



Testing for the Pareto type I distribution: a comparative study

L. Ndwandwe¹ · J. S. Allison¹ · L. Santana¹ · I. J. H. Visagie¹

Received: 21 December 2022 / Accepted: 7 July 2023 / Published online: 29 July 2023
© The Author(s) 2023

Abstract

Pareto distributions are widely used models in economics, finance and actuarial sciences. As a result, a number of goodness-of-fit tests have been proposed for these distributions in the literature. We provide an overview of the existing tests for the Pareto distribution, focussing specifically on the Pareto type I distribution. To date, only a single overview paper on goodness-of-fit testing for Pareto distributions has been published. However, the mentioned paper has a much wider scope than is the case for the current paper as it covers multiple types of Pareto distributions. The current paper differs in a number of respects. First, the narrower focus on the Pareto type I distribution allows a larger number of tests to be included. Second, the current paper is concerned with composite hypotheses compared to the simple hypotheses (specifying the parameters of the Pareto distribution in question) considered in the mentioned overview. Third, the sample sizes considered in the two papers differ substantially. In addition, we consider two different methods of fitting the Pareto Type I distribution; the method of maximum likelihood and a method closely related to moment matching. It is demonstrated that the method of estimation has a profound effect, not only on the powers achieved by the various tests, but also on the way in which numerical critical values are calculated. We show that, when using maximum likelihood, the resulting critical values are shape invariant and can be obtained using a Monte Carlo procedure. This is not the case when moment matching is employed. The paper includes an extensive Monte Carlo power study. Based on the results obtained, we recommend the use of a test based on the phi divergence together with maximum likelihood estimation.

Keywords Goodness-of-fit testing · Parametric bootstrap · Pareto distribution

1 Introduction and motivation

The Pareto distribution was first introduced by the economist Vilfredo Pareto in 1897 as a model for the distribution of income, see Pareto [41]. Since then the Pareto distribution has been widely used in a variety of fields including economics, finance, actuarial science, and reliability theory, see, e.g., Nofal and El Gebaly [38] as well as Ismaïl [24]. For an in-depth

✉ J. S. Allison
james.allison@nwu.ac.za

¹ Department of Statistics, North-West University, Potchefstroom, South Africa

discussion of the Pareto distribution the interested reader is referred to Arnold [7] where the role of this distribution in the modelling of data is discussed.

The popularity of the Pareto distribution has prompted research into several generalisations of this model. Subsequently, the originally proposed distribution became known as the Pareto type I distribution in order to distinguish this model from the variants known as the Pareto types II, III and IV as well as the so-called generalised Pareto distribution. These distributions, as well as the relationships between them, are described in detail in Arnold [7].

Due to the wide range of applications of the various types of Pareto distributions, a number of tests have been developed for the hypothesis that observed data follow a Pareto distribution. This paper provides an overview of the goodness-of-fit tests specifically developed for the Pareto type I distribution available in the literature. Although numerous overview papers are available for goodness-of-fit tests for distributions such as the normal distribution, see, e.g., Bera et al. [9], and the exponential distribution, see, e.g., Allison et al. [5], the only overview paper of this kind relating to the Pareto distribution is Chu et al. [16]. The latter investigates several existing tests for the Pareto types I and II as well as the generalised Pareto distribution. However, due to the wider scope, Chu et al. [16] does not review all of the tests available for the Pareto type I distribution; several recently proposed tests are excluded from the comparisons provided. The current paper has a narrower scope and provides an overview of existing tests specifically for the Pareto type I distribution, hereafter simply referred to as the Pareto distribution.

A further distinction between Chu et al. [16] and the study presented here is that the former considers simple hypotheses in which the parameters of the Pareto distribution are specified beforehand, whereas the current paper is concerned with the testing of the composite hypothesis that data follow a Pareto distribution with unspecified parameters. Furthermore, note that the sample sizes considered in the two papers are quite distinct; while Chu et al. [16] considers the performance of tests for larger sample sizes, our focus is on the performance of the tests in the case of smaller samples. Additionally, Chu et al. [16] employs only maximum likelihood estimation, whereas the current paper uses both maximum likelihood and the adjusted method of moments estimators. In the study presented here, we compare the powers achieved by the various tests using the two different estimation techniques and we demonstrate that the parameter estimation method, perhaps surprisingly, substantially influences the powers associated with the various tests. Lastly, the critical values used in Chu et al. [16] are obtained using a bootstrap approach; in Sect. 3, we show that it is possible to obtain critical values independent of the estimated parameters when using maximum likelihood estimation. This allows us to estimate critical values without resorting to a bootstrap procedure in the case where maximum likelihood parameter estimates are employed.

In order to proceed we introduce some notation. Let X, X_1, X_2, \dots, X_n be independent and identically distributed (i.i.d.) continuous positive random variables with an unknown distribution function F . Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ denote the order statistics of X_1, X_2, \dots, X_n . Denote the Pareto distribution function by

$$F_{\beta, \sigma}(x) = \begin{cases} 1 - \left(\frac{x}{\sigma}\right)^{-\beta}, & x \geq \sigma, \\ 0, & \text{otherwise,} \end{cases} \quad (1.1)$$

and the density function by

$$f_{\beta, \sigma}(x) = \begin{cases} \frac{\beta \sigma^\beta}{x^{\beta+1}}, & x \geq \sigma, \\ 0, & \text{otherwise,} \end{cases}$$

where $\beta > 0$ is a shape parameter and $\sigma > 0$ is a scale parameter. To indicate that the distribution of a random variable X is the Pareto distribution with shape and scale parameters β and σ , we make use of the following shorthand notation: $X \sim P(\beta, \sigma)$.

The hypothesis to be tested is that an observed data set is realised from a Pareto distribution, but we distinguish between two distinct hypothesis testing scenarios. In the first scenario, the value of σ in (1.1) is *known* while the value of β is unspecified. Note that σ determines the support of the Pareto distribution. As a result, if the support of the distribution is known, then the value of σ is also known. As a concrete example, consider the case of an insurance company. Typically, an insurance claim is subject to a so-called excess, meaning that the insurance company will only receive a claim if it exceeds a known, fixed value. A closely related example is considered in Sect. 5; here the monetary expenses (above a certain threshold) resulting from wind related catastrophes are examined. Another example is found in Arnold [7], where the lifetime tournament earnings of professional golfers are considered. However, only golfers with a total lifetime earning exceeding \$700 000 are considered. In the second hypothesis testing scenario considered, we may be interested in modelling a phenomenon for which the support of F is unknown and the values of both β and σ require estimation. In both testing scenarios, we are interested in testing the composite goodness-of-fit hypothesis

$$H_0 : F(x) = F_{\beta, \sigma}(x), \quad (1.2)$$

for some $\beta > 0$, $\sigma > 0$ and all $x > \sigma$. This hypothesis is to be tested against general alternatives.

The remainder of the paper is organised as follows. Section 2 provides an overview of a large number of tests for the Pareto distribution based on a wide range of characterisations of this distribution. Section 3 considers two types of estimators for the parameters of the Pareto distribution; the method of maximum likelihood as well as a method closely related to the method of moments. This section also details the estimation of critical values for the tests considered. An extensive Monte Carlo study is presented in Sect. 4. This section investigates and compares the finite sample performance of the tests in various of settings. Section 5 presents a practical implementation of the goodness-of-fit tests as well as the parameter estimation techniques considered. These techniques are demonstrated using a data set comprised of the monetary expenses resulting from wind related catastrophes in 40 separate instances during the year 1977. Some conclusions are presented in Sect. 6.

2 Goodness-of-fit tests for the Pareto distribution

We discuss various goodness-of-fit tests for the Pareto distribution below; tests are grouped according to the characteristic of the Pareto distribution that the tests are based on. We consider tests utilising the empirical distribution function, likelihood ratios, entropy, phi-divergence, empirical characteristic function as well as Mellin transform. Additionally, the discussion below includes tests based on the so-called inequality curve as well as various characterisations of the Pareto distribution. All tests are omnibus tests, except where stated otherwise. To simplify notation, let $U_j = F_{\beta, \sigma}(X_j)$ and $\widehat{U}_j = F_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)$, $j = 1, 2, \dots, n$, where $\widehat{\beta}_n$ and $\widehat{\sigma}_n$ are consistent estimates of the shape and scale parameters of the Pareto distribution (these estimates will be discussed in Sect. 3). Under H_0 , we have from the probability integral transform that $U_j \sim U[0, 1]$ and that \widehat{U}_j should be approximately standard uniformly distributed. Some of the tests below exploit this property.

2.1 Tests based on the empirical distribution function (edf)

Classical edf-based tests, such as the Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling tests are based on a distance measure between parametric and non-parametric estimates of the distribution function. The non-parametric estimate of the distribution function of X_1, X_2, \dots, X_n used is the edf,

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n I(X_j \leq x),$$

with $I(\cdot)$ the indicator function, while the parametric estimate of the distribution function is

$$F_{\hat{\beta}_n, \hat{\sigma}_n}(x) = 1 - \left(\frac{x}{\hat{\sigma}_n}\right)^{-\hat{\beta}_n}.$$

The Kolmogov-Smirnov test statistic, corresponding to the supremum difference between $F_{\hat{\beta}_n, \hat{\sigma}_n}$ and F_n , is

$$KS_n = \sup_{x \geq \hat{\sigma}_n} |F_n(x) - F_{\hat{\beta}_n, \hat{\sigma}_n}(x)|.$$

The remaining edf test statistics considered are (weighted) L^2 distances and have the following general form,

$$n \int_{-\infty}^{\infty} [F_n(x) - F_{\hat{\beta}_n, \hat{\sigma}_n}(x)]^2 w(x) dF_{\hat{\beta}_n, \hat{\sigma}_n}(x), \tag{2.1}$$

where $w(x)$ is some weight function. Choosing $w(x) = 1$ in (2.1), we have the Cramér–von Mises test with direct calculable form

$$CM_n = \frac{1}{12n} + \sum_{j=1}^n \left[\hat{U}_{(j)} - \frac{2j-1}{2n} \right]^2.$$

When choosing $w(x) = [F_{\hat{\beta}_n, \hat{\sigma}_n}(x)\{1 - F_{\hat{\beta}_n, \hat{\sigma}_n}(x)\}]^{-1}$, we obtain the Anderson–Darling test

$$AD_n = -n - \frac{1}{n} \sum_{j=1}^n (2j-1) [\log(\hat{U}_{(j)}) + \log(1 - \hat{U}_{(n+1-j)})].$$

Finally, setting $w(x) = [1 - F_{\hat{\beta}_n, \hat{\sigma}_n}(x)]^{-2}$, we arrive at the so-called modified Anderson–Darling test

$$MA_n = \frac{n}{2} - 2 \sum_{j=1}^n \hat{U}_j - \sum_{j=1}^n \left[2 - \frac{2j-1}{n} \right] \log(1 - \hat{U}_{(j)}).$$

While the CM_n, AD_n and MA_n tests are all weighted L^2 distances between the parametric and non-parametric estimates of the distribution function, the weight functions used vary the importance allocated to different types of deviations between these estimates. For example, when comparing the Cramér–von Mises and Anderson–Darling tests, differences in the tail of the distribution are more heavily weighted in the case of the latter than the former. For further discussions on these edf-based tests, see, Klar [27] and D’Agostino and Stephens [20]. All of the above tests reject the null hypothesis for large values of the test statistics.

2.2 Tests based on likelihood ratios

Zhang [54] proposes two general test statistics which are used to test for normality; below we adapt these tests in order to test for the Pareto distribution. The test statistics are of the form

$$T_n = \int_{-\infty}^{\infty} G_n(x)dw(x), \tag{2.2}$$

where $G_n(x)$ is the likelihood ratio statistic defined as

$$G_n(x) = 2n \left\{ F_n(x) \log \left(\frac{F_n(x)}{\widehat{U}_{(j)}} \right) + [1 - F_n(x)] \log \left(\frac{1 - F_n(x)}{1 - \widehat{U}_{(j)}} \right) \right\}.$$

The two choices of $dw(x)$ that Zhang [54] proposes, as well as the test statistics resulting from each of these choices, are presented below. The results are obtained upon setting $F_n(X_{(j)}) = (j - \frac{1}{2})/n$.

- Choosing $dw(x) = [F_n(x)\{1 - F_n(x)\}]^{-1} dF_n(x)$ leads to

$$ZA_n = - \sum_{j=1}^n \left\{ \frac{\log(\widehat{U}_{(j)})}{n - j + \frac{1}{2}} + \frac{\log(1 - \widehat{U}_{(j)})}{j - \frac{1}{2}} \right\}.$$

- Choosing $dw(x) = [F_{\widehat{\beta}_n, \widehat{\sigma}_n}(x)\{1 - F_{\widehat{\beta}_n, \widehat{\sigma}_n}(x)\}]^{-1} dF_{\widehat{\beta}_n, \widehat{\sigma}_n}(x)$ results in

$$ZB_n = \sum_{j=1}^n \left\{ \log \left(\frac{(\widehat{U}_{(j)})^{-1} - 1}{(n - \frac{1}{2})/(j - \frac{3}{4}) - 1} \right) \right\}^2.$$

Motivated by the high powers often obtained using the modified Anderson–Darling test, we also include the choice $dw(x) = \{1 - F_n(x)\}^{-2}dF_n(x)$, which leads to the test statistic

$$ZC_n = 2 \sum_{j=1}^n \left\{ \frac{n(j - \frac{1}{2})}{(n - j + \frac{1}{2})^2} \log \left(\frac{j - \frac{1}{2}}{n\widehat{U}_{(j)}} \right) + \frac{n}{n - j + \frac{1}{2}} \log \left(\frac{n - j + \frac{1}{2}}{n(1 - \widehat{U}_{(j)})} \right) \right\}.$$

All three of these tests reject the null hypothesis for large values of the test statistics.

Building on the tests for the assumption of normality that Zhang [54] proposes, Alizadeh Noughabi [2] adapts two of these test to test the assumption of exponentiality. Neither Zhang [54] nor Alizadeh Noughabi [2] derive the asymptotic properties of these tests and rather present extensive Monte Carlo studies to investigate their finite sample performances. The authors found that these tests are quite powerful compared to other tests (especially the traditional edf-based tests) against a range of alternatives.

