NEW GOODNESS-OF-FIT TESTS FOR PARETO DISTRIBUTIONS*

BY

MARIA L. RIZZO

ABSTRACT

A new approach to goodness-of-fit for Pareto distributions is introduced. Based on Euclidean distances between sample elements, the family of statistics and tests is indexed by an exponent in $(0, 2)$ on Euclidean distance. The corresponding tests are statistically consistent and have excellent performance when applied to heavy-tailed distributions. The exponent can be tailored to the particular Pareto distribution. The goodness-of-fit statistic measures all types of differences between distributions, hence it is also applicable as a minimum distance estimator. Implementation of the test statistics is developed and applied to estimation of the tail index in three well known examples of claims data, and compared with the classical EDF statistics.

KEYWORDS

Pareto, goodness-of-fit, heavy-tail, Gini, claims.

1. INTRODUCTION

The Pareto family of distributions is often applied in economics, finance, and actuarial science to measure size; for example, income, loss, or claim severity. Thus, estimation and fitting from data, and goodness-of-fit procedures that address the issue of model adequacy, are of particular interest. Pareto distributions and their properties are described in section 2.1.

In this paper we introduce and implement goodness-of-fit statistics and tests for Pareto distributions based on Euclidean distances between sample elements. These statistics have excellent empirical performance, particularly for distributions with heavy tails. Actually, we introduce a family of tests indexed by an exponent β in $(0, 2)$. The proposed test applies to univariate or multivariate data. This paper focuses on the univariate case, but there is a natural extension of the theory to multivariate loss models.

* This research was supported by the Casualty Actuarial Society through The Actuarial Foundation 2008 Individual Grants Program.

Astin Bulletin 39(2), 691-715. doi: 10.2143/AST.39.2.2044654

1.1. Existing literature

Commonly applied formal goodness-of-fit (GOF) tests for Pareto distributions are generally in the class of tests based on the empirical distribution function (EDF), such as the Kolmogorov-Smirnov (KS) test, Cramér-von Mises (CvM) test, or the Anderson-Darling (AD) test. The EDF statistics measure the distance between distributions by some function of the distance between the empirical and the hypothesized distributions. There is much literature on estimation of Pareto parameters, including Baxter (1980), Likeš (1969), Rytgaard (1990), and relevant chapters of Arnold (1983) or Kleiber and Kotz (2003). Statistics for measuring departure from a Pareto distribution are discussed in Brazauskas and Serfling (2003) and Porter et al. (1992), and procedures for fitting Pareto distributions or estimation of the tail parameter are covered in Brazauskas and Serfling (2000a,b). Recent empirical studies include Brazauskas and Serfling (2001, 2003).

Brazauskas and Serfling (2003) used the KS, CvM, and AD statistics as distance measures to rank 13 robust estimators of the tail index parameter α, as well as the unbiased maximum likelihood estimator (MLU), with the goal of obtaining a kind of consensus vote for the best estimators. That study in effect optimized each of the GOF statistics over a finite set of possible estimates. We extend and modify the study in two ways; by optimizing the goodness-of-fit statistics over the parameter space to obtain estimates, and by considering the new statistics proposed in this work.

Although the purpose of the comparison in Brazauskas and Serfling (2003) was primarily to evaluate robust estimators, one can also investigate whether the GOF statistics used to rank the estimators should be given equal weights. Indeed if one statistic is generally superior (or inferior) for the problem at hand, it is not clear how to resolve the differences in rankings.

The notion of "better" of course needs some criteria. Given that the sampled distribution is Pareto, and the location parameter is correctly specified, then in one sense an optimal estimator is the unbiased MLE. Perhaps the goal is to find the best fit for data that is only approximately Pareto. In this case the goodness-of-fit statistics measure goodness-of-fit (of the incorrect model to the data), not the goodness of the estimate. For small samples of data with large variance, robust statistics may perform better than asymptotically optimal estimates. To investigate we follow up with a cross-validation study, to determine which of the GOF statistics perform well for fitting a Pareto type I distribution. A better fit in this case corresponds to the fit with smaller error. Performance can also be compared in terms of power of the goodness-of-fit test.

1.2. Organization

The results below are organized as follows. Theoretical background and properties of the proposed statistics are presented in Section 2. Implementation including derivation of computing formulae for several Pareto goodness-of-fit statistics follows in Section 3. In Section 4, Empirical Results and Discussion, results are applied to three examples of observed claims data, and compared with existing tests. Performance of statistics as minimum distance estimators is investigated via cross-validation, and power of goodness-of-fit tests is investigated by a Monte Carlo power comparison. Our findings are summarized in Section 5.

2. Goodness-of-fit

Let $X_1, \ldots, X_n$ be a complete random sample from the distribution of X, and let F be a cumulative distribution function (CDF). Consider the goodness-of-fit problem of testing $H_0 : X \sim F$ vs $H_1 : H_0$ is false. In this paper we are interested in testing the goodness-of-fit of data to a hypothesized Pareto distribution.

2.1. Pareto Distributions

A Pareto distribution of the first type has survival function

$$\bar{F}(x) = \left(\frac{x}{\sigma}\right)^{-\alpha}, \quad x \geq \sigma > 0, \tag{2.1}$$

and density function $f(x) = \frac{\alpha\sigma^{\alpha}}{x^{\alpha+1}}$, $x \geq \sigma > 0$. Here $\sigma > 0$ is a scale parameter, and $\alpha > 0$ is a shape parameter (Pareto's index of inequality), which measures the heaviness in the upper tail. The notation $X \sim P(I)(\sigma,\alpha)$ or simply $P(\sigma,\alpha)$ indicates that X has the classical Pareto (type I) distribution given by (2.1). Pareto's second model, referred to as the Pareto type II distribution, has the survival distribution

$$\bar{F}(x) = \left[1 + \frac{x-\mu}{\sigma}\right]^{-\alpha}, \quad x \geq \mu, \tag{2.2}$$

where $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter and $\alpha > 0$ is a shape parameter. The Pareto type II model defined by (2.2) is denoted $P(II)(\mu,\sigma,\alpha)$. Pareto type I and type II models are related by a simple transformation. If $Y \sim P(II)(\mu,\sigma,\alpha)$ then $Y-(\mu-\sigma) \sim P(I)(\sigma,\alpha)$.

Pareto densities have a polynomial upper tail with index $-(\alpha+1)$. Small values of α correspond to heavier tails, and the k^{th} moments exist only if $\alpha > k$. The moments of $X \sim P(I)(\sigma,\alpha)$ are given by

$$E[X^k] = \frac{\alpha\sigma^k}{(\alpha-k)}, \quad \alpha > k; \tag{2.3}$$

in particular, $E[X] = \frac{\alpha\sigma}{\alpha-1}$, $\alpha > 1$ and $Var(X) = \frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}$, $\alpha > 2$. For P(II) (μ,σ,α) distributions, $E[X] = \frac{\alpha\sigma}{\alpha-1} + (\mu-\sigma)$, $\alpha > 1$.

Thus, theoretical results that depend on the existence of moments do not necessarily extend to Pareto distributions with arbitrary shape parameter α. However, our proposed statistics are formulated with a stability index β in $(0, \alpha)$ that can be chosen so that the corresponding moments exist.

2.2. Goodness-of-fit statistics

In the following, $\|\cdot\|$ denotes Euclidean norm, or absolute value in one dimension. The notation X' indicates that X' is an independent copy of X; that is, X and X' are independent and identically distributed (iid).

Theorem 1. *If* $X \in \mathbb{R}^d$ *and* $Y \in \mathbb{R}^d$ *are independent random vectors, and* $E(\|X\|^\beta + \|Y\|^\beta) < \infty$, *then for all* $0 < \beta < 2$

$$2E\|X-Y\|^\beta - E\|X-X'\|^\beta - E\|Y-Y'\|^\beta \geq 0, \tag{2.4}$$

with equality if and only if X and Y are identically distributed.

