Mathematics and Statistics 12(4): 388-400, 2024 http://www.hrpub.org
DOI: 10.13189/ms.2024.120411

New Classes of Tests for The Pareto Distribution Based on A Conditional Expectation Characterisation

T. Nombebe, J.S. Allison, L. Santana*, **I.J.H. Visagie**

Pure and Applied Analytics, North-West Univeristy, Potchefstroom, South Africa

Received March 6, 2024; Revised June 20, 2024; Accepted July 21, 2024

Cite This Paper in the Following Citation Styles

***(a)**: [1] T. Nombebe, J.S. Allison1, L. Santana, I.J.H. Visagie, "New Classes of Tests for The Pareto Distribution Based on A Conditional Expectation Characterisation," Mathematics and Statistics, Vol.12, No.4, pp. 388-400, 2024. DOI: 10.13189/ms.2024.120411*

***(b)**: T. Nombebe, J.S. Allison1, L. Santana, I.J.H. Visagie (2024). New Classes of Tests for The Pareto Distribution Based on A Conditional Expectation Characterisation, Mathematics and Statistics, 12(4), 388-400. DOI: 10.13189/ms.2024.120411*

Copyright ©2024 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract There are relatively few goodness-of-fit tests specifically developed for the Pareto distribution when compared to other well-known distributions like the normal or exponential distributions. This is the case even though there are a host of practical applications where it would be required to first check the assumption that the data were realised from a Pareto distribution. We propose and investigate new goodness-of-fit tests for the Pareto Type I distribution based on a specific conditional expectation that characterises the Pareto distribution. Currently, the literature contains no other tests for the Pareto distribution based on conditional expectation. We conduct a thorough Monte Carlo power study in order to assess the finite sample performance of the newly developed tests using various estimation methods. The results from the simulation study show that the newly proposed tests are competitive in terms of power performance when compared to some existing tests. It also shows that the majority of tests produce their highest powers when the unknown shape parameter is estimated by the method of moments. A practical example, where we consider the annual salaries of English Premier League football players for two consecutive seasons, is also included to illustrate the use of the newly proposed tests. We find that the salaries in the 2021–2022 season can be adequately modelled with the Pareto distribution, but not the salaries for the 2022–2023 season.

Keywords Characterisation, Conditional Expectation, Goodness-of-fit Testing, Monte Carlo, Parametric Bootstrap, Pareto Distribution, Premier League earnings

1 Introduction

Characterisations of probability distributions play an important role in understanding the properties of random variables. A characterisation of a distribution is a specific statistical or distributional attribute that *uniquely* identifies a given random variable [1, 2]. Characterisations are often exploited in order to construct goodness-of-fit tests for the distribution in question with the advantage being that such a test is often consistent.

The focus of this paper is specifically on testing for the Pareto Type I distribution, simply referred to as the Pareto distribution below. A number of tests for the Pareto distribution, based on various characterisations, are available in the literature. These include tests based on, among others, likelihood ratios, entropy, empirical characteristic functions and the Mellin transform. For a detailed treatment of these tests, see [3]. The Pareto distribution is commonly used in various fields such as economics, finance, and insurance, where the occurrence of rare events or extreme values is of interest. Reference [4] gives an extensive historical survey of its use in the context of income distribution.

We are interested in the construction of new goodness-of-fit tests for the Pareto distribution based on a characterisation involving a conditional expectation. This has not been done in the context of the Pareto distribution before, and we illustrate that the resulting tests compare favourably in terms of empirical powers to those existing in the literature.

The density and distribution functions of the Pareto distribution are

$$f_{\beta,\sigma}(x) = \frac{\beta\sigma^{\beta+1}}{x^{\beta+1}}, \text{ and } F_{\beta,\sigma}(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\beta}, \ x \geq \sigma,$$

respectively, where $\beta > 0$ is a shape parameter and $\sigma > 0$ is a scale parameter. Note that the value of σ determines the support of the distribution. Since, in many practical applications, the support of the underlying distribution is known, the value of σ is frequently known *a priori*. This situation is demonstrated using the practical examples presented in Section 5. In these cases, a simple rescaling of the data can be employed in order to reduce the density and distribution functions above to

$$f_\beta(x) = \frac{\beta}{x^{\beta+1}}, \text{ and } F_\beta(x) = 1 - x^{-\beta},\ x \geq 1. \quad (1)$$

In the remainder of this paper, we consider the case where σ is known, meaning that we define a Pareto distribution via the functions in (1).

The specific focus of this paper is a characterisation based on conditional expectation by [5], discussed in detail in Section 2. Let $X_1, \ldots, X_n$ be a random sample from some unknown distribution F. We are interested in testing the composite hypothesis that these data are realised from the Pareto distribution;

$$H_0 : F(x) = F_\beta(x), \quad (2)$$

for all $x > 1$ and some $\beta > 0$, tested against general alternatives.

The remainder of the paper is organised as follows. In Section 2, we introduce new classes of goodness-of-fit tests based on a conditional expectation characterisation proposed by [5] to test the hypothesis stated in (2). Section 3 is devoted to a discussion of the methods of parameter estimation used in the paper as well as the calculation of critical values in order to be able to test the hypothesis in (2). Results of a Monte Carlo simulation to compare the power performance of the newly proposed tests to some commonly used existing tests are presented in Section 4. Section 5 demonstrates the use of the tests applied to observed data. Finally, in Section 6 some concluding remarks are provided.

2 Newly proposed test statistics

In this section we develop classes of goodness-of-fit tests based on a conditional expectation characterisation of the Pareto distribution. This characterisation is proposed in [5] and can be restated as follows:

Characterisation. *Let X be a nonnegative continuous random variable with distribution, survival and density functions F, S and f, respectively. Then X is said to follow the Pareto distribution with parameter β if, and only if, for all $t \geq 1$*

$$E[X^s | X \geq t] = E(X^s)t^s, \quad s < \beta.$$

The proof of this characterisation can be found in [5].

It easily follows from this characterisation that X has a Pareto distribution if, and only if, for all $t \geq 1$

$$E[X^s I(X \geq t)] = S(t)E(X^s)t^s, \quad s < \beta, \quad (3)$$

where $I(\cdot)$ is the indicator function.

