On testing the assumption of the Pareto type I distribution

LM Ndwandwe

orcid.org 0000-0002-2123-8750

Thesis accepted in fulfilment of the requirements for the degree *Doctor of Philosophy in Science with Statistics* at the North-West University

Promoter: Prof JS Allison

Co-promoter: Prof IJH Visagie

Co-promoter: Prof L Santana

Graduation July 2023

33497575

Acknowledgements

I am deeply grateful to Prof James Allison, Prof Jaco Visagie and Prof Leonard Santana, whose guidance, encouragement and support throughout my research journey has been invaluable. I would like to express my sincere thanks to Statistics department and UCDP, for providing me with the necessary resources and infrastructure to complete this thesis. I am also grateful to my colleagues and friends, Dr Marius Smuts and Dr Elzanie Bothma, for their support and inspiring discussions. Special thanks to my girlfriend for her unwavering love and encouragement during the entire process. I am truly grateful for everyone who has been a part of this journey.

Abstract

The Pareto distribution is a popular model in economics, finance and actuarial sciences. Consequently, a number of goodness-of-fit tests, based on diverse characteristics, have been developed to test the goodness-of-fit hypothesis that an observed data set is realised from this distribution. This thesis reviews the existing literature on goodness-of-fit tests for the Pareto distribution and compares the power performances of these tests. Additionally, several new classes of tests for the Pareto distribution, based on various characterisations, are developed.

The first characterisation used for the development of the tests relates to the distribution of the sample minimum. We estimate the characteristic function of the sample minimum using U and V statistics and we propose two classes of tests based on these statistics. The resulting tests are shown to be consistent against fixed alternatives and the finite sample power performances of these tests are demonstrated to be competitive against those of the existing tests in the literature. A second characterisation used is a multiplicative version of the memoryless property. We base two goodness-of-fit tests on this property and we demonstrate that these tests perform well using a finite sample power study.

We also consider the setting in which random right censoring is present. We propose a new fixed point characterisation based test for the Pareto distribution in this setting. The characterisation is based on Stein's method for the approximation of integrals. The power performance of the proposed test is considered against a range of alternative distributions using various censoring proportions.

The Pareto distribution has a shape parameter, $\beta > 0$, which is typically required to be estimated in order to perform goodness-of-fit testing. A result of independent interest considered is the effect of the estimation method used on the powers achieved by various goodness-of-fit tests. That is, we consider the use of both maximum likelihood and method of moments estimators. It is found that, not only does the estimation technique used have a profound effect on the numerical powers achieved by the various tests, but it also influences the way in which critical values are approximated. Specifically, when maximum likelihood is used, the critical values are shape invariant and not a function of the estimated value of β, meaning that a fixed critical value can be obtained using Monte Carlo simulation. On the other hand, if the method of moments is used, then the critical values are required to be estimated using a parametric bootstrap procedure.

Key words: Pareto distribution, goodness-of-fit, censoring, memoryless property, fixed point characterisation.

Preface

Chapter 1 provides a broad overview and outlines the goals of the research. This thesis is presented in the article format and comprises of four pieces of writing: two have been accepted for publication (one as a journal article and one as a conference proceeding), one has been submitted for publication and the reviewers requested minor changes, and one is a draft to be submitted in the near future. The three articles as well as the proceeding are located in Chapters 2, 3, 4 and 5 with their components including abstracts, introductions, methods, results, conclusions, and references formatted according to the guidelines of the relevant peer-reviewed journals. In Chapter 6 we report the main findings along with conclusions and avenues for future research.

First article: *Testing for the Pareto type I distribution: A comparative study*, has been submitted to the journal *Metron* and the reviewers requested minor changes.

Second article: *On a new class of tests for the Pareto distribution using Fourier methods*, has been accepted for publication in the journal *Stat*.

Third article: *Revisiting the memoryless property - testing for the Pareto distribution*, will be submitted to the journal *Computational Statistics*.

Conference proceeding: *A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring*, has been published in the Proceedings of the 62$^{\text{nd}}$ Annual Conference of the South African Statistical Association.

The promoters agreed on co-authorship and gave consent for the use of these articles as part of the final thesis. The first author was solely responsible for the initial planning and proposal of the thesis, literature searches, writing all programs used in the Monte Carlo studies, interpretation of results, as well as planning and writing of the articles and the entire thesis.

STATEMENT BY CO-AUTHORS

Herewith is a statement of the co-authors giving permission that the articles may form part of this thesis.

I hereby approve the articles and give my consent that these articles may be published as part of the thesis for the degree Doctorate of Philosophy in Science with Statistics of Mr LM Ndwandwe.

Prof JS Allison	20/02/2023 Date
Prof IJH Visagie	20/02/2023 Date
Prof L Santana	20/02/2023 Date
Dr M Smuts	20/02/2023 Date

Table of Contents

Chapter 1

Introduction

1.1 Overview

Goodness-of-fit testing is a statistical technique used to determine how well a given model fits a set of observed data. This method allows researchers to assess whether the model accurately describes the data. The goal of goodness-of-fit testing is to determine the level of agreement between the observed data and the model predictions, providing valuable information for model selection, improvement, and refinement. This technique is widely used in various fields, including finance, engineering, and social sciences, to validate and improve models and to make informed decisions based on data.

In the study we investigate goodness-of-fit tests for the Pareto type I distribution (henceforth referred to as simply the Pareto distribution). In the context of the Pareto distribution, goodness-of-fit tests determine whether the observed data follows a Pareto distribution or not. The Pareto distribution is named after the Italian economist Vilfredo Pareto, who first described this distribution in his 1897 book *Cours d'économie politique* as a model for the distribution of income. Since then, several extensions of this distribution have been proposed. These extensions include a location, scale and inequality parameter, corresponding to the Pareto type II, III and IV distributions, respectively. While there are some tests available for the Pareto distribution, they are few in number when compared to the number of tests for other distributions like the normal or exponential distributions.

Given the widespread usage and many applications of the Pareto distribution, some goodness-of-fit tests have been developed to determine whether observed data is realised from this distribution. An overview of these tests is given in Chapter 2. In this thesis we introduce and implement new goodness-of-fit tests for the Pareto distributions based on specific characterisations. This includes characterisations based on order statistics and on the multiplicative memoryless property of the Pareto distribution. To evaluate the finite sample performance of the newly proposed tests, we conduct Monte Carlo studies. Furthermore, we have also developed a new test for the Pareto as a lifetime distribution in situations where random right censoring is present. This method is based on a fixed point characterisation that uses the widely recognised Stein's method for estimating distributions.

1.2 Objectives

The main goal of this thesis is to develop new classes of goodness-of-fit tests for the null hypothesis that observed data is realised from a Pareto distribution. We carry out simulation studies using Monte Carlo procedures to examine various test statistics for evaluating the goodness-of-fit of the Pareto type I distribution, and compare their performances to one another. These objectives are summarised below.

- Conduct an extensive literature review comparing previous studies and discussing unresolved issues in the field of goodness-of-fit testing for the Pareto distribution.
- Propose new classes of goodness-of-fit tests for the Pareto distribution based on two different characterisations.
- Investigate the large sample properties of the tests based on a characterisation utilizing order statistics.
- Propose a new class of tests for testing for the Pareto distribution in the presence of random right censoring, and assess how well this new class of tests performs in comparison to classical tests across a wide range of alternative distributions.
- Critically evaluate the performance of the newly proposed tests using Monte Carlo simulation.

1.3 Thesis outline

This section showcases the overarching themes, summaries, and links between the three research papers and the proceeding that make up this thesis. The aim of the three articles and the proceeding is to enhance our understanding of goodness-of-fit testing for the Pareto distribution. In Chapter 6, we summarise the final outcomes and reach conclusions, and also highlight potential areas for further exploration.

The titles and abstracts of the three articles and one proceeding, which are discussed in Chapters 2, 3, 4 and 5 respectively, are listed below.

Testing for the Pareto type I distribution: A comparative study

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.

On a new class of tests for the Pareto distribution using Fourier methods

We propose new classes of tests for the Pareto type I distribution using the empirical characteristic function. These tests are U and V statistics based on a characterisation of the Pareto distribution involving the distribution of the sample minimum. In addition to deriving simple computational forms for the proposed test statistics, we prove consistency against a wide range of fixed alternatives. A Monte Carlo study is included in which the newly proposed tests are shown to produce high powers. These powers include results relating to fixed alternatives as well as local powers against mixture distributions. The use of the proposed tests is illustrated using an observed data set.

Revisiting the memoryless property - testing for the Pareto distribution

We propose new goodness-of-fit tests for the Pareto type I distribution. These tests are based on a multiplicative version of the memoryless property which characterises this distribution. We present the results of a Monte Carlo power study demonstrating that the proposed tests are powerful compared to existing tests. The power study considers powers against fixed alternatives as well as powers against mixture distributions. Thereafter the tests considered are used in order to test the hypotheses that two sets of golfers' earnings (those of the PGA tour and LIV Golf) are realised from Pareto distributions.

A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring

We propose a new goodness-of-fit test for the Pareto Type I lifetime distribution in the presence of random right censoring. The test is based on a fixed point characterisation, which is a generalisation of the well-known Stein method for the approximation of distributions. The empirical power performance of the new test is compared to the modified Cramér-von Mises and Kolmogorov-Smirnov tests for two different censoring proportions and two alternative lifetime distributions by means of a limited Monte Carlo study.

Chapter 2

Article 1: Testing for the Pareto type I distribution: A comparative study

The first article, *Testing for the Pareto type I distribution: A comparative study*, has been submitted to *Metron*. A summary of the guidelines to authors from the journal is now presented.

Manuscript	No page limit.
Title	The title should be concise and informative.
Abstract and keywords	A maximum of 150 to 250 words and 4 to 6 keywords which can be used for indexing purposes.
Tables	For each table, please supply a table caption (title) explaining the components of the table.
References	The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text.
General formatting	A LaTeXtemplate is provided for submission.
Additional information	https://www.springer.com/journal/40300/submission-guidelines

Testing for the Pareto type I distribution: A comparative study

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.

Key words: 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 (1897). 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 & El Gebaly (2017) as well as Ismaïl (2004). For an in-depth discussion of the Pareto distribution the interested reader is referred to Arnold (2015) 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 (2015).

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. (2016), and the exponential distribution, see, e.g., Allison et al. (2017), the only overview paper of this kind relating to the Pareto distribution is Chu et al. (2019). 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. (2019) 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. (2019) 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. (2019) 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. (2019) 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. (2019) are obtained using a bootstrap approach; in Section 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 $X, X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) continuous positive random variables with an unknown distribution function F. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the order statistics of $X_1, X_2, \ldots, X_n$. Denote the Pareto distribution function by

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

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 Section 5; here the monetary expenses (above a certain threshold) resulting from wind related catastrophes are examined. Another example is found in Arnold (2015), 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

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

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 Section 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 Section 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 $\widehat{U}_j = F_{\widehat{\beta}_n,\widehat{\sigma}_n}(X_j)$, $j = 1, 2, ..., n$, 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 Section 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 $X_1, X_2, ..., X_n$ 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

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

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

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

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

$$n\int_{-\infty}^{\infty}\left[F_n(x) - F_{\widehat{\beta}_n,\widehat{\sigma}_n}(x)\right]^2 w(x)\mathrm{d}F_{\widehat{\beta}_n,\widehat{\sigma}_n}(x), \tag{2.1}$$

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

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

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

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

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

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

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 (2001) and D'Agostino & Stephens (1986). All of the above tests reject the null hypothesis for large values of the test statistics.

2.2 Tests based on likelihood ratios

Zhang (2002) 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

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

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 $\mathrm{d}w(x)$ that Zhang (2002) 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 $\mathrm{d}w(x) = [F_n(x)\{1-F_n(x)\}]^{-1}\,\mathrm{d}F_n(x)$ leads to

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

- Choosing $\mathrm{d}w(x) = \left[F_{\widehat{\beta}_n,\widehat{\sigma}_n}(x)\{1-F_{\widehat{\beta}_n,\widehat{\sigma}_n}(x)\}\right]^{-1}\mathrm{d}F_{\widehat{\beta}_n,\widehat{\sigma}_n}(x)$ results in

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

Motivated by the high powers often obtained using the modified Anderson-Darling test, we also include the choice $\mathrm{d}w(x) = \{1-F_n(x)\}^{-2}\mathrm{d}F_n(x)$, 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 (2002) proposes, Alizadeh Noughabi (2015) adapts two of these test to test the assumption of exponentiality. Neither Zhang (2002) nor Alizadeh Noughabi (2015) 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 (2002) also considers the test

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

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 (1948). The entropy of a random variable X with density and distribution functions f and F, respectively, is defined to be

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

where $F^{-1}(\cdot)$ denotes the quantile function of X. The concept of entropy has been applied in several studies, see, e.g., Kullback (1997), Kapur (1994) and Vasicek (1976), where, in particular, Vasicek (1976) proposes using

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

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 $m \le \frac{n}{2}$. 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., Kullback, 1997)

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

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

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

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

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

This test rejects the null hypothesis for large values of $KL_{n,m}$. The test statistic coincide with the one studied by Lequesne (2013), 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. (2014) 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., Jahanshahi et al., 2016)

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

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

$$HD = \frac{1}{2}\int_{0}^{1}\left(\sqrt{\left(\frac{\mathrm{d}}{\mathrm{d}p}F^{-1}(p)\right)^{-1}} - \sqrt{\frac{\beta\sigma^{\beta}}{(F^{-1}(p))^{\beta+1}}}\right)^2 \frac{\mathrm{d}}{\mathrm{d}p}F^{-1}(p)\mathrm{d}p.$$

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

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

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

Jahanshahi et al. (2016) 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. (2016) 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

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

where $\phi : [0,\infty) \longrightarrow (-\infty,\infty)$ is a convex function such that $\phi(1) = 0$ and $\phi''(1) > 0$. It is further known (see, e.g., Choi & Kim, 2006; Csiszár, 1967) 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 (2016) 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

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

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 Section 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., Silverman, 2018). 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 (2016) 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 (2016) 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 Section 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., Klar & Meintanis, 2005; Meintanis, 2009a; Betsch & Ebner, 2020). For a review of testing procedures based on the empirical characteristic functions see, e.g., Meintanis (2016).

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

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

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

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

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

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

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 $w(t) = \mathrm{e}^{-a|t|}$ and $w(t) = \mathrm{e}^{-at^2}$, 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 (2009a). 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 $\widehat{U}_1, \widehat{U}_2, ..., \widehat{U}_n$, 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 (2009a) proposes the test

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

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

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

The test rejects the null hypothesis for large values of the test statistic. Although Meintanis (2009a) 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. (2014) 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 (2009b) 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., Meintanis, 2009b)

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

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 (2009b) proposes a weighted L^2 distance between $D_n(t)$ and 0 as test statistic;

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

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

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

where

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

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

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

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 (2009b) 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

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

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., Zenga, 1984)

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

Taufer et al. (2021) proposes a test based on the constant inequality curve exhibited by the $P(\beta,\sigma)$ distribution for some $\sigma > 0$. Taufer et al. (2021) 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. (2021) 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

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

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

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

The choice $j = 1, 2, \ldots, n - \lfloor\sqrt{n}\rfloor$ ensures that $\widehat{\lambda}_j$ is a consistent estimator for λ (see Taufer et al., 2021). Setting $m = n - \lfloor\sqrt{n}\rfloor$, Theorem 2.1 states that under the null hypothesis, for any choice of p_j, $0 < p_j < 1$, $j = 1, 2, \ldots, m$, 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 $X_1, ..., X_n$, we can obtain estimators for β_0 and β_1 from the regression

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

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

The least squares estimators for β_0 and β_1 are given by

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

where $\bar{p} = \frac{m(m+1)}{2n^2}$ 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 $|\widehat{\beta}_1|$. In a finite sample study, Taufer et al. (2021) 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 Section 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 (Obradović et al., 2015)

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. (2015) provides the proof for this characterisation and proposes two test statistics based on it. In order to specify these test statistics, denoted by

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

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

$$T_n = \int_1^\infty [M_n(x) - F_n(x)]\, \mathrm{d}F_n(x)$$

and

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

Both of these tests reject the null hypothesis for large values of the test statistics. Obradović et al. (2015) 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 (Allison et al., 2022)

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

Using m as a tuning parameter, Allison et al. (2022) 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 $\min\{X_1, ..., X_m\}$ and the V-empirical distribution of $\sqrt[m]{X}$, defined as

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

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

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

$K_{n,m}$ 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. (2022) derive the limiting null distribution of all three test statistics. Upon calculating and comparing the Bahadur efficiencies, Allison et al. (2022) 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. (2022) used empirical distribution functions to construct their tests, Ndwandwe et al. (2023) propose test statistics that instead utilise empirical versions of the characteristic function. To this end, let

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

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

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

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 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. (2023) propose as test statistic that is a weighted L^2 distance between $\phi_{n,m}$ and $\xi_{n,m}$:

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

where $w_a(t)$ is a weight function (see Section 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

$$L_{n,m,a} = \frac{1}{n}\sqrt{\frac{\pi}{a}}\sum_{j=1}^{n}\sum_{k=1}^{n}\left\{\exp\left[\frac{-\left(X_{(j)}^{1/m} - X_{(k)}^{1/m}\right)^2}{4a}\right] - 2nv_{j,m}\exp\left[\frac{-\left(X_{(j)} - X_{(k)}^{1/m}\right)^2}{4a}\right]\right.$$
$$\left. + n^2 v_{j,m} v_{k,m}\exp\left[\frac{-\left(X_{(j)} - X_{(k)}\right)^2}{4a}\right]\right\},$$

where

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

The test rejects for large values of the test statistic. Ndwandwe et al. (2023) 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. (2023) also studied a test where $\xi_m(t)$ is estimated by the U-statistic

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

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 (Rossberg, 1972)

Obradović (2015) 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 $X_{(1)} \le X_{(2)} \le X_{(3)}$. 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 (2015) 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 (2015) proposes the test statistics

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

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 (2015) 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 (2015) 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 (Obradović, 2015)

In addition to the tests above, Obradović (2015) 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 (1978).