The expectation $E\|X-Y\|^\beta$ is taken with respect to the joint distribution, which by independence is $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, so that

$$E\|X-Y\|^\beta = \int\int \|x-y\|^\beta \, dF_X(x)\, dF_Y(y).$$

Theorem 1 is proved in Székely and Rizzo (2005b). For each sample observation X_j, let $E\|X_j - X\|^\beta = \int \|X_j - X\|^\beta dF_X$; that is, X_j is a constant in the integrand. Then an empirical version of the left side of inequality (2.4) is the statistic

$$Q_\beta = n\left\{\frac{2}{n}\sum_{j=1}^{n} E\left\|X_j - X\right\|^\beta - E\|X - X'\|^\beta - \frac{1}{n^2}\sum_{j,k=1}^{n}\left\|X_j - X_k\right\|^\beta\right\}, \tag{2.5}$$

which can be applied to goodness-of-fit problems and certain estimation problems. In this paper we restrict attention to univariate Pareto models, and goodness-of-fit to (or departure from) Pareto distributions is measured by the univariate statistic, where $E|X_j - X|^\beta$ and $E|X-X'|^\beta$ are computed under the hypothesized Pareto model, and exponent β is chosen to satisfy the moment condition ($\beta < \alpha/2$). The exponent β is a stability index in the sense that when $\beta < \alpha/2$ the distribution of X^β has finite variance. Expressions for $E|X_j - X|^\beta$ and $E|X-X'|^\beta$ are derived in section 3.

Alternately, for goodness-of-fit tests of $\mathrm{P}(\sigma,\alpha)$ models, it is equivalent to test the hypothesis that $T = \log(X)$ has a two-parameter exponential distribution,

$$H_0 : T \sim \mathrm{Exp}(\mu,\alpha),$$

where $\mu = \log(\sigma)$ is the location parameter and α is the rate parameter. Here $\log(X)$ always refers to the natural logarithm. The density of T is

$$f_T(t) = \alpha e^{-\alpha(t-\mu)}, \quad t \geq \mu.$$

The first and second moments of T are finite. For all $\alpha > 0$ we have $E[T] = \frac{1}{\alpha} + \mu$, and $Var(T) = \alpha^{-2}$. Hence we can alternately apply the test statistic

$$V_\beta = n\left\{\frac{2}{n}\sum_{j=1}^{n} E\left|T_j - T\right|^\beta - E|T - T'|^\beta - \frac{1}{n^2}\sum_{j,k=1}^{n}\left|T_j - T_k\right|^\beta\right\}, \tag{2.6}$$

where $T_j = \log(X_j)$, $j = 1, \ldots, n$, and $E|T_j - T|^\beta$ and $E|T - T'|^\beta$ are computed under the hypothesized exponential (log Pareto) model.

For Pareto type II samples, $X_j \sim P(II)(\mu,\sigma,\alpha)$, let $Y_j = X_j - (\mu - \sigma)$. Then $Y_j \sim P(I)(\sigma,\alpha)$. Moreover, Q and V are invariant to this transformation, as $|X_j - X_k| = |Y_j - Y_k|$, etc. Thus the statistics developed for Pareto type I distributions can be applied to the corresponding transformed Pareto type II distributions.

2.3. Properties

It can be shown that $Q_\beta \geq 0$, with large values of Q_β indicating departure from the hypothesized distribution. Similarly, $V_\beta \geq 0$ and large values of V_β support the alternative hypothesis.

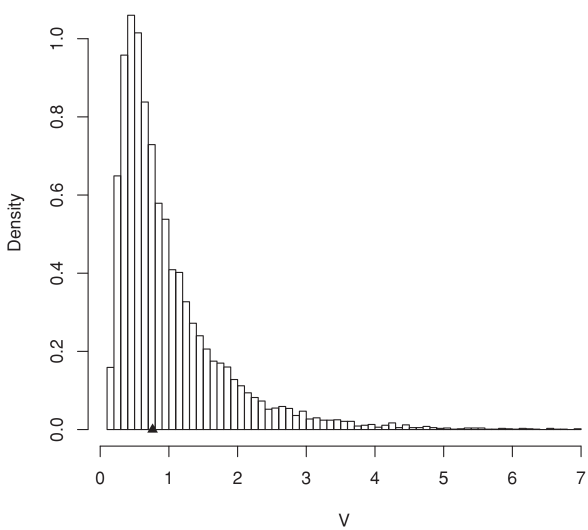

FIGURE 1: Empirical distribution of 10,000 replicates of V for the wind catastrophes data, assuming a Pareto($\sigma = 1.5$, $\alpha = 0.745$) model; the observed test statistic is marked with a triangle.

The expected values are $E[Q_\beta] = E\|X - X'\|^\beta$ and $E[V_1] = E\|\log(X) - \log(X')\|$. When the Pareto hypothesis is true and $\mathrm{Var}(X)$ is finite, Q_β converges in distribution to a quadratic form

$$\sum_{j=1}^{\infty} \lambda_j Z_j^2, \tag{2.7}$$

as sample size n tends to infinity, where λ_j are non-negative constants, and Z_j are iid standard normal random variables. Asymptotic theory of V-statistics can be applied to prove that tests based on Q_β (or V_β) are statistically consistent goodness-of-fit tests. See Székely and Rizzo (2005a) for details.

The shape of the distribution is similar to a gamma random variable. For an example, see Figure 1, a histogram of the empirical distribution of the statistic V for one of the examples considered below. The rejection region for a goodness-of-fit test based on a statistic V or Q is in the upper tail.

Another family of statistics with asymptotic distribution of the form (2.7) are the Cramér-von Mises statistics (von Mises, 1947), including the CvM and AD goodness-of-fit statistics.

3. Implementation

3.1. Statistics for the Exponential Model

Assume that $X \sim P(\sigma, \alpha)$, and $T = \log(X)$. Then $T \sim Exp(\mu, \alpha)$, where $\mu = \log \sigma$, α is the rate parameter, and $F_T(t) = 1 - e^{-\alpha(t-\mu)}$, $t \geq \mu$. Then the integrals in V_1 are

$$E|s - T| = s - \mu + \frac{1}{\alpha}(1 - 2F_T(s)), \quad s \geq \mu; \tag{3.1}$$

$$E|T - T'| = \frac{1}{\alpha}. \tag{3.2}$$

A computing formula for the corresponding test statistic is derived as follows. The first mean in the statistic $V = V_1$ is

$$\frac{1}{n}\sum_{j=1}^{n}\left(T_j - \mu + \frac{1}{\alpha}\left(1 - 2F_T\left(T_j\right)\right)\right) = \bar{T} - \left(\mu + \frac{1}{\alpha}\right) + \frac{2e^{\alpha\mu}}{\alpha n}\sum_{j=1}^{n} e^{-\alpha T_j},$$

where $\bar{T} = \frac{1}{n}\sum_{j=1}^{n} T_j$ is the sample mean.

Also, for $\beta = 1$, the last sum in (2.5) or (2.6) can be expressed as a linear function of the ordered sample. If $T_{(j)}$ denotes the j^{th} largest sample element, then

$$\sum_{j,k=1}^{n} \left|T_j - T_k\right| = 2\sum_{j=1}^{n}((2j - 1) - n)T_{(j)}.$$

Hence the computational complexity of the statistic Q_1 or V_1 is $O(n \log n)$. The statistic $V = V_1$ is given by

$$V = n\left\{2\left[\bar{T} - \mu - \frac{1}{\alpha} + \frac{2e^{\alpha\mu}}{\alpha n}\sum_{j=1}^{n} e^{-\alpha T_j}\right] - \frac{1}{\alpha} - \frac{2}{n^2}\sum_{j=1}^{n}(2j-1-n)T_{(j)}\right\}. \quad (3.3)$$

If parameters are estimated, the corresponding estimates are substituted in (3.3). (Formula (3.3) can be simplified further for computation.)

3.2. Pareto Statistics

In this section we develop the computing formula for Q_β. First we present two special cases, $\beta = 1$ and $\beta = \alpha - 1$.