From (3) we have that X follows a Pareto distribution if, and only if, for all $t \geq 1$

$$\Psi_s(t) := E[X^s I(X \geq t)] - S(t)E(X^s)t^s = 0.$$

As the first step to develop our new tests we need to estimate $\Psi_s(t)$ from the observed data $X_1, ..., X_n$. Two possible estimators for this conditional expectation characterisation are

$$\Psi^{[1]}_{n,s}(t) = \frac{1}{n}\sum_{j=1}^{n} X_j^s I(X_j > t) - S_{\widehat{\beta}}(t)\frac{t^s}{n}\sum_{j=1}^{n} X_j^s$$

and

$$\Psi^{[2]}_{n,s}(t) = \frac{1}{n}\sum_{j=1}^{n} X_j^s I(X_j > t) - \frac{t^s}{n}\sum_{j=1}^{n} X_j^s \frac{1}{n}\sum_{j=1}^{n} I(X_j > t),$$

where $S_{\widehat{\beta}}(t) = t^{-\widehat{\beta}}$, for some estimator $\widehat{\beta}$ of the shape parameter β. The only difference between these two estimators is the way in which the survival function $S(t)$ is estimated. In $\Psi^{[2]}_{n,s}$ we estimate the survival function non-parametrically, while in $\Psi^{[1]}_{n,s}$ we use a parametric estimate, $S_{\widehat{\beta}}(t)$. Both these estimators can easily be calculated from the observed data.

If the null hypothesis is true, then both $\Psi^{[1]}_{n,s}(t)$ and $\Psi^{[2]}_{n,s}(t)$ should be close to zero. Therefore, we propose the following two weighted $L2$-type test statistics:

$$L^{[1]}_{n,s,a} = \int_1^\infty \left(\Psi^{[1]}_{n,s}(t)\right)^2 w_a(t)\mathrm{d}t$$

and

$$L^{[2]}_{n,s,a} = \int_1^\infty \left(\Psi^{[2]}_{n,s}(t)\right)^2 w_a(t)\mathrm{d}t,$$

where $w_a(t)$ is a weight function ensuring the existence of the integrals above. Henceforth choose $w_a(t) = t^{-a}$, with $a > 1$ being a user-defined tuning parameter.

After some simplification we obtain the following easily calculable forms of the two test statistics

$$L^{[1]}_{n,s,a} = \sum_{j=1}^{n}\sum_{k=1}^{n} \frac{X_j^s X_k^s}{n^2}\left\{\frac{1 - \min(X_j, X_k)^{-a+1}}{a-1} - \frac{2(X_j^{s-\widehat{\beta}-a+1} - 1)}{s - \widehat{\beta} - a + 1} - \frac{1}{2s - 2\widehat{\beta} - a + 1}\right\}$$

and

$$L^{[2]}_{n,s,a} = \sum_{j=1}^{n}\sum_{k=1}^{n} \frac{X_j^s X_k^s}{n^2}\left\{\frac{1 - \min(X_j, X_k)^{-a+1}}{a-1} - \frac{2(\min(X_j, X_k)^{s-a+1} - 1)}{n(s-a+1)} + \frac{\min(X_k, X_j)^{2s-a+1} - 1}{n^2(2s-a+1)}\right\}.$$

We can obtain two additional test statistics by noting that the right-hand side of (3) can be written as

$$S(t)E(X^s)t^s = P(X > t)\frac{\beta}{\beta - s}\frac{t^{s+1}}{\beta}\frac{\beta}{t} = \frac{f_\beta(t)t^{s+1}}{\beta - s}, \quad (4)$$

using the fact that $E(X^s) = \beta/(\beta - s)$, if X follows a Pareto distribution.

Substituting (4) into (3) we now have that X follows the Pareto law, if and only if, for all $t \geq 1$

$$\Lambda_s(t) := E[X^s I(X \geq t)] - \frac{f_\beta(t)t^{s+1}}{\beta - s} = 0.$$

Again, we can estimate this quantity from the observed data $X_1, ..., X_n$. Two possible estimators for this form of the conditional expectation characterisation are

$$\Lambda_{n,s}^{[1]}(t) = \frac{1}{n}\sum_{j=1}^{n} X_j^s I(X_j > t) - \frac{f_{\widehat{\beta}}(t)t^{s+1}}{\widehat{\beta} - s}$$

and

$$\Lambda_{n,s}^{[2]}(t) = \frac{1}{n}\sum_{j=1}^{n} X_j^s I(X_j > t) - \frac{\widehat{f}_h(t)t^{s+1}}{\widehat{\beta} - s},$$

where $f_{\widehat{\beta}}(t) = \widehat{\beta}x^{-\widehat{\beta}-1}$ and $\widehat{f}_h(t)$ is a kernel density estimator with bandwidth h. In $\Lambda_{n,s}^{[1]}$ the density function is estimated parametrically, while in $\Lambda_{n,s}^{[2]}$ we make use of a non-parametric kernel density estimator.

As before, we propose the following two weighted $L2$-type test statistics

$$T_{n,s,a}^{[1]} = \int_1^\infty \left(\Lambda_{n,s}^{[1]}(t)\right)^2 w_a(t)\mathrm{d}t$$

and

$$T_{n,s,a}^{[2]} = \int_1^\infty \left(\Lambda_{n,s}^{[2]}(t)\right)^2 w_a(t)\mathrm{d}t.$$

Unfortunately, for the most common kernel functions used in $\widehat{f}_h$, the test statistic $T_{n,s,a}^{[2]}$ cannot be expressed in a simple closed-form expression. Even though it is possible to calculate the value of this statistic for any kernel using numerical integration, the results of some preliminary simulations show that this statistic exhibits erratic power, so we will not pursue it further in this paper.

The calculable form for $T_{n,s,a}^{[1]}$ is given by

$$\begin{aligned} T_{n,s,a}^{[1]} &= \\ &\frac{1}{n^2(a-1)}\sum_{j=1}^{n}\sum_{k=1}^{n} X_j^s X_k^s \left(1 - \min(X_j, X_k)^{-a+1}\right) \\ &- \frac{2\widehat{\beta}}{n(\widehat{\beta} - s)(s - \widehat{\beta} - a + 1)}\sum_{j=1}^{n} X_j^s (X_j^{s-\widehat{\beta}-a+1} - 1) \\ &- \frac{\widehat{\beta}^2}{(2s - 2\widehat{\beta} - a + 1)(\widehat{\beta} - s)^2}. \end{aligned}$$

All of the newly proposed tests can easily be calculated from observed data and reject the null hypothesis for large values of the test statistics.

3 Parameter and critical value estimation

In order to practically implement our newly proposed tests (as well as the existing tests considered in Section 4) the shape parameter, β, needs to be estimated from the observed data. Below we consider four possible estimators for the shape parameter; the maximum likelihood estimator (MLE), the method of moments estimator (MME) and two minimum distance estimators. The MLE and MME are well known and are often used in the literature to estimate β (see, for example, [6]); we include only the required formulas. The first minimum distance estimator considered minimises squared differences while the second minimises a Cramér-von Mises type distance. After considering the estimation techniques, we turn our attention to the calculation of the critical values which one needs to carry out these tests.