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 (2015) proposes based on this characterisation are

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

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 (2015) 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 (Volkova, 2016)

For a fixed k, let $X_1, ..., X_k$ 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 (2016) provides a proof of this characterisation and derives two test statistics utilizing this characterisation:

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

and

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

where

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

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

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 (2016) 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 Csörgö & Szabó (2009), a test based on Euclidean distances by Rizzo (2009), a test based on spacings by Gulati & Shapiro (2008) and a Kolmogorov-type test involving a sort of "memory-less" characterization of the Pareto distribution by Milošević & Obradović (2016).

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 a an overview of tests for some of the multitudes of tests available for exponential, see Allison et al. (2017).

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 Section 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

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

and

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

Note that if we transform $X_1, ..., X_n$ as follows:

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

then

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

and

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

As can be seen above, the transformation in (3.1) ensures that, when the Pareto distribution is fitted to $Y_1, ..., Y_n$, 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. (2021), see Section 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 $Y_j = \left(\frac{X_j}{\widehat{\sigma}_n}\right)^{\widehat{\beta}_n/2}, \ j = 1, ..., n$, 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 (1964). Instead of choosing parameter estimates so as to equate the first two population moments to the first two sample moments, Quandt (1964) 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

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

and

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

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 $\widetilde{\beta}_n$. We provide the following bootstrap algorithm for the estimation of critical values.

1. Based on data $X_1, ..., X_n$, estimate β and σ by $\widetilde{\beta}_n$ and $\widetilde{\sigma}_n$, respectively.

2. Obtain a parametric bootstrap sample $X_1^*, ..., X_n^*$ by sampling independently from $F_{\widetilde{\beta}_n, \widetilde{\sigma}_n}$.

3. Calculate $\widetilde{\beta}_n^* = \widetilde{\beta}(X_1^*, ..., X_n^*)$, $\widetilde{\sigma}_n^* = \widetilde{\sigma}(X_1^*, ..., X_n^*)$, and the value of the test statistic say $S^* = S(X_1^*, ..., X_n^*)$.

4. Repeat steps 2 and 3 B times to obtain $S_1^*, ..., S_B^*$ and obtain the order statistics $S_{(1)}^* \leq ... \leq S_{(B)}^*$.

5. The estimated critical value at a $\alpha \times 100\%$ significance level is $\widehat{C}_n = S^*_{(B\lfloor 1-\alpha \rfloor)}$, where $\lfloor x \rfloor$ 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 (Hosking, 1990), methods that involve minimising some distance criterion between distribution functions (Wolfowitz, 1953; Beran, 1977, 1978; Parr & Schucany, 1980; Boos, 1981), as well as similar minimum distance-based methed related to ϕ-diveregence (Basu et al., 1998).

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 Section 2. Section 4.1 details the simulation setting used, including the alternative distributions considered, while Section 4.2 shows the numerical results obtained together with a discussion and comparison of these results.

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 – 6 and 10 – 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

Table 1: Summary of various choices of the alternative distributions.

Alternative	Density function	Notation
Gamma	$\frac{1}{\Gamma(\theta)}(x-1)^{\theta-1}\mathrm{e}^{-(x-1)}$	$\Gamma(\theta)$
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	$\theta\mathrm{e}^{-(x-1)}(1-\mathrm{e}^{-(x-1)})^{\theta-1}$	$BE(\theta)$
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	$\frac{\sqrt{2}}{\theta\sqrt{\pi}}\exp\left(\frac{-(x-1)^2}{2\theta^2}\right)$	$HN(\theta)$

to the sample size, estimation method and the number of parameters estimated.

Table 2: Summary of power tables.

Table	n	Estimation	Parameters	Table	n	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

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. (2013) 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. (2021) as well as Ndwandwe et al. (2021). 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 R Core Team (2022).

A final remark regarding the numerical powers associated with the tests based on characterisation of the Pareto distribution is in order. These tests, see Section 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 Section 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, $I_{n2,2}$, $I_{n,3}$, $I_n^{[1]}$ 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 posses 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.

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
$LFR(0.2)$	32	39	34	41	36	35	47	22	**58**	**62**	38	40	26	39	44	32	26	20	38	53
$LFR(0.5)$	39	48	42	50	45	44	55	27	**67**	**70**	47	49	32	48	53	40	32	25	46	61
$LFR(0.8)$	43	53	47	55	50	48	61	29	**72**	**76**	52	55	36	53	59	46	36	29	50	66
$LFR(1)$	44	55	49	57	51	50	62	31	**73**	**77**	53	56	38	55	61	47	38	31	51	69
$BE(0.5)$	19	20	**42**	16	40	**47**	3	17	2	2	22	18	39	18	19	25	35	39	5	10
$BE(0.8)$	11	12	11	14	12	14	17	10	**24**	**24**	12	11	7	10	13	9	7	7	15	19
$BE(1)$	24	30	25	31	28	27	35	17	**46**	**49**	29	31	20	30	35	25	19	16	28	41
$BE(1.5)$	67	79	76	78	80	80	73	48	87	**91**	79	83	72	84	85	75	69	62	51	**88**
$TP(0.5)$	7	8	6	7	8	8	8	7	**11**	**11**	8	7	6	7	9	7	6	6	6	**11**
$TP(1,1)$	12	13	10	12	14	12	12	8	**18**	**19**	13	13	10	13	16	11	10	9	10	**18**
$TP(2,1)$	22	26	22	23	25	24	21	12	**30**	**34**	26	27	20	26	**30**	23	20	18	16	**34**
$TP(3)$	32	37	32	34	35	34	29	17	41	**47**	37	39	30	38	**43**	33	29	26	22	**47**
$D(0.2)$	12	13	10	11	16	14	11	9	17	**19**	13	14	12	14	16	12	10	10	8	**18**
$D(0.4)$	30	35	31	30	40	37	26	18	39	**44**	35	37	34	39	**41**	32	31	28	17	**44**
$D(0.6)$	52	62	58	54	66	65	44	32	62	**69**	62	66	60	67	68	58	56	51	26	**70**
$D(0.8)$	70	81	79	76	84	83	63	48	80	**87**	81	85	79	86	86	78	75	70	37	**88**
$HN(0.8)$	49	59	54	62	55	54	67	34	**77**	**80**	58	61	41	59	65	52	41	33	56	72
$HN(1)$	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 $n = 20$.

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

	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	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
$LFR(0.8)$	28	32	13	37	21	5	64	21	**66**	**69**	31	28	0	17	32	17	7	26	48	53
$LFR(1)$	29	33	14	38	22	6	64	22	**67**	**71**	32	29	0	18	33	18	8	27	49	54
$BE(0.5)$	20	21	**35**	16	**34**	**34**	3	26	4	3	22	18	32	17	20	22	28	24	7	16
$BE(0.8)$	8	8	4	10	6	5	22	11	**25**	**25**	7	5	2	2	6	3	2	7	15	16
$BE(1)$	15	15	6	19	9	3	39	13	**42**	**45**	15	13	0	7	15	8	3	14	27	31
$BE(1.5)$	34	40	17	43	25	6	68	22	**69**	**74**	39	37	0	25	41	24	11	37	49	61
$TP(0.5)$	5	4	3	5	3	4	8	6	**9**	**10**	4	3	2	2	4	3	2	6	6	7
$TP(1,1)$	7	6	2	7	4	3	15	6	**16**	**17**	6	5	1	3	6	3	2	8	9	12
$TP(2,1)$	11	11	4	12	6	2	24	8	**25**	**29**	11	9	0	5	12	6	2	15	15	22
$TP(3)$	17	18	6	18	9	3	**34**	11	**34**	**39**	17	15	0	8	18	10	4	21	21	31
$D(0.2)$	5	5	2	5	3	3	12	6	**14**	**15**	5	4	1	2	5	3	2	8	8	10
$D(0.4)$	12	12	4	12	6	2	26	8	**28**	**31**	12	10	0	6	13	7	3	17	16	23
$D(0.6)$	20	22	8	21	12	3	40	11	**41**	**47**	22	20	0	12	23	13	6	27	24	38
$D(0.8)$	29	33	14	33	19	5	**55**	16	54	**62**	32	30	0	21	34	20	11	37	35	51
$HN(0.8)$	32	37	17	43	26	6	69	24	**71**	**74**	36	32	0	21	37	21	8	29	54	58
$HN(1)$	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 $n = 20$.

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

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.

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

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. (2022) as well as Brazauskas & Serfling (2003). 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 $j \in \{1, \ldots, k\}$. 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.

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

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 Section 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 $\widehat{\beta}_n = 0.764$ while the MME is $\tilde{\beta}_n = 1.194$. In the case where both β and σ are estimated; the MLEs are $\widehat{\beta}_n = 0.796$ and $\widehat{\sigma}_n = 1.053$. The corresponding MMEs are $\tilde{\beta}_n = 1.202$ and $\widehat{\sigma}_n = 1.031$. 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 Section 3.2. In the case of MLEs, p-values are approximated using a Monte Carlo procedure; for details, see the discussion in Section 3.1. In both cases 10 000 samples are generated from the Pareto distribution. The results associated with each of the tests considered in Section 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.

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	21685
$I_n^{[2]}$	0.051	0.051	0.053	0.051	36908
$L_{n,3,2}$	0.179	0.026	0.289	0.312	22

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 of exists. As a result, we put more stock in the results obtained using the MLE than those obtained in the case of the 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 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 Kolmogov-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.

7 Statements and declarations

No financial or non-financial are directly or indirectly related to this work. None of the authors has any conflict of interest whatsoever.

8 Appendix

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

Table 10: Numerical powers when estimating 1 parameter 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	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
$LFR(0.8)$	60	73	68	77	70	66	82	38	**87**	**92**	71	75	51	73	76	65	52	41	73	83
$LFR(1)$	63	75	71	80	72	69	83	41	**88**	**93**	73	77	54	75	79	67	55	44	75	85
$BE(0.5)$	27	30	**54**	23	53	**60**	3	25	2	2	31	25	49	24	29	36	46	53	6	19
$BE(0.8)$	14	16	14	20	15	16	28	12	**30**	**34**	16	16	8	13	16	11	8	6	23	23
$BE(1)$	35	43	38	46	41	38	52	21	**60**	**68**	42	46	29	44	47	36	28	21	44	56
$BE(1.5)$	85	94	93	94	95	94	89	62	95	**98**	93	96	90	96	96	92	87	81	75	**97**
$TP(0.5)$	9	9	7	9	10	9	9	7	**12**	**13**	9	9	8	9	11	9	7	7	7	**12**
$TP(1,1)$	15	18	15	16	17	15	14	9	19	**24**	18	18	14	17	21	16	13	12	13	**23**
$TP(2,1)$	30	37	32	33	34	31	27	14	36	**44**	36	38	29	37	41	34	28	24	24	**45**
$TP(3)$	45	54	48	51	48	46	39	21	49	**60**	54	56	43	55	58	50	42	37	35	**62**
$D(0.2)$	16	19	16	16	22	19	14	10	21	**24**	19	20	18	20	**23**	18	16	15	10	**24**
$D(0.4)$	42	50	47	44	55	53	34	22	49	56	50	54	50	56	**57**	49	46	41	25	**59**
$D(0.6)$	70	81	79	76	84	83	59	41	75	84	80	85	82	**86**	**86**	79	76	71	43	**86**
$D(0.8)$	89	95	95	93	**97**	96	79	62	91	96	95	**97**	95	**97**	**97**	94	93	89	58	**97**
$HN(0.8)$	68	79	75	84	77	73	87	46	**91**	**95**	78	81	57	78	82	72	59	47	79	87
$HN(1)$	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 1 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**
$LFR(0.8)$	73	81	79	84	69	72	70	41	**85**	50	80	82	78	**85**	76	65	52	41	73	**84**
$LFR(1)$	75	83	81	**86**	72	75	74	42	**87**	58	82	84	81	**86**	79	67	55	44	75	**86**
$BE(0.5)$	15	15	42	12	46	**57**	3	25	1	0	18	11	15	4	29	36	46	**53**	6	6
$BE(0.8)$	23	26	25	**33**	16	19	18	12	26	3	26	26	18	30	16	11	8	6	23	**31**
$BE(1)$	53	61	59	**64**	46	49	32	24	55	11	59	62	53	**67**	47	36	28	21	44	**67**
$BE(1.5)$	95	98	98	98	97	98	61	70	95	48	98	**99**	98	**99**	96	92	87	81	75	**99**
$TP(0.5)$	33	38	**43**	**46**	37	40	19	15	8	0	37	39	21	38	11	9	7	7	7	32
$TP(1,1)$	59	66	**71**	**74**	66	69	26	28	14	0	65	67	42	66	21	16	13	12	13	50
$TP(2,1)$	88	93	**95**	**96**	92	94	40	59	27	0	92	93	73	92	41	34	28	24	24	76
$TP(3)$	97	**99**	**99**	**99**	**99**	**99**	55	81	40	0	98	**99**	88	**99**	58	50	42	37	35	89
$D(0.2)$	33	40	40	40	38	36	4	14	16	1	38	**41**	32	**43**	23	18	16	15	10	39
$D(0.4)$	67	75	76	72	74	73	6	31 .	42	4	74	**77**	72	**79**	57	49	46	41	25	**77**
$D(0.6)$	89	94	94	92	93	93	16	53	70	11	93	**95**	93	**95**	86	79	76	71	43	**95**
$D(0.8)$	97	**99**	**99**	**99**	**99**	**99**	30	74	89	25	**99**	**99**	**99**	**99**	97	94	93	89	58	**100**
$HN(0.8)$	77	85	84	88	75	78	79	46	**90**	64	84	86	83	**88**	82	72	59	47	79	87
$HN(1)$	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
$LFR(0.8)$	47	56	37	64	49	15	**87**	30	83	**89**	55	55	8	44	55	38	20	38	72	75
$LFR(1)$	50	60	40	67	52	17	**88**	33	84	**91**	58	58	9	47	58	40	22	40	74	77
$BE(0.5)$	27	30	**49**	23	**49**	**49**	3	33	4	3	31	26	44	24	30	32	40	39	8	25
$BE(0.8)$	12	12	7	16	12	7	**34**	13	33	**37**	12	10	2	6	9	5	3	7	23	22
$BE(1)$	25	29	15	34	22	6	**61**	17	56	**64**	28	28	2	20	27	16	8	20	43	46
$BE(1.5)$	60	70	52	73	61	24	**90**	35	85	**92**	69	71	19	63	71	55	38	59	73	83
$TP(0.5)$	6	5	3	6	4	3	**12**	6	11	**13**	5	5	1	3	5	3	2	6	7	9
$TP(1,1)$	10	10	4	10	6	3	**19**	7	17	**22**	10	9	1	6	10	6	3	11	12	16
$TP(2,1)$	19	21	10	22	13	4	**36**	10	31	**39**	21	20	2	15	21	13	7	21	23	32
$TP(3)$	31	35	19	34	21	6	**51**	14	42	**54**	34	33	4	26	34	24	14	32	35	49
$D(0.2)$	8	9	4	8	5	3	**18**	6	16	**20**	8	8	1	5	9	5	3	12	10	15
$D(0.4)$	21	24	11	23	15	4	**41**	10	35	**44**	24	24	3	18	25	16	10	26	24	36
$D(0.6)$	39	45	27	44	31	9	**64**	17	55	**67**	45	45	9	38	46	34	22	44	41	59
$D(0.8)$	54	63	44	62	49	17	**79**	25	70	**82**	62	63	18	57	63	50	36	59	56	76
$HN(0.8)$	55	65	46	72	58	20	**91**	37	88	**93**	63	63	10	52	63	45	25	44	78	81
$HN(1)$	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 $n = 30$.