Case 1. *If $X \sim P(\sigma, \alpha)$, $\alpha > 1$ and $\beta = 1$, then*

$$E|y - X| = y - E[X] + \frac{2\sigma^\alpha}{(\alpha - 1)\, y^{\alpha-1}}$$

$$= y + \frac{2\sigma^\alpha y^{1-\alpha} - \alpha\sigma}{\alpha - 1}, \quad y \geq \sigma; \quad (3.4)$$

$$E|X - X'| = \frac{2\alpha\sigma}{(\alpha-1)(2\alpha-1)} = \frac{E[X]}{\alpha - 1/2}. \quad (3.5)$$

Case 2. *If $X \sim P(\sigma, \alpha)$, $\alpha > 1$ and $\beta = \alpha - 1$, then*

$$E|y - X|^{\alpha-1} = \frac{(y-\sigma)^\alpha + \sigma^\alpha}{y}, \quad y \geq \sigma; \quad (3.6)$$

$$E|X - X'|^{\alpha-1} = \frac{2\alpha\sigma^{\alpha-1}}{\alpha+1}. \quad (3.7)$$

The statements of cases 1 and 2 can be obtained by directly evaluating the integrals.

Although the special cases above are easy to apply, in general it may be preferable to apply β that is proportional to α. For this we need case 3 below.

The Pareto type I family is closed under the power transformation. That is, if $X \sim P(\sigma, \alpha)$ and $Y = X^r$, then $Y \sim P(\sigma^r, \alpha/r)$. It is always possible to find an $r > 0$ such that the second moments of $Y = X^r$ exist, and Q_1 can be applied to measure the goodness-of-fit of Y to $P(\sigma^r, \alpha/r)$. This goodness-of-fit measure will be denoted $Q^{(r)}$.

Beta functions arise in some of the expressions below. For reference, $B_x(p, q) = \int_0^x t^{p-1}(1-t)^{q-1}\, dt$ is the incomplete beta function, and $B(p, q) = B_1(p, q)$ is the complete beta function, $B(p,q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}$, $p > 0$, $q > 0$, and $\Gamma(\cdot)$ is the complete gamma function.

Proofs of the following statements are given in Appendix A.

Case 3. *If* $X \sim P(\sigma,\alpha)$ *and* $0 < \beta < \alpha < 1$, *then*

$$E|y-X|^{\beta} = (y-\sigma)^{\beta} - \frac{\sigma^{\alpha}\left[\beta B_{y_0}(\beta, 1-\alpha) - \alpha B(\alpha-\beta, \beta+1)\right]}{y^{\alpha-\beta}}, \quad y \geq \sigma; \tag{3.8}$$

$$E|X-X'|^{\beta} = \frac{2\alpha^2\sigma^{\beta}B(\alpha-\beta, \beta+1)}{2\alpha-\beta}, \tag{3.9}$$

where $y_0 = \frac{y-\sigma}{y}$.

Case 4. *If* $X \sim P(\sigma,\alpha=1)$, $0 < \beta < 1$, *and* $y_0 = (y-\sigma)/y$, *then*

$$E|y-X|^{\beta} = (y-\sigma)^{\beta} - \sigma\beta y^{\beta-1}\left\{\frac{y_0^{\beta}}{\beta} + \frac{y_0^{\beta+1}}{\beta+1}\,{}_2F_1(1,\beta+1;\,\beta+2;\,y_0)\right\}$$
$$+\,\sigma y^{\beta-1}B(\beta+1, 1-\beta), \quad y \geq \sigma; \tag{3.10}$$

$$E|X-X'|^{\beta} = \frac{2\sigma^{\beta}}{2-\beta}B(1-\beta, \beta+1), \tag{3.11}$$

where ${}_2F_1(a,b;c;z)$ *denotes the Gauss hypergeometric function,*

$${}_2F_1(a,b;c;z) = \sum_{k=0}^{\infty}\frac{(a)_k\,(b)_k}{(c)_k}\,\frac{z^k}{k!},$$

and $(r)_k = r(r+1)\cdots(r+k-1)$ *denotes the ascending factorial.*

For $\alpha > 1$ the expressions for $E|y-X|^{\beta}$ are complicated and involve the Gauss hypergeometric function. It is simpler and more computationally efficient to apply the $Q^{(r)}$ statistics (or V) in this case.

3.3. Estimations

The proposed Pareto goodness-of-fit statistics provide a new approach to estimation of the tail index α of a $P(\sigma,\alpha)$ distribution. A minimum distance approach can be applied, where the objective is to minimize the corresponding goodness-of-fit statistic under the assumed model. For this application the statistic V_1 can be normalized to mean 1 by dividing by the mean $E|T-T'|$, and Q_{β} can be normalized by dividing by $E|X-X'|^\beta$, where $E|T-T'|$ or $E|X-X'|^\beta$ are computed under the hypothesized model.

For a formal goodness-of-fit test based on V or Q, the unknown parameters of the hypothesized $P(\sigma,\alpha)$ distribution can be estimated by a number of methods. See Arnold (1983, Ch. 5) or Kleiber and Kotz (2003, Ch. 3). Here we summarize the maximum likelihood estimators. Robust generalized median, quantile, and trimmed mean type estimators of α are described in Brazauskas and Serfling (2000a,b, 2003) and the references therein.

The joint maximum likelihood estimator (MLE) of (α,σ) is $(\hat\alpha,\hat\sigma)$, where

$$\hat\alpha = n\left[\sum \log\frac{X_j}{\hat\sigma}\right]^{-1}, \quad \hat\sigma = X_{(1)},$$

and $X_{(1)}$ is the first order statistic. By the invariance property of the MLE, substituting $\hat\alpha$ for α and $\hat\sigma$ for σ, or $\hat\mu = \log\hat\sigma$, we obtain the corresponding MLEs of the mean distances in the test statistics V or Q.

Alternately, unbiased estimators of the parameters can be derived from the MLEs. If both parameters are unknown,

$$\sigma^* = X_{1:n}\left(1-\frac{1}{(n-1)\,\hat\alpha}\right), \quad \alpha^* = \left(1-\frac{2}{n}\right)\hat\alpha$$

are unbiased estimators of the parameters (Baxter, 1980; Likeš, 1969). If one parameter is known, then

$$\alpha^* = \left(1-\frac{1}{n}\right)\hat\alpha \quad \text{or} \quad \sigma^* = X_{1:n}\left(1-\frac{1}{n\alpha}\right)$$

is unbiased for the unknown parameter.

3.4. The EDF tests

Among the formal goodness-of-fit tests applicable for this problem, the EDF tests described in Stephens (1986) are widely applied. Let $\hat F$ denote an estimate of F. The Kolmogorov-Smirnov (KS) statistic

$$D = \sup_x \left| F(X) - \hat F(X) \right|$$

measures the distance of $\hat F$ from the CDF F. The quadratic EDF statistics are based on the integrated squared distance

$$W_\psi^2 = \int \left(\hat F(x) - F(x)\right)^2 \psi(x)\, dF(x),$$

where $\psi(x)$ is a suitable weight function. If ψ is the identity function we obtain the Cramér-von Mises (CvM) statistic

$$W^2 = \int \left(\hat{F}(x) - F(x)\right)^2 dF(x),$$

while the Anderson-Darling (AD) statistic

$$A^2 = \int \frac{\left(\hat{F}(x) - F(x)\right)^2}{F(x)\,(1 - F(x))}\, dF(x),$$

is obtained by applying the weight function $\psi(x) = (F(x)\,\bar{F}(x))^{-1}$.

Using the EDF as the estimate of the CDF F, and $\hat{F}$ to denote the hypothesized model, the EDF test statistics are

$D = \max(D^+, D^-)$, where

$$D^+ = \max_{1 \le j \le n} \left\{ \frac{j}{n} - \hat{F}\left(X_{(j)}\right) \right\};\ D^- = \max_{1 \le j \le n} \left\{ \hat{F}\left(X_{(j)}\right) - \frac{j-1}{n} \right\};$$

$$W^2 = \sum_{j=1}^{n} \left[\hat{F}\left(X_{(j)}\right) - \frac{2j-1}{2n} \right]^2 + \frac{1}{12n};$$

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

For a test of fit for Pareto type I when parameters are specified or estimated by maximum likelihood, one can refer to the critical values of EDF tests given for testing the exponential model (Stephens, 1986, pp. 135-141).