Remark Zhang [54] also considers the test

$$\begin{aligned} ZD_n &= \sup_{x \in \mathbb{R}} G(x) = \max_{1 \leq j \leq n} G(X_{(j)}) \\ &= \max_{1 \leq j \leq n} \left\{ \left(j - \frac{1}{2} \right) \log \left(\frac{j - \frac{1}{2}}{n\widehat{U}_{(j)}} \right) + \left(n - j + \frac{1}{2} \right) \log \left(\frac{n - j + \frac{1}{2}}{n(1 - \widehat{U}_{(j)})} \right) \right\}. \end{aligned}$$

However, we do not include ZD_n in our Monte Carlo study as ZA_n and ZB_n proved more powerful in the papers mentioned.

2.3 Tests based on entropy

A further class of tests is based on the concept of entropy, first introduced in Shannon [47]. The entropy of a random variable X with density and distribution functions f and F , respectively, is defined to be

$$H = - \int_0^\infty f(x) \log(f(x))dx = \int_0^1 \log \left(\frac{d}{dp} F^{-1}(p) \right) dp, \tag{2.3}$$

where $F^{-1}(\cdot)$ denotes the quantile function of X . The concept of entropy has been applied in several studies, see, e.g., Kullback [29], Kapur [26] and Vasicek [50], where, in particular, Vasicek [50] proposes using

$$H_{n,m} = \frac{1}{n} \sum_{j=1}^n \log \left\{ \binom{n}{2m} (X_{(j+m)} - X_{(j-m)}) \right\} \tag{2.4}$$

as an estimator for H , where $X_{(j)} = X_{(1)}$ for $j < 1$, $X_{(j)} = X_{(n)}$ for $j > n$, and m is a window width subject to $m \leq \frac{n}{2}$. We now consider two goodness-of-fit tests based on concepts related to entropy: the Kullback–Leibler divergence and the Hellinger distance, where H is estimated by $H_{n,m}$ in the test statistic.

The Kullback–Leibler divergence between any arbitrary density function, f , and the Pareto density, $f_{\beta,\sigma}$, is defined to be [see, e.g., 29]

$$KL = - \int_\sigma^\infty f(x) \log \left(\frac{f(x)}{f_{\beta,\sigma}(x)} \right) dx.$$

It follows that the Kullback–Leibler divergence can also be expressed in terms of entropy:

$$KL = -H - \int_\sigma^\infty f(x) \log(f_{\beta,\sigma}(x))dx. \tag{2.5}$$

Estimating (2.5) by the empirical quantities mentioned above, we obtain the test statistic

$$KL_{n,m} = -H_{n,m} - \log(\hat{\beta}_n) - \hat{\beta}_n \log(\hat{\sigma}_n) + (\hat{\beta}_n + 1) \frac{1}{n} \sum_{j=1}^n \log(X_j).$$

This test rejects the null hypothesis for large values of $KL_{n,m}$. The test statistic coincide with the one studied by Lequesne [30], where the authors uses maximum likelihood estimation and a normalizing transformation to ensure the test statistic lies between 0 and 1. Alizadeh Noughabi et al. [4] uses a similar test statistic in order to test the goodness-of-fit hypothesis for the Rayleigh distribution. Authors did not derive the limiting null distribution of the test statistic, they proved that the test is consistent against general alternatives. In their simulation study the authors find that the test compared favourably to other competing tests.

The Hellinger distance between two densities f and $f_{\beta,\sigma}$ is defined as [see, e.g., 25]

$$HD = \frac{1}{2} \int_\sigma^\infty \left(\sqrt{f(x)} - \sqrt{f_{\beta,\sigma}(x)} \right)^2 dx.$$

By setting $F(x) = p$, the Hellinger distance can be expressed in terms of the quantile function as follows

$$HD = \frac{1}{2} \int_0^1 \left(\sqrt{\left(\frac{d}{dp} F^{-1}(p) \right)^{-1}} - \sqrt{\frac{\beta\sigma^\beta}{(F^{-1}(p))^{\beta+1}}} \right)^2 \frac{d}{dp} F^{-1}(p) dp.$$

From (2.3) and (2.4) it can be argued that $\frac{d}{dp}F^{-1}(p)$ can be estimated by $\{\frac{n}{2m}(X_{(j+m)} - X_{(j-m)})\}$. The resulting test statistic is given by

$$HD_{n,m} = \frac{1}{2n} \sum_{j=1}^n \frac{\left[\left\{ \frac{n}{2m}(X_{(j+m)} - X_{(j-m)}) \right\}^{-1/2} - \left(f_{\hat{\beta}_n, \hat{\sigma}_n}(X_j) \right)^{\frac{1}{2}} \right]^2}{\left\{ \frac{n}{2m}(X_{(j+m)} - X_{(j-m)}) \right\}^{-1}}$$

This test rejects the null hypothesis for large values of $HD_{n,m}$.

Jahanshahi et al. [25] uses similar arguments to propose a goodness-of-fit test for the Rayleigh distribution and proves that the test is consistent in that setting. In addition, they also propose a method for obtaining the optimum value of m by minimising bias and mean square error (MSE). In a finite sample power comparison, Jahanshahi et al. [25] finds that $HD_{n,m}$ produces the highest estimated powers against the majority of alternatives considered. In the case of alternatives considered with non-monotone hazard rates, the entropy-based tests outperform the remaining tests by some margin.

2.4 Tests based on the phi-divergence

The phi-divergence between an arbitrary density, f , and $f_{\beta,\sigma}$ is

$$D_\phi(f, f_{\beta,\sigma}) = \int_\sigma^\infty \phi \left(\frac{f(x)}{f_{\beta,\sigma}(x)} \right) f_{\beta,\sigma}(x) dx,$$

where $\phi : [0, \infty) \rightarrow (-\infty, \infty)$ is a convex function such that $\phi(1) = 0$ and $\phi''(1) > 0$. It is further known [see, e.g., 15, 18] that if ϕ is strictly convex in a neighbourhood of $x = 1$, then $D_\phi(f, f_{\beta,\sigma}) = 0$ if, and only if, $f = f_{\beta,\sigma}$. Alizadeh Noughabi & Balakrishnan [3] use this property to construct goodness-of-fit tests for a variety of different distributions. Let $E_F[\cdot]$ denote an expectation taken with respect to the distribution F . By noting that

$$D_\phi(f, f_{\beta,\sigma}) = \int_\sigma^\infty \phi \left(\frac{f(x)}{f_{\beta,\sigma}(x)} \right) \frac{f_{\beta,\sigma}(x)}{f(x)} dF(x) = E_F \left[\phi \left(\frac{f(X)}{f_{\beta,\sigma}(X)} \right) \frac{f_{\beta,\sigma}(X)}{f(X)} \right],$$

it follows that $D_\phi(f, f_{\beta,\sigma})$ can be estimated by

$$\widehat{D}_\phi(\widehat{f}_h, f_{\widehat{\beta}_n, \widehat{\sigma}_n}) = \frac{1}{n} \sum_{j=1}^n \left[\phi \left(\frac{\widehat{f}_h(X_j)}{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)} \right) \frac{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)}{\widehat{f}_h(X_j)} \right], \tag{2.6}$$

where $\widehat{f}_h(x) = \frac{1}{nh} \sum_{j=1}^n k \left(\frac{x-X_j}{h} \right)$ is the kernel density estimator with kernel function $k(\cdot)$ and bandwidth h .

In the Monte Carlo study in Sect. 4, we use the standard normal density function as kernel and choose $h = 1.06sn^{-\frac{1}{5}}$, where s is the unbiased sample standard deviation [see, e.g., 48]. We will use the following four choices of ϕ :

- The Kullback-Liebler distance (DK) with $\phi(x) = x \log(x)$.
- The Hellinger distance (DH) with $\phi(x) = \frac{1}{2}(\sqrt{x} - 1)^2$.
- The Jeffreys divergence distance (DJ) with $\phi(x) = (x - 1) \log(x)$.
- The total variation distance (DT) with $\phi(x) = |x - 1|$.

A variety of test statistics can be constructed from (2.6) using the above choices of ϕ . The test statistics corresponding to these choices are

$$\begin{aligned}
 DK_n &= \frac{1}{n} \sum_{j=1}^n \left[\log \left(\frac{\widehat{f}_h(X_j)}{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)} \right) \right], \\
 DH_n &= \frac{1}{2n} \sum_{j=1}^n \left[\left(1 - \sqrt{\frac{\widehat{f}_h(X_j)}{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)}} \right)^2 \frac{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)}{\widehat{f}_h(X_j)} \right], \\
 DJ_n &= \frac{1}{n} \sum_{j=1}^n \left[\left(1 - \frac{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)}{\widehat{f}_h(X_j)} \right) \log \left(\frac{\widehat{f}_h(X_j)}{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)} \right) \right], \text{ and} \\
 DT_n &= \frac{1}{n} \sum_{j=1}^n \left[\left| \frac{\widehat{f}_h(X_j)}{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)} - 1 \right| \frac{f_{\widehat{\beta}_n, \widehat{\sigma}_n}(X_j)}{\widehat{f}_h(X_j)} \right].
 \end{aligned}$$

All tests reject the null hypothesis for large values of the test statistics.

In addition to showing that the tests above are consistent against fixed alternatives (no derivation of the asymptotic null distribution was presented), Alizadeh Noughabi & Balakrishnan [3] also uses DK_n , DH_n , DJ_n and DT_n to test the goodness-of-fit hypothesis for the normal, exponential, uniform and Laplace distributions. The Monte Carlo study included in Alizadeh Noughabi & Balakrishnan [3] indicates that DK_n produces the highest powers amongst the phi-divergence type tests. When comparing the performance of these tests, the powers associated with DK_n were higher than the others. As a result, only DK_n is included in the Monte Carlo study presented in Sect. 4.

2.5 A test based on the empirical characteristic function

A large number of goodness-of-fit tests have been developed for a variety of distributions based on empirical characteristic functions [see, e.g., 12, 28, 31]. For a review of testing procedures based on the empirical characteristic functions see, e.g., Meintanis [33].

Recall that the characteristic function (cf) of a random variable X with distribution F_θ is given by

$$\varphi_\theta(t) = E[e^{itX}] = \int e^{itx} dF_\theta(t),$$

with $i = \sqrt{-1}$ the imaginary unit. The empirical characteristic function (ecf) is defined to be

$$\varphi_n(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j}.$$

As a general test statistic, one can use a weighted L^2 distance between the fitted cf under the null hypothesis and the ecf,

$$\int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_{\widehat{\theta}}(t)|^2 w(t) dt,$$

where $\widehat{\theta}$ represents the estimated values of the parameters of the hypothesised distribution and $w(\cdot)$ is a suitably chosen weight function ensuring that the integral is finite. Commonly used

choices for the weight function are $w(t) = e^{-a|t|}$ and $w(t) = e^{-at^2}$, respectively derived from the kernels of the Laplace and normal density functions, where $a > 0$ is a user defined tuning parameter.

The characteristic function of the Pareto distribution has a complicated closed form expression, making T_n intractable irrespective of the choice of the weight function. In order to circumvent this problem, we use the test proposed in Meintanis [31]. In order to perform this test, the data are transformed so as to approximately follow a standard uniform distribution under the null hypothesis. The test statistic used is a weighted L^2 distance between the ecf of the transformed data $\widehat{U}_1, \widehat{U}_2, \dots, \widehat{U}_n$, denoted by $\widehat{\varphi}_n(t)$, and the cf of the standard uniform distribution, given by

$$\varphi_U(t) = \frac{\sin(t) + i(1 - \cos(t))}{t}.$$

Meintanis [31] proposes the test

$$S_{n,a} = \int_{-\infty}^{\infty} |\varphi_U(t) - \widehat{\varphi}_n(t)|^2 w(t) dt.$$

Upon setting $w(t) = e^{-a|t|}$, $S_{n,a}$ simplifies to

$$S_{n,a} = \frac{1}{n} \sum_{j,k=1}^n \frac{2a}{(\widehat{U}_j - \widehat{U}_k)^2 + a^2} + 2n \left[2 \tan^{-1} \left(\frac{1}{a} \right) - a \log \left(1 + \frac{1}{a^2} \right) \right] - 4 \sum_{j=1}^n \left[\tan^{-1} \left(\frac{\widehat{U}_j}{a} \right) + \tan^{-1} \left(\frac{1 - \widehat{U}_j}{a} \right) \right].$$

The test rejects the null hypothesis for large values of the test statistic. Although Meintanis [31] does not explicitly use the resulting statistic to test for the Pareto distribution, it is demonstrated that this test is competitive when testing for the gamma, inverse Gaussian, and normal distributions. Meintanis et al. [34] considers the multivariate version of this class of tests and derives the limiting null distribution and also show that it is consistent against fixed alternatives.