	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	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
$LFR(0.8)$	69	77	77	80	77	**81**	66	31	**81**	43	76	78	75	**82**	55	38	20	38	72	71
$LFR(1)$	70	79	79	82	79	**83**	70	33	**83**	49	78	80	78	**83**	58	40	22	40	74	77
$BE(0.5)$	11	10	19	10	14	18	3	33	3	0	12	7	9	3	30	32	**40**	**39**	8	56
$BE(0.8)$	24	28	27	**34**	23	28	18	13	29	3	27	28	22	**33**	9	5	3	7	23	60
$BE(1)$	49	57	57	**61**	53	57	30	18	51	9	56	59	52	**63**	27	16	8	20	43	63
$BE(1.5)$	84	90	90	90	88	90	52	39	83	32	89	**91**	89	**92**	71	55	38	59	73	63
$TP(0.5)$	31	36	41	**44**	43	**44**	19	12	7	0	35	37	20	36	5	3	2	6	7	6
$TP(1,1)$	55	62	68	70	**71**	**71**	25	22	12	0	61	63	38	62	10	6	3	11	12	27
$TP(2,1)$	84	89	92	**93**	**93**	**93**	37	47	22	0	88	89	67	89	21	13	7	21	23	49
$TP(3)$	95	97	98	98	**99**	**99**	50	68	31	0	96	97	83	97	34	24	14	32	35	68
$D(0.2)$	28	33	**34**	33	32	32	4	8	12	1	31	**34**	25	**35**	9	5	3	12	10	21
$D(0.4)$	51	60	60	58	57	57	6	14	29	3	58	**61**	53	**63**	25	16	10	26	24	32
$D(0.6)$	70	77	78	76	75	75	13	22	48	8	76	**79**	73	**81**	46	34	22	44	41	46
$D(0.8)$	81	87	87	86	85	86	25	32	66	17	87	**88**	85	**90**	63	50	36	59	56	51
$HN(0.8)$	73	81	81	**85**	82	**85**	75	35	**87**	55	80	82	80	**85**	63	45	25	44	78	23
$HN(1)$	81	88	88	90	87	90	78	43	**91**	52	87	88	85	**91**	70	52	30	49	83	41

References

Ahsanullah, M. (1978). A characterization of the exponential distribution by spacings. *Journal of Applied Probability*, *15*(3), 650–653.

Alizadeh Noughabi, H. (2015). Testing exponentiality based on the likelihood ratio and power comparison. *Annals of Data Science*, *2*(2), 195–204.

Alizadeh Noughabi, H., & Balakrishnan, N. (2016). Tests of goodness of fit based on phi-divergence. *Journal of Applied Statistics*, *43*(3), 412–429.

Alizadeh Noughabi, R., Alizadeh Noughabi, H., & Behabadi, A. E. M. (2014). An entropy test for the rayleigh distribution and power comparison. *Journal of Statistical Computation and Simulation*, *84*(1), 151–158.

Allison, J. S., Milošević, B., Obradović, M., & Smuts, M. (2022). Distribution-free goodness-of-fit tests for the Pareto distribution based on a characterization. *Computational Statistics*, *37*(1), 403–418.

Allison, J. S., Santana, L., Smit, N., & Visagie, I. J. H. (2017). An 'apples to apples' comparison of various tests for exponentiality. *Computational Statistics*, *32*(4), 1241–1283.

Arnold, B. C. (2015). *Pareto Distributions*. CRC Press, New York.

Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, *85*, 549—-559.

Bera, A. K., Galvao, A. F., Wang, L., & Xiao, Z. (2016). A new characterization of the normal distribution and test for normality. *Econometric Theory*, *32*(5), 1216–1252.

Beran, R. J. (1977). Minimum Pareto distance estimates for parameter models. *The Annals of Statistics*, *5*, 445–463.

Beran, R. J. (1978). An efficient and robust adaptive estimator of location. *The Annals of Statistics*, *6*, 292–313.

Betsch, S., & Ebner, B. (2020). Testing normality via a distributional fixed point property in the Stein characterization. *TEST*, *29*(1), 105–138.

Boos, D. (1981). Minimum distance estimators for location and goodness of fit. *Journal of the American Statistical Association*, *76*, 663–670.

Brazauskas, V., & Serfling, R. (2003). Favorable estimators for fitting Pareto models: A study using goodness-of-fit measures with actual data. *ASTIN Bulletin: The Journal of the IAA*, *33*(2), 365–381.

Choi, B., & Kim, K. (2006). Testing goodness-of-fit for Laplace distribution based on maximum entropy. *Statistics*, *40*(6), 517–531.

Chu, J., Dickin, O., & Nadarajah, S. (2019). A review of goodness of fit tests for Pareto distributions. *Journal of Computational and Applied Mathematics*, *361*, 13–41.

Cockeran, M., Meintanis, S. G., & Allison, J. S. (2021). Goodness-of-fit tests in the Cox proportional hazards model. *Communications in Statistics - Simulation and Computation*, *50*(12), 4132–4143.

Csiszár, I. (1967). On topological properties of f-divergences. *Studia Scientiarum Mathematicarum Hungarica*, *2*, 329–339.

Csörgö, S., & Szabó, T. (2009). Weighted quantile correlation tests for Gumbel, Weibull and Pareto families. *Probab. Math. Stat.*, *29*, 227–250.

D'Agostino, R., & Stephens, M. (1986). *Goodness-of-fit Techniques*. Marcel Dekker, New York.

Giacomini, R., Politis, D. N., & White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, *29*, 567–589.

Gulati, S., & Shapiro, S. (2008). Goodness-of-fit tests for Pareto distribution. *Stat. Models Methods Biomed. Tech. Syst.*, *25*, 259–274.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistic. *Journal of the Royal Statistical Society*, *52*, 105–124.

Ismaïl, S. (2004). A simple estimator for the shape parameter of the Pareto distribution with economics and medical applications. *Journal of Applied Statistics*, *31*(1), 3–13.

Jahanshahi, S., Rad, A. H., & Fakoor, V. (2016). A goodness-of-fit test for Rayleigh distribution based on Hellinger distance. *Annals of Data Science*, *3*(4), 401–411.

Kapur, J. N. (1994). *Measures of Information and Their Applications*. Wiley-Interscience, New York.

Klar, B. (2001). Goodness-of-fit tests for the exponential and the normal distribution based on the integrated distribution function. *Annals of the Institute of Statistical Mathematics*, *53*(2), 338–353.

Klar, B., & Meintanis, S. G. (2005). Tests for normal mixtures based on the empirical characteristic function. *Computational Statistics & Data Analysis*, *49*(1), 227–242.

Kullback, S. (1997). *Information Theory and Statistics*. Dover Publications, Inc., New York.

Lequesne, J. (2013). Entropy-based goodness-of-fit test: Application to the pareto distribution. In *AIP Conference Proceedings*, vol. 1553, (pp. 155–162). American Institute of Physics.

Meintanis, S. G. (2009a). Goodness-of-fit tests and minimum distance estimation via optimal transformation to uniformity. *Journal of Statistical Planning and Inference*, *139*(2), 100–108.

Meintanis, S. G. (2009b). A unified approach of testing for discrete and continuous Pareto laws. *Statistical Papers*, *50*(3), 569–580.

Meintanis, S. G. (2016). A review of testing procedures based on the empirical characteristic function. *South African Statistical Journal*, *50*(1), 1–14.

Meintanis, S. G., Gamero, M. D. J., & Alba-Fernández, V. (2014). A class of goodness-of-fit tests based on transformation. *Communications in Statistics-Theory and Methods*, *43*(8), 1708–1735.

Milošević, B., & Obradović, M. (2016). Two-dimensional kolmogorov-type goodness-of-fit tests based on characterisations and their asymptotic efficiencies. *Journal of Nonparametric Statistics*, *28*(2), 413–427.

Ndwandwe, L., Allison, J., Smuts, M., & Visagie, I. (2023). On a new class of tests for the pareto distribution using fourier methods. *Stat*, (p. e566).

Ndwandwe, L., Allison, J. S., & Visagie, I. J. H. (2021). A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring. *Annual Proceedings of the South African Statistical Association Conference*, *1*, 17–23.

Nofal, Z. M., & El Gebaly, Y. M. (2017). New characterizations of the Pareto distribution. *Pakistan Journal of Statistics and Operation Research*, *13*, 63–74.

Obradović, M. (2015). On asymptotic efficiency of goodness of fit tests for Pareto distribution based on characterizations. *Filomat*, *29*(10), 2311–2324.

Obradović, M., Jovanović, M., & Milošević, B. (2015). Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics. *Statistics*, *49*(5), 1026–1041.

Pareto, V. (1897). *Cours d'economie Politique, Vol. II*. Lausanne: F. Rouge.

Parr, W. C., & Schucany, W. R. (1980). Minimum distance and robust estimation. *Journal of the American Statistical Association*, *75*, 616–624.

Quandt, R. E. (1964). Statistical discrimination among alternative hypotheses and some economic regularities. *Journal of Regional Science*, *5*(2), 1–23.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Rizzo, M. L. (2009). New goodness-of-fit tests for Pareto distributions. *Astin Bull.*, *39*, 691–715.

Rossberg, H. (1972). 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.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Silverman, B. W. (2018). *Density Estimation for Statistics and Data Analysis*. Routledge, London.

Taufer, E., Santi, F., Espa, G., & Dickson, M. M. (2021). Graphical representations and associated goodness-of-fit tests for Pareto and log-normal distributions based on inequality curves. *Journal of Nonparametric Statistics*, *33*(3-4), 464–481.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(1), 54–59.

Volkova, K. (2016). Goodness-of-fit tests for the Pareto distribution based on its characterization. *Statistical Methods & Applications*, *25*(3), 351–373.

Wolfowitz, J. (1953). Estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics*, *5*, 9–23.

Zenga, M. (1984). 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.

Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 281–294.

Chapter 3

Article 2: On a new class of tests for the Pareto distribution using Fourier methods

The second article, *On a new class of tests for the Pareto distribution using Fourier methods*, has been accepted for publication in *Stat*. A summary of the guidelines to authors from the journal is now presented.

Manuscript	10 pages excluding tables, figures and bibliography.
Title	The title should be concise and informative.
Abstract and keywords	An abstract of 150 to 250 words and 4 to 6 keywords which can be used for indexing purposes.
Tables	All tables are to be numbered using Arabic numerals. Tables should always be cited in text in consecutive numerical order. For each table, please supply a table caption explaining the components of the table.
References	References should be prepared according to the Publication Manual of the American Psychological Association (6th edition).
General formatting	A LaTeXtemplate is provided for submission.
Additional information	https://tinyurl.com/y4lu7f9f

Received: Added at production | Revised: Added at production | Accepted: Added at production
DOI: xxx/xxxx

ARTICLE TYPE

On a new class of tests for the Pareto distribution using Fourier methods

L. Ndwandwe | J.S. Allison | M. Smuts | I.J.H. Visagie

[1]Subject Group Statistics, North-West University, North West, South Africa

Correspondence
*Jaco Visagie, Email: jaco.visagie@nwu.ac.za
+27 18 299 2571

Present Address
1 Hoffman Street, Potchefstroom, South Africa

Summary

We propose new classes of tests for the Pareto type I distribution using the empirical characteristic function. These tests are U and V statistics based on a characterisation of the Pareto distribution involving the distribution of the sample minimum. In addition to deriving simple computational forms for the proposed test statistics, we prove consistency against a wide range of fixed alternatives. A Monte Carlo study is included in which the newly proposed tests are shown to produce high powers. These powers include results relating to fixed alternatives as well as local powers against mixture distributions. The use of the proposed tests is illustrated using an observed data set.

KEYWORDS:
Empirical characteristic function Goodness-of-fit testing Pareto distribution V and U statistics.

1 | INTRODUCTION

The Pareto distribution, nowadays commonly known as the Pareto type I distribution, was originally introduced by Pareto (1897). Since then several extensions of this distribution have been proposed. These extensions are achieved via the inclusion of a location, scale and inequality parameter, corresponding to the Pareto type II, III and IV distributions, respectively. Additionally, a so-called generalised Pareto distribution has been introduced. For an in depth discussion of the various types of Pareto distributions as well as the relationships between them, the interested reader is referred to Arnold (2015).

The Pareto distribution is a popular model in engineering, economics, finance and actuarial science, especially where phenomena characterised by heavy tails are studied, see, e.g. Fisk (1961), Ismaïl (2004), Nofal and El Gebaly (2017). Concrete examples of the use of the Pareto distribution include the modelling of failure times of mechanical components, see Bourguignon, Saulo, and Fernandez (2016), as well as the modelling of excess of losses in insurance claims, see Rytgaard (1990). Due to its heavy tail, this distribution also plays a pivotal role in extreme value theory see, Beirlant, Goegebeur, Segers, and Teugels (2004). Further examples of the use of the Pareto distribution can be found in Amin (2007) and Soliman (2000). A number of characterisations for the Pareto distribution have been developed in the literature, see, e.g. Gupta (1973) as well as Pareto (1897).

Due to the popularity of this distribution as well as its wide range of applications, goodness-of-fit tests have been developed in order to test the hypothesis that an observed data set is compatible with the assumption of being realised from this distribution. For recent overview papers and discussions of some of these tests, see Chu, Dickin, and Nadarajah (2019) and Ndwandwe, Allison, Santana, and Visagie (2022) as well as the references therein. Chu et al. (2019) review tests for the generalised Pareto distribution as well as the Pareto types I and II, whereas Ndwandwe et al. (2022) investigates several existing tests specifically for the Pareto type I distribution. Although tests exist for the Pareto type I distribution, they are few in number when compared to those for other distributions such as, for example, the normal or exponential distributions. In the remainder of this paper, we refer to the Pareto type I distribution simply as the Pareto distribution.

We propose new classes of goodness-of-fit tests for the Pareto distribution based on a characterisation involving the distribution of the sample minimum. In order to proceed, we introduce some notation. Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) realisations from a continuous random variable, X, with unknown distribution function F. X is said to follow the Pareto distribution with shape parameter β, denoted by $X \sim P(\beta)$, if it has distribution function

$$F(x) = 1 - x^{-\beta}, \quad x \geq 1,$$

for some $\beta > 0$. The composite null hypothesis to be tested is

$$H_0 : X \sim P(\beta), \tag{1.1}$$

for some unspecified $\beta > 0$. This hypothesis is tested against general alternatives. Throughout this paper, the value of β is estimated by its method of moments estimator $\widehat{\beta}_n = \overline{X}_n/(\overline{X}_n - 1)$, where $\overline{X}_n$ is the sample mean. We opt to use the method of moments since Ndwandwe et al. (2022) find that the competing tests considered in the Monte Carlo study of Section 4 generally produce higher powers when using this estimator than is the case when using maximum likelihood estimation.

The remainder of this paper is structured as follows. In Section 2, new classes of tests are proposed for the Pareto distribution, whereas Section 3 contains theoretical results pertaining to the asymptotic behaviour of the tests. A Monte Carlo study is presented in Section 4, while Section 5 contains an example pertaining to observed data. The paper concludes in Section 6.

2 | A NEW CLASS OF TESTS FOR THE PARETO DISTRIBUTION

The proposed tests are based on a characterisation of the Pareto distribution via the distribution of the sample minimum. This characterisation is discussed in Allison, Milošević, Obradović, and Smuts (2022) and is as follows:

Theorem 1. Let $X, X_1, X_2 \dots$ be i.i.d. random variables from a continuous distribution with distribution function F. Let $m \geq 2$ be an integer. $X^{1/m}$ and $\min(X_1, \dots, X_m)$ have the same distribution if, and only if,

$$F(x) = 1 - x^{-\beta}, \quad x \geq 1,$$

for some $\beta > 0$.

Since a random variable is characterised by its Fourier transform, we can base a test on the V and U empirical characteristic functions of $X^{1/m}$ and $\min(X_1, ..., X_m)$. Tests based on these quantities have been shown to produce high powers in finite sample settings; the interested reader is referred to Meintanis (2016).

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 $\min(X_1, \ldots, X_m)$, respectively. Denote the empirical versions of ϕ_m and ξ_m by

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

and

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

where $X_{1:n} < \cdots < X_{n:n}$ denote the order statistics of $X_1, \ldots, X_n$. Theorem 1 implies that, for all $t \in \mathbb{R}$ and $m \le n$, $\phi_m(t) = \xi_m(t)$ if, and only if, $X \sim P(\beta)$ for some $\beta > 0$. We propose a class of tests for the hypothesis in (1.1) based on a weighted L2 distance between $\phi_{n,m}$ and $\xi_{n,m}$:

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

where $w_a(t)$ is a weight function that ensures the existence of the integral and a is an user defined tuning parameter. Popular choices of w_a include the Laplace and Gaussian kernels.