4. Empirical Results and Discussion

Three data sets are described and analyzed below: *Wind Catastrophes (1977)*, *OLT Bodily Injury Liability Claims (1976)*, and *Norwegian Fire Claims (1975)*. These three examples were chosen for comparison with results by Brazauskas and Serfling (2003) and to extend a study that compared and ranked 14 estimators of the tail index of Pareto type I models. The data sets described below are given in the appendix of this paper for easy reference.

The statistics applied in this paper were implemented in the R statistical computing software, which is available by general public license.

4.1. Wind catastrophes data

The wind-catastrophes data shown in Table 6 is from an example in Hogg and Klugman (1984, p. 64). Losses due to wind-related catastrophes were recorded to the nearest million dollars; the data comprise 40 rounded loss amounts of \$2 million or more.

Although the wind catastrophe losses are assumed to arise from a continuous model, the data have been discretized by grouping. To be consistent with Brazauskas and Serfling (2003) and for reproducible results, we de-group by equally spacing the data, according to the method outlined in Brazauskas and Serfling (2003), letting

$$X_k = (1 - k/(m+1))\,A + k/(m+1)\,B, \quad k = 1, \ldots, m, \tag{4.1}$$

where interval (A, B) contains exactly m grouped sample observations. After de-grouping, the scale parameter is $\sigma = 1.5$ rather than $\sigma = 2$. The maximum likelihood estimate of α is 0.764 and the unbiased estimator is $\alpha^* = 0.745$.

Remark 1. For the wind catastrophes data, the scale parameter $\sigma = 1.5$ was applied by Brazauskas and Serfling (2003) and Hogg and Klugman (1984); Philbrick (1981) applied $\sigma = 2.0$. For comparison with results of Brazauskas and Serfling (2003), we apply scale parameter $\sigma = 1.5$.

Table 1 illustrates the statistics V and Q_β applied as a goodness-of-fit measure to compare and rank several estimators of α. The estimators include α^* (MLU), three quantile estimators (Q1-Q3), five trimmed mean estimators (TM1-TM5), and five generalized median estimators (GM1-GM5).

The ranks in Table 1 can be compared with Table 4.1 in Brazauskas and Serfling (2003), which includes only the EDF statistics. In this example, ranks

TABLE 1

SUMMARY OF GOODNESS-OF-FIT ANALYSIS OF ESTIMATES OF α FOR THE WIND CATASTROPHE DATA.

	$\hat{\alpha}$	CvM	AD	KS	stat	V	stat	$Q_{\alpha/4}$	stat	$Q_{\alpha/3}$
MLU	0.745	12	12	6	0.763	7	0.947	12	0.959	12
Q1	0.605	13	13	14	1.244	14	0.954	13	0.975	13
Q2	0.731	10	10	2.5	0.730	5	0.931	10	0.939	10
Q3	0.791	14	14	13	0.980	13	1.030	14	1.060	14
TM1	0.707	7	5	4	0.713	2	0.911	7	0.916	6
TM2	0.677	2	2	8	0.765	8	0.903	1	0.908	1
TM3	0.664	4	6	11	0.813	11	0.905	4	0.911	5
TM4	0.667	3	4	10	0.800	10	0.904	3	0.910	4
TM5	0.673	1	3	9	0.778	9	0.903	2	0.908	2
GM1	0.653	6	8	12	0.867	12	0.909	6	0.917	7
GM2	0.692	5	1	7	0.729	4	0.905	5	0.909	3
GM3	0.714	8	7	2.5	0.713	1	0.916	8	0.922	8
GM4	0.723	9	9	1	0.719	3	0.923	9	0.930	9
GM5	0.744	11	11	5	0.760	6	0.946	11	0.958	11

based on V are similar to those obtained by the KS test. Two versions of the Q_β statistic are considered; $\beta = \hat{\alpha}/3$ and $\beta = \hat{\alpha}/4$. Corresponding ranks of the estimates are very much in consensus with the CvM and AD statistics, perhaps most closely aligned with CvM.

In Figure 2 each of the six statistics is plotted against the parameter estimates α. For comparison purposes, the statistics have been scaled to a common range by dividing each by its respective maximum over the interval. The graphs reveal that although each statistic achieves its minimum at approximately the same estimate, the shapes of the curves differ. The minima of the KS, CvM, AD, V, $Q_{\alpha/3}$ and $Q_{\alpha/4}$ statistics are achieved at approximately 0.724, 0.673, 0.686, 0.711, 0.680, and 0.678, respectively.

Figure 1 is a density histogram of the replicates V for the $P(\sigma = 1.5, \alpha = 0.745)$ model, with the value of the observed test statistic $V = 0.762$ marked by a triangle. The histogram is a large sample approximation to the asymptotic distribution of V under the null hypothesis. The median of the empirical distribution is 0.743 and the observed statistic is at the 51.3 percentile, clearly non-significant.

4.2. OLT Bodily Injury Liability Claims (1976)

This data from Patrik (1980, p. 99) is shown in Table 7, the grouped losses (in thousands) for the $500,000 policy limit for 1976 Owners, Landlords and Tenants (OLT) bodily liability claims. A Pareto model is fit to losses at least $25,000.

TABLE 2.

SUMMARY OF GOODNESS-OF-FIT ANALYSIS OF ESTIMATES OF α FOR THE OLT BODILY INJURY LIABILITY CLAIMS DATA (RANKS).

	$\hat{\alpha}$	CvM	AD	KS	V_1	$Q_{\alpha-1}$	$Q^{(3)}$
MLU	1.140	11	12	12	9	12	8
Q1	1.172	14	14	14	14	14	14
Q2	1.111	2	5	6	2	3	6
Q3	1.161	13	13	13	13	13	12
TM1	1.098	4	1	4	8	1	9
TM2	1.093	8	3	2	11	6	11
TM3	1.110	2	4	5	4	2	7
TM4	1.125	5	8	8	3	8	1
TM5	1.127	6	9	9	5	9	2
GM1	1.133	9.5	10.5	10.5	6.5	10.5	3.5
GM2	1.082	12	7	1	12	7	13
GM3	1.094	7	2	3	10	5	10
GM4	1.113	2	6	7	1	4	5
GM5	1.133	9.5	10.5	10.5	6.5	10.5	3.5

The data is de-grouped for analysis using (4.1) described above. The hypothesized model is $P(\sigma = 25, \alpha)$, where $\hat{\alpha} = 1.153$ is the MLE and $\alpha^* = 1.140$. The ranks of the estimates are shown in Table 2, corresponding to Table 4.2 in Brazauskas and Serfling (2003), and estimates are also compared in Figure 3.

For the OLT liability data, it appears that V and $Q^{(3)}$ rank the estimates similarly. Other $Q^{(r)}$ statistics (not shown) produce essentially the same ranks as $Q^{(3)}$.

It is easier to interpret the rankings from the plots in Figure 3, where each statistic is plotted against the estimates $\hat{\alpha}$. For ease of interpretation, in the plots each statistic is scaled by dividing it by its maximum value over the interval shown. Here we see that each statistic achieves its minimum at a value within the range of estimates in Table 2. The minimum values of the KS, CvM, AD, V, $Q_{\alpha-1}$ and $Q^{(3)}$ statistics occur at 1.084, 1.111, 1.099, 1.118, 1.104, and 1.123, respectively.

4.3. Norwegian Fire Claims (1975)

This data is from Beirlant et al. (1996, Appendix I). The part of the data analyzed here comprise the total damage by 142 fires in Norway for the year 1975, for claims above 500,000 Norwegian krones. The losses shown in Table 8 are recorded in 1000's of Norwegian krones.

Again, the data is de-grouped for analysis using (4.1) described above. The hypothesized model is $P(\sigma = 500, \alpha)$. The MLE is $\hat{\alpha} = 1.218$ and $\alpha^* = 1.209$ is MLU.