2.6 A test based on the Mellin transform

Meintanis [32] introduces a test based on the moments of the reciprocal of the random variable X . If X follows a Pareto distribution, then $E(X^t)$, $t > 0$, only exists when $t < \beta$. On the other hand, the Mellin transform of X , given by

$$M(t) = E(X^{-t}), t > 0,$$

exists for all $t > 0$ if X is a Pareto random variable. Given an observed sample, the empirical Mellin transform is defined to be

$$\widehat{M}_n(t) = \frac{1}{n} \sum_{j=1}^n X_j^{-t}.$$

If X is a $P(\beta, \sigma)$ random variable, then $M(t)$ satisfies [see, e.g.,32]

$$D(t) = (\beta + t)\sigma^t M(t) - \beta = 0, \quad t > 0.$$

Based on a random sample, $D(t)$ can be estimated by

$$D_n(t) = (\widehat{\beta}_n + t)\widehat{M}_n(t) - \widehat{\beta}_n.$$

Meintanis [32] proposes a weighted L^2 distance between $D_n(t)$ and 0 as test statistic;

$$G_{n,a} = n \int_0^\infty D_n^2(t)w(t)dt,$$

where $w(t)$ is a suitable weight function, depending on a user-defined parameter a . After some algebra $G_{n,a}$ simplifies to

$$G_{n,a} = \frac{1}{n} \left[(\widehat{\beta}_n + 1)^2 \sum_{j,k=1}^n I_w^{(0)}(X_j X_k) + \sum_{j,k=1}^n I_w^{(2)}(X_j X_k) + 2(\widehat{\beta}_n + 1) \sum_{j,k=1}^n I_w^{(1)}(X_j X_k) \right] + \widehat{\beta}_n \left[n\widehat{\beta}_n I_w^{(0)}(1) - 2(\widehat{\beta}_n + 1) \sum_{j=1}^n I_w^{(0)}(X_j) - 2 \sum_{j=1}^n I_w^{(1)}(X_j) \right],$$

where

$$I_w^{(m)}(t) = \int_0^\infty (t - 1)^m \frac{1}{x^t} w(t)dt, \quad m = 0, 1, 2.$$

Choosing $w(t) = e^{-at}$, one has

$$I_a^{(0)}(t) = (a + \log t)^{-1},$$

$$I_a^{(1)}(t) = \frac{1 - a - \log t}{(a + \log t)^2},$$

and

$$I_a^{(2)}(t) = \frac{2 - 2a + a^2 + 2(a - 1) \log t + \log^2 t}{(a + \log t)^3},$$

culminating in an easily calculable test statistic.

The test rejects the null hypothesis for large values of the test statistic. Meintanis [32] proves the consistency of the test against fixed alternatives and uses a Monte Carlo study to demonstrate that the power performance of the test compares favourably to that of the classical goodness-of-fit tests.

2.7 A test based on an inequality curve

Let X be a positive random variable with distribution function F and finite mean μ . Let $L(p) = Q(F^{-1}(p))$, with

$$F^{-1}(p) = \inf\{x : F(x) \geq p\},$$

the generalised inverse of F and

$$Q(x) = \frac{1}{\mu} \int_0^x t dF(t),$$

the first incomplete moment of X . Using this notation, the inequality curve $\lambda(p)$, $p \in (0, 1)$ is defined to be [see, e.g., 53]

$$\lambda(p) = 1 - \frac{\log(1 - L(p))}{\log(1 - p)}.$$

Taufer et al. [49] proposes a test based on the constant inequality curve exhibited by the $P(\beta, \sigma)$ distribution for some $\sigma > 0$. Taufer et al. [49] proves the following characterisation for the Pareto distribution based on $\lambda(p)$.

Theorem 2.1 *The inequality curve $\lambda(p)$ is equal to $\frac{1}{\beta}$ over all values of p , $p \in (0, 1)$ if, and only if, F is the Pareto distribution function, $F_{\beta, \sigma}$.*

In order to use this characterisation to develop goodness-of-fit tests, Taufer et al. [49] uses the following approach. Defining the empirical version of $Q(x)$ as

$$Q_n(x) = \frac{\sum_{j=1}^n X_j I(X_j \leq x)}{\sum_{j=1}^n X_j},$$

the estimator for $L(p)$ becomes

$$L_n(p) = Q_n(F_n^{-1}(p)) = \frac{\sum_{j=1}^i X_{(j)}}{\sum_{k=1}^n X_{(k)}}, \quad \frac{i-1}{n} \leq p \leq \frac{i}{n}, \quad i = 1, 2, \dots, n,$$

where $F_n^{-1}(p) = \inf\{x : F_n(x) \leq p\}$. Finally, an estimator for $\lambda(p)$ is given by

$$\hat{\lambda}_j = 1 - \frac{\log(1 - L_n(p_j))}{\log(1 - p_j)}, \quad j = 1, 2, \dots, n - \lfloor \sqrt{n} \rfloor, \quad p_j = \frac{j}{n}.$$

The choice $j = 1, 2, \dots, n - \lfloor \sqrt{n} \rfloor$ ensures that $\hat{\lambda}_j$ is a consistent estimator for λ [see49]. Setting $m = n - \lfloor \sqrt{n} \rfloor$, Theorem 2.1 states that under the null hypothesis, for any choice of $p_j, 0 < p_j < 1, j = 1, 2, \dots, m$, and $\beta > 1$, we have the linear equation

$$\lambda_j = \beta_0 + \beta_1 p_j,$$

with $\beta_0 = \frac{1}{\beta}$ and $\beta_1 = 0$. Now, based on the data X_1, \dots, X_n , we can obtain estimators for β_0 and β_1 from the regression

$$\hat{\lambda}_j = \beta_0 + \beta_1 p_j + \varepsilon_j,$$

where $\varepsilon_j = \hat{\lambda}_j - \lambda_j$.

The least squares estimators for β_0 and β_1 are given by

$$\hat{\beta}_0 = \frac{1}{m} \sum_{j=1}^m \hat{\lambda}_j \quad \text{and} \quad \hat{\beta}_1 = \sum_{j=1}^m \frac{\hat{\lambda}_j (p_j - \bar{p})}{S_p^2},$$

where $\bar{p} = \frac{m+1}{2n}$ and $S_p^2 = \frac{m(m^2-1)}{12n^2}$.

Testing the hypothesis in (1.2) is equivalent to testing the hypothesis

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_A : \beta_1 \neq 0,$$

where the null hypothesis is rejected for large values of $|\hat{\beta}_1|$. In a finite sample study, Taufer et al. [49] finds that this test is oversized in the case of small sample sizes ($n = 20$), but achieves the nominal significance level for larger samples ($n = 100$). The results indicate that the test compares favourably against the traditional tests in terms of estimated powers. Since the focus of the current research is on the finite sample performance of tests in the case of small samples, we do not include this test in the Monte Carlo study presented in Sect. 4.

2.8 Tests based on various characterisations of the Pareto distribution

A wide range of characterisations of the Pareto distribution is available and several have been used to develop goodness-of-fit tests. In what follows, we state these characterisations and discuss the associated test in each case. It should be noted that, although the tests below are equally useful in the situation where both parameters of the Pareto distribution are required to be estimated, the asymptotic theory was developed in the setting where the scale parameter is known. Furthermore, the majority of the tests are consistent. The exceptions are those tests employing an integral expression in which the integrand is a linear difference.

Each of the subsections below are dedicated to a characterisations and contains a brief discussion on the associated test.

2.8.1 Characterisation 1 [40]

Let X and Y be i.i.d. positive absolutely continuous random variables. The random variable X and $\max\left\{\frac{X}{Y}, \frac{Y}{X}\right\}$ have the same distribution if, and only if, X follows a Pareto distribution.

Obradović et al. [40] provides the proof for this characterisation and proposes two test statistics based on it. In order to specify these test statistics, denoted by

$$M_n(x) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n I \left\{ \max \left(\frac{X_i}{X_j}, \frac{X_j}{X_i} \right) \leq x \right\}, \quad x \geq 1,$$

the U -empirical distribution function of the random variable $\max\{X/Y, Y/X\}$. The test statistics are specified to be

$$T_n = \int_1^\infty [M_n(x) - F_n(x)] dF_n(x)$$

and

$$V_n = \sup_{x \geq 1} |M_n(x) - F_n(x)|.$$

Both of these tests reject the null hypothesis for large values of the test statistics. Obradović et al. [40] calculates Bahadur efficiencies for selected alternative distributions and also determines some of the locally optimal alternatives. The mentioned paper also derives the null distribution of T_n and shows that $\sqrt{n}T_n$ converges to a centered normal random variable with variance $\frac{5}{108}$. A limited Monte Carlo study shows that the tests T_n and V_n are competitive against the traditional KS_n and CM_n tests.

2.8.2 Characterisation 2 [6]

Let X, X_1, \dots, X_n be i.i.d. positive absolutely continuous random variables from some distribution function F . The random variables $\sqrt[m]{X}$ and $\min\{X_1, \dots, X_m\}$ have the same distribution if, and only if, F is the Pareto distribution, for all integers $2 \leq m \leq n$.

Using m as a tuning parameter, Allison et al. [6] proposes three classes of tests for the Pareto distribution based on the characterisation above. The test statistics used are discrepancy measures between the empirical distribution of $\min\{X_1, \dots, X_m\}$ and the V -empirical distribution of $\sqrt[m]{X}$, defined as

$$\Delta_{n,m}(x) = \frac{1}{n} \sum_{j=1}^n I \left\{ X_j^{\frac{1}{m}} \leq x \right\} - \frac{1}{n^m} \sum_{j_1, \dots, j_m=1}^n I \{ \min(X_{j_1}, \dots, X_{j_m}) \leq x \}.$$

Based on $\Delta_{n,m}$, the authors propose the following test statistics

$$\begin{aligned}
 I_{n,m} &= \int_1^\infty \Delta_{n,m}(x) dF_n(x), \\
 K_{n,m} &= \sup_{x \geq 1} |\Delta_{n,m}(x)|, \\
 M_{n,m} &= \int_1^\infty \Delta_{n,m}^2(x) dF_n(x).
 \end{aligned}$$

$K_{n,m}$ and $M_{n,m}$ reject the null hypothesis for large values of the test statistics, while $I_{n,m}$ rejects for large values of $|I_{n,m}|$. Allison et al. [6] derive the limiting null distribution of all three test statistics. Upon calculating and comparing the Bahadur efficiencies, Allison et al. [6] found that the test $I_{n,m}$ has the best performance among the three in terms of local efficiency. This result is reinforced by a finite sample power study which results in the recommendation of choosing $I_{n,m}$ with $m = 2$.

Where Allison et al. [6] used empirical distribution functions to construct their tests, Ndwandwe et al. [37] propose test statistics that instead utilise empirical versions of the characteristic function. To this end, let

$$\phi_m(t) = E \left[e^{itX^{1/m}} \right] \quad \text{and} \quad \xi_m(t) = E \left[e^{it \min(X_1, \dots, X_m)} \right]$$

be the characteristic functions of $X^{1/m}$ and $\min(X_1, \dots, X_m)$, respectively. Denote the empirical versions of ϕ_m and ξ_m by

$$\phi_{n,m}(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j^{1/m}}$$

and

$$\xi_{n,m}(t) = \frac{1}{n^m} \sum_{k_1=1}^n \dots \sum_{k_m=1}^n e^{it \min(X_{k_1}, \dots, X_{k_m})}.$$

The characterisation implies that, for all $t \in \mathbb{R}$ and $0 \leq m \leq n$, $\phi_m(t) = \xi_m(t)$ if, and only if, $X \sim P(\beta)$ for some $\beta > 0$. As is usually the case in characteristic function based tests, Ndwandwe et al. [37] propose as test statistic that is a weighted L^2 distance between $\phi_{n,m}$ and $\xi_{n,m}$:

$$L_{n,m,a} = n \int_{-\infty}^\infty |\phi_{n,m}(t) - \xi_{n,m}(t)|^2 w_a(t) dt,$$

where $w_a(t)$ is a weight function (see Sect. 2.5 for some more detail on the weight function). Setting $w_a(t) = e^{-at^2}$, the test statistic $L_{n,m,a}$ simplifies to

$$\begin{aligned}
 L_{n,m,a} &= \frac{1}{n} \sqrt{\frac{\pi}{a}} \sum_{j=1}^n \sum_{k=1}^n \left\{ \exp \left[\frac{-\left(X_j^{1/m} - X_k^{1/m}\right)^2}{4a} \right] \right. \\
 &\quad \left. - 2nv_{j,m} \exp \left[\frac{-\left(X_j - X_k^{1/m}\right)^2}{4a} \right] \right\}
 \end{aligned}$$

$$+ n^2 v_{j,m} v_{k,m} \exp \left[\frac{-(X_{(j)} - X_{(k)})^2}{4a} \right] \Bigg\},$$

where

$$v_{j,m} := \frac{1}{n^m} [(n - j + 1)^m - (n - j)^m].$$

The test rejects for large values of the test statistic. Ndwandwe et al. [37] comment on the limiting null distribution of the test statistic and also demonstrate that the test is consistent against a wide range of fixed alternative distributions. A Monte Carlo study also revealed that $L_{n,m,a}$ performed better in terms of empirical powers than the majority of the other tests that were evaluated. For implementing the test the authors recommend choosing the values of the tuning parameters as $m = 3$ and $a = 2$.

Remark: Ndwandwe et al. [37] also studied a test where $\xi_m(t)$ is estimated by the U -statistic

$$\psi_{n,m}(t) = \binom{n}{m}^{-1} \sum_{1 \leq k_1 < \dots < k_m \leq n} e^{it \min(X_{k_1}, \dots, X_{k_m})}.$$

However, this test was found to be less powerful than $L_{n,m,a}$, and will not be discussed further in this paper.

2.8.3 Characterisation 3 [46]

Obradović [39] uses the following special case of Rossberg’s characterisation of the Pareto distribution to construct a goodness-of-fit test:

Let $X_1, X_2,$ and X_3 be i.i.d. positive absolutely continuous random variables and denote the corresponding order statistics by $X_{(1)} \leq X_{(2)} \leq X_{(3)}$. If $X_{(2)}/X_{(1)}$ and $\min(X_1, X_2)$ are identically distributed, then X_1 follows a Pareto distribution.

In order to base a test on this characterisation, Obradović [39] suggests estimating the distribution of $X_{(2)}/X_{(1)}$ by

$$G_n(x) = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n I\{\text{median}(X_i, X_j, X_k) / \min(X_i, X_j, X_k) \leq x\}, \quad x \geq 1,$$

and the distribution of $\min(X_1, X_2)$ by

$$H_n(x) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n I\{\min(X_i, X_j) \leq x\}, \quad x \geq 1.$$

Tests can be based on the discrepancy between G_n and H_n ; Obradović [39] proposes the test statistics

$$I_n^{[1]} = \int_1^\infty (G_n(x) - H_n(x)) dF_n(x),$$

and

$$D_n^{[1]} = \sup_{x \geq 1} |G_n(x) - H_n(x)|.$$

Both tests reject the null hypothesis for large values of the test statistics. Bahadur efficiencies for these tests are presented in Obradović [39] where the results show that, while no test

outperforms all others, each test is found to be locally optimal against certain classes of alternatives. Obradović [39] also shows that the asymptotic null distribution of $\sqrt{n}I_n^{[1]}$ is normal with mean 0 and variance $\frac{52}{1125}$.

