability assigned to  $X_{(j)}$  by the model  $P(\sigma, \hat{\alpha})$ , for  $j = 1, \dots, n$ . Note that  $F_n(X_{(j)}) = j/n$ , for  $j = 1, \dots, n$ . The goodness of fit statistics are then defined as follows.

The KS statistic  $D_n$ :

$$D_n^+ = \max_{1 \leq j \leq n} \left( \frac{j}{n} - \hat{F}(X_{(j)}) \right), \quad D_n^- = \max_{1 \leq j \leq n} \left( \hat{F}(X_{(j)}) - \frac{j-1}{n} \right),$$

$$D_n = \max\{D_n^+, D_n^-\}.$$

The CvM statistic  $W_n^2$ :

$$W_n^2 = \sum_{j=1}^n \left( \hat{F}(X_{(j)}) - \frac{2j-1}{2n} \right)^2 + \frac{1}{12n}.$$

The AD statistic  $A_n^2$ :

$$A_n^2 = -n - \frac{1}{n} \sum_{j=1}^n \left\{ (2j-1) \log \left( \hat{F}(X_{(j)}) \right) + (2n+1-2j) \log \left( 1 - \hat{F}(X_{(j)}) \right) \right\}.$$

When the parameter  $\alpha$  is estimated by  $\hat{\alpha}_{\text{ML}}$ , critical values and formulas for significance levels for the statistics  $D_n$ ,  $W_n^2$ , and  $A_n^2$  are available in D'Agostino and Stephens (1986), Tables 4.11 and 4.12, pp. 135–136. Actually, the results developed there are for the model  $E(\mu, \theta)$  but, due to the equivalence relation between (1) and (2), they can also be directly applied to the model  $P(\sigma, \alpha)$ .

## 1.2 Robustness Criterion: Breakdown Point

A popular and effective criterion for robustness of an estimator is its *breakdown point* (BP), loosely characterized as the largest proportion of corrupted sample observations that the estimator can cope with. In other words, the BP of an estimator measures the degree of resistance of the estimator to the influence of outlying observations which possibly (but not necessarily) represent contamination of a data set rather than merely unusually extreme observations generated by the target parametric model.

Brazauskas and Serfling (2000a,b) considered two types of contamination – upper and lower contamination – and, consequently, defined separate versions of BP:

*Lower (Upper) Breakdown Point* (LBP/UBP): 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.

For modeling insurance loss data, however, contamination of the lower type is of lesser concern because the lower limit of losses is usually pre-defined by a contract. (For example, the lower limit can be represented as a deductible.) Thus, in the present treatment we favor estimators which have *nonzero* UBP.

## 1.3 Efficiency Criterion: Variance

If sample observations follow the postulated parametric model, then it is well-known that, for large data sets, the MLE attains (in its approximating normal distribution) the minimum possible variance among a large class of competing estimators. Therefore, it can be regarded as a quantitative benchmark for efficiency considerations. In particular, for the model  $P(\sigma, \alpha)$  with  $\sigma$  known, the MLE of  $\alpha$  is readily derived in Arnold (1983), and given by

$$\hat{\alpha}_{\text{ML}} = \frac{n}{\sum_{i=1}^n \log(X_i/\sigma)}.$$

Its exact distribution theory is described by the statement that

$$\frac{2n\alpha}{\hat{\alpha}_{\text{ML}}} \text{ has cdf } \chi_{2n}^2,$$

where  $\chi_\nu^2$  denotes the chi-square distribution with  $\nu$  degrees of freedom. This implies that  $\hat{\alpha}_{\text{ML}}$  is a biased estimator of  $\alpha$ , but multiplication by the factor  $(n-1)/n$  yields an unbiased version,

$$\hat{\alpha}_{\text{MLU}} = \frac{n-1}{n} \hat{\alpha}_{\text{ML}} = \frac{n-1}{\sum_{i=1}^n \log(X_i/\sigma)}.$$

For further details on exact distribution theory of  $\hat{\alpha}_{\text{ML}}$  see Rytgaard (1990).

Following techniques in Brazauskas and Serfling (2000a,b), it can be shown that for large sample size  $n$ ,  $\hat{\alpha}_{\text{MLU}}$  is approximately normal with mean  $\alpha$  and variance  $\alpha^2/n$ . Moreover, other competing estimators  $\hat{\alpha}$  for  $\alpha$  considered here are approximately normal with mean  $\alpha$  and variance  $c\alpha^2/n$  for some constant  $c > 1$  and large  $n$ . This means that confidence intervals for the parameter  $\alpha$  based on the competing estimators will be  $\sqrt{c}$  times wider than those based on the MLU. Such optimal precision of the MLU, however, is achieved at the price of robustness, which becomes crucial when the actual data departs from the assumed parametric model. Hence, the MLU is most *efficient* but is *nonrobust* with  $\text{UBP} = 0$ .

## 2 The Estimators

The MLE and MLU were given in Section 1.3. Here we introduce the other methods considered in this study for estimation of the parameter  $\alpha$ . In particular, we present quantile, trimmed mean, and generalized median type estimators. For further details and discussion see Brazauskas and Serfling (2000a,b).

### 2.1 Quantile Type Estimators

Quantile type estimators of  $\alpha$  are completely unaffected by additional information about  $\sigma$ . For this reason and for compatibility with the existing literature, we describe this approach here for the case when  $\sigma$  is treated as an unknown nuisance parameter.