TABLE 3.

SUMMARY OF GOODNESS-OF-FIT ANALYSIS OF ESTIMATES OF α FOR THE NORWEGIAN FIRE CLAIMS DATA (RANKS).

	$\hat{\alpha}$	CvM	AD	KS	V_1	$Q_{\alpha-1}$	$Q^{(3)}$
MLU	1.209	11.5	8	13	6	13	3
Q1	1.234	9.5	10.5	3.5	11.5	3.5	11.5
Q2	1.232	8	9	5	10	5	10
Q3	1.203	13	13	14	8	14	7
TM1	1.221	1	1.5	8	5	8	6
TM2	1.229	5.5	7	6	9	6	9
TM3	1.234	9.5	10.5	3.5	11.5	3.5	11.5
TM4	1.235	11.5	12	2	13	2	13
TM5	1.226	3.5	5	7	7	7	8
GM1	1.242	14	14	1	14	1	14
GM2	1.220	2	1.5	9	4	9	5
GM3	1.217	3.5	3	10	3	10	4
GM4	1.215	5.5	4	11	1	11	2
GM5	1.214	7	6	12	2	12	1

Table 3 extends the analysis as summarized in Table 4.3 of Brazauskas and Serfling (2003), showing the ranks of the estimates according to each of the goodness-of-fit measures. Figure 4 summarizes the same analysis graphically, where the minimum values of the KS, CvM, AD, V, $Q_{\alpha-1}$ and $Q^{(3)}$ statistics are at 1.255, 1.222, 1.220, 1.215, 1.251, and 1.212, respectively.

Note that the rankings of the KS statistic and $Q_{\alpha-1}$ match in Table 3; a ranking which orders the estimates in decreasing order. That is, the KS and $Q_{\alpha-1}$ statistics achieve their respective minimums outside of the range of the estimates in the table. The QQ plot in Figure 5 suggests that the $P(\sigma = 500, \alpha = 1.209)$ model is a very good fit to the fire claims data. Considering the evidence of the QQ plot in Figure 5, it seems more reasonable that the parameter α should be within the range of estimates in Table 3. In this example we can also observe that V and $Q^{(3)}$ are in approximate agreement with each other, ranking 1.214 and 1.215 in first and second, while the CvM and AD statistics rank 1.221 and 1.220 at the top.

4.4. Hypothesis test results

Goodness-of-fit tests based on V and Q statistics can easily be applied using Monte Carlo methods to obtain the critical values of the test statistics or significance probabilities. We tested the null hypothesis $H_0 : X \sim P(\sigma, \alpha)$ using simulation size 10,000. The results are summarized in Table 4. The p-values for KS, CvM, and AD tests are reported in Brazauskas and Serfling (2003).

TABLE 4.

GOODNESS-OF-FIT TESTS FOR FITTED PARETO MODELS BASED ON MAXIMUM LIKELIHOOD ESTIMATES OF TAIL INDEX α AND SPECIFIED σ IN THREE EXAMPLES (p-VALUES BASED ON SIMULATION SIZE 10,000).

Data	σ	MLE	KS	CvM	AD	V	$Q_{\alpha/3}$
Wind	1.5	0.764	0.51	0.27	0.24	0.44	0.39
OLT	25	1.152	0.35	0.42	0.26	0.35	0.60
Fire	500	1.218	0.70	0.89	0.71	0.99	0.99

In each case the Pareto hypothesis is retained when α is estimated by the MLE in the fitted model. Note that the minimum distance estimate of α using the statistic V_1 is 1.215, which is almost exactly equal to the MLE, 1.218. This fact is reflected in the high p-value.

The quadratic statistics, V, and Q, represented in Figures 2-4 have similar shapes. For comparison, we plotted the statistics together in Figure 6, where it is more obvious that the statistics are not equivalent.

4.5. Cross-validation

Using the goodness-of-fit statistics to rank the estimates in Examples 1-3 implicitly supposes that each of the goodness-of-fit statistics is comparable in terms of the ability to measure departure from a Pareto(I) model. Power of a test depends on the alternative, and no GOF test is uniformly most powerful against all alternatives. The test statistics can be compared via cross-validation. Choose an integer k less than sample size n. Then

1. For each replicate, $j = 1, \ldots, m$ and each GOF statistic $\ell = 1, \ldots, r$:
   (a) Randomly select a training sample of size k from the full sample, reserving the remaining $n - k$ observations for the test set.
   (b) Using the j^{th} training set find the value of $\hat{\alpha}_{j\ell}$ that minimizes the ℓ^{th} GOF statistic.
   (c) Compute the squared error $e_{j\ell}^2$ when the $P(\sigma, \hat{\alpha}_{j\ell})$ model is fit to the j^{th} test set.
2. Compute the mean of the replicates $e_{j\ell}^2$ for the ℓ^{th} statistic, $\ell = 1, \ldots, r$, which is an estimate of the expected squared error when the statistic is applied as a minimum distance estimator.

Remark 2. Estimated squared error for the test set $\sum(F_n(x_i) - \hat{F}(x_i))^2$ is related to the CvM statistic. Other criteria than an L_2 distance for measuring error are possible, but we use the squared error here for its ease of interpretation.

The cross-validation experiment was replicated 40,000 times using a training sample size $k = 20$. The results are summarized in Table 5. For convenient interpretation, values in each row are divided by the result obtained for the MLU estimate. Thus, a value greater than 1 indicates that the mean squared error of the fit corresponding to the GOF distance statistic is higher than that of MLU, while values lower than 1 indicate better fit on average. In addition to the three data sets above, a simulated Pareto data set is included in the analysis for comparison.

Cross-validation suggests that each of the goodness-of-fit statistics performs reliably well, and it is reasonable to use any of them to rank and compare

TABLE 5.

CROSS-VALIDATION ESTIMATES FOR SQUARED ERROR OF FITS FOR THREE DATA SETS, RELATIVE TO MLU ESTIMATES.

Data	n	k	Error	MLU	MLE	KS	CvM	AD	V
Wind	40	20	Mean	1.00	1.03	1.03	1.05	0.98	0.93
			SD	1.00	1.02	1.11	1.10	0.96	0.88
OLT	90	20	Mean	1.00	1.01	0.73	0.77	0.72	0.73
			SD	1.00	1.02	0.56	0.58	0.55	0.56
Fire	142	20	Mean	1.00	1.00	1.06	1.04	0.92	0.84
			SD	1.00	1.01	1.15	1.16	0.98	0.87
P(500, 1.2)†	142	20	Mean	1.00	1.00	1.05	1.03	0.92	0.83
			SD	1.00	1.01	1.15	1.15	0.97	0.86

† Simulated Pareto(σ, α) data.

estimates or to fit the distribution. The AD statistic and proposed statistic V had better performance in the three examples than KS or CvM. In two of three examples, our statistic V achieves the best result in terms of estimated squared error for the fit on the test data, and can be considered best overall for the set of three examples. On the simulated Pareto data, the statistic V has the lowest average squared error.

Remark 3. If some prior information is available about α, we can also apply other statistics as minimum distance estimators. The statistic $Q_{\alpha/3}$ performed even better than V in some preliminary cross-validation studies. In this optimization problem crossing integer boundaries in the parameter space must be handled carefully, but the statistic V can be applied across the entire parameter space $\alpha > 0$.

4.6. Monte Carlo power comparison

Although a comprehensive Monte Carlo power study is beyond the scope of this paper, we compared the power of the EDF tests with the new test based on V. In each of these tests, the equivalent two parameter exponential distribution is the null distribution.

The null hypothesis is $H_0 : X \sim \text{Pareto}(\sigma,\alpha)$ [$\log X \sim \text{Exp}(\log(\sigma), \alpha)$]. In case (i) $\alpha = 1.2$, $\sigma = 1$ and case (ii) $\alpha = 0.7$, $\sigma = 1$. Each test for V applies parametric simulation of the null distribution with 199 replicates. Power is estimated as the proportion of significant tests in 2000 simulated data sets at 10% significance.