2.8.4 Characterisation 4 [39]

In addition to the tests above, Obradović [39] also proposes tests for the Pareto distribution based on the following characterisation which is linked to a characterisation of the exponential distribution due to Ahsanullah [1].

Let X_1, X_2 and X_3 be i.i.d. positive absolutely continuous random variables with strictly monotone distribution function and monotonically increasing or decreasing hazard function and denote the order statistics by $X_{(1)} \leq X_{(2)} \leq X_{(3)}$. The random variable $X_{(3)}/X_{(2)}$ and $X_{(2)}/X_{(1)}$ have the same distribution if, and only if, the distribution of X follows a Pareto distribution.

The test statistics that Obradović [39] proposes based on this characterisation are

$$I_n^{[2]} = \int_1^\infty (J_n(x) - K_n(x)) dF_n(x),$$

and

$$D_n^{[2]} = \sup_{x \geq 1} |J_n(x) - K_n(x)|,$$

where

$$J_n(x) = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n I \{ \max(X_i, X_j, X_k) / \text{median}(X_i, X_j, X_k) \leq x \}, \quad x \geq 1,$$

and

$$K_n(x) = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n I \{ \text{median}(X_i, X_j, X_k) / \{ \min(X_i, X_j, X_k) \}^2 \leq x \}, \quad x \geq 1.$$

Both tests reject the null hypothesis for large values of the test statistic. As with the tests given in Section 2.8.3, Obradović [39] concludes that, while neither of the tests was dominant against all alternatives in terms of local efficiency, they are both locally optimal for certain classes of alternatives. It is again showed that $I_n^{[2]}$ has an asymptotic normally distributed null distribution.

2.8.5 Characterisation 5 [51]

For a fixed k , let X_1, \dots, X_k be i.i.d. non negative and bounded random variables having absolutely continuous distribution function F . The random variable X_1 and $X_{(k)}/X_{(k-1)}$ have the same distribution if, and only if, F is the Pareto distribution.

Volkova [51] provides a proof of this characterisation and derives two test statistics utilizing this characterisation:

$$I_n^{(k)} = \int_1^\infty [H_n(t) - F_n(t)] dF_n(t),$$

and

$$D_n^{(k)} = \sup_{x \geq 1} |H_n(t) - F_n(t)|,$$

where

$$H_n(t) = \binom{n}{k}^{-1} \sum_{1 \leq j_1 < \dots < j_k \leq n} I \{X_{(k, \{j_1, \dots, j_k\})} / X_{(k-1, \{j_1, \dots, j_k\})} < t\}, \quad t \geq 1,$$

and $X_{(k, \{j_1, \dots, j_k\})}$ denotes the k^{th} order statistic of the subsample X_{j_1}, \dots, X_{j_k} .

Both tests reject the null hypothesis for large values of the test statistics. In addition to deriving the conditions for local optimality of the tests, Volkova [51] also derives the null distribution of $I_n^{(k)}$ for $k = 3$ and $k = 4$. It is shown that $\sqrt{n}I_n^{(3)}$ and $\sqrt{n}I_n^{(4)}$ converge, under the null, to zero mean normal random variables with variances $\frac{11}{120}$ and $\frac{271}{2100}$, respectively. Due to its computationally expensive nature and the large number of tests already included, we opted to exclude this test from the Monte Carlo study.

2.9 Other tests

While we tried to consider the majority of tests available for the Pareto distribution, we will now only mention four others which are outside the scope of the paper. These are a weighted quantile correlation test by Csörgő and Szabó [19], a test based on Euclidean distances by Rizzo [45], a test based on spacings by Gulati and Shapiro [22] and a Kolmogorov-type test involving a sort of “memory-less” characterization of the Pareto distribution by Milošević and Obradović [35].

In addition, it should be noted that the Pareto distribution is closely linked to the exponential distribution; if $X \sim P(\beta, \sigma)$, then $Y = \log(X/\sigma)$ follows an exponential distribution with mean $1/\beta$. As a result, we can use this transformation, with estimated values of β and σ , in order to transform the data, and then we can use goodness-of-fit tests for the exponential distribution in order to test the hypothesis in (1.2). For an overview of tests for some of the multitudes of tests available for exponential, see Allison et al. [5].

3 Parameter and critical value estimation

In this section we discuss two popular methods for the estimation of the parameters of the Pareto distribution: the method of maximum likelihood as well as a method closely related to moment matching. The empirical results in Sect. 4 demonstrate that the choice of estimation method used has a profound effect on the powers achieved by the tests considered. As a result, it is necessary to discuss the procedures in some detail.

We consider parameter estimation in the setting where both β and σ are required to be estimated. In the testing scenario where σ is known, the estimated value of σ can simply be replaced by this known value.

For each estimation method we also discuss how the critical values are estimated.

3.1 Maximum likelihood estimators (MLEs)

In the case where both σ and β are unknown, the MLEs of σ and β are respectively given by

$$\widehat{\sigma}_n := \widehat{\sigma}(X_1, \dots, X_n) = X_{(1)},$$

and

$$\widehat{\beta}_n := \widehat{\beta}(X_1, \dots, X_n) = \frac{n}{\sum_{j=1}^n \log\left(\frac{X_j}{\widehat{\sigma}_n}\right)}.$$

Note that if we transform X_1, \dots, X_n as follows:

$$Y_j = \left(\frac{X_j}{\widehat{\sigma}_n}\right)^{\widehat{\beta}_n}, \quad j = 1, \dots, n, \tag{3.1}$$

then

$$\widehat{\sigma}(Y_1, \dots, Y_n) = 1,$$

and

$$\begin{aligned} \widehat{\beta}(Y_1, \dots, Y_n) &= \frac{n}{\sum_{j=1}^n \log\left(\frac{Y_j}{Y_{(1)}}\right)} = \frac{n}{\widehat{\beta}(X_1, \dots, X_n) \sum_{j=1}^n \log\left(\frac{X_j}{X_{(1)}}\right)} \\ &= \frac{\widehat{\beta}(X_1, \dots, X_n)}{\widehat{\beta}(X_1, \dots, X_n)} = 1. \end{aligned}$$

As can be seen above, the transformation in (3.1) ensures that, when the Pareto distribution is fitted to Y_1, \dots, Y_n , the resulting parameter estimates are fixed at $\widehat{\sigma}_n = \widehat{\beta}_n = 1$. This enables us to approximate fixed critical values by Monte Carlo simulations not depending on $\widehat{\sigma}_n$ or $\widehat{\beta}_n$. As a result, the limit null distribution is independent of the values of σ and β if the data are transformed as in (3.1). This result essentially renders the critical values for tests for the Pareto distribution shape invariant in the case where estimation is performed using MLE.

It should be noted that, if the transformation in (3.1) is used, then the sample minimum is $Y_{(1)} = 1$. This leads to computational issues for several of the tests discussed above. Specifically, the calculation of AD_n, ZA_n, ZB_n and ZC_n break down. In order to circumvent these numerical problems, we set $Y_{(1)} = 1.0001$ when computing these test statistics.

Remark The test proposed in Taufer et al. [49], see Sect. 2.7, assumes that the mean of the Pareto distribution fitted to the transformed values is finite. Let $\widehat{\mu}_n := \widehat{\beta}_n / (\widehat{\beta}_n - 1)$ denote the mean of the fitted Pareto distribution; $\widehat{\mu}_n$ is finite if, and only if, $\widehat{\beta}_n > 1$. As a result, the transformation in (3.1) leads to numerical problems with the implementation of this test. In order to obtain critical values for this test, we recommend using the transformation $Y_j = \left(\frac{X_j}{\widehat{\sigma}_n}\right)^{\widehat{\beta}_n/2}, j = 1, \dots, n$, which results in $\widehat{\mu} := 2$.

3.2 Adjusted method of moments estimators (MMEs)

The traditional implementation of the method of moments requires that both the mean and the variance of the distribution be finite. In the case of the Pareto distribution, this implies that $\beta > 2$. As a result, the traditional method of moments estimators are not consistent when estimating the parameters of a $P(\beta, \sigma)$ distribution when $\beta < 2$.

A partial solution to the problem explained above is found when using the so-called adjusted method of moments estimators proposed in Quandt [43]. Instead of choosing parameter estimates so as to equate the first two population moments to the first two sample moments, Quandt [43] equates the first population and sample moments as well as equating the observed minimum to the expected value of the sample minimum. The resulting estimators are

$$\tilde{\beta}_n := \tilde{\beta}(X_1, \dots, X_n) = \frac{n\bar{X} - X_{(1)}}{n(\bar{X} - X_{(1)})},$$

and

$$\tilde{\sigma}_n := \tilde{\sigma}(X_1, \dots, X_n) = \frac{\bar{X}\tilde{\beta}_n - \bar{X}}{\tilde{\beta}_n}.$$

Note that this method only requires the assumption that the population mean is finite, meaning that we assume only that $\beta > 1$. When analysing a data set in practice, we recommend rather using the MLE in cases where the MME for β is close to 1.

Unlike the case of maximum likelihood, we are unable to obtain fixed critical values; the critical values are functions of the estimated shape parameter $\tilde{\beta}_n$. We provide the following bootstrap algorithm for the estimation of critical values.

1. Based on data X_1, \dots, X_n , estimate β and σ by $\tilde{\beta}_n$ and $\tilde{\sigma}_n$, respectively.
2. Obtain a parametric bootstrap sample X_1^*, \dots, X_n^* by sampling independently from $F_{\tilde{\beta}_n, \tilde{\sigma}_n}$.
3. Calculate $\tilde{\beta}_n^* = \tilde{\beta}(X_1^*, \dots, X_n^*)$, $\tilde{\sigma}_n^* = \tilde{\sigma}(X_1^*, \dots, X_n^*)$, and the value of the test statistic say $S^* = S(X_1^*, \dots, X_n^*)$.
4. Repeat steps 2 and 3 B times to obtain S_1^*, \dots, S_B^* and obtain the order statistics $S_{(1)}^* \leq \dots \leq S_{(B)}^*$.
5. The estimated critical value at a $\alpha \times 100\%$ significance level is $\hat{C}_n = S_{(B[1-\alpha])}^*$, where $[x]$ denotes the floor of x .

We now turn our attention to the numerical powers of the tests obtained using the two estimation methods discussed above.

3.3 Other estimation methods

While the numerical result presented later in this paper are based on the MLE and MME estimators mentioned above, one can also consider alternative methods of estimation. For the sake of completeness, we note that there are many other alternative methods of obtaining these estimators such as the L -moment estimator [23], methods that involve minimising some distance criterion between distribution functions [10, 11, 13, 42, 52], as well as similar minimum distance-based method related to ϕ -divergence [8].

4 Monte Carlo results

In this section we present a Monte Carlo study in which we examine the empirical sizes as well as the empirical powers achieved by the various tests discussed in Sect. 2. Section 4.1 details the simulation setting used, including the alternative distributions considered, while Sect. 4.2 shows the numerical results obtained together with a discussion and comparison of these results.

Table 1 Summary of various choices of the alternative distributions

Alternative	Density function	Notation
Gamma	$\frac{1}{\Gamma(\theta)} (x - 1)^{\theta-1} e^{-(x-1)}$	$\Gamma(\theta)$
Weibull	$\theta(x - 1)^{\theta-1} \exp(-(x - 1)^\theta)$	$W(\theta)$
Log-normal	$\exp\left\{-\frac{1}{2}(\log(x - 1)/\theta)^2\right\} / \left\{\theta(x - 1)\sqrt{2\pi}\right\}$	$LN(\theta)$
Linear failure rate	$(1 + \theta(x - 1)) \exp(-(x - 1) - \theta(x - 1)^2/2)$	$LF(\theta)$
Beta exponential	$\theta e^{-(x-1)} (1 - e^{-(x-1)})^{\theta-1}$	$BE(\theta)$
Tilted Pareto	$\frac{1+\theta}{(x+\theta)^2}$	$TP(\theta)$
Dhillon	$\frac{\theta+1}{x+1} \exp\left\{-\log(x + 1)^{\theta+1}\right\} (\log(x + 1))^\theta$	$DH(\theta)$
Half-normal	$\frac{\sqrt{2}}{\theta\sqrt{\pi}} \exp\left(\frac{-(x-1)^2}{2\theta^2}\right)$	$HN(\theta)$

Table 2 Summary of power tables

Table	<i>n</i>	Estimation	Parameters	Table	<i>n</i>	Estimation	Parameters
3	20	MLE	1 parameter	10	30	MLE	1 Parameter
4	20	MME	1 parameter	11	30	MME	1 Parameter
5	20	MLE	2 parameters	12	30	MLE	2 Parameters
6	20	MME	2 parameters	13	30	MME	2 Parameters

4.1 Simulation setting

We consider four different Monte Carlo settings. In the first two of these we consider the case in which only the shape parameter of the Pareto distribution requires estimation, while both the shape and scale parameters are estimated in the third and fourth settings. Furthermore, in the first and third settings we use maximum likelihood estimation in order to obtain parameter estimates, while the adjusted method of moments is used in the second and fourth settings.

We calculate empirical sizes and powers for samples of size $n = 20$ and $n = 30$. The empirical powers are calculated against the range of alternative distributions given in Table 1. Traditionally, these alternatives have support $(0, \infty)$. In order to ensure that the simulated data have the same support as the Pareto distribution, these alternatives are shifted by 1.

The powers obtained against these alternative distributions are displayed in Table 3, 4, 5, 6 and 10, 11, 12, 13. The highest two powers in each row (including ties) are highlighted. For ease of reference, the entries in Table 2 below gives a brief summary of the settings used in these power tables with respect to the sample size, estimation method and the number of parameters estimated.