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

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

with

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

and

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

where  $u_i = -\log(1 - p_i)$ ,  $1 \leq i \leq k$ , and  $[x]$  denotes the least integer  $\geq x$ . Such estimators were introduced and studied for the Pareto problem by Quandt (1966) for  $k = 2$  and by Koutrouvelis (1981) for general  $k \geq 2$ .

Choosing the minimum of the determinant of the asymptotic covariance matrix of the estimators of  $\sigma$  and  $\alpha$  as an optimality criterion, Koutrouvelis (1981) found that the optimal choice of  $p_1$  is always

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

and the remaining optimal quantile levels  $p_2^\circ, \dots, p_k^\circ$  are:

- For  $k = 2$ , take  $p_2^\circ = 1 - (1 - p_1^\circ) e^{-1.5936} \approx .80$ .
- For  $k = 5$ , take  $p_2^\circ = 1 - (1 - p_1^\circ) e^{-0.6003} \approx .45$ ,  $p_3^\circ = 1 - (1 - p_1^\circ) e^{-1.3544} \approx .74$ ,  $p_4^\circ = 1 - (1 - p_1^\circ) e^{-2.3721} \approx .91$ , and  $p_5^\circ = 1 - (1 - p_1^\circ) e^{-3.9657} \approx .98$ .

We denote the optimal estimators of  $\alpha$  by  $\hat{\alpha}_Q^{\text{opt},k}$ . We also consider a nonoptimal case (denoted by  $\hat{\alpha}_Q^*$ ):

- For  $k = 5$ , take  $p_1 = .13$ ,  $p_2 = .315$ ,  $p_3 = .50$ ,  $p_4 = .685$ , and  $p_5 = .87$ .

**Remark** When the number  $k$  of quantiles is chosen to equal the number of unknown parameters of the model, this method corresponds to what is called *percentile matching* by Klugman, Panjer, and Willmot (1998).

## 2.2 Trimmed Mean Estimators

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

$$\hat{\alpha}_T = \left( \sum_{i=1}^n c_{ni} \log(X_{(i)}/\sigma) \right)^{-1},$$

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

$$d(\beta_1, \beta_2, n) = \sum_{j=[n\beta_1]+1}^{n-[n\beta_2]} \sum_{i=0}^{j-1} (n-i)^{-1}.$$

These estimators correspond to the trimmed mean estimators introduced and studied by Kimber (1983a,b) for the equivalent problem of estimation of  $\theta = \alpha^{-1}$  in the model  $E(\mu, \theta)$  with  $\mu$  known. The above  $c_{ni}$ 's are a choice making  $\hat{\theta}_T = \hat{\alpha}_T^{-1}$  mean-unbiased for  $\theta = \alpha^{-1}$ .

### 2.3 Generalized Median Estimators

Generalized median (GM) statistics are defined by taking the median of the  $\binom{n}{k}$  evaluations of a given kernel  $h(x_1, \dots, x_k)$  over all  $k$ -sets of the data. See Serfling (1984, 2000) for general discussion. In Brazauskas and Serfling (2000a), such estimators were considered for the parameter  $\alpha$  in the case of  $\sigma$  known:

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

with a particular kernel  $h(x_1, \dots, x_k)$ :

$$h(x_1, \dots, x_k; \sigma) = \frac{1}{C_k} \frac{k}{\sum_{j=1}^k \log(x_j/\sigma)},$$

where  $C_k$  is a multiplicative median-unbiasing factor, i.e., chosen so that the distribution of  $h(X_{i_1}, \dots, X_{i_k}; \sigma)$  has *median*  $\alpha$ . Values of  $C_k$ , for  $k = 2 : 10$ , are provided in the following table. (For  $k > 10$ ,  $C_k$  is given by a very accurate approximation,  $C_k \approx k/(k - 1/3)$ .)

$k$	2	3	4	5	6	7	8	9	10
$C_k$	1.1916	1.1219	1.0893	1.0705	1.0582	1.0495	1.0431	1.0382	1.0343

## 3 Data Sets and Preliminary Diagnostics

We choose three data sets for analysis in this study: Wind Catastrophes (1977), OLT Bodily Injury Liability Claims (1976), and Norwegian Fire Claims (1975). These data sets are of interest because they have been analyzed extensively in the actuarial literature. In this section we first present the data and briefly mention methods of analysis proposed in the literature. Then we describe a data de-grouping technique which we apply for the wind data, the liability data, and the Norwegian data. Finally, for an initial assessment of the validity of distributional assumptions, we provide histograms and *quantile-quantile* plots (QQ-plots) for each data set.

### 3.1 Wind Catastrophes (1977)

The Wind Catastrophes (1977) data set is taken from Hogg and Klugman (1984), p. 64. It represents 40 losses that occurred in 1977 due to wind-related catastrophes. The data were recorded to the nearest \$1,000,000 and include only those losses of \$2,000,000 or more. The following display provides the losses (in millions of dollars):

2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 4 4 4 5  
5 5 5 6 6 6 6 8 8 9 15 17 22 23 24 24 25 27 32 43

In Hogg and Klugman (1984) two parametric models were used to fit the wind data: truncated exponential (with the truncation point 1.5) and two-parameter Pareto. Derrig,

Ostaszewski, and Rempala (2000) also studied this data set and, in addition to the above parametric models, used the empirical nonparametric and the bootstrap approaches to estimate the probability that a wind loss will exceed 29.5 (that is, \$29,500,000). Further, Philbrick (1985), among several applications of the single-parameter Pareto distribution to real data, investigated a  $P(\sigma, \alpha)$  fit to the wind data with the truncation point  $\sigma = 2$ . He advocated the use of the MLE for estimation of  $\alpha$  but apparently was unaware that this estimator is biased.