Results are summarized in Figures 7-10. In results of Figures 7 and 8 the sample size is $n = 30$ and the alternative α_1 varies in increments of 0.1 for cases (i) and (ii), respectively. In Figures 9 and 10, $\alpha = 1.2$ and the alternatives are fixed at $\alpha_1 = 1.4$ and $\alpha_1 = 1$, with sample size n on the horizontal axis. This comparison suggests that the V test is somewhat more powerful than the EDF tests for the examples investigated.

5. Summary

We have introduced and implemented several new statistics for measuring goodness-of-fit in Pareto type I and type II models, and illustrated their application in estimation and tests on three examples of claims data.

The empirical studies presented above suggest that the proposed statistic V, which measures goodness-of-fit of $\log X$ to the $\text{Exp}(\log\sigma, \alpha)$ model, is easy to apply, universally applicable, and a good measure of fit. The statistic $Q_{1-\alpha}$ is easy to apply, but may not be the best minimum distance statistic for estimation purposes. The statistics Q_β, for $0 < \beta < \alpha < 1$, perform well but require evaluation of beta functions and do not have a simple form when $\alpha \geq 1$. Finally, the statistics $Q^{(r)}$, which measure fit of a power transformation X^r of the data to a Pareto distribution, are easy to apply, universally applicable, with similar performance as V. Both V and $Q^{(r)}$ have computational complexity $O(n \log n)$, as do the EDF tests. Cross-validation on three examples suggests that the V statistic is a better measure of fit to Pareto(I) distribution than the EDF tests.

Although each of our proposed statistics has desirable statistical properties including statistical consistency, V or $Q^{(r)}$ could be recommended for simplicity and universal applicability. Comparative power studies against Pareto and non-Pareto alternatives are planned for future research.

Application to multivariate loss distributions is a promising extension. The univariate Pareto goodness-of-fit statistics were given as special cases of multivariate statistics, hence the statistics introduced in this paper have a natural extension to testing goodness-of-fit of multivariate loss models. Such an extension is not possible with the EDF statistics because multivariate observations cannot be ranked. Theoretical properties of our proposed statistics including statistical consistency will hold in the multivariate case under the same assumptions; that is, no distributional properties other than existence of second moments are assumed for inference.

APPENDIX

A. Proof of Statements

Lemma 1. *If* $X \sim P(\sigma,\alpha)$ *and* $0 < \beta < \alpha$, *then*

$$\int_y^\infty (x-y)^\beta \, dF_X(x) = \frac{\alpha\sigma^\alpha B(\alpha-\beta,\, \beta+1)}{y^{\alpha-\beta}}, \quad y \ge \sigma; \tag{A.1}$$

and

$$\int_\sigma^\infty \int_y^\infty (x-y)^\beta \, dF_X(x)\, dF_X(y) = \frac{\alpha^2\sigma^\beta B(\alpha-\beta,\, \beta+1)}{2\alpha-\beta}. \tag{A.2}$$

Proof. After a change of variables $t = x - y$ we obtain

$$\int_y^\infty (x-y)^\beta x^{-\alpha-1} dx = \int_0^\infty t^\beta (t+y)^{-\alpha-1} dt = \int_0^\infty \frac{t^\beta dt}{(y+t)^{\alpha+1}}$$
$$= \int_0^\infty \frac{t^\beta dt}{y^{\alpha+1}[1+t/y]^{\alpha+1}} = \frac{1}{y^q}\int_0^\infty \frac{t^{p-1} dt}{y^p[1+t/y]^{p+q}},$$

where $p = \beta + 1$ and $q = \alpha - \beta$. The integrand above is proportional to a beta density of the second kind (see e.g. Kleiber and Kotz (2003, 6.1.1)), which has density function

$$f(t) = \frac{t^{p-1}}{b^p B(p,q)[1 + t/b]^{p+q}}, \quad t > 0,$$

where $b > 0$, $p > 0$, and $q > 0$. Hence

$$\alpha\sigma^\alpha \int_y^\infty (x - y)^\beta x^{-\alpha-1} dx = \frac{\alpha\sigma^\alpha B(\alpha - \beta, \beta + 1)}{y^{\alpha-\beta}}.$$

Equation (A.2) follows directly from (A.1) by the power rule of integration. □

Case 3. $[0 < \beta < \alpha < 1]$

Proof. Make the substitution $t = (y - x)/y$ and set $y_0 = (y - \sigma)/y$. Then using integration by parts we obtain

$$\begin{aligned} \int_\sigma^y (y - x)^\beta dF_X(x) &= \alpha\sigma^\alpha y^{\beta-\alpha} \int_0^{y_0} t^\beta (1 - t)^{-\alpha-1} dt \\ &= \alpha\sigma^\alpha y^{\beta-\alpha} \left\{ \frac{y_0^\beta (1 - y_0)^{-\alpha}}{\alpha} - \frac{\beta}{\alpha} \int_0^{y_0} t^{\beta-1} (1 - t)^{-\alpha} dt \right\} \\ &= (y - \sigma)^\beta - \frac{\beta\sigma^\alpha}{y^{\alpha-\beta}} B_{y_0}(\beta, 1 - \alpha). \qquad \text{(A.3)} \end{aligned}$$

Combining (A.3) and equation (A.1) of Lemma 1 we obtain equation (3.8). Statement (3.9) follows from (A.2) in Lemma 1. □

Case 4. $[\alpha = 1 \text{ and } 0 < \beta < 1]$

Proof. Applying integration by parts and change of variables $t = (y - x)/y$, we obtain

$$\begin{aligned} \int_\sigma^y (y - x)^\beta dF(x) &= (y - \sigma)^\beta - \sigma\beta \int_\sigma^y (y - x)^{\beta-1} x^{-1} dx \\ &= (y - \sigma)^\beta - \sigma\beta \int_0^{y_0} t^{\beta-1} (1 - t)^{-1} dt \qquad \text{(A.4)} \\ &= (y - \sigma)^\beta - \sigma\beta y^{\beta-1} \left\{ \frac{y_0^\beta}{\beta} - \frac{y_0^{\beta+1}}{\beta + 1} {}_2F_1(\beta + 1, 1; \beta + 2; y_0) \right\}. \end{aligned}$$

In the last step a known result is applied (see e.g. Prudnikov, et al., 1990, pp. 29-30). Combining (A.4) and (A.1) from Lemma 1, we obtain (3.10). Finally, (3.11) can be obtained by integration, or more simply as the limit as α approaches 1 from below of (3.9). □

B. Data Sets

TABLE 6

WIND CATASTROPHE LOSSES (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

TABLE 7.

OLT BODILY INJURY LIABILITY CLAIMS (1976) IN $1000'S.

loss	*n*	loss	*n*	loss	*n*	loss	*n*
25-30	11	50-55	3	120-130	2	240-250	2
30-35	18	55-60	2	140-150	3	260-270	1
35-40	9	70-75	9	190-200	1	280-290	1
40-45	4	75-80	1	200-210	2	290-300	2
45-50	11	95-100	4	220-230	1	340-350	1
						410-420	2

TABLE 8.

NORWEGIAN FIRE CLAIMS (1975) (1000 NORWEGIAN KRONES).

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

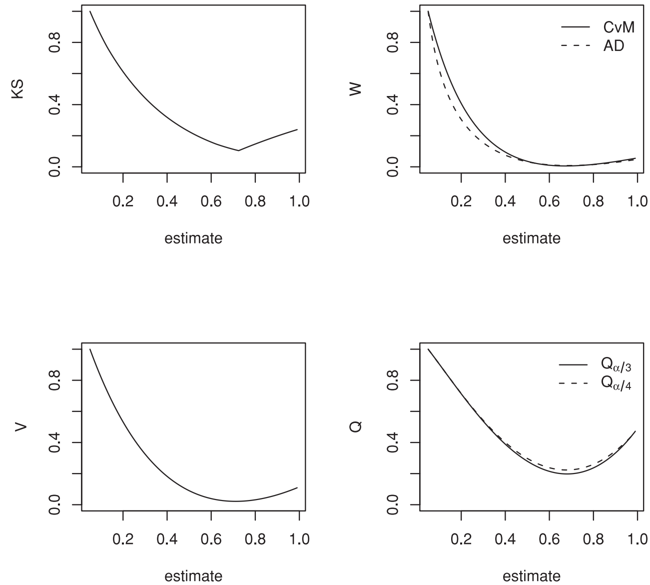

FIGURE 2: Value of goodness-of-fit statistics vs estimates for the wind catastrophes data, scaled to (0, 1]. The minima are achieved at 0.724 (KS), 0.673 (CvM), 0.686 (AD), 0.711 (V), 0.680 ($Q_{\alpha/3}$), and 0.678 ($Q_{\alpha/4}$).