Where MLE is used for parameter estimation, we approximate critical values using 100,000 Monte Carlo replications. Thereafter we generate 10,000 samples from each alternative distribution considered and we calculate the empirical powers as the percentages (rounded to the nearest integers) of these samples that resulted in the rejection of H_0 in (1.2). In the case where MME is used in order to perform parameter estimation we are unable to calculate fixed critical values. As a result, we use the warp-speed bootstrap method proposed in Giacomini et al. [21] in order to arrive at empirical critical values and powers in

this case. This technique entails the following: each Monte Carlo sample is not subject to a large number of time-consuming bootstrap replications since only one bootstrap sample is taken for each Monte Carlo replication. The warp-speed method has been used in numerous studies to evaluate the power performances of the tests, see, e.g. Cockeran et al. [17] as well as Ndwandwe et al. [36]. In this setting, we make use of 50,000 Monte Carlo samples (which then also imply 50,000 bootstrap replications). All calculations were done in R v4.2.2 [44].

A final remark regarding the numerical powers associated with the tests based on characterisation of the Pareto distribution is in order. These tests, see Sect. 2.8, are typically much more computationally expensive to evaluate than the other tests considered. As a result, it is simply not feasible to calculate numerical powers for these tests using the warp-speed bootstrap. However, note that these tests do not require parameter estimation (the test statistics are not functions of the estimated parameter values) and we simply treat these tests as if parameter estimation is performed using MLE. Consequently, we are able to, once more, compute fixed critical values. The numerical powers reported in the tables are obtained based on these fixed critical values. In order to appreciate the large difference between the computational times required for the various tests, see Table 9 in Sect. 5.

4.2 Simulation results and discussion

We begin our discussion of the performance of the tests with the remark that the powers generally increase with sample size; meaning the powers associated with samples of size $n = 30$ are higher than those associated with $n = 20$. In the discussion below, we consider the results obtained using samples of size $n = 20$, before turning our attention to the cases where $n = 30$.

Before turning our attention to a general discussion of the empirical power results or a comparison between the results associated with the various settings considered, we discuss the results obtained using maximum likelihood estimation. The numerical results shown in Tables 3 and 5 indicate that each of the tests closely maintains the specified nominal significance level of 5%. When considering the numerical powers, it is clear that the DK_n test generally outperforms all of the competing tests against the majority of alternatives considered. In the case of samples of size $n = 20$, this impressive power performance is followed closely by that of $KL_{n,10}$, which provides power close to those achieved by DK_n . In the case where $n = 30$, DK_n still produces the highest powers, followed by ZC_n .

We now turn our attention to the results in Tables 4 and 6 (as well as those in Tables 11 and 13), obtained when using the method of moments to perform parameter estimation. The tests generally fail to achieve the specified nominal significance level of 5% against the $P(1, 1)$ distribution. Of course, the tests for which parameter estimation are not required (T_n , $I_{n,2,2}$, $I_{n,3}$, $I_n^{[1]}$ and $I_n^{[2]}$) do not exhibit this shortcoming. The general lack of adherence to the specified significance level can be ascribed to the fact that the first moment of the $P(1, 1)$ distribution does not exist. For the remaining Pareto distributions considered, all of which possess a finite first moment, the sizes of the tests closely coincide with the specified nominal significance level. The results presented in Table 4 indicate that the tests that generally exhibit the highest levels of statistical power are $G_{n,2}$, MA_n , and $L_{n,m,a}$ when $n = 20$. Turning our attention to the case where $n = 30$ (see Tables 12 and 13), we see that the most powerful test is still $G_{n,2}$ followed by MA_n . However, in this case the performances of these tests is closely followed by that of ZB_n .

One striking feature of the reported empirical results is the noticeably poor performance of the DK_n test when using MME. While this test achieved the highest power against the

majority of the alternatives when employing MLEs for parameter estimation, it frequently produces the lowest power when using the MMEs. This illustrates the importance of the choice of the estimation method when testing the assumption of the Pareto distribution.

When considering the effects of sample size and the number of parameters to be estimated, the powers are influenced in the expected way. An increase in sample size generally results in an increase in empirical power, while the settings in which a single parameter requires estimation generally produce higher numerical powers than is the case for settings in which both parameters are estimated. When comparing the results obtained using MLE and MME, we see that one estimation method does not increase the powers associated with all of the tests. That is, changing the estimation method used from MLE to MME results in an increase in the powers of some of the tests while the other tests experience a decrease in power. The most striking example is the DK_n test which shows excellent powers when using MLE while exhibiting dismal powers when using MME.

5 Practical application

We now employ the various tests considered in order to ascertain whether or not an observed data set is compatible with the assumption of being realised from a Pareto distribution. The data set is comprised of the monetary expenses incurred as a result of wind related catastrophes in 40 separate instances during 1977, rounded to the nearest million US dollars. The data are provided in Table 7.

The rounding of the recorded values in Table 7 causes unrealistic clustering in the data which may lead to problems when testing for the Pareto distribution. In order to circumvent the associated problems, we use the de-grouping algorithm discussed in Allison et al. [6] as well as Brazauskas and Serfling [14]. This algorithm replaces the values in each group of tied observations with the expected value of the order statistics of the uniform distribution with the same range. That is, if one observes k identical integer values, x , in an interval ($l = x - 1/2, u = x + 1/2$), we replace these values by

$$\left(\frac{k + 1 - j}{k + 1}\right)l + \left(\frac{j}{k + 1}\right)u,$$

for $j \in \{1, \dots, k\}$. We emphasise that this de-grouping algorithm does not change the mean of the data set. The de-grouped data can be found in Table 8.

When testing the hypothesis of the Pareto distribution for the data set, we consider each of the four settings used in the Monte Carlo study presented in Sect. 4. That is, we test the hypothesis in both the one and two parameter cases and we use MLE as well as MME in order to arrive at parameter estimates. Note that, when fitting a one parameter distribution, the support of the distribution is assumed known. Since the observed minimum is rounded to 2, we conclude that no value less than 1.5 is possible. As a result, we fix $\sigma = 1.5$ in the cases where the one parameter distribution is considered. No such assumption is necessary for the two parameter case; in this case, the value of σ is simply estimated from the data.

When assuming that $\sigma = 1.5$, the MLE of β is calculated to be $\hat{\beta}_n = 0.764$ while the MME is $\tilde{\beta}_n = 1.194$. In the case where both β and σ are estimated; the MLEs are $\hat{\beta}_n = 0.796$ and $\hat{\sigma}_n = 1.053$. The corresponding MMEs are $\tilde{\beta}_n = 1.202$ and $\tilde{\sigma}_n = 1.031$. The empirical p -values associated with each of these four instances are shown in Table 9. When using MMEs, the empirical p -values are obtained via a parametric bootstrap procedure employing a modified version of the algorithm presented in Sect. 3.2. In the case of MLEs, p -values are approximated using a Monte Carlo procedure; for details, see the discussion in Sect. 3.1. In

Table 3 Numerical powers when estimating 1 parameter using MLE with $n = 20$

	KS_n	CM_n	AD_n	MA_n	ZA_n	ZB_n	ZC_n	$KL_{n,1}$	$KL_{n,10}$	DK_n	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(2,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(5,1)$	5	5	6	5	5	5	5	5	5	5	6	5	6	6	5	5	5	5	5	5
$P(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	22	24	47	18	44	50	3	17	1	1	25	23	44	23	25	30	40	44	6	11
$\Gamma(0.8)$	10	11	10	12	12	13	15	9	20	20	10	10	7	9	12	8	7	7	13	16
$\Gamma(1)$	25	30	25	31	29	28	35	17	46	49	29	32	20	31	35	25	20	16	27	42
$\Gamma(1.2)$	45	56	51	56	55	54	58	30	71	76	55	59	45	60	63	50	43	36	43	69
$W(0.5)$	35	38	62	30	58	63	4	26	1	1	39	36	59	36	36	43	56	58	9	19
$W(0.8)$	9	9	10	11	10	12	14	8	17	17	9	9	7	8	10	8	7	7	12	14
$W(1.2)$	50	62	57	62	61	60	62	34	75	80	61	65	50	65	69	56	48	40	46	74
$W(1.5)$	82	92	90	92	92	91	89	66	96	98	91	93	85	94	95	88	83	77	67	96
$LN(1)$	56	66	64	55	80	78	39	37	59	65	66	70	75	73	73	64	66	62	19	73
$LN(1.2)$	26	30	27	23	43	39	21	19	34	37	30	33	35	35	37	29	28	27	11	37
$LN(1.5)$	7	7	6	7	10	9	9	9	13	13	8	8	7	8	9	7	6	6	7	11
$LN(2.5)$	27	30	43	24	28	31	3	14	1	0	30	28	45	30	27	31	42	41	10	13

Table 3 continued

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.2)	32	39	34	41	36	35	47	22	58	62	38	40	26	39	44	32	26	20	38	53
<i>LFR</i> (0.5)	39	48	42	50	45	44	55	27	67	70	47	49	32	48	53	40	32	25	46	61
<i>LFR</i> (0.8)	43	53	47	55	50	48	61	29	72	76	52	55	36	53	59	46	36	29	50	66
<i>LFR</i> (1)	44	55	49	57	51	50	62	31	73	77	53	56	38	55	61	47	38	31	51	69
<i>BE</i> (0.5)	19	20	42	16	40	47	3	17	2	2	22	18	39	18	19	25	35	39	5	10
<i>BE</i> (0.8)	11	12	11	14	12	14	17	10	24	24	12	11	7	10	13	9	7	7	15	19
<i>BE</i> (1)	24	30	25	31	28	27	35	17	46	49	29	31	20	30	35	25	19	16	28	41
<i>BE</i> (1.5)	67	79	76	78	80	80	73	48	87	91	79	83	72	84	85	75	69	62	51	88
<i>TP</i> (0.5)	7	8	6	7	8	8	8	7	11	11	8	7	6	7	9	7	6	6	6	11
<i>TP</i> (1, 1)	12	13	10	12	14	12	12	8	18	19	13	13	10	13	16	11	10	9	10	18
<i>TP</i> (2, 1)	22	26	22	23	25	24	21	12	30	34	26	27	20	26	30	23	20	18	16	34
<i>TP</i> (3)	32	37	32	34	35	34	29	17	41	47	37	39	30	38	43	33	29	26	22	47
<i>D</i> (0.2)	12	13	10	11	16	14	11	9	17	19	13	14	12	14	16	12	10	10	8	18
<i>D</i> (0.4)	30	35	31	30	40	37	26	18	39	44	35	37	34	39	41	32	31	28	17	44
<i>D</i> (0.6)	52	62	58	54	66	65	44	32	62	69	62	66	60	67	68	58	56	51	26	70
<i>D</i> (0.8)	70	81	79	76	84	83	63	48	80	87	81	85	79	86	86	78	75	70	37	88
<i>HN</i> (0.8)	49	59	54	62	55	54	67	34	77	80	58	61	41	59	65	52	41	33	56	72
<i>HN</i> (1)	54	65	59	68	60	59	73	39	83	85	63	66	45	64	70	57	45	37	62	78

Table 4 Numerical powers when estimating 1 parameter using MME with $n = 20$

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1,1)$	9	10	12	15	10	11	14	6	4	1	10	11	6	11	5	5	5	5	5	13
$P(2,1)$	5	5	5	5	5	5	4	5	5	2	5	5	5	5	5	5	5	5	5	5
$P(5,1)$	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	4
$P(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	4
$\Gamma(0.5)$	15	15	37	11	38	47	2	17	1	0	17	13	16	5	25	30	40	44	6	3
$\Gamma(0.8)$	15	17	16	20	11	14	9	9	17	3	16	17	12	20	12	8	7	7	13	20
$\Gamma(1)$	38	44	42	47	33	35	17	18	41	8	43	45	37	51	35	25	20	16	27	50
$\Gamma(1.2)$	65	74	73	75	64	66	26	35	67	16	72	76	69	80	63	50	43	36	43	77
$W(0.5)$	16	15	43	11	45	56	2	26	0	0	19	11	26	6	36	43	56	58	9	15
$W(0.8)$	15	17	17	21	11	14	5	9	13	1	16	17	10	19	10	8	7	7	12	21
$W(1.2)$	66	75	73	76	65	68	38	37	73	29	73	77	72	80	69	56	48	40	46	80
$W(1.5)$	91	96	96	96	93	94	70	68	96	73	95	96	96	97	95	88	83	77	67	97
$LN(1)$	73	82	83	75	90	87	10	43	54	13	80	83	85	85	73	64	66	62	19	87
$LN(1.2)$	42	51	52	47	59	55	3	23	28	4	48	53	51	57	37	29	28	27	11	53
$LN(1.5)$	17	21	22	25	21	19	2	11	9	1	19	22	15	25	9	7	6	6	7	21
$LN(2.5)$	11	9	26	26	22	34	36	19	0	0	12	7	9	5	27	31	42	41	10	37
$LFR(0.2)$	45	53	50	57	40	43	28	23	55	15	52	55	47	59	44	32	26	20	38	58
$LFR(0.5)$	51	59	56	63	47	49	39	28	65	25	58	60	54	65	53	40	32	25	46	64

Table 4 continued

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.8)	54	63	59	66	50	53	45	30	70	32	61	64	59	68	59	46	36	29	50	68
<i>LFR</i> (1)	57	66	63	69	53	56	49	32	73	38	64	67	63	70	61	47	38	31	51	69
<i>BE</i> (0.5)	12	11	32	10	34	44	2	17	1	0	13	9	12	4	19	25	35	39	5	3
<i>BE</i> (0.8)	17	19	18	24	13	16	10	10	19	3	19	20	14	23	13	9	7	7	15	24
<i>BE</i> (1)	38	45	42	48	34	35	17	19	41	8	43	46	38	51	35	25	19	16	28	51
<i>BE</i> (1.5)	83	91	91	90	86	88	35	53	84	30	90	92	90	94	85	75	69	62	51	93
<i>TP</i> (0.5)	27	31	35	39	30	33	19	13	7	0	30	32	16	32	9	7	6	6	6	29
<i>TP</i> (1, 1)	46	53	58	62	52	56	25	22	11	0	51	54	29	54	16	11	10	9	10	44
<i>TP</i> (2, 1)	76	82	86	88	81	84	38	44	19	0	80	82	55	82	30	23	20	18	16	67
<i>TP</i> (3)	90	94	96	97	94	95	50	64	28	0	93	94	72	94	43	33	29	26	22	80
<i>D</i> (0.2)	26	31	32	32	29	27	4	12	13	2	30	32	24	34	16	12	10	10	8	33
<i>D</i> (0.4)	50	59	59	57	57	55	3	23	31	4	57	61	53	64	41	32	31	28	17	60
<i>D</i> (0.6)	73	82	82	79	80	79	8	40	55	9	80	83	79	86	68	58	56	51	26	85
<i>D</i> (0.8)	89	94	94	93	93	93	16	58	75	17	93	95	93	96	86	78	75	70	37	96
<i>HN</i> (0.8)	59	68	65	72	55	58	55	35	76	42	66	69	65	72	65	52	41	33	56	73
<i>HN</i> (1)	67	76	74	80	63	66	57	40	82	40	75	77	71	81	70	57	45	37	62	78