Note that

$$S_{n,m,a} = \frac{n}{n^{2m}}\sum_{k_1=1}^{n}\cdots\sum_{k_{2m}=1}^{n} h(X_{k_1}, \ldots, X_{k_{2m}}; a), \tag{2.1}$$

where

$$\begin{aligned} h(X_{k_1}, \ldots, X_{k_{2m}}; a) \; = & \int_{-\infty}^{\infty} \Big[\cos\Big(t(X_{k_1}^{1/m} - X_{k_2}^{1/m})\Big) - 2\cos\Big(t(X_{k_1}^{1/m} - \min(X_{k_1}, \ldots, X_{k_m}))\Big) \\ & + \; \cos\Big(t(\min(X_{k_1}, \ldots, X_{k_m}) - \min(X_{k_{m+1}}, \ldots, X_{k_{2m}}))\Big)\Big] w_a(t)\mathrm{d}t. \end{aligned}$$

Therefore, $S_{n,m,a}$ is a V statistic of order $2m$ with kernel h. The form of $S_{n,m,a}$ specified in (2.1) is computationally expensive (e.g., if $m = 4$, then computing $S_{n,m,a}$ requires the evaluation of an eight fold summation). However, after some combinatorics and algebraic manipulation, $\xi_{n,m}(t)$ can be expressed as a single sum in terms of the order statistics:

$$\xi_{n,m}(t) = \sum_{j=1}^{n} v_{j,m} e^{itX_{j:n}},$$

with

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

When using a Laplace kernel as weight function, $w_a(t) = e^{-a|t|}$, we denote the resulting test statistic by $S_{n,m,a}^{(1)}$;

$$\begin{aligned} S_{n,m,a}^{(1)} = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} & \left\{ \frac{2a}{a^2 + \left(X_{j:n}^{1/m} - X_{k:n}^{1/m}\right)^2} - nv_{j,m}\frac{4a}{a^2 + \left(X_{j:n} - X_{k:n}^{1/m}\right)^2} \right. \\ & \left. + n^2 v_{j,m} v_{k,m}\frac{2a}{a^2 + \left(X_{j:n} - X_{k:n}\right)^2} \right\}. \end{aligned}$$

Upon setting the weight function equal to a Gaussian kernel, $\widetilde{w}_a(t) = e^{-at^2}$, we obtain $S_{n,m,a}^{(2)}$;

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

Above we consider $S_{n,m,a}$, based on V statistics. We now turn our attention to the situation where the empirical characteristic functions are estimated using U statistics. Denote the difference between the U empirical characteristic functions of $X^{1/m}$ and $\min(X_1, \ldots, X_m)$ by

$$D_{n,m}(t) = \phi_{n,m}(t) - \psi_{n,m}(t),$$

where

$$\psi_{n,m}(t) = \binom{n}{m}^{-1} \sum_{1 \le k_1 < \cdots < k_m \le n} \mathrm{e}^{it\min(X_{k_1},\ldots,X_{k_m})}$$

is the empirical characteristic function of the $\binom{n}{m}$ random variables $\min(X_{k_1}, \ldots, X_{k_m})$, $1 \le k_1 < \cdots < k_m \le n$.

After some algebra it follows that $\psi_{n,m}(t)$ can be expressed as a single summation;

$$\psi_{n,m}(t) = \binom{n}{m}^{-1} \sum_{j=1}^{n-m+1} u_{j,m} \mathrm{e}^{itX_{j:n}},$$

where

$$u_{j,m} = \binom{n-j}{m-1}.$$

From Theorem 1 it follows that, if $X_1, \ldots, X_n$ is a random sample from the Pareto distribution, then the difference between $\phi_{n,m}(t)$ and $\psi_{n,m}(t)$ should be close to zero. We thus suggest the test statistic

$$T_{n,m,a} = n \int_{-\infty}^{\infty} |\phi_{n,m}(t) - \psi_{n,m}(t)|^2 w_a(t)\mathrm{d}t. \tag{2.2}$$

After some algebra, we obtain the following easily calculable expression for the test statistic based on the choices $w_a(t) = e^{-a|t|}$ and $\widetilde{w}_a(t) = e^{-at^2}$, respectively:

$$\begin{aligned} T^{(1)}_{n,m,a} = {} & \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{2a}{a^2 + \left(X_{j:n}^{1/m} - X_{k:n}^{1/m}\right)^2} - \sum_{j=1}^{n-m+1} \sum_{k=1}^{n} \binom{n}{m}^{-1} u_{j,m} \frac{4a}{a^2 + \left(X_{j:n} - X_{k:n}^{1/m}\right)^2} \\ & + n \sum_{j=1}^{n-m+1} \sum_{k=1}^{n-m+1} \binom{n}{m}^{-2} u_{j,m} u_{k,m} \frac{2a}{a^2 + \left(X_{j:n} - X_{k:n}\right)^2}, \end{aligned}$$

and

$$\begin{aligned} T^{(2)}_{n,m,a} = \sqrt{\frac{\pi}{a}} \Bigg\{ & \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \exp\left[\frac{-\left(X_{j:n}^{1/m} - X_{k:n}^{1/m}\right)^2}{4a}\right] \\ & -2 \sum_{j=1}^{n-m+1} \sum_{k=1}^{n} \binom{n}{m}^{-1} u_{j,m} \exp\left[\frac{-\left(X_{j:n} - X_{k:n}^{1/m}\right)^2}{4a}\right] \\ & +n \sum_{j=1}^{n-m+1} \sum_{k=1}^{n-m+1} \binom{n}{m}^{-2} u_{j,m} u_{k,m} \exp\left[\frac{-\left(X_{j:n} - X_{k:n}\right)^2}{4a}\right] \Bigg\}. \end{aligned}$$

Since both the tests $T_{n,m,a}$ and $S_{n,m,a}$ are based on a weighted L2 distance, they reject the null hypothesis for large values of the test statistic.

3 | CONSISTENCY OF THE TESTS

In this section we only present the results pertaining to $T_{n,m,a}$; the derivations relating to $S_{n,m,a}$ follows from analogous arguments and are therefore omitted for the sake of brevity. Before proceeding to prove the consistency of $T_{n,m,a}$, some comments about the asymptotic null distribution of the test statistic are in order.

$T_{n,m,a}$ is formulated as a weighted L2-type statistic involving empirical characteristic functions. The asymptotic null distribution of these classes of statistics are studied in, amongst others, Feuerverger and Mureika (1977), Baringhaus and Henze (1988), Klar and Meintanis (2005) as well as Baringhaus, Ebner, and Henze (2017). The asymptotic null distribution of $T_{n,m,a}$ will typically correspond to that of $\int_{-\infty}^{\infty} |V(t)|^2 w_a(t)dt =: T_{m,a}$, where $V(\cdot)$ is a Gaussian process with zero-mean. $T_{m,a}$ has the same distribution as $\sum_{j=1}^{\infty} \lambda_j \chi_j^2$, where χ_j^2 are i.i.d random variables following a chi-squared distribution with one degree of freedom. However, the covariance matrix of $V(\cdot)$ as well as the eigenvalues λ_j depend on the unknown underlying distribution F, usually in a complicated way.

The following theorem is concerned with the asymptotic behaviour of $T_{n,m,a}$ under fixed alternative distributions.

Theorem 2. Let $X_1, ..., X_n$ be independent copies of a continuous random variable X with finite mean, then

$$\frac{T_{n,m,a}}{n} \xrightarrow{p} \Delta_{m,w} := \int_{-\infty}^{\infty} |\phi_m(t) - \psi_m(t)|^2 w_a(t)\mathrm{d}t \geq 0, \tag{3.1}$$

as $n \to \infty$, with $\Delta_{m,w} = 0$ if, and only if, $X \sim P(\beta)$.

Proof. Recall from (2.2) that

$$\frac{T_{n,m,a}}{n} = \int_{-\infty}^{\infty} D^2_{n,m}(t) w_a(t)\mathrm{d}t,$$

where $D^2_{n,m}(t) = |\phi_{n,m}(t) - \psi_{n,m}(t)|^2$. We have, by the law of large numbers, that $\phi_{n,m}(t) \xrightarrow{p} \phi_m(t)$ and that $\psi_{n,m}(t) \xrightarrow{p} \psi_m(t)$. By the continuous mapping theorem it follows that $D^2_{n,m}(t) \xrightarrow{p} D^2_m(t) := |\phi_m(t) - \psi_m(t)|^2$. Now, $D^2_{n,m}(t) \leq 4$, hence an application of Lebesgue's theorem of dominated convergence yields (3.1). In view of the characterisation given in Theorem 1, it follows that $\Delta_{m,w}$ is zero if, and only if $X \sim P(\beta)$ □

Under the alternative hypothesis we have that $\Delta_{m,w} > 0$. As a result $T_{n,m,a} \xrightarrow{p} \infty$ as $n \longrightarrow \infty$, which proves the consistency of the proposed test.

4 | MONTE CARLO STUDY

In this section Monte Carlo simulations are used to compare the finite sample performance of the newly proposed tests to the following seven existing goodness-of-fit tests for the Pareto distribution:

- The traditional Kolmogorov-Smirnov (KS_n), Cramér-von Mises (CM_n) and Anderson-Darling (AD_n) tests.

- A test, proposed by Zhang (2002), based on the likelihood ratio. The test statistic is given by

$$ZA_n = \int_{-\infty}^{\infty} G^2_n(t)\left[F_n(t)\left(1 - F_n(t)\right)\right]^{-1} dF_n(t),$$

  where

$$G^2_n(t) = 2n\left\{F_n(t)\log\left(\frac{F_n(t)}{F(t,\hat{\beta}_n)}\right) + [1 - F_n(t)]\log\left(\frac{1 - F_n(t)}{1 - F(t,\hat{\beta}_n)}\right)\right\}$$

  is the likelihood ratio statistic and $F_n(t) = \frac{1}{n}\sum_{j=1}^{n} I(X_j \leq t)$. The computational form of the test statistic is

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

- A test based on Mellin transform proposed by Meintanis (2009). The test statistic is given by

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

where

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

Choosing $\mathsf{w(x) = e^{-ax}}$, one has

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

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

and

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

The value of the tuning parameter a is set to $\mathsf{1}$ in order to obtain the numerical results presented.

- A test based on the empirical characterisation function proposed by Meintanis, Gamero, and Alba-Fernández (2014). The test is a weighted $\mathsf{L2}$ distance between the empirical characteristic function of transformed data and the characteristic function of the standard uniform distribution. The simplified version of the test statistic is given by

$$\begin{aligned} \mathsf{M_{n,a}} \; = \; & \frac{1}{\mathsf{n}} \sum_{\mathsf{j,k=1}}^{\mathsf{n}} \frac{\mathsf{2a}}{(\widehat{\mathsf{U}}_\mathsf{j} - \widehat{\mathsf{U}}_\mathsf{k})^2 + \mathsf{a}^2} + \mathsf{2n}\left[2\tan^{-1}\left(\frac{1}{\mathsf{a}}\right) - \mathsf{a}\log\left(1+\frac{1}{\mathsf{a}^2}\right)\right] \\ & - 4\sum_{\mathsf{j=1}}^{\mathsf{n}}\left[\tan^{-1}\left(\frac{\widehat{\mathsf{U}}_\mathsf{j}}{\mathsf{a}}\right) + \tan^{-1}\left(\frac{1-\widehat{\mathsf{U}}_\mathsf{j}}{\mathsf{a}}\right)\right], \end{aligned}$$

  where

$$\widehat{U}_j = F_{\widehat{\beta}_n}(X_j), j = 1, ..., n.$$

  We implement the test for $\mathsf{a = 1}$.

- A test proposed by Allison et al. (2022), which is also based on the characterisation given in Theorem 1, but makes use of empirical distribution functions instead of empirical characteristic functions. The test statistic is given by

$$I_{n,m} = \int_1^\infty \Delta_{n,m}(x)\mathrm{d}F_n(x),$$

  where

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

  We implement the test for m=2.

All tests rejects the null hypothesis for large values of the test statistic.

4.1 | Simulation setting

Power (and size) estimates are calculated at a significance level of 5% for sample sizes $\mathsf{n = 20}$ and $\mathsf{n = 30}$ using 50 000 independent Monte Carlo replications. In order to avoid the use of asymptotic critical values and since the null distribution of the test statistic depend on the unknown value of the shape parameter β (even asymptotically), we use a parametric bootstrap procedure to calculate numerical critical values. For computational efficiency, we employ the warp-speed bootstrap methodology proposed by Giacomini, Politis, and White (2013) to calculate estimated powers. This methodology is outlined in the following algorithm:

1. Draw a sample of size n, say $\mathsf{X_1, \ldots, X_n}$ from an alternative distribution and estimate the parameter β by $\widehat{\beta}_\mathsf{n} = \overline{\mathsf{X}}_\mathsf{n}/(\overline{\mathsf{X}}_\mathsf{n} - 1)$.

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

TABLE 1 Alternative distributions used

2. Calculate the value of the test statistic say $S = S_n(X_1, \ldots, X_n)$.

3. Generate a bootstrap sample $X_1^*, \ldots, X_n^*$ by independently sampling from a $P(\widehat{\beta}_n)$ distribution. Calculate the value of the test statistic using the bootstrap sample, $S^* = S_n(X_1^*, \ldots, X_n^*)$.

4. Repeat steps 1–3, MC times to obtain $S_1, \ldots, S_{MC}$ and $S_1^*, \ldots, S_{MC}^*$, where S_j and S_j^* denote the values of the test statistic for the j^{th} sample generated in Steps 2 and 3, respectively.

5. The power estimate is given by $\frac{1}{MC}\sum_{j=1}^{MC} I\left(S_j > S^*_{\lfloor MC(1-\alpha)\rfloor:MC}\right)$ for $j = 1, \ldots, MC$, where $S^*_{j:MC}$ denotes the j^{th} order statistic of $S_1^*, \ldots, S_{MC}^*$, where $\lfloor\cdot\rfloor$ denotes the floor function and $I(\cdot)$ denotes the indicator function.

The simulation study presented considers two sets of power results. The first is concerned with powers against the fixed alternative distributions specified in Table 1. The resulting empirical powers for sample sizes of $n = 20$ and $n = 30$ can be found in Tables 2 and 3, respectively. Second, we consider some local power estimates where we simulate data from two families of mixture distributions. In the first of the mixture distributions used, we simulate from a $LN(1)$ with probability p, and from a Pareto distribution (with the same mean as the $LN(1)$) with probability $1 - p$; the empirical powers obtained are reported in Table 4. The second family of mixture distributions is obtained upon replacing the $LN(1)$ distribution by the exponential distribution with mean 0.5; the calculated powers can be found in Table 5. The results shown in Tables 4 and 5 include two powers for each listed distribution; the first is associated with a sample size of 20, while the second is the estimated power based on a sample of size 30.

The reported empirical powers of the new tests were obtained by setting $m = 3$ and $a = 2$ in all instances. Several other values for these parameters were considered when performing an extensive Monte Carlo study (over a wide range of sample sizes and numerous alternative distributions); however, these specified choices generally resulted in high powers. For the sake of brevity, we omit the results pertaining to other parameter configurations and only display those associated with $m = 3$ and $a = 2$. All calculations were performed in R; see R Core Team (2020).

4.2 | Simulation results

The power estimates in Tables 2 to 5 are the percentages (rounded to one decimal place) of the number of samples resulting in a rejection of the null hypothesis. For ease of comparison the highest two powers in each row (including ties) are printed in bold.