### 3.2 OLT Bodily Injury Liability Claims (1976)

The complete OLT Bodily Injury Liability (1976) data set is available in Patrik (1980), p. 99. It is prepared by the Insurance Services Office and represents Owners, Landlords and Tenants (OLT) bodily injury liability losses for the policy limit \$500,000 for policy year 1976 evaluated as of March 31, 1978. Patrik (1980) described general principles of selection, estimation, and testing of loss models for casualty insurance claims. For illustrative purposes he used the two-parameter Pareto distribution to fit the entire range of claims, including the OLT Bodily Injury Liability (1976) claims.

Here we follow Philbrick's (1985) approach and fit the single-parameter Pareto distribution only to the claims that are greater than \$25,000. The grouped losses (recorded in thousands of dollars) in exceedance of this threshold are presented in Appendix, Table A.1.

### 3.3 Norwegian Fire Claims (1975)

This data set is one among 20 sets of Norwegian Fire Claims, for years 1972-1984 and 1986-1992, presented in Appendix I of Beirlant, Teugels, and Vynckier (1996). It represents the total damage done by 142 fires in Norway for the year 1975. (For this year a single-parameter Pareto distribution seems to provide a reasonably good fit to these data. Fits of similar quality are observed for several other years as well.) A priority of 500,000 Norwegian kroner was in force, thus no claims below this limit were recorded. Actual losses (in thousands of Norwegian kroner) are provided in Appendix, Table A.2.

The Norwegian fire claims for various years have been extensively analyzed by Beirlant, Teugels, and Vynckier (1996). Their approach is based on extreme value theory, which concentrates exclusively on the upper tail of the data. We will not pursue that approach here. However, we will follow techniques developed there for data diagnostics, namely, methods of Section 1.5 which describe how to construct specific QQ-plots.

### 3.4 A Method for Data De-grouping

We start with a motivational example based on the wind data. Losses recorded there are rounded to the nearest million which suggests that *actual* losses corresponding to 2, for example, were not exactly 2 but rather somewhere between 1.5 and 2.5. (This seems to be one of the reasons why Hogg and Klugman (1984) considered the left-hand endpoint 1.5 for the truncated exponential model.) To avoid ties and inappropriate clustering of claims due to such rounding, we apply a simple data de-grouping method as follows.

Let us continue with the wind data and, in particular, losses of size 2. It is reasonable to assume that actual observations that correspond to 2 are *equally spaced* (or, equivalently, *uniformly distributed*) on the interval (1.5, 2.5). Thus, for the wind data, instead of the 12 observations “2” we shall use 1.58, 1.65, 1.73, 1.81, 1.88, 1.96, 2.04, 2.12, 2.19, 2.27, 2.35, 2.42, as the actual data. More formally,

if  $(A, B)$  is an interval of losses and  $m$  is the number of losses within  $(A, B)$ , then  $m$  uniformly distributed losses  $x_1, \dots, x_m$  in that interval are found according to the formula:

$$x_k = \left(1 - \frac{k}{m+1}\right) A + \frac{k}{m+1} B, \quad k = 1, \dots, m.$$

We emphasize that such an approach neither distorts the original grouping nor changes the total loss amount within a group. It is easy to implement in practice and, most importantly, it makes the data continuous, thus allowing methods of estimation and goodness-of-fit to be applied directly. We apply this technique to all three data sets.

Finally, one may also consider more sophisticated data de-grouping schemes by employing, for example, the beta family of distributions instead of uniform. In that case, however, additional information, such as mean and variance of losses within the interval, is required.

### 3.5 Preliminary Diagnostics

In Figure 1, we illustrate the results of preliminary diagnostics for the three data sets at hand. Three plots in the first column correspond to the Wind Catastrophes (1977), in the second column – OLT Bodily Injury Liability Claims (1976), and in the third column – Norwegian Fire Claims (1975) data. The following conclusions are quite evident:

- Histograms for all the data sets exhibit a similar shape for the underlying distribution. Two *one-parameter* models seem to be appropriate candidates: the truncated exponential, and the Pareto.
- The exponential QQ-plots clearly reveal that a truncated exponential model does not fit any of the three data sets (a good fit corresponds to a 45° line).
- In comparison with the exponential case, the Pareto QQ-plot shows *mild improvement* for the wind data, *significant improvement* for the liability data, and *nearly perfect fit* for the Norwegian fire data.

**Remark** Plotting of lognormal or Weibull QQ-plots, for example, reveals significant improvements over the one-parameter Pareto model for the wind and the liability data. These distributions, however, are *two-parameter* models (thus, less parsimonious) and are not considered as competitors to the one-parameter Pareto model.  $\square$