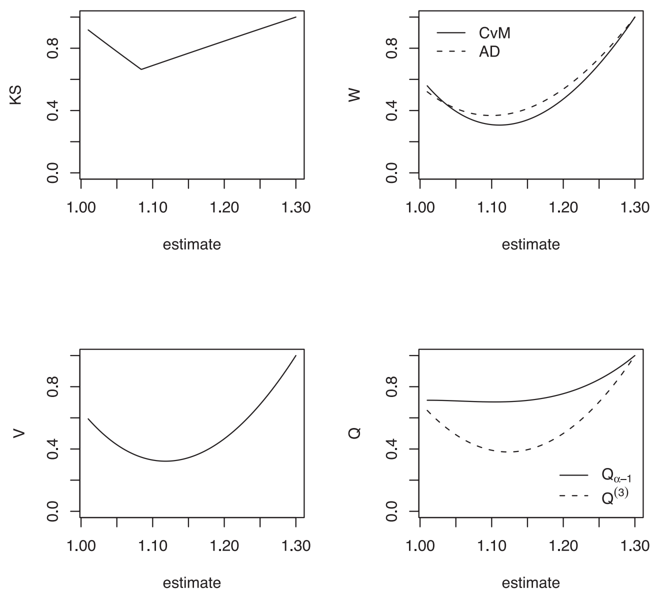

FIGURE 3: Value of goodness-of-fit statistics vs estimates for the OLT bodily injury liability data, scaled to (0,1]. The minima are achieved at 1.084 (KS), 1.111 (CvM), 1.099 (AD), 1.118 (V), 1.104 ($Q_{\alpha-1}$), and 1.123 ($Q^{(3)}$).

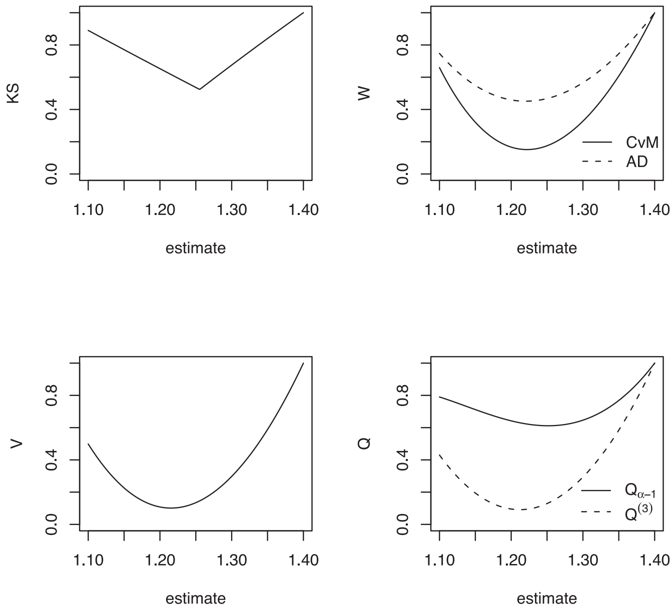

FIGURE 4: Value of goodness-of-fit statistics vs estimates for the Norwegian fire claims data, scaled to (0,1]. The minima are achieved at 1.255 (KS), 1.222 (CvM), 1.220 (AD), 1.215 (V), 1.251 ($Q_{\alpha-1}$), and 1.212 ($Q^{(3)}$).

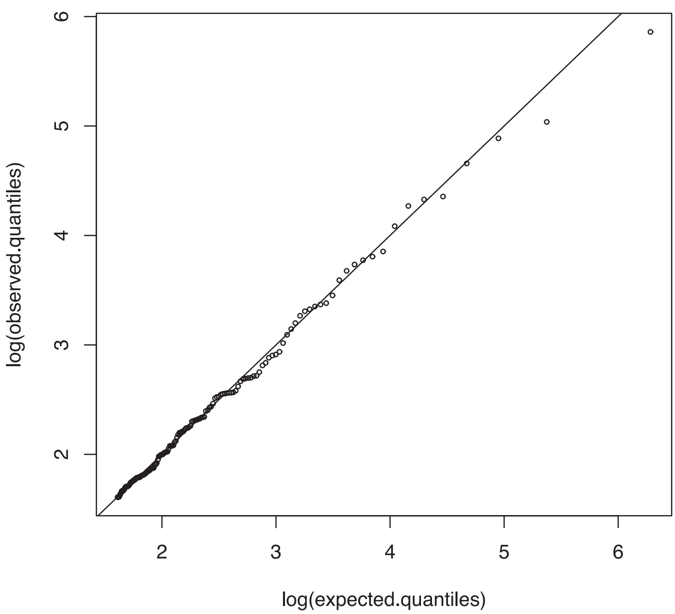

FIGURE 5: QQ plot of fire data on log-log scale, assuming a Pareto(500, 1.209) model.

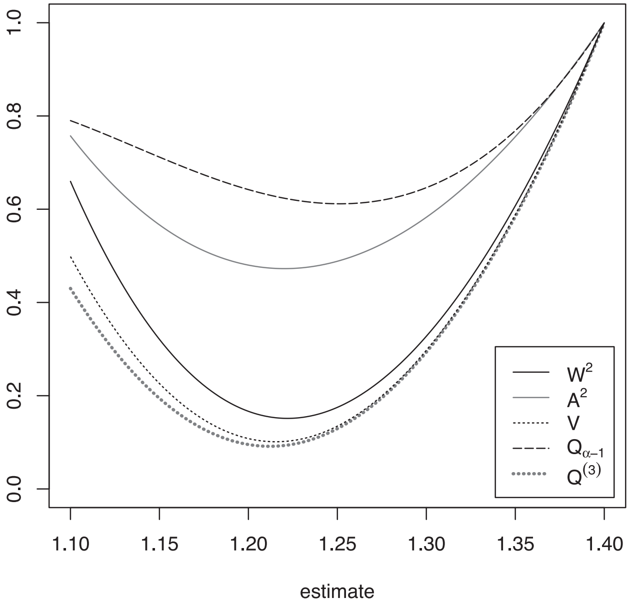

FIGURE 6: Goodness-of-fit statistics vs estimates for the Norwegian fire claims data, scaled to (0,1].

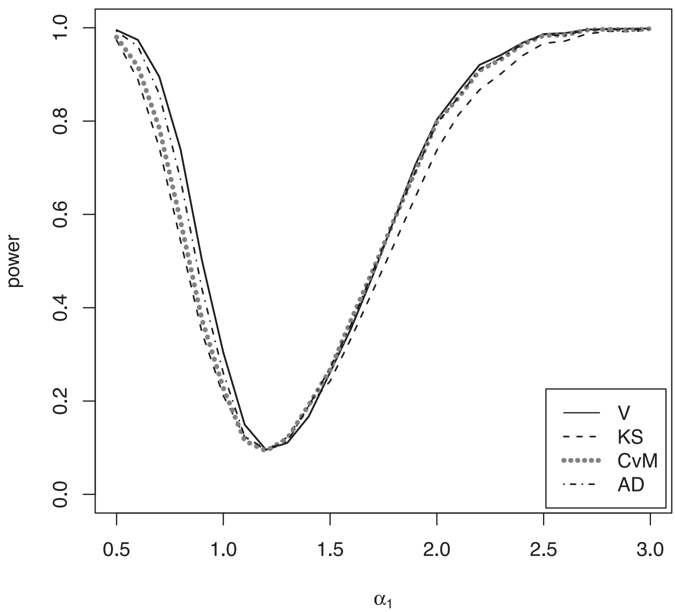

FIGURE 7: Power comparison of V and EDF tests for case (i) $\alpha = 1.2$, alternative α_1, $n = 30$, at 10% significance.