3.1 Parameter estimation

Denote the observed data by $X_1, X_2, \ldots, X_n$. The MLE of β is given by

$$\widehat{\beta} = \frac{n}{\sum_{j=1}^{n} \log X_j}.$$

The MME is obtained by equating the first moment of the fitted distribution to the first sample moment. Of course, this implicitly assumes that the first moment exists or, equivalently, that $\beta > 1$. The MME is then

$$\widetilde{\beta} = \frac{\bar{X}_n}{\bar{X}_n - 1},$$

where $\bar{X}_n = \sum_{j=1}^{n} X_j/n$.

We now consider the two minimum distance estimators. These estimators are obtained by minimising some discrepancy measures between $F_{\widehat{\beta}}$ and the empirical distribution function,

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

The Cramér–von Mises (CV) distance measure is based on the squared integral difference between F_n and F_β, and is given by

$$D_n^{CV}(\beta) = \int_1^\infty \left(F_n(x) - F_\beta(x)\right)^2 f_\beta(x)\mathrm{d}x, \quad (5)$$

with calculable form

$$D_n^{CV}(\beta) = \frac{1}{12n} + \sum_{j=1}^{n}\left(1 - X_{(j)}^{-\beta} - \frac{2j-1}{2n}\right)^2,$$

with $X_{(1)}, \ldots, X_{(n)}$ denoting the order statistics. The resulting estimator is

$$\widehat{\beta}_{CV} = \arg\min_\beta D_n^{CV}(\beta).$$

The final estimator considered is the so-called squared difference estimator. The discrepancy measure to be minimised is similar to (5), but the integrand is now weighted by F_n;

$$\begin{aligned} D_n^{SD}(\beta) &= \int_1^\infty \left(F_n(x) - F_\beta(x)\right)^2 \mathrm{d}F_n(x) \\ &= \sum_{j=1}^{n}\left(1 - X_{(j)}^{-\beta} - \frac{j}{n+1}\right)^2. \end{aligned}$$

The corresponding estimator is

$$\widehat{\beta}_{SD} = \arg\min_{\beta} D_n^{SD}(\beta).$$

When computing MDEs, we use the optim function in R together with the "BFGS" optimisation method [?, see]]Bro1970,Fle1970,Gol1970,Sha1970. These calculations require starting values for the optimisation procedure, the MMEs are used for this purpose.

We only consider two minimum distance estimators above. For a full treatment of the subject, the interested reader is referred to [7], [8, 9], [10], as well as [11]. A recent reference on minimum distance estimators, specifically in the context of the Lomax distribution, is [12].

3.2 Estimation of critical values

To determine whether one should reject the null hypothesis, it is necessary to first compute the critical values for each test. We therefore turn our attention to explaining the calculation of these numerical critical values. When using MLEs together with the transformed data

$$Y_j = X_j^{\widehat{\beta}},\ j = 1, ..., n,$$

the critical values are independent of the value of $\widehat{\beta}$. In this case, we obtain the critical values via Monte Carlo simulation. However, when using the other estimation techniques, the critical value depends on the value of $\widehat{\beta}$ and so a parameteric bootstrap procedure should be employed. Both the Monte Carlo procedure used for the MLEs and the parametric bootstrap procedure required for the other estimation techniques are outlined in [3].

4 Monte Carlo study

A Monte Carlo study is employed to examine the empirical power performance of the tests discussed in the previous sections against various fixed alternative distributions. The aim of this simulation study is to compare the finite-sample power performance of the newly proposed tests to the following existing goodness-of-fit tests for the Pareto distribution.

- The traditional Kolmogorov-Smirnov (KS_n), Cramer-von Mises (CV_n), Anderson-Darling (AD_n) and modified Anderson-Darling (MA_n) tests.
- A test proposed by [13] based on the likelihood ratio. The test statistic has calculable form

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

- A test based on the Mellin transform proposed by [14]. The test statistic is

$$\begin{aligned}
&G_{n,a} = \\
&\frac{1}{n}\Bigg[(\widehat{\beta}_n+1)^2\sum_{j,k=1}^{n} I_w^{(0)}(X_jX_k) \\
&\quad + \sum_{j,k=1}^{n} I_w^{(2)}(X_jX_k) + 2(\widehat{\beta}_n+1)\sum_{j,k=1}^{n} I_w^{(1)}(X_jX_k)\Bigg] \\
&\quad + \widehat{\beta}_n\Bigg[n\widehat{\beta}_n I_w^{(0)}(1) - 2(\widehat{\beta}_n+1)\sum_{j=1}^{n} I_w^{(0)}(X_j) \\
&\quad - 2\sum_{j=1}^{n} I_w^{(1)}(X_j)\Bigg],
\end{aligned}$$

where

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

For the choice $w(t) = e^{-at}$, we have that

$$\begin{aligned}
I_a^{(0)}(t) &= (a+\log t)^{-1}, \\
I_a^{(1)}(t) &= \frac{1-a-\log t}{(a+\log t)^2}, \\
I_a^{(2)}(t) &= \frac{2-2a+a^2+2(a-1)\log t+\log^2 t}{(a+\log t)^3}.
\end{aligned}$$

We show powers obtained using $a = 2$ in the numerical results below as this has been shown to produce good powers.

- A test based on the difference between two empirical characteristic functions proposed by [15]. The test statistic is

$$NA_{n,m,a} = n\int_{-\infty}^{\infty}|\phi_{n,m}(t) - \xi_{n,m}(t)|^2 e^{-at^2}\mathrm{d}t,$$

where

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

and

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

The test statistic has the following simplified form

$$\begin{aligned}
&NA_{n,m,a} = \\
&\frac{1}{n}\sqrt{\frac{\pi}{a}}\sum_{j=1}^{n}\sum_{k=1}^{n}\Bigg\{\exp\left[\frac{-\left(X_{(j)}^{1/m} - X_{(k)}^{1/m}\right)^2}{4a}\right] \\
&\quad - 2nv_{j,m}\exp\left[\frac{-\left(X_{(j)} - X_{(k)}^{1/m}\right)^2}{4a}\right] \\
&\quad + n^2 v_{j,m}v_{k,m}\exp\left[\frac{-\left(X_{(j)} - X_{(k)}\right)^2}{4a}\right]\Bigg\},
\end{aligned}$$

where

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

For implementing the test, [15] recommend choosing $m = 3$ and $a = 2$; results for these choices are therefore shown in the tables below.

All tests reject for large values of the test statistics.

4.1 Monte Carlo setup

Our aim is to compare the numerical sizes and powers achieved by the newly proposed tests to those of the existing tests for the Pareto distribution discussed above. We consider the powers achieved against a range of fixed alternative distributions with support $x > 1$ in various settings. Two different sample sizes, $n = 20$ and $n = 30$, are used, as well as different estimation methods for β. The alternative distributions considered, together with their density functions and the notation used for these distributions, are specified in Table 1.