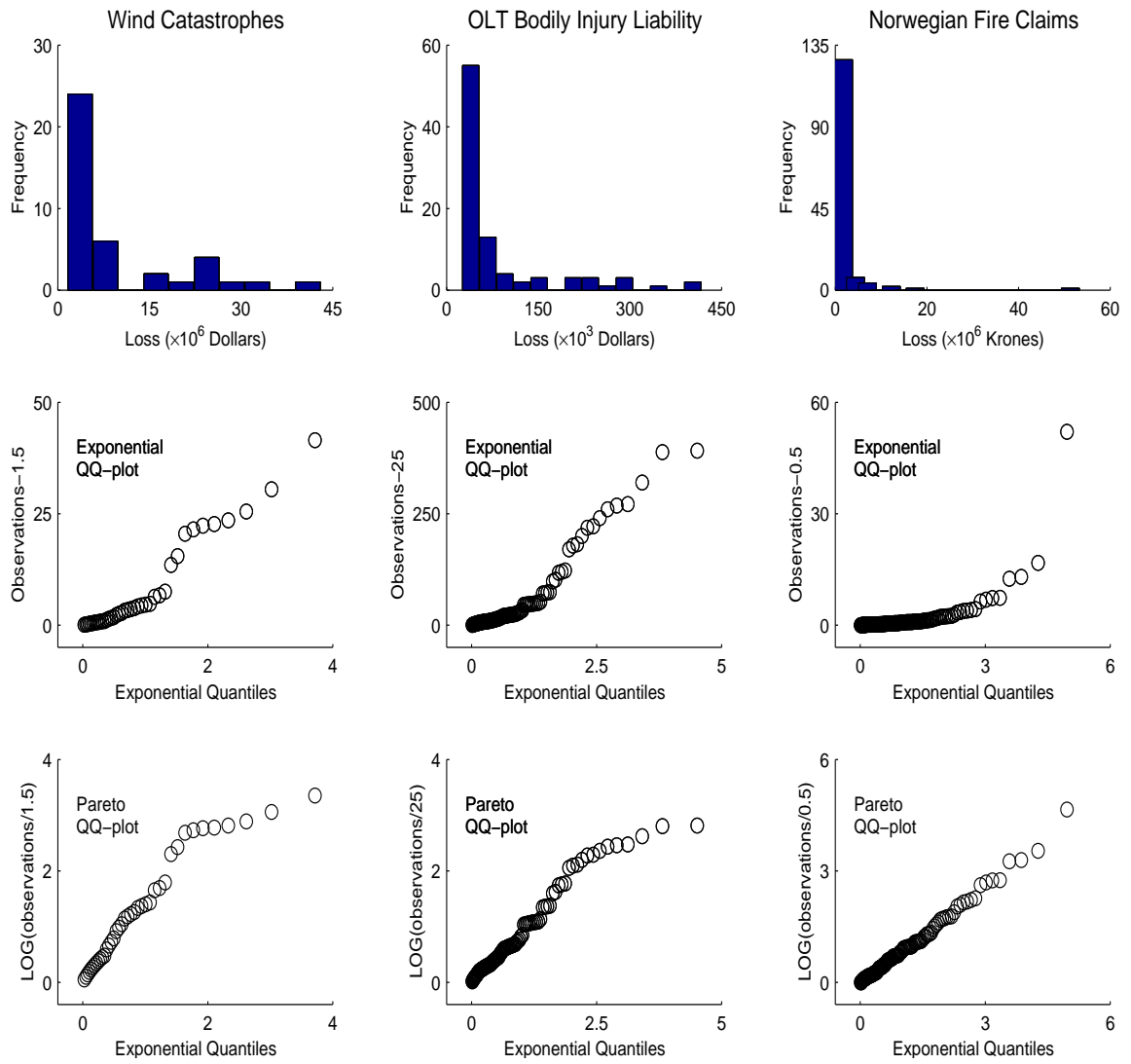


FIGURE 1. Preliminary diagnostics for all three data sets.

## 4 Comparisons and Conclusions

In Section 4.1 we present summarizing tables for each data set, showing values of the estimates of  $\alpha$ , values of the KS, CvM, and AD statistics, and ranks of the estimators based on these goodness-of-fit measures. In Section 4.2, conclusions are drawn and recommendations are provided. Performances of all estimators under consideration are compared by simultaneously examining their ranks (for all three data sets), their UBP's, and their variances.

**Remarks** (i) Ranks to estimators are assigned as follows. The estimator with the lowest value for a selected goodness-of-fit measure receives rank 1, the estimator with the second lowest value (for the same measure) – receives rank 2, etc. The idea of ranking estimators or models based on a certain criterion is not new. It has been suggested and quite extensively discussed by Klugman, Panjer, and Willmot (1998), Section 2.9.2.

(ii) For situations when the number  $\binom{n}{k}$  of kernel evaluations needed for computation of  $\hat{\alpha}_{\text{GM}}$  becomes extremely large, we reduce the computational burden by randomly choosing  $10^7$  kernel evaluations if  $\binom{n}{k}$  exceeds  $10^7$ . Such an approach maintains a high degree of numerical accuracy (up to 3 decimal places) and renders the computational burden negligible. For instance, for the Norwegian data ( $n = 142$ ), it requires only 150 seconds to compute (simultaneously) all estimators used in this study on a Pentium II 400MHz laptop computer. Further discussion on computational aspects of the GM estimators is available in Brazauskas and Serfling (2000a).  $\square$