The results shown in Tables 2 and 3 indicate that all the tests maintain the nominal significance level of 0.05. Furthermore, the results demonstrate that the proposed test $S^{(2)}_{n,3,2}$ outperforms all the other tests for the majority of alternatives considered, closely followed by

TABLE 2 Empirical powers against fixed alternatives for $n = 20$

Dist	KS_n	CV_n	AD_n	ZA_n	$G_{n,2}$	$M_{n,1}$	$I_{n,2}$	$S^{(1)}_{n,3,2}$	$S^{(2)}_{n,3,2}$	$T^{(1)}_{n,3,2}$	$T^{(2)}_{n,3,2}$
P(2)	5.2	5.3	5.2	5.2	5.5	5.4	4.9	5.3	5.2	5.2	5.2
P(5)	5.1	5.1	5.0	4.9	5.2	5.1	4.7	4.3	4.3	4.2	4.3
P(10)	5.0	5.0	5.0	4.9	4.7	5.0	4.6	4.1	4.1	4.2	4.2
Γ(0.5)	15.0	15.3	**37.6**	**38.2**	9.1	12.8	30.6	4.4	2.9	11.9	9.9
Γ(0.8)	14.9	16.2	15.3	11.1	16.4	16.5	8.2	**20.5**	**20.4**	15.6	15.1
Γ(1)	38.4	45.0	42.5	34.0	47.1	46.2	24.9	**50.6**	**51.3**	42.6	42.4
Γ(1.2)	64.4	73.7	72.5	63.6	76.2	75.5	50.7	**76.6**	**77.9**	70.2	70.6
W(0.5)	16.2	15.1	42.4	**45.1**	12.6	11.2	**42.9**	18.9	15.1	29.7	25.2
W(0.8)	14.8	16.2	16.2	10.4	15.0	16.5	7.0	**19.8**	**20.0**	15.7	14.4
W(1.2)	65.1	74.7	72.7	64.9	78.2	76.3	55.1	**78.8**	**79.1**	72.6	72.4
W(1.5)	91.3	96.1	95.6	93.1	**97.4**	96.7	88.7	**97.0**	**97.0**	95.3	95.1
LN(1)	71.8	81.8	82.3	**90.2**	**85.9**	83.2	63.7	84.0	85.8	79.5	80.1
LN(1.2)	42.1	51.0	51.6	**59.6**	**55.2**	53.2	28.4	51.2	53.8	45.3	46.4
LN(1.5)	17.2	20.9	**22.2**	21.2	20.0	**21.8**	6.1	19.2	21.0	16.6	16.9
LN(2.5)	11.3	9.4	25.9	22.0	4.3	7.3	31.6	31.6	35.7	**43.5**	**47.3**
LFR(0.2)	45.3	52.9	50.2	40.5	55.5	54.3	33.1	**59.1**	**59.3**	51.2	50.9
LFR(0.5)	51.4	59.3	56.2	46.7	62.3	60.7	40.3	**64.8**	**64.5**	57.2	56.6
LFR(0.8)	54.6	63.3	59.9	50.6	66.6	64.5	45.5	**68.6**	**68.4**	61.1	60.6
LFR(1)	56.5	65.6	62.4	53.4	68.9	66.9	47.9	**70.6**	**70.2**	63.3	62.8
BE(0.5)	12.0	11.4	**31.7**	**33.6**	6.5	8.7	24.6	4.8	3.2	9.5	7.1
BE(0.8)	17.7	19.4	18.5	13.0	19.1	19.6	9.2	**23.7**	**23.8**	18.2	17.6
BE(1)	37.6	44.3	41.7	33.3	46.4	45.7	24.7	**50.0**	**50.8**	42.1	41.9
BE(1.5)	84.2	91.3	90.9	87.0	**93.3**	92.3	74.9	92.3	**93.0**	89.2	89.4
D(0.2)	25.7	30.6	**31.1**	29.0	29.4	**31.8**	12.2	29.6	31.0	25.6	25.6
D(0.4)	51.0	60.0	59.9	58.6	60.9	**61.7**	32.6	58.8	**62.0**	53.0	54.3
D(0.6)	73.9	82.7	82.7	80.9	**84.6**	84.1	57.4	82.1	**84.3**	77.3	78.8
D(0.8)	88.2	94.1	94.2	93.1	**95.4**	94.9	78.3	94.3	**95.5**	91.8	92.8
HN(0.8)	58.0	66.8	63.7	55.1	70.5	68.2	50.5	**72.1**	**71.8**	65.0	64.3
HN(1)	68.5	76.7	74.3	62.9	78.5	77.4	57.3	**79.8**	**79.5**	73.6	72.8

$S^{(1)}_{n,3,2}$ and $G_{n,2}$. The powers of $T^{(1)}_{n,3,2}$ and $T^{(2)}_{n,3,2}$ do not compare favorably to those of $S^{(1)}_{n,3,2}$ and $S^{(2)}_{n,3,2}$ but are still competitive in terms of power against the traditional tests; i.e., KS_n, CV_n and AD_n. We also note that the tests $T^{(1)}_{n,3,2}$ and $T^{(2)}_{n,3,2}$ produce the highest powers

TABLE 3 Empirical powers against fixed alternatives for $n = 30$

Dist	KS_n	CV_n	AD_n	ZA_n	$G_{n,2}$	$M_{n,1}$	$I_{n,2}$	$S^{(1)}_{n,3,2}$	$S^{(2)}_{n,3,2}$	$T^{(1)}_{n,3,2}$	$T^{(2)}_{n,3,2}$
P(2)	5.3	5.2	5.2	5.0	5.5	5.3	5.1	5.0	5.0	4.9	4.9
P(5)	5.0	5.1	5.3	5.0	5.0	5.1	4.9	4.5	4.6	4.6	4.5
P(10)	5.0	5.0	5.1	5.2	5.0	5.1	5.0	4.7	4.7	4.5	4.5
Γ(0.5)	20.4	20.6	**48.0**	**50.9**	10.4	16.4	43.5	9.9	6.6	20.2	16.6
Γ(0.8)	19.3	21.4	20.1	13.2	20.2	21.1	9.9	**25.0**	**25.1**	19.9	19.1
Γ(1)	52.3	60.6	58.6	45.2	62.1	61.9	36.8	**64.4**	**65.6**	57.5	58.1
Γ(1.2)	81.9	89.4	88.9	81.6	**90.8**	90.3	71.7	90.1	**91.4**	86.5	87.4
W(0.5)	21.9	20.2	53.7	**60.0**	16.2	13.9	**60.7**	34.3	27.5	46.4	40.2
W(0.8)	19.9	21.7	21.6	12.3	18.7	21.5	7.8	**24.5**	**24.6**	19.5	18.2
W(1.2)	82.6	90.1	89.5	83.1	**92.5**	91.2	76.1	92.0	**92.4**	88.9	89.1
W(1.5)	98.5	99.7	99.7	99.1	**99.8**	99.7	98.2	99.7	**99.8**	99.5	99.5
LN(1)	88.6	94.2	95.2	**98.5**	96.0	94.6	86.0	95.3	**96.3**	93.5	94.3
LN(1.2)	56.0	66.6	69.1	**78.7**	**71.6**	68.1	44.7	65.6	69.2	60.2	62.4
LN(1.5)	21.1	26.3	**28.7**	**29.3**	25.8	27.5	8.0	22.2	24.4	19.7	20.3
LN(2.5)	12.8	9.6	30.4	30.1	3.5	6.6	48.5	47.5	51.5	**58.6**	**61.6**
LFR(0.2)	62.2	70.7	68.3	55.6	72.4	71.5	48.5	**74.4**	**74.8**	67.7	67.3
LFR(0.5)	68.8	77.8	75.9	65.0	80.2	78.9	58.9	**81.2**	**81.4**	75.5	75.1
LFR(0.8)	72.3	81.7	79.6	69.3	83.9	82.7	64.7	**84.5**	**84.3**	79.7	79.2
LFR(1)	75.0	83.4	81.5	71.9	**85.5**	84.2	68.6	**85.6**	85.4	80.6	80.3
BE(0.5)	15.8	15.1	**41.0**	**46.0**	6.8	10.6	35.8	8.9	5.0	16.6	11.6
BE(0.8)	23.5	26.7	25.2	16.4	25.3	26.5	11.9	**30.9**	**31.0**	24.9	24.1
BE(1)	52.5	60.5	58.5	46.1	62.3	61.6	37.0	**64.5**	**65.5**	57.8	57.8
BE(1.5)	95.0	98.3	98.3	96.6	**98.9**	98.5	92.1	98.5	**98.8**	97.6	97.9
D(0.2)	33.3	39.9	**40.5**	37.7	38.8	**41.0**	17.7	36.6	38.8	32.2	33.3
D(0.4)	66.8	75.6	76.2	73.9	**77.0**	**77.3**	50.8	73.3	76.7	68.2	70.9
D(0.6)	88.8	94.3	94.6	93.5	**95.3**	94.9	79.3	93.5	**95.1**	91.2	92.7
D(0.8)	97.5	99.2	99.3	99.0	**99.5**	99.3	95.0	99.1	**99.4**	98.6	99.0
HN(0.8)	77.6	85.6	83.9	75.1	**87.5**	86.1	71.0	**87.7**	**87.5**	83.3	82.9
HN(1)	84.8	91.0	89.9	81.2	**92.1**	91.5	77.1	**92.3**	**92.1**	89.1	88.7

against the LN(2.5) alternative by a substantial margin. The results obtained for the mixture distributions are in accordance with those associated with the fixed alternatives.

TABLE 4 Empirical powers against lognormal mixtures for $n = 20$ (top line) and $n = 30$ (bottom line)

p	KS_n	CV_n	AD_n	ZA_n	$G_{n,2}$	$M_{n,1}$	$I_{n,2}$	$S^{(1)}_{n,3,2}$	$S^{(2)}_{n,3,2}$	$T^{(1)}_{n,3,2}$	$T^{(2)}_{n,3,2}$
0.0	5.2	5.2	5.0	5.0	5.3	5.1	4.7	4.6	4.6	4.4	4.4
	5.2	5.1	5.2	5.1	5.3	5.3	5.0	4.8	4.7	4.8	4.8
0.1	4.9	4.9	4.8	5.1	**5.2**	5.1	4.8	**5.3**	**5.2**	4.7	4.5
	5.1	5.0	4.8	5.0	**5.2**	5.0	4.9	**5.1**	5.0	4.8	4.6
0.2	5.7	5.7	5.2	5.6	6.0	5.7	4.8	**6.7**	**6.4**	5.5	5.1
	5.8	5.8	5.5	5.5	6.1	5.9	5.1	**7.2**	**7.0**	6.2	5.7
0.3	7.3	7.8	7.0	7.2	8.4	8.1	5.3	**9.6**	**9.4**	7.5	7.1
	7.7	8.1	7.6	7.8	8.9	8.4	6.1	**10.3**	**10.4**	8.3	7.9
0.4	10.1	11.2	10.2	9.6	11.9	11.6	6.8	**13.6**	**13.5**	10.6	10.0
	12.2	13.3	12.3	11.2	14.2	13.9	8.6	**16.6**	**16.7**	13.0	12.5
0.5	15.0	16.9	15.4	14.8	18.2	17.5	9.5	**20.3**	**20.6**	16.1	15.6
	18.1	21.3	19.9	18.3	23.1	22.2	13.0	**25.0**	**25.7**	20.3	20.2
0.6	21.5	24.9	23.4	22.2	26.7	26.2	14.0	**28.9**	**29.2**	23.2	22.4
	28.1	32.9	31.8	27.9	34.6	34.1	20.6	**37.3**	**38.9**	31.6	32.0
0.7	31.0	36.2	34.7	33.1	38.6	37.8	21.6	**40.3**	**41.4**	33.9	33.3
	42.7	50.2	49.0	44.1	53.0	52.1	32.7	**53.8**	**56.1**	47.4	48.3
0.8	44.0	51.2	50.2	48.5	54.0	53.1	32.0	**55.0**	**56.6**	48.5	48.1
	58.1	67.1	67.0	63.2	69.9	68.7	48.9	**70.1**	**72.5**	65.0	65.9
0.9	57.9	67.1	66.7	68.3	**70.7**	68.7	47.3	69.9	**71.8**	64.3	64.2
	74.5	82.7	83.4	83.5	**85.8**	83.9	68.9	85.0	**86.7**	81.1	82.5

5 | PRACTICAL APPLICATION

Below, we apply each of the tests considered to an observed data set. The data concerned is the the lifetime tournament earnings, up to 1980, of all professional golfers whose earnings exceeded \$700 000, as reported in the Golf magazine, 1981 yearbook. This data set was also discussed and analysed by Arnold (2015). The reported salaries, in thousands of dollars, can be found in Table 6.

The support of the data in Table 6 is, per definition, $(700, \infty)$. As a result, we rescale the data by dividing each number in the table by 700 and then we test the hypothesis that the resulting data are realised from a Pareto distribution. When fitting a Pareto distribution to the data using the method of moments, we obtain $\widehat{\beta}_n = 2.495$. Before proceeding to the results pertaining to the formal testing procedures,

TABLE 5 Empirical powers against exponential mixtures for $n = 20$ (top line) and $n = 30$ (bottom line)

p	KS_n	CV_n	AD_n	ZA_n	$G_{n,2}$	$M_{n,1}$	$I_{n,2}$	$S^{(1)}_{n,3,2}$	$S^{(2)}_{n,3,2}$	$T^{(1)}_{n,3,2}$	$T^{(2)}_{n,3,2}$
0.0	5.2	5.2	5.1	4.9	5.4	5.2	4.7	4.8	4.8	4.9	4.9
	5.1	5.2	5.2	5.3	5.4	5.3	5.2	5.1	5.0	5.0	5.1
0.1	5.5	5.5	5.3	**5.7**	**5.9**	5.6	4.8	5.4	5.2	5.0	5.0
	5.5	5.7	5.6	5.6	**5.9**	**5.8**	4.8	5.2	5.3	4.9	4.9
0.2	5.9	5.9	5.6	**6.1**	**6.2**	5.9	5.1	5.6	5.5	5.0	5.1
	6.2	**6.4**	6.0	6.2	**6.6**	**6.4**	5.6	6.3	6.2	5.6	5.5
0.3	6.4	6.7	6.3	**7.0**	**7.2**	6.9	5.5	6.6	6.5	5.9	5.9
	6.7	6.9	6.5	**7.1**	**7.1**	6.9	6.1	7.0	6.8	5.7	5.7
0.4	7.5	7.6	6.8	7.5	**7.8**	**7.7**	5.9	7.6	7.3	6.5	6.2
	7.5	**7.8**	7.2	**7.8**	**8.2**	**7.8**	6.6	**7.8**	7.5	6.5	6.4
0.5	7.8	8.6	7.6	8.2	**9.1**	8.6	6.8	**8.8**	8.4	7.2	7.1
	8.8	9.4	8.4	8.9	**9.7**	9.4	7.5	**10.0**	9.6	8.0	7.8
0.6	9.2	9.8	8.7	9.3	**10.5**	9.8	7.5	**10.5**	10.1	8.2	8.2
	10.5	11.3	10.3	10.5	**11.9**	11.5	9.4	**12.3**	11.8	10.0	9.7
0.7	10.7	11.4	9.8	10.5	12.2	11.6	8.4	**12.5**	**12.4**	9.6	9.5
	12.3	13.8	12.4	12.3	14.6	13.7	10.5	**15.8**	**15.3**	12.5	12.1
0.8	12.0	13.3	11.5	12.3	14.6	13.7	9.7	**16.5**	**16.2**	12.5	12.3
	15.1	16.8	15.1	14.6	18.2	17.5	12.5	**20.2**	**19.9**	16.0	15.6
0.9	13.9	15.4	13.4	13.2	17.4	16.0	10.6	**19.8**	**19.7**	15.1	15.0
	17.5	20.1	18.0	16.9	22.3	21.1	14.9	**25.6**	**25.2**	20.1	19.9

we consider visual tests of fit for the Pareto distribution. Figure 1 shows the empirical distribution function of the rescaled data, together with the fitted Pareto distribution function. Figure 2 shows a quantile-quantile plot comparing the empirical quantiles to those of the fitted distribution. Both figures indicate a close correspondence between the empirical properties of the data and those expected under the null hypothesis of the Pareto distribution.

We now turn our attention to the results obtained using the goodness-of-fit tests discussed in Section 4. Table 7 contains the test statistic values with the corresponding p-values of the tests. These p-values were calculated based on 10 000 samples of size 50 simulated from a Pareto distribution with parameter $\beta = 2.495$ (corresponding to the parameter of the fitted distribution).

708	712	729	746	753	759	769	771	778	778
814	816	820	825	841	844	849	871	878	883
912	944	965	1001	1005	1016	1031	1051	1056	1066
1092	1095	1109	1171	1184	1208	1338	1374	1410	1433
1519	1537	1627	1684	1690	1829	1858	2202	2474	3581

TABLE 6 The golfer data set.

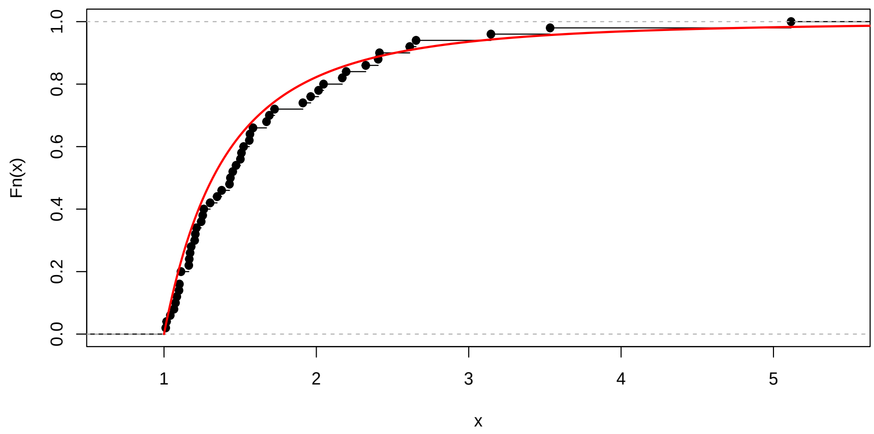

FIGURE 1 Empirical and fitted distribution functions.