A nominal significance level of 5% is used throughout and all calculations are performed using R v4.2.2 [16]. The estimated powers, reported in Tables 2 to 5, are the percentages of rejections of the null hypothesis in 10 000 independent Monte Carlo replications (rounded to the nearest integer). In an attempt to aid comparison, the highest and second highest powers against each alternative distribution are printed in bold and highlighted in black and grey, respectively. When β is estimated by its MLE, the empirical critical values of all the tests are obtained using $MC = 50\,000$ independent Monte Carlo replications. For all other estimation techniques, the critical values are obtained using the 'warp-speed' bootstrap procedure of [17] over each of the $MC = 10\,000$ Monte Carlo replications.

For both sample sizes considered, we calculate the numerical powers obtained when estimating β using the four different estimators discussed in Section 3. However, the minimum distance estimators are omitted from the power tables due to their inferior power performance compared to those associated with the MLE and MME. In order to illustrate the differences between the numerical powers achieved by the different estimation techniques, consider Figure 1. This figure illustrates the empirical powers against the TP, LW and BN alternative distributions, calculated when $n = 20$, for each of the four estimation methods for the tests $L^{[2]}_{n,0.1,3}$, $T^{[1]}_{n,0.1,3}$ and AD_n. These three distributions have the Pareto distribution as a special case when $\theta = 0$, with deviations from the null hypothesis for $\theta > 0$. That is, they can be used to illustrate contiguous alternatives to the Pareto distribution as θ increases above zero. When considering the numerical powers shown in Figure 1, it is clear that MME produces the highest power, while the second highest powers are obtained using MLE. The two minimum distance estimators are associated with the lowest powers in each of the distributions considered. This trend is also observed when considering other alternative distributions. As a result, we concern ourselves only with the powers achieved using the MLE and the MME in the power tables to follow.

Each of the newly proposed tests contains two tuning parameters: a and s. The construction of the tests implicitly assumes that at least the s^{th} moment of the underlying distribution exists and so we recommend using fairly small values for s. Specifically, s should be smaller than β, and the Pareto distribution has a finite mean if, and only if, $\beta \geq 1$. As a result, we recommend choosing $s < 1$. A preliminary power study indicated that the combination of $a = 3$ and $s = 0.1$ generally provide high powers against a wide range of alternative distributions. As a result, we recommend this combination of tuning parameters when using the test. The empirical powers reported in Tables 2 to 5 were obtained using these parameters.

4.2 Discussion of numerical results

In the discussion presented below we first consider those results associated with the MLE before turning our attention to the empirical powers obtained using the MME. Thereafter we compare these two sets of results to one another.

4.2.1 Sizes and powers obtained using MLE

Consider the numerical sizes and powers obtained using MLE, see Tables 2 and 4.

All of the tests considered largely maintain the specified nominal significance level of 5% for both of the sample sizes used. As is to be expected, the powers of the tests generally increase with sample size.

The tests that generally achieve the highest powers against the specified alternative distributions are the $L^{[2]}_{n,0.1,3}$ and $NA_{n,3,2}$ tests. Specifically, the $L^{[2]}_{n,0.1,3}$ test is shown to produce very high powers against LN, TP, LW and BN for the majority of the choices of the parameters of these alternative distributions. This test either outperforms all competitors outright, or produces powers close to the best performing competitor against these distributions for both of the sample sizes considered. However, we note that while the test has excellent performance in the majority of cases investigated, it produces no power for the alternatives $\Gamma(0.4)$, $LN(2.5)$, and $BE(0.5)$.

4.2.2 Sizes and powers obtained using MME

Interestingly, the two tests with the best performance when using the MLE, namely $L^{[2]}_{n,0.1,3}$ and $NA_{n,3,2}$, are undersized when using the MME for both sample sizes. This phenomenon is especially pronounced in the case of the $L^{[2]}_{n,0.1,3}$. However, both of these tests achieve high powers against a wide range of alternatives. In fact, these two tests, together with MA_n and $T^{[1]}_{n,0.1,3}$ generally provide the highest power against the widest range of alternative distributions. The most powerful of the newly proposed tests remains $L^{[2]}_{n,0.1,3}$ test which is shown to be especially powerful against the TP and LW distributions.

4.2.3 A comparison between the results obtained using MLE and MME

For each of the tests considered, the average power against all of the alternatives considered is higher when using MME

Table 1. Probability density functions for choices of the alternative distributions.

Alternative	Density function	Notation
Pareto (θ)	$\frac{\theta}{x^{\theta+1}}$	$P(\theta)$
Gamma (θ)	$\frac{1}{\Gamma(\theta)}(x-1)^{\theta-1}\mathrm{e}^{-(x-1)}$	$\Gamma(\theta)$
Log-normal(θ)	$\exp\left\{-\frac{1}{2}\left(\log(x-1)/\theta\right)^2\right\} / \left\{\theta(x-1)\sqrt{2\pi}\right\}$	$LN(\theta)$
Linear failure rate(θ)	$(1+\theta(x-1))\exp(-(x-1)-\theta(x-1)^2/2)$	$LF(\theta)$
Beta exponential(θ)	$\theta\mathrm{e}^{-(x-1)}(1-\mathrm{e}^{-(x-1)})^{\theta-1}$	$BE(\theta)$
Tilted Pareto(θ)	$\frac{1+\theta}{(x+\theta)^2}$	$TP(\theta)$
Log-Weibull(θ)	$\frac{(1+\theta)\log^{\theta} x}{x^{2+\theta}}$	$LW(\theta)$
Benini(θ)	$\frac{\exp(-\theta\log^2 x)}{x^2}(1+2\theta\log x)$	$BN(\theta)$