## 4.1 Summary of Pareto Fits

### DISCUSSION OF TABLE 4.1

The fitted model is  $P(\sigma = 1.5, \hat{\alpha})$  with values of  $\hat{\alpha}$  ranging from 0.605 (for  $\hat{\alpha}_Q^{\text{opt},2}$ ) to 0.791 (for  $\hat{\alpha}_Q^{\text{opt},5}$ ). This range differs somewhat from the findings of Philbrick (1985), where the MLE value of 0.976 (for grouped data) is reported and compared to a typical parameter value of 1.0 for the property insurance line. The main reason for such a difference is the choice of the truncation point. In Philbrick (1985),  $\sigma = 2$  is used. However, in view of our discussion in Section 3.4 and Example 1 in Hogg and Klugman (1984), p. 64, we believe that the choice of 1.5 is more natural. Moreover, all three goodness-of-fit tests very strongly support the appropriateness of the  $P(\sigma = 1.5, \hat{\alpha}_{\text{ML}} = 0.764)$  model with the goodness-of-fit values .1071 (KS), .1106 (CvM), .7329 (AD), and corresponding *p-values*: .51 (KS), .27 (CvM), .24 (AD). While the corresponding *p-values* for the  $P(\sigma = 2.0, \hat{\alpha}_{\text{ML}} = 0.945)$  model (for the de-grouped data) are comparable for the CvM and AD statistics, the *p-value* for the KS statistic is substantially lower: .33 (KS), .30 (CvM), .23 (AD). Thus, based on this discussion, we choose the model  $P(\sigma = 1.5, \hat{\alpha})$ .

Table 4.1 suggests that, although the  $P(\sigma = 1.5, \hat{\alpha}_{\text{ML}})$  model is accepted by all three tests, additional improvements of the fit are possible if we use the unbiased version MLU, which in turn can be even further improved by robust estimators. For example, the estimators  $\hat{\alpha}_T$  (with  $\beta_1 = 0, \beta_2 = .05$ ),  $\hat{\alpha}_{\text{GM}}$  (with  $k = 4, k = 5$ , and  $k = 10$ ), and  $\hat{\alpha}_Q^*$ , all have *uniformly* smaller ranks than  $\hat{\alpha}_{\text{MLU}}$ .  $\square$

TABLE 4.1. Values of  $\hat{\alpha}$ , goodness-of-fit statistics, and ranks for the wind data.

<i>Estimator</i>	$\hat{\alpha}$	<i>KS</i>	<i>rank</i>	<i>CvM</i>	<i>rank</i>	<i>AD</i>	<i>rank</i>
MLU	.745	.0980	6	.0911	12	.6484	12
$Q^{\text{opt},2}$	.605	.1320	14	.0956	13	.7939	13
$Q^*, k = 5$	.731	.0911	$2\frac{1}{2}$	.0792	10	.5999	10
$Q^{\text{opt},5}$	.791	.1198	13	.1445	14	.8881	14
T, $\beta_1 = 0, \beta_2 = .05$	.707	.0932	4	.0642	7	.5457	5
T, $\beta_1 = 0, \beta_2 = .10$	.677	.1031	8	.0562	2	.5335	2
T, $\beta_1 = 0, \beta_2 = .15$	.664	.1077	11	.0568	4	.5487	6
T, $\beta_1 = 0, \beta_2 = .20$	.667	.1066	10	.0564	3	.5441	4
T, $\beta_1 = 0, \beta_2 = .25$	.673	.1045	9	.0561	1	.5368	3
GM, $k = 2$	.653	.1118	12	.0594	6	.5720	8
GM, $k = 3$	.692	.0981	7	.0587	5	.5316	1
GM, $k = 4$	.714	.0911	$2\frac{1}{2}$	.0679	8	.5576	7
GM, $k = 5$	.723	.0884	1	.0734	9	.5777	9
GM, $k = 10$	.744	.0975	5	.0901	11	.6445	11

DISCUSSION OF TABLE 4.2

The fitted model is  $P(\sigma = 25,000, \hat{\alpha})$  with values of  $\hat{\alpha}$  ranging from 1.082 (for  $\hat{\alpha}_{GM}$  with  $k = 3$ ) to 1.172 (for  $\hat{\alpha}_Q^{opt,2}$ ). Philbrick (1985) reports the MLE value of 1.108. This is slightly below the industry values of 1.245 (all classes liability; truncation point 25,000) and 1.159 (high severity liability; truncation point 30,000), which are available in Reichle and Yonkunas (1985), Appendix E. In our case, the range of  $\hat{\alpha}$  is in closer agreement with the industry values. A minor discrepancy between Philbrick and our MLE's is due to different data de-grouping approaches.

The goodness-of-fit tests show strong evidence that the  $P(\sigma = 25,000, \hat{\alpha}_{ML} = 1.153)$  model is appropriate with the goodness-of-fit values (*p-values*): .0755 (.35) for KS, .0843 (.42) for CvM, .7153 (.26) for AD. Nonetheless, the fit based on the MLE is again uniformly improved by MLU, which is further improved by all T and GM estimators and by  $\hat{\alpha}_Q^*$ .  $\square$

TABLE 4.2. Values of  $\hat{\alpha}$ , goodness-of-fit statistics, and ranks for the liability data.

<i>Estimator</i>	$\hat{\alpha}$	<i>KS</i>	<i>rank</i>	<i>CvM</i>	<i>rank</i>	<i>AD</i>	<i>rank</i>
MLU	1.140	.0735	12	.0794	11	.6795	12
$Q^{opt,2}$	1.172	.0784	14	.0944	14	.7843	14
$Q^*, k = 5$	1.111	.0690	6	.0748	2	.6343	5
$Q^{opt,5}$	1.161	.0767	13	.0881	13	.7420	13
T, $\beta_1 = 0, \beta_2 = .05$	1.098	.0670	4	.0757	4	.6302	1
T, $\beta_1 = 0, \beta_2 = .10$	1.093	.0662	2	.0766	8	.6314	3
T, $\beta_1 = 0, \beta_2 = .15$	1.110	.0689	5	.0748	2	.6336	4
T, $\beta_1 = 0, \beta_2 = .20$	1.125	.0712	8	.0759	5	.6500	8
T, $\beta_1 = 0, \beta_2 = .25$	1.127	.0715	9	.0762	6	.6532	9
GM, $k = 2$	1.133	.0724	$10\frac{1}{2}$	.0775	$9\frac{1}{2}$	.6641	$10\frac{1}{2}$
GM, $k = 3$	1.082	.0656	1	.0795	12	.6395	7
GM, $k = 4$	1.094	.0664	3	.0764	7	.6310	2
GM, $k = 5$	1.113	.0693	7	.0748	2	.6359	6
GM, $k = 10$	1.133	.0724	$10\frac{1}{2}$	.0775	$9\frac{1}{2}$	.6641	$10\frac{1}{2}$