Test	Statistic	p-value	Test	Statistic	p-value
$\mathsf{KS_n}$	0.125	0.3223	$\mathsf{I}_{n,2}$	6.7×10^{-3}	0.7368
$\mathsf{CV_n}$	0.158	0.2897	$S^{(1)}_{n,3,2}$	3.7×10^{-4}	0.2276
$\mathsf{AD_n}$	0.926	0.2885	$S^{(2)}_{n,3,2}$	2.7×10^{-4}	0.1924
$\mathsf{ZA_n}$	3.433	0.0956	$T^{(1)}_{n,3,2}$	2.4×10^{-4}	0.3334
$G_{n,2}$	0.792	0.1770	$T^{(2)}_{n,3,2}$	1.7×10^{-3}	0.2863
$M_{n,1}$	0.2842	0.2085			

TABLE 7 Summary results for the lifetime tournament earnings through golf.

It is clear from the reported p-values in Table 7 that none of the tests considered reject the assumption that the data are realised from a Pareto distribution at a 5% level of significance. These results are similar to those obtained by Arnold (2015) where, using a different array of tests, the author does not find evidence to reject the hypothesis of the Pareto distribution.

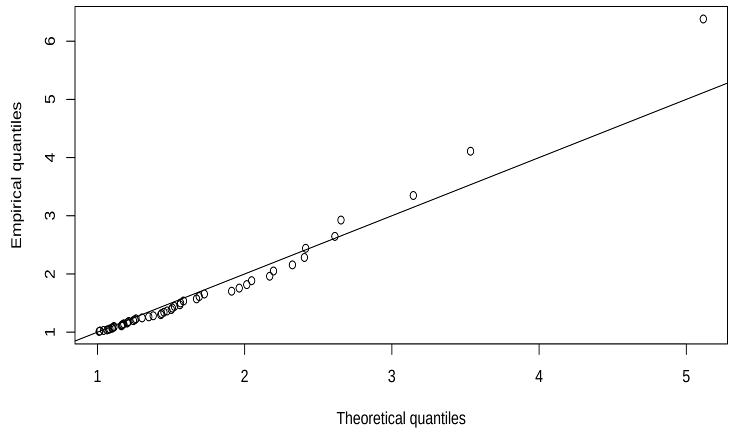

FIGURE 2 Quantile-quantile plot.

6 | CONCLUDING REMARKS

In this study, we introduced new classes of goodness-of-fit tests for the Pareto distribution based on a characterisation involving the sample minimum. The proposed tests are based on empirical characteristic functions estimated via V and U statistics, and include two tuning parameters. We compared the performances of our tests to those of some commonly used goodness-of-fit tests and the results presented suggest that the newly proposed tests are competitive, often outperforming the existing tests against the alternative distributions considered. Based on the numerical performance of the tests, we recommend using the tests based on V statistics (using a Gaussian kernel and setting the tuning parameters to $\mathsf{m} = 3$ and $\mathsf{a} = 2$; this test is denoted $\mathsf{S}^{(2)}_{\mathsf{n},3,2}$ in the text). Note the marked difference between the finite sample performances of $\mathsf{S}_{\mathsf{n},\mathsf{m},\mathsf{a}}$ and $\mathsf{T}_{\mathsf{n},\mathsf{m},\mathsf{a}}$. The asymptotic theory relating to U and V statistics implies that, for large samples, the performances of the two tests will be similar. However, no theoretical results in this regard are available for small samples. In fact we were unable to find papers where the same tests, based on U and V statistics, were both included and numerically compared. The relative performance of these tests can be investigated as part of future research.

We mentioned in the introduction that the Pareto distribution has found applications in the field of survival analysis and reliability theory. Censoring often occurs in these fields due to the nature of the study. A possible avenue for future research is to modify the newly proposed statistics to test for censored Pareto distributions. Some work in this regards has been done by Fernández and Rivera (2020), where they consider Kaplan-Meier U and V statistics, and by Cuparić and Milošević (2022), where they derive asymptotic results for a new characterisation based test for exponentiality in the presence of random right censored data. Further details regarding the development of tests for the censored exponential distribution can be found in Bothma, Allison, Cockeran, and Visagie (2021), while Bothma, Allison, and Visagie (2022) considers tests for the censored Weibull distribution. For results pertaining specifically to tests for the censored Pareto distribution, the interested reader is referred to Ndwandwe, Allison, and Visagie (2021).

References

Allison, J. S., Milošević, B., Obradović, M., & Smuts, M. (2022). Distribution-free goodness-of-fit tests for the Pareto distribution based on a characterization. Computational Statistics, 37(1), 403–418.

Amin, Z. H. (2007). Tests for the validity of the assumption that the underlying distribution of life is Pareto. Journal of Applied Statistics, 34(2), 195–201.

Arnold, B. C. (2015). Pareto Distributions. New York: CRC Press.

Baringhaus, L., Ebner, B., & Henze, N. (2017). The limit distribution of weighted L^2-goodness-of-fit statistics under fixed alternatives, with applications. Annals of the Institute of Statistical Mathematics, 69(5), 969–995.

Baringhaus, L., & Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. Metrika, 35(1), 339–348.

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). Statistics of Extremes: Theory and Applications. Chichester: John Wiley and Sons.

Bothma, E., Allison, J. S., Cockeran, M., & Visagie, I. J. H. (2021). Characteristic function and Laplace transform-based tests for exponentiality in the presence of random right censoring. Stat, 10(1), e394.

Bothma, E., Allison, J. S., & Visagie, I. J. H. (2022). New classes of tests for the Weibull distribution using Stein's method in the presence of random right censoring. Computational Statistics. doi: https://doi.org/10.1007/s00180-021-01178-0

Bourguignon, A., Saulo, B., & Fernandez, R. N. (2016). A new Pareto-type distribution with applications in reliability and income data. Physica A: Statistical Mechanics and its Applications, 457, 166–175.

Chu, J., Dickin, S., & Nadarajah, S. (2019). A review of goodness of fit tests for Pareto distributions. Journal of Computational and Applied Mathematics, 361, 13-41.

Cuparić, M., & Milošević, B. (2022). New characterization-based exponentiality tests for randomly censored data. Test, 31(2), 461–487.

Fernández, T., & Rivera, N. (2020). Kaplan-Meier V-and U-statistics. Electronic Journal of Statistics, 14(1), 1872–1916.

Feuerverger, A., & Mureika, R. A. (1977). The empirical characteristic function and its applications. The Annals of Statistics, 5(1), 88–97.

Fisk, P. R. (1961). The graduation of income distributions. Econometrica, 29(2), 171–185.

Giacomini, R., Politis, D. N., & White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Econometric Theory, 29(3), 567–589.

Gupta, R. C. (1973). A characteristic property of the exponential distribution. Sankhyā: The Indian Journal of Statistics, Series B, 365–366.

Ismaïl, S. (2004). A simple estimator for the shape parameter of the Pareto distribution with economics and medical applications. Journal of Applied Statistics, 31(1), 3–13.

Klar, B., & Meintanis, S. G. (2005). Tests for normal mixtures based on the empirical characteristic function. Computational Statistics & Data Analysis, 49(1), 227–242.

Meintanis, S. G. (2009). A unified approach of testing for discrete and continuous Pareto laws. Statistical Papers, 50(3), 569–580.

Meintanis, S. G. (2016). A review of testing procedures based on the empirical characteristic function. The South African Statistical Journal, 50(1), 1–14.

Meintanis, S. G., Gamero, M. D. J., & Alba-Fernández, V. (2014). A class of goodness-of-fit tests based on transformation. Communications in Statistics-Theory and Methods, 43(8), 1708–1735.

Ndwandwe, L., Allison, J. S., Santana, L., & Visagie, I. J. H. (2022). Testing for the Pareto type I distribution: A comparative study. arXiv preprint arXiv:2211.10088.

Ndwandwe, L., Allison, J. S., & Visagie, I. J. H. (2021). A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring. Proceedings of the 62nd Annual Conference of SASA(1), 17–23.

Nofal, Z. M., & El Gebaly, Y. M. (2017). New characterizations of the Pareto distribution. Pakistan Journal of Statistics and Operation Research, 13(1), 63–74.

Pareto, V. (1897). Cours d'economie politique, vol. ii. Lausanne: F. Rouge.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rytgaard, M. (1990). Estimation in the Pareto distribution. ASTIN Bulletin: The Journal of the IAA, 20(2), 201–216.

Soliman, A. A. (2000). Bayes prediction in a Pareto lifetime model with random sample size. Journal of the Royal Statistical Society: Series D (The Statistician), 49(1), 51–62.

Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(2), 281–294.

$\square$

Chapter 4

Article 3: Revisiting the memoryless property - testing for the Pareto distribution

The third article, *Revisiting the memoryless property - testing for the Pareto distribution*, is a draft article and has therefore not yet been submitted to any journal. We plan to submit this paper to *Computational Statistics*.

Manuscript	No page limit.
Title	The title should be concise and informative.
Abstract and keywords	An abstract of 150 to 250 words and 4 to 6 keywords which can be used for indexing purposes.
Tables	All tables are to be numbered using Arabic numerals. Tables should always be cited in text in consecutive numerical order. For each table, please supply a table caption explaining the components of the table.
References	Cite references in the text by name and year in parentheses. The list of references should only include works that are cited in the text and that have been published or accepted for publication. Reference list entries should be alphabetized by the last names of the first author of each work.
General formatting	A LaTeXtemplate is provided for submission.
Additional information	https://tinyurl.com/486kbupt

Revisiting the memoryless property - testing for the Pareto distribution

L. Ndwandwe, J.S. Allison, L. Santana and I.J.H. Visagie

Abstract

We propose new goodness-of-fit tests for the Pareto type I distribution. These tests are based on a multiplicative version of the memoryless property which characterises this distribution. We present the results of a Monte Carlo power study demonstrating that the proposed tests are powerful compared to existing tests. The power study considers powers against fixed alternatives as well as powers against mixture distributions. Thereafter the tests considered are used in order to test the hypotheses that two sets of golfers' earnings (those of the PGA tour and LIV Golf) are realised from Pareto distributions.

Key words: Memoryless property, Goodness-of-fit testing, Pareto distribution.

1 Introduction

The memoryless property of the exponential distribution is well-known and characterises this distribution among the continuous laws. For example, if the exponential distribution is used to model waiting times, then the probability of an event occurring within the next unit of time is independent of the amount of time that has passed. Let Y be a random variable denoting an exponential waiting time. The memoryless property of the exponential distribution can be expressed as

$$P(Y \geq s + t | Y > s) = P(Y > t), \quad t, s \geq 0. \tag{1.1}$$

For more details regarding this and other characterisations, see Tomy et al. (2020) as well as Galambos & Kotz (2006). The memoryless property has been used to develop several goodness-of-fit tests for the exponential distribution, see Ahmad & Alwasel (1999) as well as Alwasel (2001). Below, we use a related characterisation to propose new goodness-of-fit tests for the Pareto distribution.

The Pareto distribution is named after the Italian economist Vilfredo Pareto, who first described this law in his 1897 book *Cours d'économie politique*, see Pareto (1897). Several generalisations of this distribution have been proposed, see Arnold (2015), and the originally proposed distribution has subsequently become known as the Pareto type I distribution. Below, we refer to the Pareto type I distribution simply as the Pareto distribution. This distribution is closely related to the exponential distribution. If Y is exponential with mean $1/\beta$, then $X = \mathrm{e}^Y$ is a Pareto distributed random variable with shape parameter $\beta > 0$, denoted by $X \sim P(\beta)$. In this case, the distribution function of X is

$$F(x) = 1 - x^{-\beta}, \quad \beta > 0 \quad x > 1.$$

The link between the exponential and Pareto distributions means that the memoryless property of the exponential distribution has implications for the Pareto distribution as well. The Pareto distribution exhibits a multiplicative version of the memoryless property, which can be expressed as follows. $X \sim P(\beta)$ for some $\beta > 0$ if, and only if,

$$P(X \geq st|X > s) = P(X > t), \quad t, s \geq 1. \tag{1.2}$$

The Pareto distribution is widely used to model phenomena exhibiting heavy tails. Examples of the use of the Pareto distribution include the modelling of excess losses in insurance claims, see Rytgaard (1990), as well as the modelling of failure times of mechanical components, see Bourguignon & Saulo (2018). The Pareto is also a popular model found in survival analysis and reliability theory, see, Ndwandwe et al. (2021) as well as Nofal & El Gebaly (2017). For a recent review of goodness-of-fit tests for the Pareto distribution, see Ndwandwe et al. (2022). Furthermore, there exists a close link between Lomax and Pareto distributions; For a recent paper regarding parameter estimation for the Lomax distribution, see Nombebe et al. (2022).

The remainder of the paper is structured as follows. In Section 2, we introduce two new classes of goodness-of-fit tests based on a characterization that uses the multiplicative memoryless property. Section 3 includes a Monte Carlo study comparing the performances of the newly proposed tests to those of existing tests. The finite sample performances are considered against a wide range of fixed alternative distributions. Additionally, the performances of the tests against a range of mixture distributions are also compared in this section. In Section 4, we demonstrate the use of the tests in order to determine whether or not observed data sets, relating to the earning of golfers, are realised from a Pareto distribution. Finally, Section 5 presents some conclusions.

2 The test statistics based on the memoryless property

In this section we develop two classes of goodness-of-fit tests for the Pareto distribution based on the characterisation given in (1.2). Given a random sample $X, X_1, X_2, ..., X_n$, from an unknown probability distribution F with density f, we wish to test the composite null hypothesis

$$H_0 : X \sim P(\beta), \tag{2.1}$$

for some $\beta > 0$, against general alternatives.

To estimate β, we use two different estimators; the maximum likelihood estimator (MLE) and the method of moments estimator (MME). The MLE for the shape parameter of the Pareto distribution is

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

while the MME is

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

with $X_{(1)}$ the minimum of $X_1, ..., X_n$ and $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$.

Upon setting $t = s$ in (1.2), we obtain that $X \sim P(\beta)$ if, and only if,

$$P(X > t^2) = [P(X > t)]^2, \quad t \geq 1. \tag{2.2}$$

We now propose a test statistic based on (2.2). We estimate the left hand side of (2.2) nonparametrically using the empirical survival function, $S_n(t) = \frac{1}{n}\sum_{j=1}^{n} I(X_j > t)$, and we estimate the right hand side parametrically by $\left[S_{\hat{\beta}}(t)\right]^2 = t^{-2\hat{\beta}}$. Accordingly, we suggest the following test statistic:

$$\begin{aligned} MP_n^{(1)} &= \int_1^\infty \left(S_n(t^2) - \left[S_{\hat{\beta}}(t)\right]^2\right)^2 dF_{\hat{\beta}}(t) \\ &= \int_1^\infty \left(\frac{1}{n}\sum_{j=1}^{n} I(X_j > t^2) - t^{-2\hat{\beta}}\right)^2 \hat{\beta} t^{-(\hat{\beta}+1)} dt. \end{aligned}$$

From some straightforward calculation we have that

$$MP_n^{(1)} = \frac{2}{3n}\sum_{j=1}^{n} X_j^{-3\hat{\beta}/2} - \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} \min(X_j, X_k)^{-\hat{\beta}/2} + \frac{8}{15}.$$

The double summation in the expression for $MP_n^{(1)}$ can be reduced to a single summation based on the order statistics, $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$;

$$MP_n^{(1)} = \frac{2}{3n}\sum_{j=1}^{n} X_j{}^{-3\hat{\beta}/2} - \frac{1}{n^2}\sum_{j=1}^{n}[(n-j+1)^2 - (n-j)^2]X_{(j)}^{-\hat{\beta}/2} + \frac{8}{15}.$$

For the second test statistic, note that (1.2) implies that X has a Pareto distribution if, and only if,

$$P(X > st) = P(X > s)P(X > t), \quad s,t \geq 1. \tag{2.3}$$

Again, estimating the left hand side of (2.3) nonparametrically by the empirical survival function and the right hand side by the parametric survival function with estimated shape parameter, we have the test statistic

$$\begin{aligned} MP_n^{(2)} &= \int_1^{\infty}\int_1^{\infty}\left(\frac{1}{n}\sum_{i=1}^{n} I(X_j > st) - s^{-\hat{\beta}}t^{-\hat{\beta}}\right)^2 dF_{\hat{\beta}}(s)dF_{\hat{\beta}}(t) \\ &= \int_1^{\infty}\int_1^{\infty}\left(\frac{1}{n}\sum_{i=1}^{n} I(X_j > st) - s^{-\hat{\beta}}t^{-\hat{\beta}}\right)^2 \hat{\beta}^2 s^{-\hat{\beta}-1}t^{-\hat{\beta}-1}dsdt. \end{aligned}$$

After some algebraic manipulation we have that

$$\begin{aligned} MP_n^{(2)} = \frac{10}{9} &- \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\min(X_j, X_k)^{-\hat{\beta}} - \frac{\hat{\beta}}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\min(X_j, X_k)^{-\hat{\beta}}\log\bigl(\min(X_j, X_k)\bigr) \\ &- \frac{\hat{\beta}}{n}\sum_{j=1}^{n}\left[\frac{1 - X_j{}^{-2\hat{\beta}}}{2\hat{\beta}} - X_j{}^{-2\hat{\beta}}\log(X_j)\right]. \end{aligned}$$

Again this test statistic simplifies to a single sum involving order statistic

$$\begin{aligned} MP_n^{(2)} = \frac{10}{9} &- \frac{1}{n^2}\sum_{j=1}^{n}\bigl[(n-j+1)^2 - (n-j)^2\bigr]X_{(j)}^{-\hat{\beta}} - \frac{\hat{\beta}}{n^2}\sum_{j=1}^{n}\bigl[(n-j+1)^2 - (n-j)^2\bigr]X_{(j)}^{-\hat{\beta}}\log(X_{(j)}) \\ &- \frac{\hat{\beta}}{n}\sum_{j=1}^{n}\left[\frac{1 - X_j{}^{-2\hat{\beta}}}{2\hat{\beta}} - X_j{}^{-2\hat{\beta}}\log(X_{(j)})\right]. \end{aligned}$$

Both $MP_n^{(1)}$ and $MP_n^{(2)}$ reject the hypothesis in (2.1) for large values.