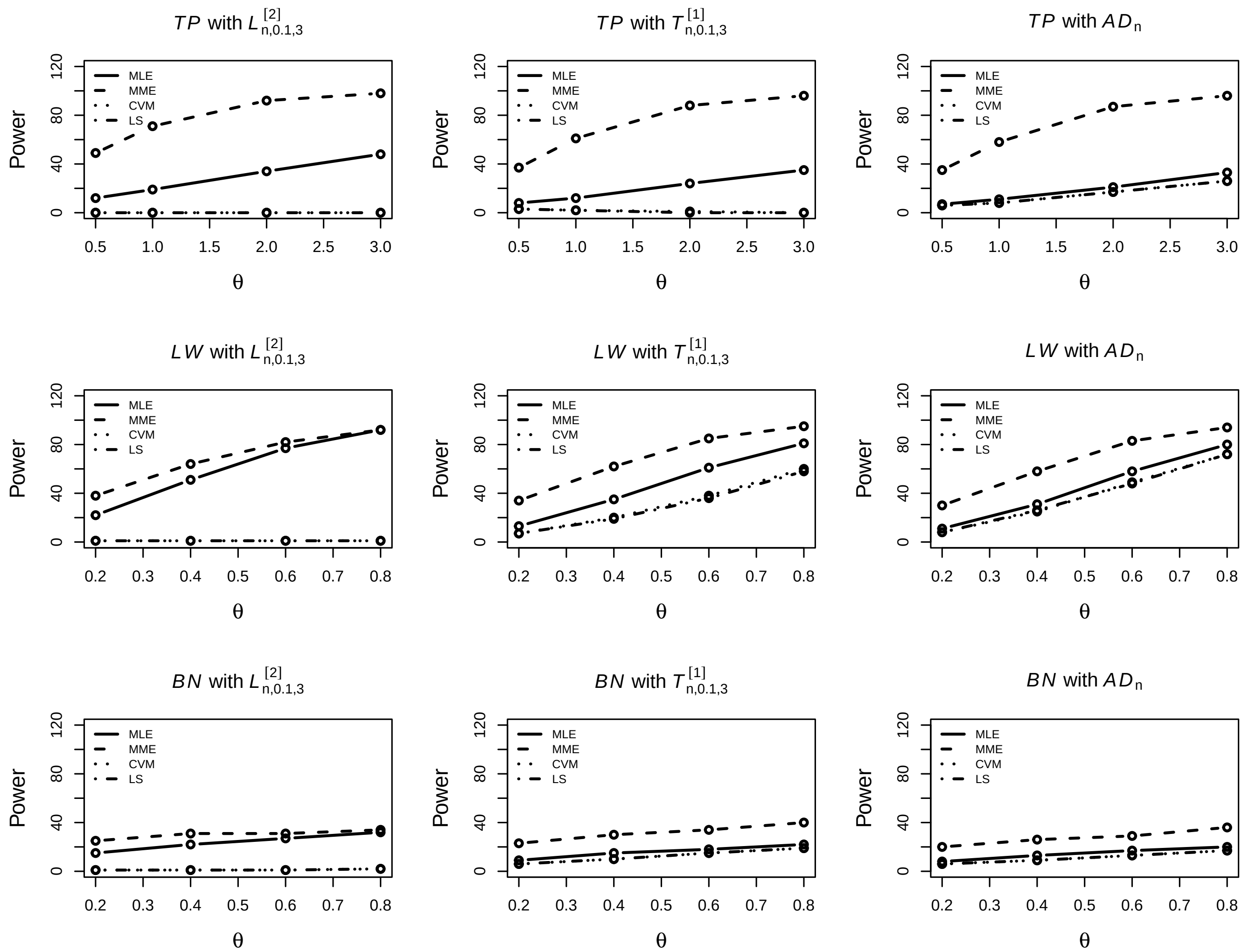

Figure 1. Numerical powers achieved using the different estimation techniques for three of the tests considered

.

Table 2. The approximated sizes and powers of the tests under MLE estimation, $n = 20$.

	KS_n	CV_n	AD_n	MA_n	ZA_n	$G_{n,2}$	$NA_{n,3,2}$	$L^{[1]}_{n,0.1,3}$	$L^{[2]}_{n,0.1,3}$	$T^{[1]}_{n,0.1,3}$
P(4)	5	5	5	5	5	5	5	5	5	5
P(10)	5	5	5	5	5	5	5	5	5	5
Γ(0.4)	49	52	**77**	42	**75**	48	26	59	0	60
Γ(0.8)	9	10	10	**12**	11	9	**16**	9	11	9
Γ(1.0)	24	30	25	30	29	32	**41**	26	**36**	25
Γ(1.3)	57	68	64	67	68	73	**79**	65	**77**	64
LN(1.0)	56	66	64	54	**80**	69	72	68	**86**	67
LN(1.2)	24	30	27	23	**43**	33	37	31	**50**	30
LN(1.6)	5	5	4	5	6	5	**7**	5	**8**	5
LN(2.5)	26	29	**44**	24	27	26	13	**34**	0	**34**
LFR(0.3)	34	42	36	45	39	44	**55**	37	**46**	36
LFR(0.6)	41	50	44	53	47	53	**63**	44	**54**	44
LFR(0.9)	44	54	48	56	50	56	**67**	48	**57**	48
LFR(1.2)	47	57	52	**60**	54	**60**	**70**	51	**60**	51
BE(0.5)	18	20	**42**	16	**39**	15	10	24	0	25
BE(0.8)	11	12	11	**14**	12	11	**19**	10	**14**	10
BE(1.0)	24	30	25	30	29	32	**41**	26	**36**	26
BE(1.5)	69	80	77	79	81	85	**89**	78	**88**	77
TP(0.5)	7	8	7	7	9	8	**11**	8	**12**	8
TP(1.0)	12	13	11	12	14	13	**18**	12	**19**	12
TP(2.0)	22	26	21	23	25	26	**34**	24	**34**	24
TP(3.0)	32	38	33	35	36	39	**48**	36	**48**	35
LW(0.2)	12	13	11	11	16	13	**18**	13	**22**	13
LW(0.4)	30	36	31	30	40	38	**44**	35	**51**	35
LW(0.6)	51	62	58	54	66	65	**70**	61	**77**	61
LW(0.8)	72	82	80	77	85	86	**88**	81	**92**	81
BN(0.2)	9	10	8	9	11	10	**14**	9	**15**	9
BN(0.4)	14	16	13	15	17	16	**23**	15	**22**	15
BN(0.6)	17	20	17	19	20	21	**29**	18	**27**	18
BN(0.8)	20	24	20	23	23	25	**33**	22	**32**	22

Table 3. The approximated sizes and powers of the tests under MME estimation, $n = 20$.

	KS_n	CV_n	AD_n	MA_n	ZA_n	$G_{n,2}$	$NA_{n,3,2}$	$L^{[1]}_{n,0.1,3}$	$L^{[2]}_{n,0.1,3}$	$T^{[1]}_{n,0.1,3}$
P(4)	5	6	5	5	5	5	4	5	2	5
P(10)	5	5	5	5	5	5	4	5	2	5
Γ(0.4)	41	43	**72**	31	**73**	34	10	34	0	34
Γ(0.8)	14	16	15	**20**	11	16	**19**	17	15	17
Γ(1.0)	37	45	42	**47**	33	**47**	**50**	46	39	**47**
Γ(1.3)	76	84	83	84	76	**87**	**87**	85	79	85
LN(1.0)	71	81	83	74	**90**	84	**86**	83	82	85
LN(1.2)	41	49	50	45	**58**	55	53	51	**58**	53
LN(1.6)	13	16	17	**21**	14	17	16	16	**26**	18
LN(2.5)	11	9	**27**	26	22	6	**36**	9	13	10
LFR(0.3)	46	55	52	**58**	44	**58**	**62**	56	46	56
LFR(0.6)	51	60	56	**63**	47	62	**66**	62	47	62
LFR(0.9)	57	65	63	**69**	53	67	**70**	66	51	66
LFR(1.2)	58	66	63	**70**	54	69	**71**	68	50	68
BE(0.5)	13	12	**32**	10	**34**	7	3	11	2	10
BE(0.8)	18	20	19	**25**	14	21	**25**	22	18	23
BE(1.0)	37	44	41	**46**	31	45	**50**	45	38	**46**
BE(1.5)	84	91	90	90	86	**93**	**93**	91	86	92
TP(0.5)	27	32	35	**40**	30	33	28	33	**49**	37
TP(1.0)	46	54	58	**63**	52	55	44	55	**71**	61
TP(2.0)	76	82	87	**88**	82	84	67	83	**92**	**88**
TP(3.0)	90	94	**96**	**96**	93	94	80	94	**98**	**96**
LW(0.2)	25	30	30	32	30	**34**	31	32	**38**	**34**
LW(0.4)	50	57	58	57	55	**62**	59	60	**64**	**62**
LW(0.6)	73	82	83	80	80	**85**	84	84	82	**85**
LW(0.8)	88	94	94	92	93	**95**	**95**	**95**	92	**95**
BN(0.2)	18	20	20	**23**	16	22	22	21	**25**	**23**
BN(0.4)	24	28	26	29	22	30	**32**	29	**31**	30
BN(0.6)	26	32	29	32	25	**34**	**38**	33	31	**34**
BN(0.8)	32	38	36	38	31	**40**	**44**	39	34	**40**