DISCUSSION OF TABLE 4.3

The fitted model is  $P(\sigma = 500,000, \hat{\alpha})$  with values of  $\hat{\alpha}$  ranging from 1.203 (for  $\hat{\alpha}_Q^{\text{opt},5}$ ) to 1.242 (for  $\hat{\alpha}_{GM}$  with  $k = 2$ ). The narrowness of the range points to a very good fit between the data and Pareto model, which was initially suggested above by the QQ-plot. Further, the three tests show extremely strong evidence in support of the  $P(\sigma = 500,000, \hat{\alpha}_{ML} = 1.218)$  model with the goodness-of-fit values (*p-values*): .0500 (.70) for KS, .0343 (.89) for CvM, .3647 (.71) for AD. Therefore, it is not surprising that in this case the MLE fit is among the best, improved upon by only the  $\hat{\alpha}_{GM}$  (with  $k = 3$ ) and  $\hat{\alpha}_T$  (with  $\beta_1 = 0, \beta_2 = .05$ ) fits.  $\square$

TABLE 4.3. Values of  $\hat{\alpha}$ , goodness-of-fit statistics, and ranks for the Norwegian data.

<i>Estimator</i>	$\hat{\alpha}$	<i>KS</i>	<i>rank</i>	<i>CvM</i>	<i>rank</i>	<i>AD</i>	<i>rank</i>
MLU	1.209	.0517	13	.0353	$11\frac{1}{2}$	.3693	8
$Q^{\text{opt},2}$	1.234	.0470	$3\frac{1}{2}$	.0351	$9\frac{1}{2}$	.3717	$10\frac{1}{2}$
$Q^*, k = 5$	1.232	.0473	5	.0348	8	.3698	9
$Q^{\text{opt},5}$	1.203	.0529	14	.0367	13	.3759	13
T, $\beta_1 = 0, \beta_2 = .05$	1.221	.0494	8	.0341	1	.3645	$1\frac{1}{2}$
T, $\beta_1 = 0, \beta_2 = .10$	1.229	.0479	6	.0345	$5\frac{1}{2}$	.3674	7
T, $\beta_1 = 0, \beta_2 = .15$	1.234	.0470	$3\frac{1}{2}$	.0351	$9\frac{1}{2}$	.3717	$10\frac{1}{2}$
T, $\beta_1 = 0, \beta_2 = .20$	1.235	.0468	2	.0353	$11\frac{1}{2}$	.3728	12
T, $\beta_1 = 0, \beta_2 = .25$	1.226	.0485	7	.0343	$3\frac{1}{2}$	.3658	5
GM, $k = 2$	1.242	.0454	1	.0369	14	.3825	14
GM, $k = 3$	1.220	.0496	9	.0342	2	.3645	$1\frac{1}{2}$
GM, $k = 4$	1.217	.0502	10	.0343	$3\frac{1}{2}$	.3649	3
GM, $k = 5$	1.215	.0506	11	.0345	$5\frac{1}{2}$	.3655	4
GM, $k = 10$	1.214	.0508	12	.0346	7	.3659	6

## 4.2 Comparisons, Conclusions, and Recommendations

### DISCUSSION OF TABLE 4.4

“Robustness versus efficiency” comparisons show that GM-type estimators dominate the competition. In particular, for a fixed variance (or UBP), any Q or T-type estimator can be improved upon by a GM estimator with as good variance (UBP) and larger UBP (smaller variance). For example,  $\hat{\alpha}_Q^{\text{opt},5}$  with UBP = .019 and variance = 1.079 is improved upon by  $\hat{\alpha}_{\text{GM}}(k = 5)$ , with UBP = .129 and variance = 1.061 and  $\hat{\alpha}_{\text{GM}}(k = 10)$ , with UBP = .067 and variance = 1.019. Similarly,  $\hat{\alpha}_T(\beta_1 = 0, \beta_2 = .15)$  with UBP = .150 and variance = 1.277 is improved upon by  $\hat{\alpha}_{\text{GM}}(k = 3)$ , with UBP = .206 and variance = 1.141, and  $\hat{\alpha}_{\text{GM}}(k = 4)$ , with UBP = .159 and variance = 1.088.

For the goodness-of-fit comparisons, if an estimator has at least 2 out of 3 ranks of corresponding statistics smaller than another estimator, then its performance is considered better. For the wind data, for example,  $\hat{\alpha}_T(\beta_1 = 0, \beta_2 = .10)$  with ranks (8, 2, 2) is better than  $\hat{\alpha}_T(\beta_1 = 0, \beta_2 = .05)$  with ranks (4, 7, 5) but worse than  $\hat{\alpha}_{\text{GM}}(k = 3)$  with ranks (7, 5, 1). This approach suggests that, for all data sets,  $\hat{\alpha}_Q^*$  demonstrates the strongest performance among Q-type estimators but is outperformed by the best T-type ( $\hat{\alpha}_T$  with  $\beta_1 = 0, \beta_2 = .05$  and  $\beta_1 = 0, \beta_2 = .10$ ) and by the best GM-type ( $\hat{\alpha}_{\text{GM}}$  with  $k = 3$  and  $k = 4$ ) estimators.  $\square$

TABLE 4.4. Comparisons based on ranks, UBP, and variance of estimators.