3 Monte Carlo study

In this section, we use Monte Carlo simulation to compare the performances of our newly proposed tests to those of existing tests found in the statistical literature. The classical goodness-of-fit tests included are the Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling tests. Additionally, we include a test based on likelihood ratios as well as a test based on the Mellin transform. The calculable forms of the test statistics are specified below:

- The Kolmogorov-Smirnov test:

$$KS_n = \max\left\{\max_{1\le j\le n}\left[\frac{j}{n} - F_{\widehat{\beta}_n}(X_{(j)})\right], \max_{1\le j\le n}\left[F_{\widehat{\beta}_n}(X_{(j)}) - \frac{j-1}{n}\right]\right\}.$$

- The Cramér-von Mises test:

$$CM_n = \frac{1}{12n} + \sum_{j=1}^{n}\left[F_{\widehat{\beta}_n}(X_{(j)}) - \frac{2j-1}{2n}\right]^2.$$

- The Anderson-Darling test:

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

- The test proposed by Zhang (2002), based on likelihood ratios:

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

- The test proposed by Meintanis (2009), based on the Mellin transform:

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

where

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

Choosing $w(x) = e^{-ax}$, one has

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

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

and

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

To obtain the numerical results presented later in this section, the tuning parameter a has been assigned a value of 1.

All of the tests above reject the null hypothesis in (2.1) for large values of the test statistics.

3.1 Simulation setting

Note that, when using MLE to estimate β, the critical values of the tests considered do not depend on the value of β. If $X \sim P(\beta)$, then $Y := X^{\beta} \sim P(1)$. Consider the transformed sample $Y_j = X_j^{\widehat{\beta}},\ j = 1, \ldots, n$, and note that

$$\widehat{\beta}(Y_1, ..., Y_n) = \frac{n}{\sum_{j=1}^{n} \log(Y_j)} = \frac{n}{\widehat{\beta}(X_1, ..., X_n) \sum_{j=1}^{n} \log(X_j)} = \frac{\widehat{\beta}(X_1, ..., X_n)}{\widehat{\beta}(X_1, ..., X_n)} = 1. \qquad (3.1)$$

For more details regarding this transformation and the estimation of critical values using Monte Carlo simulation, see Section 3.1 of Ndwandwe et al. (2022). We simulate 100 000 Monte Carlo samples (from a $P(1)$ distribution) to estimate these critical values. Next, we obtain 10 000 samples from an alternative distribution in order to obtain the empirical power against the distribution in question. Below, we calculate empirical powers against the distributions listed in Table 1.

In the case of MME, fixed critical values cannot be obtained because $\widetilde{\beta}(Y_1, ..., Y_n) \neq 1$. Instead we employ a parametric bootstrap approach to obtain these empirical critical values. Due to the time consuming nature of the bootstrap procedure required, we use the warp-speed bootstrap technique proposed by Giacomini et al. (2013) to obtain empirical powers. To this end, we generate 50 000 Monte Carlo samples and for each Monte Carlo sample we obtain one parametric bootstrap sample. The estimated empirical power is then the proportion of these 50 000 Monte Carlo samples that lead to the rejection of the null hypothesis in (2.1), see Ndwandwe et al. (2022) for an algorithm for the implementation of the warp-speed bootstrap.

3.2 Powers against fixed alternatives

For each of the tests considered, power estimates were calculated for samples of size $n = 20$ and $n = 30$. These estimates are calculated against the alternative distributions given in Table 1 for various parameter values. The resulting numerical powers are calculated to be the percentage, rounded to the nearest integer, of the samples that lead to the rejection of the null hypothesis. These powers are shown in Tables 2 and 3 for samples sizes 20 and 30, respectively. These tables include the results obtained using both MLE and MME. For ease of comparison, we circle the highest two of the numerical powers associated with MME and we indicate the two highest powers achieved using MLE with a rectangle (including ties). In each case, the best performing test overall is indicated using shading as well. All calculations were performed in R, see R Core Team (2020).

Table 1: Summary of various choices of the alternative distributions.

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

We discuss the powers reported in Tables 2 and 3 below. Both tables indicate that all of the tests considered achieves the specified size of 5% against each of the Pareto distributions included in the Monte Carlo study. Unsurprisingly, the powers reported in Table 3 are generally higher than those reported in Table 2, indicating that an increase in sample size results in an increase in power against the alternatives considered.

Consider the result obtained using MLE. The tables indicate that, using this estimation technique, $MP_n^{(1)}$ and $G_{n,1}$ perform best, closely followed by $MP_n^{(2)}$. The ZA_n and CV_n tests also generally perform well in this setting. When turning our attention to the results obtained using the MME, we see that $MP_n^{(2)}$ and $G_{n,1}$ exhibit the highest powers, followed by $MP_n^{(1)}$ and AD_n.

We now compare the results obtained using MLE and MME. Although the power achieved by a given test using the MLE is higher than that achieved using MME in a few cases, the reverse is generally true. In the vast majority of cases, the results obtained using MME is superior to those associated with MLE. In some cases this outperformance is substantial. For example, the observed power of the AD_n against $TP(2)$ distribution is 86% when using MME, compared to a mere 21% using MLE.

Taking all of the observations made above into consideration, our recommendation when testing for the Pareto distribution based on observed data is to use either the $MP_n^{(2)}$ or $G_{n,1}$ test. We recommend that these tests be used while estimating the value of β using MME.

Table 2: Numerical powers against fixed alternatives with $n = 20$

	KS_n		CV_n		AD_n		ZA_n		$MP_n^{(1)}$		$MP_n^{(2)}$		$G_{n,1}$	
Distr	MME	MLE	MME	MLE	MME	MLE	MME	MLE	MME	MLE	MME	MLE	MME	MLE
$P(2)$	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
$P(5)$	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
P(10)	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	4	4
$\Gamma(0.8)$	15	9	16	10	15	10	11	11	18	**12**	**21**	**14**	**19**	10
$\Gamma(1)$	37	24	44	29	42	24	33	28	47	31	**48**	**32**	**51**	**32**
$\Gamma(1.2)$	64	45	74	56	72	51	64	55	**75**	**57**	**75**	**57**	**80**	**60**
$W(0.8)$	15	9	17	9	17	9	11	**10**	19	**10**	**22**	**12**	**20**	8
$W(1.2)$	65	50	75	62	73	57	65	60	**76**	**63**	**76**	62	**80**	**65**
$W(1.5)$	91	82	96	92	96	90	93	91	**97**	**93**	96	92	**98**	**94**
$LN(1)$	72	55	82	66	83	64	**90**	**80**	80	62	72	51	**85**	**72**
$LN(1.5)$	17	6	20	7	22	6	21	**10**	22	7	**23**	7	**24**	**8**
$LN(2.5)$	11	27	9	29	**26**	**43**	**22**	27	14	26	17	17	5	**32**
$HN(0.5)$	44	37	53	47	49	41	43	43	**56**	**50**	**57**	**51**	55	47
$HN(1)$	67	54	76	65	74	59	62	60	78	**68**	**80**	**70**	**80**	66
$HN(1.2)$	74	58	82	70	80	64	69	66	84	**73**	**86**	**76**	**86**	71
$LFR(0.2)$	45	31	52	38	49	33	40	35	55	**41**	**57**	**43**	**59**	**41**
$LFR(0.8)$	54	42	64	53	60	47	51	49	66	**56**	**68**	**57**	**69**	55
$LFR(1)$	56	45	65	55	62	50	53	52	68	**58**	**69**	**60**	**70**	57
$BE(0.8)$	17	11	19	12	18	11	13	12	22	**13**	**25**	**16**	**23**	12
$BE(1)$	39	25	46	31	43	25	34	28	48	32	**50**	**34**	**52**	**33**
$BE(1.5)$	84	67	**91**	**80**	**91**	75	86	**80**	**91**	**80**	90	77	**94**	**83**
$TP(1)$	46	12	53	**13**	**58**	10	51	**13**	57	**13**	**58**	**13**	54	**13**
$TP(2)$	76	22	82	**26**	**86**	21	82	25	85	**26**	**86**	24	82	25
$TP(3)$	90	32	94	**37**	**96**	32	94	35	**96**	**38**	**96**	36	94	36
$D(0.4)$	50	29	**60**	34	**60**	30	57	**39**	59	33	57	30	**64**	**38**
$D(0.6)$	73	51	82	61	**83**	56	80	**65**	82	59	79	54	**86**	**67**
$D(0.8)$	89	70	**95**	81	94	79	93	**84**	94	81	92	76	**96**	**87**

Table 3: Numerical powers against fixed alternatives with $n = 30$

	KS_n		CV_n		AD_n		ZA_n		$MP_n^{(1)}$		$MP_n^{(2)}$		$G_{n,1}$	
Distr	MME	MLE	MME	MLE	MME	MLE	MME	MLE	MME	MLE	MME	MLE	MME	MLE
$P(2)$	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
$P(5)$	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
$P(10)$	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
$\Gamma(0.8)$	19	12	21	14	21	13	14	14	**25**	**16**	**29**	**19**	**25**	14
$\Gamma(1)$	53	36	61	44	59	39	47	42	64	47	**65**	**48**	**67**	**49**
$\Gamma(1.2)$	82	64	89	76	89	73	82	76	**90**	**78**	**90**	77	**93**	**80**
$W(0.8)$	20	10	21	11	21	11	13	12	**25**	**13**	**29**	**16**	23	10
$W(1.2)$	83	69	90	81	90	78	83	81	**91**	**83**	90	81	**93**	**85**
$W(1.5)$	99	95	**100**	**99**	**100**	**99**	99	**99**	**100**	**99**	**100**	**99**	**100**	**99**
$LN(1.0)$	88	75	94	85	**95**	87	**98**	**95**	92	82	87	71	**95**	**90**
$LN(1.5)$	21	8	26	8	**28**	7	**28**	**14**	**28**	8	**28**	8	**30**	**10**
$LN(2.5)$	13	38	10	42	**31**	**58**	**30**	39	16	39	19	26	4	**44**
$HN(0.5)$	61	54	71	66	68	61	61	63	**74**	**69**	**75**	**70**	73	68
$HN(1)$	85	73	91	84	90	81	82	83	92	**87**	**93**	**88**	**93**	86
$HN(1.2)$	90	78	95	88	94	86	87	88	**96**	**91**	**96**	**92**	**96**	90
$LFR(0.2)$	63	47	71	57	69	52	56	54	74	60	**75**	**62**	**77**	**61**
$LFR(0.8)$	73	60	82	73	80	69	70	71	84	**76**	**85**	**78**	**85**	**76**
$LFR(1)$	74	63	83	75	81	71	72	73	85	**78**	**86**	**79**	**86**	**78**
$BE(0.8)$	24	15	27	17	26	15	16	17	30	**19**	**34**	**23**	**31**	17
$BE(1)$	52	36	60	44	59	39	46	42	63	46	**64**	**48**	**67**	**48**
$BE(1.5)$	95	86	98	94	**99**	93	97	**95**	98	**95**	98	93	**99**	**96**
$TP(1)$	58	15	66	**17**	**70**	14	64	**17**	69	**17**	**70**	16	65	**17**
$TP(2)$	88	30	93	**37**	**95**	32	93	33	**94**	**37**	**94**	34	92	**37**
$TP(3)$	97	45	**99**	**54**	**99**	49	**99**	47	**99**	**53**	**99**	51	**99**	**53**
$D(0.4)$	66	41	75	51	**76**	48	73	**56**	74	49	71	43	**78**	**57**
$D(0.6)$	89	70	**94**	81	**94**	80	93	**84**	**94**	80	92	73	**96**	**86**
$D(0.8)$	97	89	**99**	95	**99**	95	**99**	**96**	**99**	95	98	92	**99**	**97**

3.3 Powers against mixture distributions

In addition to the powers obtained against fixed alternatives, we also consider two families of mixture distributions. Consider a mixture distribution with mixing proportion p, contaminated by some distribution G. When simulating from this distribution, we generate from a realisation from G with probability p and we generate a Pareto distributed random variate with probability $1 - p$. The shape parameter of the Pareto distribution used is chosen so that the means of the two distributions under consideration are equal. The two contaminating distributions considered are the exponential and half-normal distributions, each with mean 2. In order to ensure that the support of these distributions coincide with that of the Pareto distribution, both of these distributions are translated by 1 (meaning that the mean of both of these distributions are changed to 3). The calculated numerical powers can be found in Tables 4 and 5. The same conventions are used in order to indicate the best performing tests among those using MLE and MME, respectively, as well as for the overall most powerful test as was the case in Tables 2 and 3. Note that, in Tables 4 and 5, the results associated with sample sizes $n = 20$ and $n = 30$ are included in the same table.

When considering the powers against mixture distributions reported in Tables 4 and 5, similar results are in order as was the case for the fixed alternatives considered. Again see that the powers increase with sample size and that the powers associated with MME is typically higher than those associated with MLE.

Consider the powers against the exponential mixture, shown in Table 4. When using MLE, the best performing tests are ZA_n and $G_{n,1}$, followed by $MP_n^{(2)}$. When turning our attention to the results obtained using MME, the $MP_n^{(2)}$ test is nearly uniformly most powerful against the mixtures considered, but it is closely followed by $G_{n,1}$. Overall, the most powerful test is seen to be $MP_n^{(2)}$ using MME.

The powers shown in Table 5 indicate that the best performing test using MLE is $MP_n^{(2)}$, very closely followed by $MP_n^{(1)}$. Turning our attention to the results pertaining to MME, we see that $MP_n^{(2)}$ is not outperformed by any test against any of the mixtures considered. However, we again note that the $G_{n,1}$ test achieves powers close to those of $MP_n^{(2)}$.

The results above serve to support our findings against fixed alternatives. That is, we recommend the use of MME in order to perform parameter estimation, and we recommend using either the $MP_n^{(2)}$ or $G_{n,1}$ to test the hypothesis of the Pareto distribution.