Table 4. The approximated sizes and powers of the tests under MLE estimation, $n = 30$.

	KS_n	CV_n	AD_n	MA_n	ZA_n	$G_{n,2}$	$NA_{n,3,2}$	$L^{[1]}_{n,0.1,3}$	$L^{[2]}_{n,0.1,3}$	$T^{[1]}_{n,0.1,3}$
P(4)	5	5	5	5	5	5	5	5	5	5
P(10)	5	5	5	5	5	5	5	5	5	5
Γ(0.4)	66	70	**89**	60	**88**	64	49	76	0	76
Γ(0.8)	12	14	13	**17**	14	13	**20**	12	14	12
Γ(1.0)	35	45	39	47	43	48	**56**	39	**49**	39
Γ(1.3)	76	87	85	88	87	**91**	**93**	84	**91**	83
LN(1.0)	77	86	87	77	**96**	89	88	88	**97**	88
LN(1.2)	36	44	44	35	**63**	48	50	46	**68**	46
LN(1.6)	6	6	5	5	**8**	6	**8**	5	**9**	6
LN(2.5)	39	43	**58**	35	40	36	26	48	0	**49**
LFR(0.3)	49	61	56	**65**	59	**65**	**73**	55	63	54
LFR(0.6)	57	69	64	**73**	67	**73**	**80**	62	69	61
LFR(0.9)	61	74	69	**78**	71	**78**	**84**	67	73	66
LFR(1.2)	66	78	74	**81**	75	**81**	**86**	72	77	71
BE(0.5)	27	30	**55**	23	**53**	20	19	35	0	36
BE(0.8)	15	17	15	**21**	17	16	**25**	14	16	14
BE(1.0)	35	43	38	47	42	**48**	**55**	39	**48**	38
BE(1.5)	86	94	93	94	95	96	**97**	93	**97**	92
TP(0.5)	9	9	8	8	10	9	**12**	9	**14**	9
TP(1.0)	15	17	15	16	17	18	**23**	16	**25**	16
TP(2.0)	31	37	32	34	34	38	**44**	35	**47**	34
TP(3.0)	45	54	49	51	48	57	**63**	52	**64**	52
LW(0.2)	16	19	16	16	22	20	**24**	18	**30**	18
LW(0.4)	42	51	48	44	56	55	**59**	51	**68**	50
LW(0.6)	70	81	80	76	84	86	**87**	81	**91**	80
LW(0.8)	89	95	95	93	96	**97**	**97**	95	**99**	95
BN(0.2)	11	12	10	12	13	13	**17**	11	**17**	11
BN(0.4)	18	22	18	22	21	24	**29**	20	**29**	19
BN(0.6)	25	30	25	29	29	33	**38**	27	**38**	27
BN(0.8)	30	36	32	35	34	39	**46**	33	**45**	33

Table 5. The approximated sizes and powers of the tests under MME estimation, $n = 30$.

	KS_n	CV_n	AD_n	MA_n	ZA_n	$G_{n,2}$	$NA_{n,3,2}$	$L^{[1]}_{n,0.1,3}$	$L^{[2]}_{n,0.1,3}$	$T^{[1]}_{n,0.1,3}$
P(4)	5	5	5	5	5	5	4	5	1	5
P(10)	5	5	5	5	5	5	4	5	2	5
Γ(0.4)	56	58	**85**	45	**86**	47	27	49	0	49
Γ(0.8)	19	21	21	**27**	13	21	**26**	23	21	24
Γ(1.0)	53	61	59	**65**	48	63	**65**	62	54	63
Γ(1.3)	90	**96**	95	**96**	91	**96**	**96**	**96**	92	**96**
LN(1.0)	89	95	**96**	90	**98**	95	**96**	95	95	**96**
LN(1.2)	55	66	68	61	**78**	70	69	69	**74**	70
LN(1.6)	16	20	22	**26**	20	21	18	20	**32**	23
LN(2.5)	13	10	**31**	**31**	**31**	5	**53**	9	12	9
LFR(0.3)	65	74	73	**78**	61	76	**77**	75	61	74
LFR(0.6)	69	79	76	**81**	65	79	**82**	79	65	79
LFR(0.9)	75	82	79	**85**	70	83	**84**	82	66	82
LFR(1.2)	76	84	82	**86**	73	85	**86**	85	68	85
BE(0.5)	14	13	**40**	11	**44**	7	4	11	2	10
BE(0.8)	23	26	25	**33**	15	26	**30**	28	24	28
BE(1.0)	52	60	58	**65**	45	63	**66**	61	54	62
BE(1.5)	95	98	98	98	97	**99**	**99**	98	96	98
TP(0.5)	33	37	42	**47**	38	40	32	39	**60**	45
TP(1.0)	58	66	70	**73**	63	67	50	67	**83**	**73**
TP(2.0)	89	93	95	**96**	93	93	77	93	**98**	**96**
TP(3.0)	97	**99**	**99**	**99**	**99**	**99**	88	**99**	**100**	**99**
LW(0.2)	33	39	40	39	37	42	39	41	**49**	**43**
LW(0.4)	67	75	76	73	74	78	77	78	**82**	**79**
LW(0.6)	89	94	94	92	93	**95**	**95**	**95**	**95**	**95**
LW(0.8)	98	99	99	99	99	**100**	99	99	99	**100**
BN(0.2)	23	28	28	**31**	22	29	29	29	**33**	**31**
BN(0.4)	33	39	38	**41**	28	**41**	**43**	40	37	**41**
BN(0.6)	37	43	41	44	32	**45**	**49**	44	39	**45**
BN(0.8)	41	47	46	48	38	**50**	**55**	49	44	49

than is the case when using MLE. That is, both the existing and new tests produce higher powers when estimating the shape parameter of the Pareto distribution using MME rather than MLE. This finding is supported by the empirical results shown in Figure 1. As a result, we recommend using MME in order to do parameter estimation regardless of the test used.