<i>Estimator</i>	Wind Catastrophes			OLT Liability			Norwegian Claims			UBP Variance	
	KS, CvM, AD	KS, CvM, AD	KS, CvM, AD	KS, CvM, AD	KS, CvM, AD	KS, CvM, AD	KS, CvM, AD	KS, CvM, AD	( $\times \alpha^2/n$ )		
MLU	6	12	12	12	11	12	13	$11\frac{1}{2}$	8	0	1
$Q^{\text{opt},2}$	14	13	13	14	14	14	$3\frac{1}{2}$	$9\frac{1}{2}$	$10\frac{1}{2}$	.203	1.541
$Q^*, k = 5$	$2\frac{1}{2}$	10	10	6	2	5	5	8	9	.130	1.383
$Q^{\text{opt},5}$	13	14	14	13	13	13	14	13	13	.019	1.079
T, $\beta_1 = 0, \beta_2 = .05$	4	7	5	4	4	1	8	1	$1\frac{1}{2}$	.050	1.089
T, $\beta_1 = 0, \beta_2 = .10$	8	2	2	2	8	3	6	$5\frac{1}{2}$	7	.100	1.180
T, $\beta_1 = 0, \beta_2 = .15$	11	4	6	5	2	4	$3\frac{1}{2}$	$9\frac{1}{2}$	$10\frac{1}{2}$	.150	1.277
T, $\beta_1 = 0, \beta_2 = .20$	10	3	4	8	5	8	2	$11\frac{1}{2}$	12	.200	1.383
T, $\beta_1 = 0, \beta_2 = .25$	9	1	3	9	6	9	7	$3\frac{1}{2}$	5	.250	1.501
GM, $k = 2$	12	6	8	$10\frac{1}{2}$	$9\frac{1}{2}$	$10\frac{1}{2}$	1	14	14	.293	1.280
GM, $k = 3$	7	5	1	1	12	7	9	2	$1\frac{1}{2}$	.206	1.141
GM, $k = 4$	$2\frac{1}{2}$	8	7	3	7	2	10	$3\frac{1}{2}$	3	.159	1.088
GM, $k = 5$	1	9	9	7	2	6	11	$5\frac{1}{2}$	4	.129	1.061
GM, $k = 10$	5	11	11	$10\frac{1}{2}$	$9\frac{1}{2}$	$10\frac{1}{2}$	12	7	6	.067	1.019

## Conclusions

Based on the comparisons in Table 4.4, the following conclusions emerge:

- The GM-type estimators offer the best trade-offs between robustness and efficiency, which translates into an excellent performance in terms of goodness-of-fit. The best fits are provided by the  $\hat{\alpha}_{\text{GM}}$  ( $k = 3$  and  $k = 4$ ) estimators, which offer moderate to high protection against contamination (UBP) and low to moderate sacrifice in accuracy (variance).
- The T-type estimators are slightly less competitive in terms of “robustness versus efficiency” comparisons. However, their goodness-of-fit performance is as good as that of the GM-type estimators. The best fits are provided by the  $\hat{\alpha}_{\text{T}}$  ( $\beta_1 = 0, \beta_2 = .05$  and  $\beta_1 = 0, \beta_2 = .10$ ) estimators, which offer low to moderate protection against contamination (UBP) and low to moderate sacrifice in accuracy (variance).
- The Q-type estimators are outperformed with respect to both criteria, “robustness versus efficiency” and goodness-of-fit, by the T and GM-type estimators and, thus, are less competitive.
- The nonrobust but most efficient MLU neither can improve nor be improved by any other estimator with respect to the “robustness versus efficiency” criterion, because it has the best variance and the worst UBP. However, its performance with respect to goodness-of-fit is *consistently* among the worst, implying that for “robustness versus efficiency” comparisons the robustness should be given a higher priority.

## Practical Recommendations

When fitting Pareto models to loss data, the following steps are necessary:

1. If data are grouped or ties are present, de-group it using methods of Section 3.4; otherwise, go to the next step.
2. Use diagnostic tools – histogram and QQ-plots – to visually determine whether a Pareto model is appropriate.
3. Compute  $\hat{\alpha}_{\text{ML}}$  and apply the KS, CvM, and AD statistics to formally test if the Pareto model provides an adequate fit to the data. Note that tables with the critical values of these statistics are only available when  $\alpha$  is estimated by MLE. (Tables are presented in D’Agostino and Stephens (1986), pp. 135–136, and Durbin (1975), Table 3.)
4. Compute  $\hat{\alpha}$  using the MLU and Q, T, and GM-type estimators. If the range of  $\hat{\alpha}$ ’s is narrow (as in the case of the Norwegian fire data), then the fit is very good and even the MLE can be relied on. However, if the range is relatively wide (as in the case of the wind data), then ranking of robust estimators has to be used to refine the fit.
5. In situations when all three goodness-of-fit tests support the Pareto model, the T-type estimators with 5%–10% trimming and the GM-type estimators with  $k = 3$  and  $k = 4$  perform the best. Estimators with high UBP should be applied if one of the tests

rejects the Pareto model or if the QQ-plots are not satisfactory, or if the range of  $\hat{\alpha}$ 's is very wide.