Table 5 Numerical powers when estimating 2 parameters using MLE with $n = 20$

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(2,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(5,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	23	25	39	19	38	37	2	26	2	2	25	23	37	23	26	27	33	28	9	18
$\Gamma(0.8)$	7	7	5	8	7	5	19	10	21	22	7	5	2	2	6	3	2	7	13	14
$\Gamma(1)$	15	15	5	18	9	3	38	13	41	44	15	13	0	7	15	7	3	14	26	31
$\Gamma(1.2)$	26	29	11	33	18	4	58	18	60	64	28	26	0	16	30	16	7	26	41	48
$W(0.5)$	35	38	54	30	52	51	3	36	2	1	39	36	52	35	37	39	47	41	13	27
$W(0.8)$	6	6	5	8	7	6	17	10	19	20	7	4	3	2	6	3	3	7	12	13
$W(1.2)$	28	32	13	36	20	4	61	19	63	68	32	29	0	18	33	18	8	29	44	52
$W(1.5)$	48	56	30	60	41	11	82	32	82	87	55	53	1	39	58	38	19	50	66	75
$LN(1)$	14	14	4	14	7	2	28	8	30	34	14	12	0	7	15	8	4	20	16	26
$LN(1.2)$	7	7	2	7	4	2	17	6	20	21	7	6	0	3	7	3	2	11	10	14
$LN(1.5)$	4	4	3	4	3	3	8	6	10	10	4	3	2	2	4	3	2	5	6	7
$LN(2.5)$	29	32	43	26	38	34	2	21	1	1	33	31	43	32	32	33	39	30	13	22
$LFR(0.2)$	20	22	9	26	14	4	50	17	53	57	21	18	0	11	22	11	4	19	37	40
$LFR(0.5)$	25	28	11	33	18	5	59	19	61	64	27	24	0	14	28	15	6	23	44	48

Table 5 continued

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.8)	28	32	13	37	21	5	64	21	66	69	31	28	0	17	32	17	7	26	48	53
<i>LFR</i> (1)	29	33	14	38	22	6	64	22	67	71	32	29	0	18	33	18	8	27	49	54
<i>BE</i> (0.5)	20	21	35	16	34	34	3	26	4	3	22	18	32	17	20	22	28	24	7	16
<i>BE</i> (0.8)	8	8	4	10	6	5	22	11	25	25	7	5	2	2	6	3	2	7	15	16
<i>BE</i> (1)	15	15	6	19	9	3	39	13	42	45	15	13	0	7	15	8	3	14	27	31
<i>BE</i> (1.5)	34	40	17	43	25	6	68	22	69	74	39	37	0	25	41	24	11	37	49	61
<i>TP</i> (0.5)	5	4	3	5	3	4	8	6	9	10	4	3	2	2	4	3	2	6	6	7
<i>TP</i> (1, 1)	7	6	2	7	4	3	15	6	16	17	6	5	1	3	6	3	2	8	9	12
<i>TP</i> (2, 1)	11	11	4	12	6	2	24	8	25	29	11	9	0	5	12	6	2	15	15	22
<i>TP</i> (3)	17	18	6	18	9	3	34	11	34	39	17	15	0	8	18	10	4	21	21	31
<i>D</i> (0.2)	5	5	2	5	3	3	12	6	14	15	5	4	1	2	5	3	2	8	8	10
<i>D</i> (0.4)	12	12	4	12	6	2	26	8	28	31	12	10	0	6	13	7	3	17	16	23
<i>D</i> (0.6)	20	22	8	21	12	3	40	11	41	47	22	20	0	12	23	13	6	27	24	38
<i>D</i> (0.8)	29	33	14	33	19	5	55	16	54	62	32	30	0	21	34	20	11	37	35	51
<i>HN</i> (0.8)	32	37	17	43	26	6	69	24	71	74	36	32	0	21	37	21	8	29	54	58
<i>HN</i> (1)	36	42	19	48	29	8	75	27	76	79	40	37	0	24	42	24	10	32	60	63

Table 6 Numerical powers when estimating 2 parameters using MME with $n = 20$

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1,1)$	9	11	12	15	13	14	14	6	3	1	10	11	6	11	5	5	5	5	5	9
$P(2,1)$	5	5	5	5	5	5	4	5	5	2	5	5	5	5	5	5	5	5	5	4
$P(5,1)$	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	3
$P(10,1)$	5	5	5	5	5	5	5	5	5	4	5	5	5	4	5	5	5	5	5	3
$\Gamma(0.5)$	10	10	15	8	10	11	2	26	2	0	10	8	9	4	26	27	33	28	9	3
$\Gamma(0.8)$	16	18	17	21	15	18	9	10	18	3	17	19	15	23	6	3	2	7	13	16
$\Gamma(1)$	34	40	39	43	36	40	15	13	36	7	39	42	36	47	15	7	3	14	26	30
$\Gamma(1.2)$	53	61	61	63	58	62	22	19	55	12	60	63	56	68	30	16	7	26	41	39
$W(0.5)$	12	10	18	10	14	16	1	37	1	0	12	8	12	5	37	39	47	41	13	17
$W(0.8)$	17	19	19	23	15	18	5	10	14	1	18	19	13	22	6	3	3	7	12	17
$W(1.2)$	53	61	61	63	58	63	31	19	59	20	60	64	60	68	33	18	8	29	44	38
$W(1.5)$	71	79	79	80	78	82	54	32	81	46	78	80	79	82	58	38	19	50	66	42
$LN(1)$	33	40	39	38	38	39	7	9	25	7	39	42	38	45	15	8	4	20	16	22
$LN(1.2)$	23	28	28	28	27	27	3	7	15	3	26	29	24	32	7	3	2	11	10	17
$LN(1.5)$	14	16	18	21	16	16	2	6	6	1	16	17	10	19	4	3	2	5	6	11
$LN(2.5)$	12	11	24	30	31	37	37	28	1	0	12	9	4	8	32	33	39	30	13	39
$LFR(0.2)$	41	48	48	51	45	50	25	16	49	12	47	50	45	56	22	11	4	19	37	34
$LFR(0.5)$	46	54	53	57	51	56	35	19	58	19	52	55	51	60	28	15	6	23	44	36

Table 6 continued

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.8)	48	57	56	60	54	60	41	20	63	25	55	59	56	63	32	17	7	26	48	38
<i>LFR</i> (1)	51	59	58	62	58	63	44	22	66	29	58	61	58	64	33	18	8	27	49	38
<i>BE</i> (0.5)	8	8	13	8	8	10	3	25	3	0	9	6	7	4	20	22	28	24	7	4
<i>BE</i> (0.8)	18	21	21	25	18	21	10	11	21	3	20	22	17	26	6	3	2	7	15	19
<i>BE</i> (1)	35	41	40	43	36	41	16	13	37	6	39	42	36	47	15	8	3	14	27	30
<i>BE</i> (1.5)	63	72	72	72	69	72	28	23	64	20	71	74	69	77	41	24	11	37	49	40
<i>TP</i> (0.5)	25	29	33	37	36	36	19	10	6	0	28	30	15	30	4	3	2	6	6	17
<i>TP</i> (1, 1)	42	48	54	57	56	57	24	16	9	0	46	49	25	49	6	3	2	8	9	22
<i>TP</i> (2, 1)	68	75	80	82	82	82	34	32	15	0	74	76	47	75	12	6	2	15	15	28
<i>TP</i> (3)	82	88	91	92	92	92	44	47	20	0	86	88	62	88	18	10	4	21	21	28
<i>D</i> (0.2)	21	24	25	25	24	23	4	7	10	1	23	25	18	27	5	3	2	8	8	16
<i>D</i> (0.4)	35	41	42	41	39	39	3	10	21	3	40	43	34	46	13	7	3	17	16	24
<i>D</i> (0.6)	49	57	57	56	55	55	7	14	34	6	56	59	51	62	23	13	6	27	24	31
<i>D</i> (0.8)	61	69	68	67	66	67	13	19	48	11	67	70	64	73	34	20	11	37	35	35
<i>HN</i> (0.8)	53	62	60	65	60	66	49	23	69	33	60	63	61	67	37	21	8	29	54	39
<i>HN</i> (1)	60	69	69	73	67	73	51	27	75	31	68	71	66	75	42	24	10	32	60	43

Table 7 Wind catastrophes original data set

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 8 Wind catastrophes de-grouped data set

1.58	1.65	1.73	1.81	1.88	1.96	2.04	2.12
2.19	2.27	2.35	2.42	2.70	2.90	3.10	3.30
3.75	4.00	4.25	4.70	4.90	5.10	5.30	5.70
5.90	6.10	6.30	7.83	8.17	9.00	15.00	17.00
22.00	23.00	23.83	24.17	25.00	27.00	32.00	43.00

Table 9 p -values for the wind catastrophes data set

Test	MLE 1	MME 1	MLE 2	MME 2	Time
KS_n	0.509	0.013	0.547	0.013	1
CM_n	0.271	0.004	0.403	0.004	1
AD_n	0.242	0.001	0.000	0.001	1
MA_n	0.114	0.000	0.153	0.000	1
ZA_n	0.075	0.006	0.000	0.002	1
ZB_n	0.078	0.002	0.000	0.001	1
ZC_n	0.009	0.226	0.000	0.229	1
$KL_{n,1}$	0.395	0.107	0.460	0.152	2
$KL_{n,10}$	0.009	0.009	0.010	0.012	2
DK_n	0.013	0.537	0.014	0.511	7
$S_{n,0.5}$	0.299	0.006	0.412	0.006	2
$S_{n,1}$	0.171	0.003	0.291	0.003	2
$G_{n,0.5}$	0.217	0.023	0.604	0.034	12
$G_{n,2}$	0.133	0.002	0.278	0.002	13
T_n	0.265	0.266	0.616	0.724	138
$I_{n,2}$	0.632	0.632	0.917	0.900	112
$I_{n,3}$	0.425	0.425	0.912	0.863	4746
$I_n^{[1]}$	0.301	0.303	0.354	0.968	21, 685
$I_n^{[2]}$	0.051	0.051	0.053	0.051	36, 908
$L_{n,3,2}$	0.179	0.026	0.289	0.312	22

both cases 10 000 samples are generated from the Pareto distribution. The results associated with each of the tests considered in Sect. 4 are shown. The column headings used indicate the estimation method used as well as the number of parameters estimated. The final column in the table shows the time required in order to arrive at the reported p -values in seconds. The reported results are obtained using a 64 bit Windows 10 operating system with an AMD Ryzen 7 5800U CPU @ 1.90 GHz with 8 GB of RAM. Note the substantial computational times associated with the tests based on characterisations of the Pareto distribution.

In the interpretation of the p -values, we use a nominal significance level of 5%. When assuming a known σ of 1.5 and using MLE to estimate the value of β , the majority of the

test statistics do not reject the null hypothesis. The exceptions, which reject the hypothesis of the Pareto distribution, are ZC_n , $KL_{n,10}$, and DK_n . In the case where parameter estimation is performed using the MME, the situation is reversed and 12 of the 20 tests considered reject the Pareto assumption. The tests not rejecting the null hypothesis in this case are ZC_n , $KL_{n,1}$, DK_n , T_n , $I_{n,2}$, $I_{n,3}$, $I_n^{[1]}$ and $I_n^{[2]}$.

We now turn our attention to the case where both β and σ require estimation. We start by considering the results obtained using MLE. In this case the majority of the tests do not reject the null hypothesis. Only AD_n , ZA_n , ZB_n , ZC_n , $KL_{n,10}$ and DK_n reject the null hypothesis while the remaining 14 tests do not reject the null hypothesis. Finally, when considering the results associated with MME, we observe that the majority of the tests reject the hypothesis of the Pareto distribution. The exceptions to this are the ZC_n , $KL_{n,1}$, DK_n , T_n , $I_{n,2}$, $I_{n,3}$, $I_n^{[1]}$, $I_n^{[2]}$ and $L_{n,3,2}$ tests.

When comparing the p -values associated with the practical example, some further remarks are in order. It should be noted that the estimated value of β is close to 1 when using the MME, while a value of less than 1 is obtained when using MLE. This raises some doubt as to the assumption implicit in the MME that the first moment exists. As a result, we put more stock in the results obtained using the MLE than those obtained using MME. When using the MLE, in both the one and two parameter cases, the majority of the tests do not reject the Pareto assumption, providing evidence in favour of the null hypothesis. As a result, we conclude that the Pareto distribution is likely an appropriate model for the data considered.

6 Concluding remarks

The goal of this study is to review the existing goodness-of-fit tests for the Pareto type I distribution based on a wide range of characteristics of this distribution. Below we provide brief descriptions of these characteristics and the tests related to them. The tests based on the edf, commonly known as the traditional tests, are Kolmogorov-Smirnov (KS_n), Cramér-von Mises (CV_n), Anderson-Darling (AD_n) and modified Anderson-Darling (MA_n) tests. We also consider tests based on likelihood ratios. These tests are either weighted by some function of the edf (ZA_n and ZC_n) or by the distribution function under the null hypothesis with estimated parameters (ZB_n).