Based on all of the numerical results presented, we would recommend using either $L^{[2]}_{n,0.1,3}$ or $NA_{n,3,2}$, with β estimated using MME, when testing the hypothesis that data are realised from a Pareto distribution.

5 Practical data application

As was mentioned in the introduction, the Pareto distribution is used as a model for a wide variety of phenomena. In the field of extreme value theory, this distribution is often used to model the tails of a distribution, especially in cases where we expect *a priori* that the underlying distribution will exhibit a heavy tail.

Below we consider the observed annual salaries of English Premier League (EPL) football players for the seasons 2021–2022 as well as 2022–2023. Specifically, we are interested in determining whether or not the Pareto distribution is an appropriate model for the distribution of the salaries of players in excess of 10 million GBP. As a result, the hypothesis of interest is whether or not the reported salaries, rescaled by a factor of 10 million, are realised from the Pareto distribution. We consider the salaries from the two seasons separately. Both data sets used are obtained from [18]. These salaries (reported in millions of GBP) can be found in Table 6. The sample sizes are 20 and 28, respectively.

Table 6. Practical data example: EPL earnings above 10 million GBP for the 2021-2022 and 2022-2023 seasons (values are reported in millions).

2021–2022				
10.31	10.40	10.40	10.40	10.40
10.40	11.44	13.00	13.00	14.14
15.08	15.08	15.60	15.60	16.90
17.68	18.20	19.50	20.80	26.80
2022–2023				
10.40	10.40	10.40	10.40	10.40
10.40	10.40	10.40	11.44	11.44
11.70	12.48	13.00	13.00	13.00
13.78	15.08	15.34	15.47	15.60
15.60	16.90	17.68	18.20	18.20
19.50	19.50	20.80		

We use each of the tests discussed in Section 4, including the newly proposed tests based on the conditional expectation characterisation, to test

whether each of these two data sets is realised from a Pareto distribution. The code for this application can be found here:

`https://tinyurl.com/CondExp2023`.

However, before turning our attention to these formal hypothesis testing procedures, we present some graphical tests below. The empirical distribution functions (EDF) of each of the data sets are plotted in Figure 2. These distributions are overlaid with two fitted Pareto distributions in each case; these correspond to the fitted distributions obtained using MLE and MME for the parameter β featuring in the test statistics.

The distributions showcased in Figure 2 indicate that the salaries in the 2021–2022 season correspond quite closely to what would be expected under the Pareto distribution. While no glaringly obvious departure from this distribution is visible in the comparison for the 2022–2023 salaries, the tail of the empirical distribution seems to be lighter than is the case for the fitted distributions. Furthermore, around the 15 million GBP mark (indicated by 1.5 on the x-axis), the empirical distribution seems to be systematically lower than the fitted distributions.

Another graphical test that we employ is side-by-side violin plots (with overlaid boxplots). These are shown, for both data sets considered, in Figure 3. One can see from these plots that both data sets have a pronounced positive third central sample moment, which is indicative of a right-skewed underlying distribution. Again, it should be noted that the salaries associated with 2021–2022 seem to be realised from a distribution with a heavier tail than is the case for the salaries in the 2022–2023 season.

We now formally test the hypothesis that each of the data sets considered is realised from the Pareto distribution. That is, we compute the p-values for each test considered. Table 7 shows the estimated p-values obtained by using a parametric bootstrap with $B = 10\,000$ bootstrap replications, that is, the bootstrap samples are obtained from a Pareto distribution with the parameter estimated using either the MLE or MME.

Table 7. p-values (multiplied by 100) for testing the Pareto null hypothesis for the earnings above 10 million GBP for the EPL's 2021–2022 and 2022–2023 seasons (highlighted values are significant at a $\alpha = 0.1$ level).

	2021–2022		2022–2023	
	MLE	MME	MLE	MME
KS_n	17	15	16	12
CV_n	10	11	9	8
AD_n	10	11	6	6
MA_n	11	9	5	4
ZA_n	20	21	2	2
$G_{n,2}$	41	32	30	20
$NA_{n,3,2}$	35	31	25	21
$L^{[1]}_{n,0.1,3}$	9	9	6	6
$L^{[2]}_{n,0.1,3}$	38	27	32	20
$T^{[1]}_{n,0.1,3}$	9	10	6	6

The p-values from the first data set show that the majority of the tests do not reject the null hypothesis at a nominal significance level of 10%. With the second data set, the majority of tests, with the exception of KS_n, $G_{n,2}$, $NA_{n,3,2}$, and $L^{[2]}_{n,0.1,3}$, reject the null hypothesis that the data are realised from a Pareto distribution. To summarise these results, we

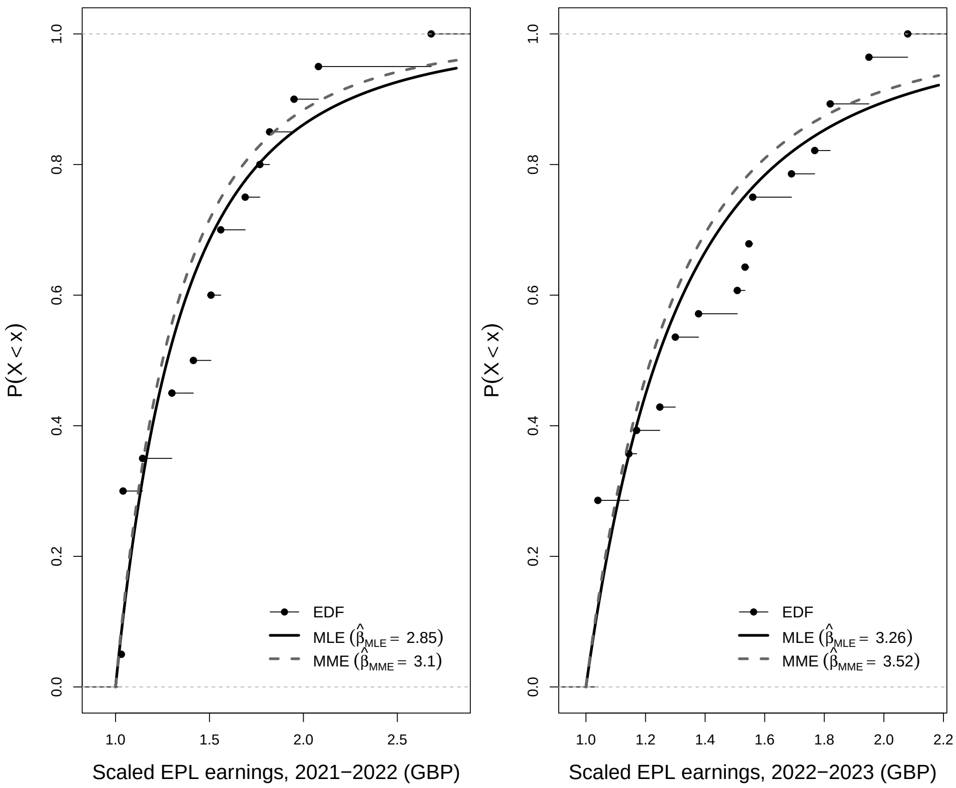

Figure 2. EDF plots of the EPL earnings above 10 million GBP for the two seasons; two theoretical Pareto CDFs are overlaid using MLE and MME for the parameter.