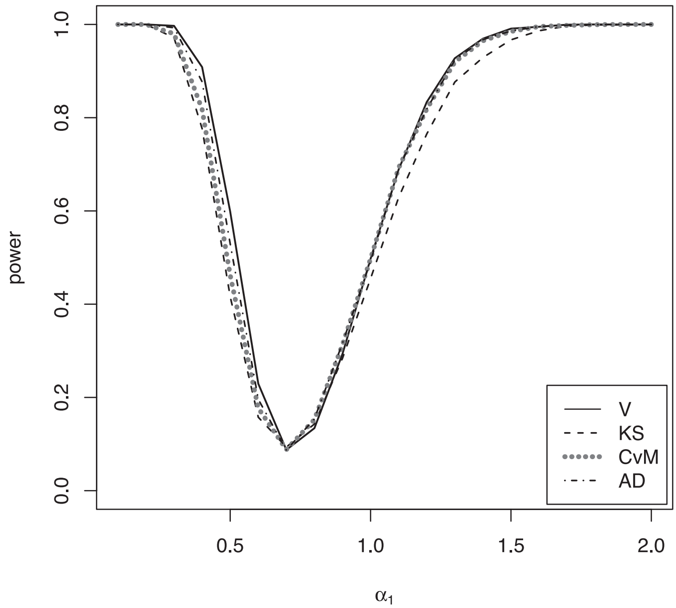

FIGURE 8: Power comparison of V and EDF tests for case (ii) $\alpha = 0.7$, alternative α_1, $n = 30$, at 10% significance.

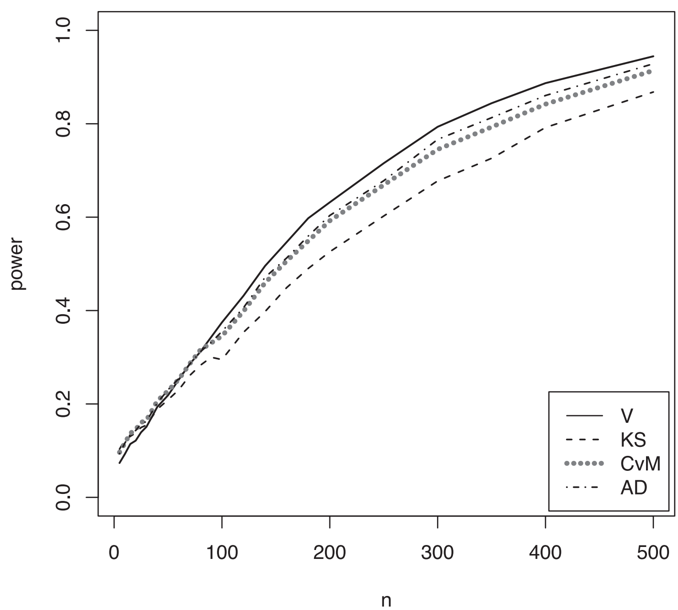

FIGURE 9: Power comparison of V and EDF tests for case (i) $\alpha = 1.2$, alternative $\alpha_1 = 1.4$, at 10% significance.

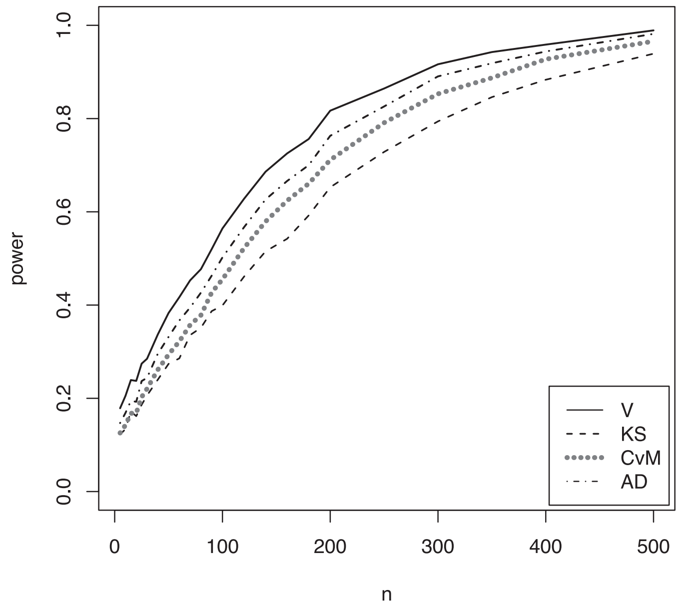

FIGURE 10: Power comparison of V and EDF tests for case (i) $\alpha = 1.2$, alternative $\alpha_1 = 1.0$, at 10% significance.

REFERENCES

ARNOLD, B.C. (1983) *Pareto Distributions*. International Co-operative Publishing House, Fairland, MD.

BAXTER, M.A. (1980) Minimum variance unbiased estimation of the parameters of the Pareto distribution. *Metrika*, **27**: 133-138.

BEIRLANT, J., TEUGELS, J.L. and VYNCKIER, P. (1996) *Practical Analysis of Extreme Values*. Leuven University Press, Leuven, Belgium.

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.

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.

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.

BRAZAUSKAS, V. and SERFLING, R. (2003) Favorable estimators for fitting Pareto models: A study using goodness-of-fit measures with actual data. *ASTIN Bulletin*, **33(2)**: 365-381. DOI: 10.2143/AST.33.2.503698.

HOEFFDING, W. (1948) A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, **19**: 293-325.

HOGG, R.V. and KLUGMAN, S.A. (1984) *Loss Distributions*. Wiley, New York.

KLEIBER, C. and KOTZ, S. (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley.

LIKEŠ, J. (1969) Minimum variance unbiased estimation of the parameters of power-function and Pareto's distribution. *Statistische Hefte*, **10**: 104-110.

Patrik, G. (1980) Estimating casualty insurance loss amount distributions. *Proceedings of the Casualty Actuarial Society*, **LXXVII**: 57-109.

Philbrick, S.W. and Jurschak, J. (1981) Discussion of "Estimating casualty insurance loss amount distributions". *Proceedings of the Casualty Actuarial Society*, **LXVIII**: 101-106.

Porter III, J.E., Coleman, J.W. and Moore, A.H. (1992) Modified KS, AD, and C-vM tests for the Pareto distribution with unknown location & scale parameters. *IEEE Transactions on Reliability*, **41(1)**: 112-117.

Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. (1990) *Integrals and series* (Integraly i riady, *translated from the Russian by N.M. Queen*). Gordon and Breach Science Publishers, New York.

Rytgaard, M. (1990) Estimation in the Pareto distribution. *ASTIN Bulletin*, **20(2)**: 201-216.

Stephens, M.A. (1986) Tests based on EDF statistics. In R.B. D'Agostino and M.A. Stephens, editors, *Goodness-of-Fit Techniques*, pages 97-193. Marcel Dekker, New York.

Székely, G.J. and Rizzo, M.L. (2005a) A new test for multivariate normality. *Journal of Multivariate Analysis*, **93(1)**: 58-80.

Székely, G.J. and Rizzo, M.L. (2005b) Hierarchical clustering via joint between-within distances: extending Ward's minimum variance method. *Journal of Classification*, **22(2)**: 151-183.

von Mises, R. (1947) On the asymptotic distributions of differentiable statistical functionals. *Annals of Mathematical Statistics*, **2**: 209-348.

Maria L. Rizzo
Dept. of Mathematics & Statistics
Bowling Green State University
Bowling Green, OH 43403
E-Mail: mrizzo@bgsu.edu
Phone: 419-372-7474
Fax: 419-372-6092