Next we consider the Hellinger distance ($M_{m,n}$) and Kullback-Leibler divergence ($KL_{n,m}$) tests which are based on the concept of entropy. Furthermore, we review tests based on phi-divergence. These tests are based on four distance measures; the Kullback-Leibler distance (DK_n), the Hellinger distance (DH_n), the Jeffreys divergence distance (DJ_n) as well as the Total variation distance (DT_n).

Although the Pareto distribution does not have a closed form expression for its characteristic function, we include a test, $S_{n,a}$, utilising the characteristic function of the uniform distribution. We also discuss a test involving the Mellin transform (G_n) as well as a test based on the fact that the Pareto distribution has a constant inequality curve (TS_n). Finally, we consider a number of tests utilising different characterisations of the Pareto distribution (T_n , $I_{n,2}$, $I_{n,3}$, $I_n^{[1]}$, $I_n^{[2]}$ and $L_{n,3,2}$).

For the Monte Carlo simulation, we consider eight different distributions (with various parameter settings) under the alternative hypothesis. Some of the tests utilised require parameter estimation. To this end, we consider the maximum likelihood estimators (MLE) and the adjusted method of moments estimators (MMEs). The power performance of the tests are

considered in the case where only the shape parameter of the Pareto distribution requires estimation as well as in the case where both the shape and scale parameters are unknown.

The numerical powers of the various test statistics are investigated and compared using a Monte Carlo study. This study shows that $KL_{n,10}$ and DK_n produces impressive power results against a range of alternative distributions when using MLE in order to estimate the parameters of the Pareto distribution. In the case where MMEs are used to perform parameter estimation, the $G_{n,2}$ test produces the highest powers followed by MA_n . It should, however, be noted that $G_{n,2}$ produces the lowest powers against $W(0.5)$, $LN(2.5)$ and the tilted Pareto distribution. When taking all of the above into account, we recommend using DK_n together with MLE when testing for the Pareto distribution in practice.

Funding Open access funding provided by North-West University. No financial or non-financial benefits are directly or indirectly related to this work.

Data availability Not available.

Declarations

Conflict of interest None of the authors has any conflict of interest whatsoever. The work of the last three authors are based on research supported by the National Research Foundation (NRF). Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Appendix

This appendix contains the numerical results pertaining to samples of size $n = 30$.

Table 10 Numerical powers when estimating λ parameter using MLE with $n = 30$

	$K\hat{S}_n$	$C\hat{M}_n$	AD_n	MA_n	ZA_n	ZB_n	ZC_n	$KL_{n,1}$	$KL_{n,10}$	DK_n	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(2,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(5,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	32	35	60	27	56	63	2	25	1	1	36	32	57	32	37	45	55	59	9	22
$\Gamma(0.8)$	12	14	13	17	14	15	24	11	26	30	14	14	8	12	14	10	8	7	20	21
$\Gamma(1)$	35	44	39	47	42	39	52	21	60	68	43	47	29	45	48	37	29	21	44	56
$\Gamma(1.2)$	64	76	73	77	76	74	76	40	84	91	75	80	64	80	81	71	62	52	64	85
$W(0.5)$	48	52	75	43	72	77	2	39	1	0	53	49	74	48	54	61	71	75	14	35
$W(0.8)$	10	11	11	14	13	14	20	11	22	24	11	10	6	8	11	8	6	7	16	16
$W(1.2)$	69	81	78	82	81	79	81	45	88	94	80	84	69	84	85	76	68	57	69	89
$W(1.5)$	95	99	99	99	99	99	98	81	99	100	99	99	97	99	99	98	96	93	89	100
$LN(1)$	77	86	87	77	96	94	52	50	74	79	85	88	94	91	90	86	87	84	28	89
$LN(1.2)$	36	44	44	35	64	59	28	23	45	49	43	48	54	53	52	44	44	41	15	51
$LN(1.5)$	8	9	8	8	15	12	10	9	16	16	9	10	10	10	11	8	7	7	7	12
$LN(2.5)$	39	43	57	34	38	42	2	20	1	0	44	41	60	41	43	49	58	58	16	26
$LFR(0.2)$	46	56	51	60	56	52	67	28	74	80	54	59	38	58	60	48	37	28	57	69
$LFR(0.5)$	54	66	61	71	65	61	77	35	83	89	64	69	46	67	70	58	45	36	67	79

Table 10 continued

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.8)	60	73	68	77	70	66	82	38	87	92	71	75	51	73	76	65	52	41	73	83
<i>LFR</i> (1)	63	75	71	80	72	69	83	41	88	93	73	77	54	75	79	67	55	44	75	85
<i>BE</i> (0.5)	27	30	54	23	53	60	3	25	2	2	31	25	49	24	29	36	46	53	6	19
<i>BE</i> (0.8)	14	16	14	20	15	16	28	12	30	34	16	16	8	13	16	11	8	6	23	23
<i>BE</i> (1)	35	43	38	46	41	38	52	21	60	68	42	46	29	44	47	36	28	21	44	56
<i>BE</i> (1.5)	85	94	93	94	95	94	89	62	95	98	93	96	90	96	96	92	87	81	75	97
<i>TP</i> (0.5)	9	9	7	9	10	9	9	7	12	13	9	9	8	9	11	9	7	7	7	12
<i>TP</i> (1, 1)	15	18	15	16	17	15	14	9	19	24	18	18	14	17	21	16	13	12	13	23
<i>TP</i> (2, 1)	30	37	32	33	34	31	27	14	36	44	36	38	29	37	41	34	28	24	24	45
<i>TP</i> (3)	45	54	48	51	48	46	39	21	49	60	54	56	43	55	58	50	42	37	35	62
<i>D</i> (0.2)	16	19	16	16	22	19	14	10	21	24	19	20	18	20	23	18	16	15	10	24
<i>D</i> (0.4)	42	50	47	44	55	53	34	22	49	56	50	54	50	56	57	49	46	41	25	59
<i>D</i> (0.6)	70	81	79	76	84	83	59	41	75	84	80	85	82	86	86	79	76	71	43	86
<i>D</i> (0.8)	89	95	95	93	97	96	79	62	91	96	95	97	95	97	97	94	93	89	58	97
<i>HN</i> (0.8)	68	79	75	84	77	73	87	46	91	95	78	81	57	78	82	72	59	47	79	87
<i>HN</i> (1)	73	85	81	88	83	79	92	52	94	97	83	86	63	84	87	77	64	53	84	91

Table 11 Numerical powers when estimating λ parameter using MME with $n = 30$

	KS_n	CM_n	AD_n	MA_n	ZA_n	ZB_n	ZC_n	$KL_{n,1}$	$KL_{n,10}$	DK_n	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1,1)$	10	11	12	14	10	12	13	6	4	0	10	11	6	11	5	5	5	5	5	13
$P(2,1)$	5	5	5	5	5	5	5	5	5	2	5	5	5	5	5	5	5	5	5	4
$P(5,1)$	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	4
$P(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4
$\Gamma(0.5)$	21	21	48	15	52	61	2	25	1	0	24	17	21	6	37	45	55	59	9	7
$\Gamma(0.8)$	19	22	20	27	13	17	16	11	22	3	21	21	15	25	14	10	8	7	20	26
$\Gamma(1)$	53	61	59	65	47	50	32	24	55	11	60	62	54	67	48	37	29	21	44	65
$\Gamma(1.2)$	82	89	89	90	82	84	48	47	83	25	88	90	86	93	81	71	62	52	64	91
$W(0.5)$	22	20	54	14	61	71	1	38	0	0	25	14	35	6	54	61	71	75	14	27
$W(0.8)$	20	21	21	28	12	16	10	11	18	1	21	21	12	24	11	8	6	7	16	23
$W(1.2)$	83	90	89	91	83	85	60	48	87	43	89	91	89	93	85	76	68	57	69	92
$W(1.5)$	98	100	100	100	99	99	91	85	99	92	100	100	100	100	99	98	96	93	89	100
$LN(1)$	89	94	95	89	99	98	18	59	72	17	93	95	97	95	90	86	87	84	28	97
$LN(1.2)$	55	67	69	61	78	75	6	29	39	4	63	68	71	72	52	44	44	41	15	69
$LN(1.5)$	22	27	29	31	29	28	2	13	12	1	25	28	21	31	11	8	7	7	7	24
$LN(2.5)$	13	10	31	31	31	45	39	27	0	0	14	7	10	4	43	49	58	58	16	52
$LFR(0.2)$	61	70	69	74	56	59	50	31	71	22	69	72	65	76	60	48	37	28	57	75
$LFR(0.5)$	69	77	76	81	64	68	63	37	81	38	76	78	73	82	70	58	45	36	67	81

Table 11 continued

	$K\hat{S}_n$	CM_n	AD_n	MA_n	ZA_n	ZB_n	ZC_n	$KL_{n,1}$	$KL_{n,10}$	DK_n	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.8)	73	81	79	84	69	72	70	41	85	50	80	82	78	85	76	65	52	41	73	84
<i>LFR</i> (1)	75	83	81	86	72	75	74	42	87	58	82	84	81	86	79	67	55	44	75	86
<i>BE</i> (0.5)	15	15	42	12	46	57	3	25	1	0	18	11	15	4	29	36	46	53	6	6
<i>BE</i> (0.8)	23	26	25	33	16	19	18	12	26	3	26	26	18	30	16	11	8	6	23	31
<i>BE</i> (1)	53	61	59	64	46	49	32	24	55	11	59	62	53	67	47	36	28	21	44	67
<i>BE</i> (1.5)	95	98	98	98	97	98	61	70	95	48	98	99	98	99	96	92	87	81	75	99
<i>TP</i> (0.5)	33	38	43	46	37	40	19	15	8	0	37	39	21	38	11	9	7	7	7	32
<i>TP</i> (1, 1)	59	66	71	74	66	69	26	28	14	0	65	67	42	66	21	16	13	12	13	50
<i>TP</i> (2, 1)	88	93	95	96	92	94	40	59	27	0	92	93	73	92	41	34	28	24	24	76
<i>TP</i> (3)	97	99	99	99	99	99	55	81	40	0	98	99	88	99	58	50	42	37	35	89
<i>D</i> (0.2)	33	40	40	40	38	36	4	14	16	1	38	41	32	43	23	18	16	15	10	39
<i>D</i> (0.4)	67	75	76	72	74	73	6	31.	42	4	74	77	72	79	57	49	46	41	25	77
<i>D</i> (0.6)	89	94	94	92	93	93	16	53	70	11	93	95	93	95	86	79	76	71	43	95
<i>D</i> (0.8)	97	99	99	99	99	99	30	74	89	25	99	99	99	99	97	94	93	89	58	100
<i>HN</i> (0.8)	77	85	84	88	75	78	79	46	90	64	84	86	83	88	82	72	59	47	79	87
<i>HN</i> (1)	85	91	90	93	82	85	82	55	93	62	90	92	87	93	87	77	64	53	84	93

Table 12 Numerical powers when estimating 2 parameters using MLE with $n = 30$

	KS_n	CM_n	AD_n	MA_n	ZA_n	ZB_n	ZC_n	$KL_{n,1}$	$KL_{n,10}$	DK_n	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1, 1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	5	5	5
$P(2, 1)$	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(5, 1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$P(10, 1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	33	36	55	27	53	52	2	34	2	1	37	33	53	32	38	40	48	45	11	28
$\Gamma(0.8)$	10	11	7	14	11	7	30	12	29	32	11	8	2	5	8	5	3	7	20	20
$\Gamma(1)$	26	30	16	35	23	6	61	17	57	66	29	28	3	21	28	17	8	21	43	47
$\Gamma(1.2)$	45	54	34	58	44	13	82	26	76	86	52	53	9	44	53	37	22	40	63	71
$W(0.5)$	49	53	72	43	70	68	2	49	2	1	54	49	70	47	54	56	64	62	17	42
$W(0.8)$	9	9	7	12	10	8	25	12	25	27	9	7	3	4	7	4	3	6	17	16
$W(1.2)$	49	59	40	64	50	16	86	29	80	88	58	59	11	50	59	42	26	45	68	76
$W(1.5)$	77	85	72	88	79	43	97	52	95	98	85	85	33	79	86	73	55	74	88	94
$LN(1)$	26	30	16	28	20	5	46	12	40	50	30	30	5	25	31	21	14	34	25	42
$LN(1.2)$	12	13	6	12	8	2	28	7	26	31	13	13	1	9	13	8	5	17	14	22
$LN(1.5)$	5	4	2	5	3	3	11	5	11	13	4	4	2	3	4	3	2	5	7	7
$LN(2.5)$	42	45	58	36	51	45	1	27	1	1	47	44	60	43	48	49	56	48	18	35
$LFR(0.2)$	34	41	24	47	34	9	74	22	70	77	40	39	4	30	39	25	13	27	55	59
$LFR(0.5)$	42	51	31	58	43	13	83	27	79	86	49	49	6	38	49	32	17	33	66	69

Table 12 continued

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.8)	47	56	37	64	49	15	87	30	83	89	55	55	8	44	55	38	20	38	72	75
<i>LFR</i> (1)	50	60	40	67	52	17	88	33	84	91	58	58	9	47	58	40	22	40	74	77
<i>BE</i> (0.5)	27	30	49	23	49	49	3	33	4	3	31	26	44	24	30	32	40	39	8	25
<i>BE</i> (0.8)	12	12	7	16	12	7	34	13	33	37	12	10	2	6	9	5	3	7	23	22
<i>BE</i> (1)	25	29	15	34	22	6	61	17	56	64	28	28	2	20	27	16	8	20	43	46
<i>BE</i> (1.5)	60	70	52	73	61	24	90	35	85	92	69	71	19	63	71	55	38	59	73	83
<i>TP</i> (0.5)	6	5	3	6	4	3	12	6	11	13	5	5	1	3	5	3	2	6	7	9
<i>TP</i> (1, 1)	10	10	4	10	6	3	19	7	17	22	10	9	1	6	10	6	3	11	12	16
<i>TP</i> (2, 1)	19	21	10	22	13	4	36	10	31	39	21	20	2	15	21	13	7	21	23	32
<i>TP</i> (3)	31	35	19	34	21	6	51	14	42	54	34	33	4	26	34	24	14	32	35	49
<i>D</i> (0.2)	8	9	4	8	5	3	18	6	16	20	8	8	1	5	9	5	3	12	10	15
<i>D</i> (0.4)	21	24	11	23	15	4	41	10	35	44	24	24	3	18	25	16	10	26	24	36
<i>D</i> (0.6)	39	45	27	44	31	9	64	17	55	67	45	45	9	38	46	34	22	44	41	59
<i>D</i> (0.8)	54	63	44	62	49	17	79	25	70	82	62	63	18	57	63	50	36	59	56	76
<i>HN</i> (0.8)	55	65	46	72	58	20	91	37	88	93	63	63	10	52	63	45	25	44	78	81
<i>HN</i> (1)	60	71	52	78	65	25	94	42	92	95	69	69	14	58	70	52	30	49	83	86