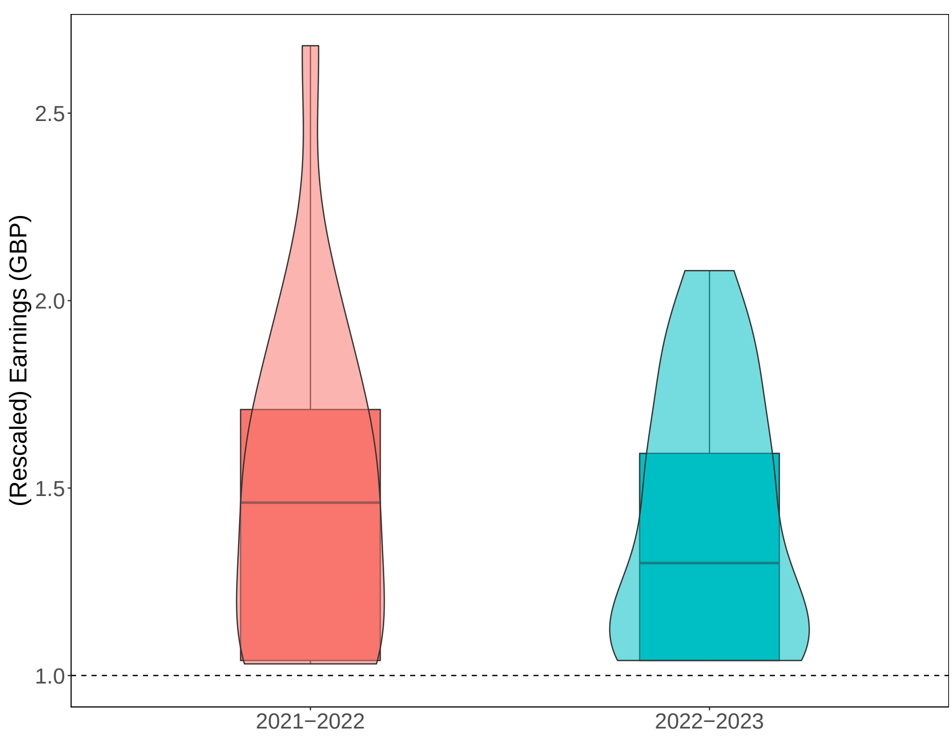

Figure 3. Side-by-side violin plots (with overlaid boxplots) of the EPL earnings above 10 million GBP for the two seasons.

conclude that the Pareto distribution is an appropriate model for the salaries in the 2021–2022 season, but not necessarily for the 2022–2023 season. Another model should be used for the latter data set.

6 Conclusions

In this paper we proposed and investigated new tests for the hypothesis that observed data are realisations from a Pareto distribution. These tests are based on a characterisation involving a conditional expectation. Using a simulation study, the finite sample power performance of the newly proposed tests are compared to that of the existing tests under a range of alternative distributions and overall the newly proposed test, $L^{[2]}_{n,s,a}$, performs very well when compared to the other tests. From extensive Monte Carlo simulations we recommend using $a = 3$ and $s = 0.1$ as this choice produced good powers for the majority of the alternatives considered.

REFERENCES

[1] A. E. Koudou and C. Ley, "Characterizations of GIG laws: A survey," *Probability Surveys*, vol. 11, pp. 161–176, 2014.

[2] H. Nagaraja, "Characterizations of probability distributions," in *Springer Handbook of Engineering Statistics*, pp. 79–95, Springer, 2006.

[3] L. Ndwandwe, J. S. Allison, L. Santana, and I. J. H. Visagie, "Testing for the Pareto type I distribution: a comparative study," *Metron*, vol. 81, pp. 215–256, 2023.

[4] B. C. Arnold, *Pareto Distributions*. CRC Press, Taylor and Francis Group, 2015.

[5] Z. M. Nofal and Y. M. El Gebaly, "New characterizations of the Pareto distribution," *Pakistan Journal of Statistics and Operation Research*, vol. 13, pp. 63–74, 2017.

[6] J. Ngatchou–Wandji, T. Nombebe, L. Santana, and J. Allison, "On classes of consistent tests for the Type I Pareto distribution based on a characterization involving order statistics," *Statistics*, vol. 0, no. 0, pp. 1–31, 2024.

[7] J. Wolfowitz, "Estimation by the minimum distance method," *Annals of the Institute of Statistical Mathematics*, vol. 5, pp. 9–23, 1953.

[8] R. J. Beran, "Minimum Pareto distance estimates for parameter models," *The Annals of Statistics*, vol. 5, pp. 445–463, 1977.

[9] R. J. Beran, "An efficient and robust adaptive estimator of location," *The Annals of Statistics*, vol. 6, pp. 292–313, 1978.

[10] D. Boos, "Minimum distance estimators for location and goodness of fit," *Journal of the American Statistical Association*, vol. 76, pp. 663–670, 1981.

[11] W. C. Parr and W. R. Schucany, "Minimum distance and robust estimation," *Journal of the American Statistical Association*, vol. 75, pp. 616–624, 1980.

[12] T. Nombebe, J. S. Allison, L. Santana, and I. J. H. Visagie, "On fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques," *Computation*, vol. 44, no. 11, p. https://doi.org/10.3390/computation11030044, 2023.

[13] J. Zhang, "Powerful goodness-of-fit tests based on the likelihood ratio," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 2, pp. 281–294, 2002.

[14] S. G. Meintanis, "A unified approach of testing for discrete and continuous Pareto laws," *Statistical Papers*, vol. 50, no. 3, pp. 569–580, 2009.

[15] L. Ndwandwe, J. S. Allison, M. Smuts, and I. J. H. Visagie, "On a new class of tests for the Pareto distribution using Fourier methods," *Stat*, p. e566, 2023.

[16] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[17] R. Giacomini, D. N. Politis, and H. White, "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators," *Econometric Theory*, vol. 29, pp. 567–589, 2013.

[18] `www.spotrac.com`, "EPL salary rankings," 2023. https://www.spotrac.com/epl/rankings/year/2021/ (Accessed 2023-09-18).