## 5 Acknowledgments

The authors are very appreciative of constructive remarks provided by an anonymous referee. These have led to many improvements in the paper. Also, support of the first author by a grant from the Actuarial Education and Research Fund and 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-0103698, are gratefully acknowledged.

## References

- [1] Arnold, B. C. (1983). *Pareto Distributions*. International Cooperative Publishing House, Fairland, Maryland.
- [2] Beirlant, J., Teugels, J. L., and Vynckier, P. (1996). *Practical Analysis of Extreme Values*. Leuven University Press, Leuven, Belgium.
- [3] Brazauskas, V. and Serfling, R. (2000a). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal*, **4**(4), 12–27.
- [4] Brazauskas, V. and Serfling, R. (2000b). Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes*, **3**(3), 231–249.
- [5] Brazauskas, V. and Serfling, R. (2001). Small sample performance of robust estimators of tail parameters for Pareto and exponential models. *Journal of Statistical Computation and Simulation*, **70**(1), 1–19.
- [6] D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker, New York.
- [7] Derrig, R. A., Ostaszewski, K. M., and Rempala, G. A. (2000). Applications of resampling methods in actuarial practice. *Proceedings of the Casualty Actuarial Society*, **LXXXVII**, 322–364.
- [8] Durbin, J. (1975). Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. *Biometrika*, **62**, 5–22.
- [9] Hogg, R. V. and Klugman, S. A. (1984). *Loss Distributions*. Wiley, New York.

- [10] Kimber, A. C. (1983a). Trimming in gamma samples. *Applied Statistics*, **32**, 7–14.
- [11] Kimber, A. C. (1983b). Comparison of some robust estimators of scale in gamma samples with known shape. *Journal of Statistical Computation and Simulation*, **18**, 273–286.
- [12] Klugman, S.A., Panjer, H.H., and Willmot, G.E. (1998). *Loss Models: From Data to Decisions*. Wiley, New York.
- [13] Koutrouvelis, I. A. (1981). Large-sample quantile estimation in Pareto laws. *Communications in Statistics, Part A—Theory and Methods*, **10**, 189–201.
- [14] Patrik, G. (1980). Estimating casualty insurance loss amount distributions. *Proceedings of the Casualty Actuarial Society*, **LXVII**, 57–109.
- [15] Philbrick, S. W. (1985). A practical guide to the single parameter Pareto distribution. *Proceedings of the Casualty Actuarial Society*, **LXXII**, 44–84.
- [16] Philbrick, S. W. and Jurschak, J. (1981). Discussion of “Estimating casualty insurance loss amount distributions.” *Proceedings of the Casualty Actuarial Society*, **LXVIII**, 101–106.
- [17] Quandt, R. E. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika*, **10**, 55–82.
- [18] Reichle, K. A. and Yonkunas, J. P. (1985). Discussion of “A practical guide to the single parameter Pareto distribution.” *Proceedings of the Casualty Actuarial Society*, **LXXII**, 85–123.
- [19] Rytgaard, M. (1990). Estimation in the Pareto distribution. *ASTIN Bulletin*, **20**(2), 201–216.
- [20] Serfling, R. (1984) Generalized  $L$ -,  $M$ - and  $R$ -statistics. *Annals of Statistics*, **12**, 76–86.
- [21] Serfling, R. (2000). “Robust and nonparametric estimation via generalized  $L$ -statistics: theory, applications, and perspectives,” In: *Advances in Methodological and Applied Aspects of Probability and Statistics*, Balakrishnan, N. (Ed.), pp. 197–217. Gordon & Breach.

## Appendix

TABLE A.1. OLT Bodily Injury Liability Claims (1976) Data ( $\times 1000$  dollars).

<i>Loss Amount</i>	<i>Number of Losses</i>	<i>Loss Amount</i>	<i>Number of Losses</i>	<i>Loss Amount</i>	<i>Number of Losses</i>
25-30	11	70- 75	9	220-230	1
30-35	18	75- 80	1	240-250	2
35-40	9	95-100	4	260-270	1
40-45	4	120-130	2	280-290	1
45-50	11	140-150	3	290-300	2
50-55	3	190-200	1	340-350	1
55-60	2	200-210	2	410-420	2

SOURCE: Patrik (1980), Appendix F, Part 1.

TABLE A.2. Norwegian Fire Claims (1975) Data ( $\times 1000$  Norwegian kroner).

500	552	600	650	798	948	1180	1479	2497	7371
500	557	605	672	800	957	1243	1485	2690	7772
500	558	610	674	800	1000	1248	1491	2760	7834
502	570	610	680	800	1002	1252	1515	2794	13000
515	572	613	700	826	1009	1280	1519	2886	13484
515	574	615	725	835	1013	1285	1587	2924	17237
528	579	620	728	862	1020	1291	1700	2953	52600
530	583	622	736	885	1024	1293	1708	3289	
530	584	632	737	900	1033	1298	1820	3860	
530	586	635	740	900	1038	1300	1822	4016	
540	593	635	748	910	1041	1305	1848	4300	
544	596	640	752	912	1104	1327	1906	4397	
550	596	650	756	927	1108	1387	2110	4585	
550	600	650	756	940	1137	1455	2251	4810	
551	600	650	777	940	1143	1475	2362	6855	

SOURCE: Beirlant, Teugels, and Vynckier (1996), Appendix I.