Table 13 Numerical powers when estimating 2 parameters using MME with $n = 30$

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
$P(1,1)$	10	11	12	14	13	14	13	6	4	0	10	11	6	11	5	4	4	5	5	10
$P(2,1)$	5	5	5	5	5	5	5	5	5	2	5	5	5	5	5	5	5	5	5	4
$P(5,1)$	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	4
$P(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	3
$\Gamma(0.5)$	15	15	25	11	18	21	2	34	2	0	16	11	14	4	38	40	48	45	11	8
$\Gamma(0.8)$	21	23	23	28	19	23	17	12	25	3	23	24	18	28	8	5	3	7	20	23
$\Gamma(1)$	50	58	57	61	53	58	30	18	51	9	56	59	52	64	28	17	8	21	43	49
$\Gamma(1.2)$	73	81	82	83	79	82	43	29	73	20	80	83	78	86	53	37	22	40	63	64
$W(0.5)$	16	14	28	12	25	30	1	48	1	0	18	9	21	4	54	56	64	62	17	30
$W(0.8)$	21	24	24	30	19	23	11	12	21	1	23	24	16	28	7	4	3	6	17	22
$W(1.2)$	74	82	81	83	80	83	54	30	77	33	80	83	80	86	59	42	26	45	68	64
$W(1.5)$	90	95	95	95	95	96	81	53	95	70	94	95	95	96	86	73	55	74	88	72
$LN(1)$	50	58	58	54	57	58	12	13	35	9	56	59	58	62	31	21	14	34	25	39
$LN(1.2)$	32	39	40	39	39	39	5	9	20	3	37	41	36	44	13	8	5	17	14	29
$LN(1.5)$	17	20	22	25	20	21	2	7	8	1	19	21	13	23	4	3	2	5	7	14
$LN(2.5)$	14	11	27	33	38	46	39	36	0	0	14	8	5	6	48	49	56	48	18	54
$LFR(0.2)$	58	67	66	71	64	69	47	24	67	19	65	68	63	73	39	25	13	27	55	71
$LFR(0.5)$	65	74	73	77	72	77	60	29	77	32	72	75	70	79	49	32	17	33	66	65

Table 13 continued

	$K S_n$	$C M_n$	$A D_n$	$M A_n$	$Z A_n$	$Z B_n$	$Z C_n$	$K L_{n,1}$	$K L_{n,10}$	$D K_n$	$S_{n,0.5}$	$S_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	T_n	$I_{n,2}$	$I_{n,3}$	$I_n^{[1]}$	$I_n^{[2]}$	$L_{n,3,2}$
<i>LFR</i> (0.8)	69	77	77	80	77	81	66	31	81	43	76	78	75	82	55	38	20	38	72	71
<i>LFR</i> (1)	70	79	79	82	79	83	70	33	83	49	78	80	78	83	58	40	22	40	74	77
<i>BE</i> (0.5)	11	10	19	10	14	18	3	33	3	0	12	7	9	3	30	32	40	39	8	56
<i>BE</i> (0.8)	24	28	27	34	23	28	18	13	29	3	27	28	22	33	9	5	3	7	23	60
<i>BE</i> (1)	49	57	57	61	53	57	30	18	51	9	56	59	52	63	27	16	8	20	43	63
<i>BE</i> (1.5)	84	90	90	90	88	90	52	39	83	32	89	91	89	92	71	55	38	59	73	63
<i>TP</i> (0.5)	31	36	41	44	43	44	19	12	7	0	35	37	20	36	5	3	2	6	7	6
<i>TP</i> (1, 1)	55	62	68	70	71	71	25	22	12	0	61	63	38	62	10	6	3	11	12	27
<i>TP</i> (2, 1)	84	89	92	93	93	93	37	47	22	0	88	89	67	89	21	13	7	21	23	49
<i>TP</i> (3)	95	97	98	98	99	99	50	68	31	0	96	97	83	97	34	24	14	32	35	68
<i>D</i> (0.2)	28	33	34	33	32	32	4	8	12	1	31	34	25	35	9	5	3	12	10	21
<i>D</i> (0.4)	51	60	60	58	57	57	6	14	29	3	58	61	53	63	25	16	10	26	24	32
<i>D</i> (0.6)	70	77	78	76	75	75	13	22	48	8	76	79	73	81	46	34	22	44	41	46
<i>D</i> (0.8)	81	87	87	86	85	86	25	32	66	17	87	88	85	90	63	50	36	59	56	51
<i>HN</i> (0.8)	73	81	81	85	82	85	75	35	87	55	80	82	80	85	63	45	25	44	78	23
<i>HN</i> (1)	81	88	88	90	87	90	78	43	91	52	87	88	85	91	70	52	30	49	83	41

References

1. Ahsanullah, M.: A characterization of the exponential distribution by spacings. *J. Appl. Prob.* **15**(3), 650–653 (1978)
2. Alizadeh Noughabi, H.: Testing exponentiality based on the likelihood ratio and power comparison. *Ann. Data Sci* **2**(2), 195–204 (2015)
3. Alizadeh Noughabi, H., Balakrishnan, N.: Tests of goodness of fit based on phi-divergence. *J. Appl. Stat.* **43**(3), 412–429 (2016)
4. Alizadeh Noughabi, R., Alizadeh Noughabi, H., Behabadi, A.E.M.: An entropy test for the rayleigh distribution and power comparison. *J. Stat. Comput. Simul.* **84**(1), 151–158 (2014)
5. Allison, J.S., Santana, L., Smit, N., Visagie, I.J.H.: An ‘apples to apples’ comparison of various tests for exponentiality. *Comput. Stat.* **32**(4), 1241–1283 (2017)
6. Allison, J.S., Milošević, B., Obradović, M., Smuts, M.: Distribution-free goodness-of-fit tests for the Pareto distribution based on a characterization. *Comput. Stat.* **37**(1), 403–418 (2022)
7. Arnold, B.C.: *Pareto Distributions*. CRC Press, New York (2015)
8. Basu, A., Harris, I.R., Hjort, N.L., Jones, M.C.: Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85**, 549–559 (1998)
9. Bera, A.K., Galvao, A.F., Wang, L., Xiao, Z.: A new characterization of the normal distribution and test for normality. *Economet. Theory* **32**(5), 1216–1252 (2016)
10. Beran, R.J.: Minimum Pareto distance estimates for parameter models. *Ann. Stat.* **5**, 445–463 (1977)
11. Beran, R.J.: An efficient and robust adaptive estimator of location. *Ann. Stat.* **6**, 292–313 (1978)
12. Betsch, S., Ebner, B.: Testing normality via a distributional fixed point property in the Stein characterization. *TEST* **29**(1), 105–138 (2020)
13. Boos, D.: Minimum distance estimators for location and goodness of fit. *J. Am. Stat. Assoc.* **76**, 663–670 (1981)
14. Brazauskas, V., Serfling, R.: Favorable estimators for fitting Pareto models: a study using goodness-of-fit measures with actual data. *ASTIN Bull. J. IAA* **33**(2), 365–381 (2003)
15. Choi, B., Kim, K.: Testing goodness-of-fit for Laplace distribution based on maximum entropy. *Statistics* **40**(6), 517–531 (2006)
16. Chu, J., Dickin, O., Nadarajah, S.: A review of goodness of fit tests for Pareto distributions. *J. Comput. Appl. Math.* **361**, 13–41 (2019)
17. Cockeran, M., Meintanis, S.G., Allison, J.S.: Goodness-of-fit tests in the Cox proportional hazards model. *Commun. Stat. Simul. Comput.* **50**(12), 4132–4143 (2021)
18. Csizsár, I.: On topological properties of f-divergences. *Studia Scientiarum Mathematicarum Hungarica* **2**, 329–339 (1967)
19. Csörgő, S., Szabó, T.: Weighted quantile correlation tests for Gumbel, Weibull and Pareto families. *Probab. Math. Stat.* **29**, 227–250 (2009)
20. D’Agostino, R., Stephens, M.: *Goodness-of-fit Techniques*. Marcel Dekker, New York (1986)
21. Giacomini, R., Politis, D.N., White, H.: A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econom. Theory* **29**, 567–589 (2013)
22. Gulati, S., Shapiro, S.: Goodness-of-fit tests for Pareto distribution. *Stat. Models Methods Biomed. Tech. Syst.* **25**, 259–274 (2008)
23. Hosking, J.R.M.: *L*-moments: analysis and estimation of distributions using linear combinations of order statistic. *J. R. Stat. Soc.* **52**, 105–124 (1990)
24. İsmail, S.: A simple estimator for the shape parameter of the Pareto distribution with economics and medical applications. *J. Appl. Stat.* **31**(1), 3–13 (2004)
25. Jahanshahi, S., Rad, A.H., Fakoor, V.: A goodness-of-fit test for Rayleigh distribution based on Hellinger distance. *Ann. Data Sci.* **3**(4), 401–411 (2016)
26. Kapur, J.N.: *Measures of Information and Their Applications*. Wiley-Interscience, New York (1994)
27. Klar, B.: Goodness-of-fit tests for the exponential and the normal distribution based on the integrated distribution function. *Ann. Inst. Stat. Math.* **53**(2), 338–353 (2001)
28. Klar, B., Meintanis, S.G.: Tests for normal mixtures based on the empirical characteristic function. *Comput. Stat. Data Anal.* **49**(1), 227–242 (2005)
29. Kullback, S.: *Information Theory and Statistics*. Dover Publications Inc, New York (1997)
30. Lequesne, J.: Entropy-based goodness-of-fit test: Application to the pareto distribution. In: *AIP Conference Proceedings*, vol. 1553, (pp. 155–162). American Institute of Physics (2013)
31. Meintanis, S.G.: Goodness-of-fit tests and minimum distance estimation via optimal transformation to uniformity. *J. Stat. Plan. Inference* **139**(2), 100–108 (2009)
32. Meintanis, S.G.: A unified approach of testing for discrete and continuous Pareto laws. *Stat. Pap.* **50**(3), 569–580 (2009)

33. Meintanis, S.G.: A review of testing procedures based on the empirical characteristic function. *South Afr. Stat. J.* **50**(1), 1–14 (2016)
34. Meintanis, S.G., Gamero, M.D.J., Alba-Fernández, V.: A class of goodness-of-fit tests based on transformation. *Commun. Stat. Theory Methods* **43**(8), 1708–1735 (2014)
35. Milošević, B., Obradović, M.: Two-dimensional kolmogorov-type goodness-of-fit tests based on characterizations and their asymptotic efficiencies. *J. Nonparamet. Stat.* **28**(2), 413–427 (2016)
36. Ndwandwe, L., Allison, J.S., Visagie, I.J.H.: A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring. *Annu. Proc. South Afri. Stat. Assoc. Conf.* **1**, 17–23 (2021)
37. Ndwandwe, L., Allison, J., Smuts, M., Visagie, I.: On a new class of tests for the pareto distribution using fourier methods. *Stat.* (p. e566) (2023)
38. Nofal, Z.M., El Gebaly, Y.M.: New characterizations of the Pareto distribution. *Pak. J. Stat. Oper. Res.* **13**, 63–74 (2017)
39. Obradović, M.: On asymptotic efficiency of goodness of fit tests for Pareto distribution based on characterizations. *Filomat* **29**(10), 2311–2324 (2015)
40. Obradović, M., Jovanović, M., Milošević, B.: Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics. *Statistics* **49**(5), 1026–1041 (2015)
41. Pareto, V.: *Cours d'économie Politique*, vol. II. F. Rouge, Lausanne (1897)
42. Parr, W.C., Schucany, W.R.: Minimum distance and robust estimation. *J. Am. Stat. Assoc.* **75**, 616–624 (1980)
43. Quandt, R.E.: Statistical discrimination among alternative hypotheses and some economic regularities. *J. Region. Sci.* **5**(2), 1–23 (1964)
44. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (2022). <https://www.R-project.org/>
45. Rizzo, M.L.: New goodness-of-fit tests for Pareto distributions. *Astin Bull.* **39**, 691–715 (2009)
46. Rossberg, H.: Characterization of the exponential and the Pareto distributions by means of some properties of the distributions which the differences and quotients of order statistics are subject to. *Statistics* **3**(3), 207–216 (1972)
47. Shannon, C.E.: A mathematical theory of communication. *Bell. Syst. Tech. J.* **27**(3), 379–423 (1948)
48. Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. Routledge, London (2018)
49. Taufer, E., Santi, F., Espa, G., Dickson, M.M.: Graphical representations and associated goodness-of-fit tests for Pareto and log-normal distributions based on inequality curves. *J. Nonparamet. Stat.* **33**(3–4), 464–481 (2021)
50. Vasicek, O.: A test for normality based on sample entropy. *J. R. Stat. Soc. Ser. B (Methodol.)* **38**(1), 54–59 (1976)
51. Volkova, K.: Goodness-of-fit tests for the Pareto distribution based on its characterization. *Stat. Methods Appl.* **25**(3), 351–373 (2016)
52. Wolfowitz, J.: Estimation by the minimum distance method. *Ann. Inst. Stat. Math.* **5**, 9–23 (1953)
53. Zenga, M.: Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito. *Giornale degli economisti e Annali di Economia* **5**(6), 301–326 (1984)
54. Zhang, J.: Powerful goodness-of-fit tests based on the likelihood ratio. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **64**(2), 281–294 (2002)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.