Table 4: Numerical powers against exponential mixture distributions

Proportion	n	KS_n MME	KS_n MLE	CV_n MME	CV_n MLE	AD_n MME	AD_n MLE	ZA_n MME	ZA_n MLE	$MP_n^{(1)}$ MME	$MP_n^{(1)}$ MLE	$MP_n^{(2)}$ MME	$MP_n^{(2)}$ MLE	$G_{n,1}$ MME	$G_{n,1}$ MLE
0.1	20	**6**	**5**	**6**	**5**	5	**5**	5	**5**	**6**	5	**6**	5	**5**	**5**
	30	**5**	**5**	**5**	**5**	**5**	**5**	**6**	**6**	**5**	**5**	**5**	**5**	**5**	**5**
0.2	20	**6**	5	**6**	5	5	5	**6**	**6**	**6**	5	**6**	6	**6**	**5**
	30	**6**	**6**	**6**	**6**	**6**	5	**6**	**6**	**6**	**6**	**6**	**6**	**6**	5
0.3	20	6	**6**	6	**6**	6	5	**7**	**6**	**7**	6	**7**	6	**7**	**5**
	30	6	6	**7**	6	6	6	**7**	**7**	**7**	6	**7**	**7**	**7**	6
0.4	20	7	6	7	6	6	5	7	**7**	**8**	7	**8**	7	**7**	**6**
	30	**8**	7	**8**	7	7	6	**8**	**8**	**8**	7	**9**	**8**	**8**	7
0.5	20	8	6	8	**7**	7	6	8	**7**	8	7	**9**	8	**9**	**6**
	30	9	8	**10**	**9**	9	7	9	**9**	**10**	**9**	**11**	**9**	**10**	8
0.6	20	9	8	**10**	8	8	7	9	**9**	**10**	8	**11**	9	**10**	**8**
	30	11	9	**12**	10	10	8	11	**11**	**12**	10	**13**	**11**	**12**	10
0.7	20	10	9	12	**10**	10	8	11	**10**	12	10	**13**	10	**13**	**9**
	30	13	11	15	12	13	11	13	**13**	**16**	**13**	**17**	**14**	**16**	**13**
0.8	20	12	10	13	11	11	9	12	**15**	**14**	11	**14**	12	**15**	**11**
	30	15	12	17	14	15	12	15	**15**	18	**15**	**19**	**15**	**19**	**15**
0.9	20	14	11	16	**13**	14	10	13	**13**	**17**	13	**17**	14	**18**	**13**
	30	18	14	21	17	19	14	17	17	22	**18**	**23**	**19**	**23**	**18**

Table 5: Numerical powers against lognormal mixture distributions

Proportion	n	KS_n MME	KS_n MLE	CV_n MME	CV_n MLE	AD_n MME	AD_n MLE	ZA_n MME	ZA_n MLE	$MP_n^{(1)}$ MME	$MP_n^{(1)}$ MLE	$MP_n^{(2)}$ MME	$MP_n^{(2)}$ MLE	$G_{n,1}$ MME	$G_{n,1}$ MLE
0.1	20	**6**	**6**	**6**	5	5	5	**6**	**6**	**6**	**6**	**6**	**6**	**6**	5
	30	5	5	**6**	5	5	5	**6**	**6**	**6**	**6**	**6**	**6**	**6**	5
0.2	20	7	6	**8**	**7**	7	6	7	**7**	**8**	**7**	**8**	**7**	**8**	6
	30	7	7	8	7	7	6	8	**8**	**9**	**8**	**9**	**8**	8	7
0.3	20	10	8	10	**9**	9	7	9	**9**	**11**	**9**	**12**	**10**	**11**	8
	30	11	10	13	**11**	11	9	11	**11**	13	**11**	**14**	**13**	13	**11**
0.4	20	13	11	14	12	12	10	12	12	**15**	**13**	**16**	**14**	**15**	12
	30	16	13	17	15	16	13	14	15	19	**16**	**20**	**18**	18	15
0.5	20	16	13	19	15	16	12	15	15	**20**	**17**	**21**	**18**	**20**	15
	30	22	18	26	22	23	18	20	20	27	**23**	**28**	**25**	26	22
0.6	20	21	17	25	20	21	16	19	19	**26**	**22**	**28**	**24**	**26**	20
	30	29	24	34	29	31	25	26	27	36	**32**	**38**	**34**	35	30
0.7	20	27	22	32	27	29	23	25	24	**34**	**29**	**35**	**31**	**34**	27
	30	39	32	46	40	42	35	35	36	47	**42**	**49**	**45**	47	41
0.8	20	35	28	41	35	37	30	32	32	43	**37**	**44**	**38**	**44**	36
	30	48	40	56	50	53	45	45	47	58	**25**	**60**	**25**	58	51
0.9	20	42	35	50	43	46	37	40	39	52	**45**	**54**	**48**	**53**	44
	30	59	50	68	62	65	56	57	58	**70**	**65**	**71**	**66**	70	64

\$35 862 767	\$12 435 714	\$10 374 499	\$9 042 785	\$8 859 666	\$8 292 500	\$8 137 000
\$8 024 167	\$7 378 500	\$6 933 000	\$6 135 314	\$6 130 785	\$5 566 000	\$5 438 500
\$4 611 500	\$4 543 367	\$4 535 000	\$4 524 285	\$4 334 964	\$4 289 314	\$4 248 667
\$4 226 000	\$3 959 500	\$3 700 000	\$3 548 666	\$3 507 583		

Table 6: Season earnings of LIV golf players earning in excess of \$3.5 million.

4 Practical application

We now apply all the tests mentioned in the Monte Carlo section to observed data sets; the 2022 earnings of male golf players on the PGA (Professional Golfers' Association of America) and LIV Golf tours, respectively. LIV Golf is a professional golf tour started in Saudi Arabia and financed by their sovereign wealth oil fund. LIV represents the number 54 in Roman numerals and refers to both the number of holes played at each LIV tour event, as well as the number of shots played on a standard par 72 course if every hole played results in a birdie. The inaugural season started on 9 June 2022. To make the LIV Golf tour more enticing, the organisers offer prize earnings that are much higher than those offered by the more established PGA tour.

Table 6 shows the earnings of the 26 LIV Golf tour players who earned more than \$3.5 million in the 2022 season. The average number of events played by these 26 players on the LIV Golf tour is only 6.9, whilst the average earnings is \$7 255 386 per player and the average player earnings per tournament is \$1 051 505. Table 7 shows the earnings of the 28 PGA tour players who earned more than \$3.5 million in the 2022 season. The average number of events played by these 28 players on the PGA tour is 23.18, where this average decreases to 17.89 if one only counts the tournaments in which a player made the cut (i.e., played on the final two days of the tournament). The average earnings is \$6 098 395 per player and the average player earnings per tournament (based on 17.89 events) is \$340 882.9. From these summary statistics it is clear that the earnings for the top earners is much higher for those playing in the LIV tour as compared to those playing in the PGA tour.

Figure 1 shows side by side violin plots with overlayed box plots for the earnings data given in Tables 6 and 7. Both density estimates shown indicate that the distributions of the earnings are positively skewed. Furthermore, both plots show evidence of the presence of outliers.

The Pareto distribution was initially introduced as a model for income. We test the

$14 046 909	$10 107 897	$9 405 081	$9 369 605	$8 654 566	$7 427 299	$7 073 986
$7 012 672	$6 829 575	$6 520 597	$6 117 886	$5 776 298	$5 567 974	$5 289 842
$5 248 220	$5 076 060	$5 018 443	$4 940 600	$4 868 461	$4 837 271	$4 722 433
$4 310 047	$3 940 513	$3 876 590	$3 757 425	$3 718 990	$3 623 137	$3 616 679

Table 7: Season earnings of PGA players earning in excess of $3.5 million.

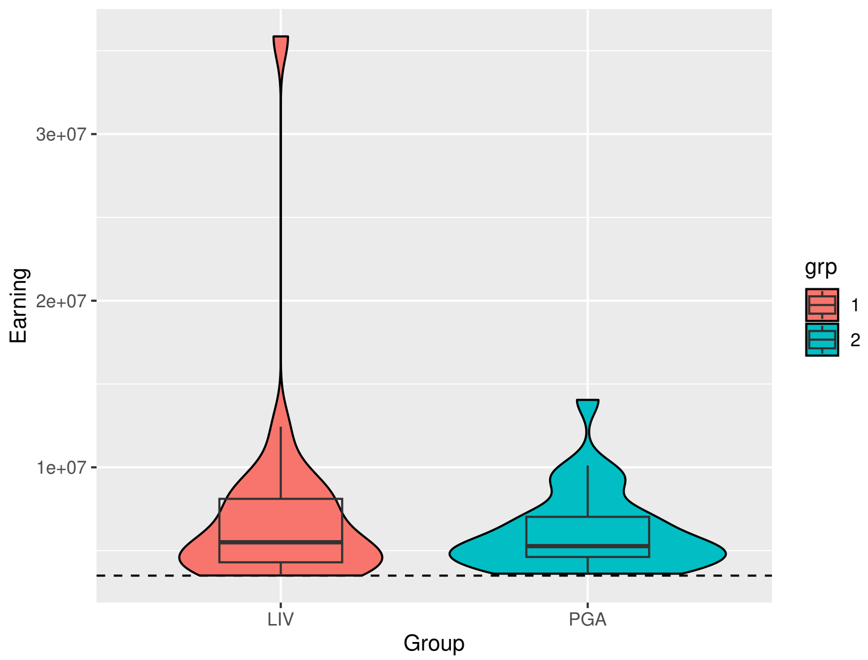

Figure 1: An image of a violin plot for PGA and LIV above 3.5m

	MME			MLE
Test	Statistic	p-value	Statistic	p-value
KS_n	0.255	0.0243	0.206	0.1284
CV_n	0.356	0.0505	0.177	0.2668
AD_n	1.655	0.0741	0.891	0.3313
ZA_n	3.484	0.1302	3.440	0.2273
$MP_n^{(1)}$	0.009	0.0489	0.005	0.2849
$MP_n^{(2)}$	0.009	0.0448	0.004	0.3135
$G_{n,1}$	0.225	0.0412	0.045	0.4480

Table 8: Summary results for the PGA Results and Ranking.

	MME			MLE
Test	Statistic	p-value	Statistic	p-value
KS_n	0.151	0.4989	0.131	0.5095
CV_n	0.112	0.4823	0.069	0.5375
AD_n	0.641	0.5331	0.457	0.5504
ZA_n	3.369	0.5501	3.357	0.5469
$MP_n^{(1)}$	0.003	0.4922	0.002	0.5219
$MP_n^{(2)}$	0.003	0.4747	0.002	0.4735
$G_{n,1}$	0.047	0.4515	0.008	0.5009

Table 9: Summary results for the LIV Golf Results and Ranking.

hypotheses that the earnings of the top players of the PGA tour and LIV Golf, respectively, are realised from a Pareto distribution. Tables 8 and 9 show the values of the test statistics as well as the corresponding p-values for the PGA and LIV Golf earnings, respectively.

Table 8 indicates that the null hypothesis of the Pareto distribution is rejected by KS_n, $MP_n^{(1)}$, $MP_n^{(2)}$ and $G_{n,1}$ at a 5% significance level when using the MME to estimate β. It should be noted that CV_n fails to reject the null hypothesis by the smallest of margins. If we set the nominal significance level of to 10%, then all of the tests considered, with the exception of ZA_n, reject the null hypothesis. On the other hand, none of the tests reject the hypothesis of the Pareto distribution when using MLE in order to perform parameter estimation. In the Monte Carlo power study, we recommend using $MP_n^{(2)}$ or $G_{n,1}$ together with MME. As a result, the rejection of these tests in this setting casts serious doubt as to the hypothesis that the earnings associated with the PGA tour are realised from a Pareto distribution.

We now turn our attention to the results associated with the LIV Golf results shown in Table 9. In this case all seven tests considered fail to reject the null hypothesis, regardless of how β is estimated, for any reasonable level of significance. We conclude that there is little doubt as to the compatibility of these data with the Pareto assumption. One possible reason for the discrepancy in the results is the mentioned higher prize money associated with LIV Golf as compared to the PGA tour. The heavy tail of the Pareto seems to provide a reasonable model for these extreme earnings.

5 Conclusions

We propose two new goodness-of-fit tests for the Pareto distribution based on a multiplicative version of the memoryless property which characterises this distribution. The resulting tests are computationally inexpensive as both test statistics require only the calculation of single summations.

A Monte Carlo study investigates the empirical powers associated with the new tests and compares these results to those of existing tests in the literature. We consider powers against fixed alternatives as well as mixture distributions. In both cases, the newly proposed tests are found to be competitive, often outperforming the existing tests against the alternative distributions considered.

A practical example relating to the earnings of golfers is included. The seasons' earnings of golfers earning in excess of \$3.5 million in both the PGA and LIV Golf are considered separately. In each case, we test the hypothesis that the observed earnings are realised from a Pareto distribution. Interestingly, we find that the Pareto distribution provides an accurate model for the earnings from LIV Golf, while there is evidence suggesting that the earnings of the PGA are not realised from this distribution.

References

Ahmad, I. A., & Alwasel, I. A. (1999). A goodness-of-fit test for exponentiality based on the memoryless property. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(3), 681–689.

Alwasel, I. (2001). On goodness of fit testing of exponenttality using the memoryless property. *Journal of Nonparametric Statistics*, *13*(4), 569–581.

Arnold, B. C. (2015). *Pareto Distributions*. CRC Press, New York.

Bourguignon, M., & Saulo, H. (2018). On a new approach to estimate the shape parameter of the inverse gaussian distribution. *South African Statistical Journal*, *52*(1), 15–27.

Galambos, J., & Kotz, S. (2006). *Characterizations of Probability Distributions.: A Unified Approach with an Emphasis on Exponential and Related Models.*, vol. 675. Springer.

Giacomini, R., Politis, D. N., & White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, *29*, 567–589.

Meintanis, S. G. (2009). A unified approach of testing for discrete and continuous Pareto laws. *Statistical Papers*, *50*(3), 569–580.

Ndwandwe, L., Allison, J., Santana, L., & Visagie, I. J. H. (2022). Testing for the Pareto type I distribution: A comparative study. *arXiv preprint arXiv:2211.10088*.

Ndwandwe, L. M., Allison, J. S., & Visagie, I. J. H. (2021). A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring. *Proceedings of the South African Statistical Association*, *62*, 17–23.

Nofal, Z. M., & El Gebaly, Y. M. (2017). New characterizations of the Pareto distribution. *Pakistan Journal of Statistics and Operation Research*, *13*, 63–74.

Nombebe, T., Allison, J. S., Santana, L., & Visagie, I. J. H. (2022). On fitting the Lomax distribution: a comparison between minimum distance estimators and other estimation techniques. *Computation, https://doi.org/10.3390/computation11030044*, *11*(44).

Pareto, V. (1897). *Cours d'economie Politique, Vol. II*. Lausanne: F. Rouge.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org/`

Rytgaard, M. (1990). Estimation in the Pareto distribution. *ASTIN Bulletin: The Journal of the IAA*, *20*(2), 201–216.

Tomy, L., Jose, M., & Veena, G. (2020). A review on recent generalizations of exponential distribution. *Biom. Biostat. Int. J*, *9*, 152–156.

Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 281–294.

Chapter 5

Conference proceeding: A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring

The conference proceeding, *A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring*, has been published as part of the proceedings of the 2021 South African Statistical Association's conference. A summary of the guidelines to authors from the journal is now presented.

Manuscript	Authors must limit the submission to eight pages or less.
Title	The title should be concise and informative.
Abstract and keywords	The aim of the abstract is to provide a concise description of your article. It should preferably be no more than 300 words and contain a minimum of symbols and references. Keywords should be listed below the abstract and should appear in alphabetical order, each keyword starting with a capital letter and separated from the previous keyword by a comma.
Tables	Table names and descriptions are placed directly above the table. Tables are numbered using Arabic numerals in the order they are cited. Table labels start with the word "Table" and the number is followed by a full stop, all in bold. Then a description follows in normal font, terminated by a full stop.
References	Only references cited in the text should be included. No section number is used. Use the small caps font for the author names. The titles of books are typeset in title case whereas the titles of journal articles employ normal sentence case. Journal titles are written out in full and not abbreviated.
General formatting	A LaTeX template is provided for submission.
Additional information	https://tinyurl.com/3pcwnk7u

Proceedings of the 62nd Annual Conference of SASA, pp. 17–23

A NEW FIXED POINT CHARACTERISATION BASED TEST FOR THE PARETO DISTRIBUTION IN THE PRESENCE OF RANDOM CENSORING

L. Ndwandwe, J. S. Allison and I. J. H. Visagie

School of Mathematical and Statistical Sciences, North-West University, South Africa

We propose a new goodness-of-fit test for the Pareto Type I lifetime distribution in the presence of random right censoring. The test is based on a fixed point characterisation, which is a generalisation of the well-known Stein method for the approximation of distributions. The empirical power performance of the new test is compared to the modified Cramér-von Mises and Kolmogorov-Smirnov tests for two different censoring proportions and two alternative lifetime distributions by means of a limited Monte Carlo study.

Keywords: Fixed point characterisation, Goodness-of-fit testing, Pareto distribution, Random censoring.

1. Introduction

The Pareto distribution, nowadays commonly known as the Pareto Type I distribution, was first introduced by Pareto (1897). It has became a popular model to use in economics, finance and actuarial science, especially where phenomena characterised by heavy tails are studied (see, Nofal and El Gebaly, 2017, Ismaïl, 2004, Dyer, 1981, Malik, 1970 and Fisk, 1961). Due to the popularity of this distribution, goodness-of-fit tests have been developed in order to test the hypothesis that an observed dataset is realised from the Pareto distribution. For a recent overview and discussion of some of these tests, see Chu et al. (2019) and the references therein.

The Pareto distribution is also used to model lifetimes in survival analysis and reliability theory; see, e.g., Amin (2007), Ouyang and Wu (1994) and Davis and Feldstein (1979). In many of these type of applications random right censoring is present and one would like to test the hypothesis that the lifetime distribution follows the Pareto Type I distribution. Apart from the traditional Kolmogorov-Smirnov and Cramér-von Mises tests, no other tests are available in the statistical literature to test the goodness-of-fit of the Pareto distribution when random censoring is present. In this paper we propose a new fixed point characterisation based test for the Pareto distribution in the presence of random right censoring.

Before proceeding, we introduce some notation. Let $X, X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) positive random variables with unknown continuous distribution function G. Let $C_1, \dots, C_n$ be i.i.d. censoring variables with unknown continuous distribution function H. We assume non-informative censoring throughout. The observed values are the pairs (T_j, δ_j), $j = 1, \dots, n$,

Corresponding author: I. J. H. Visagie (jaco.visagie@nwu.ac.za)

MSC2020 subject classifications: 62G09, 62G10, 62N03

where $T_j = \min(X_j, C_j)$ and $\delta_j = I(X_j < C_j)$, with $I(\cdot)$ denoting the indicator function. The order statistics of $T_1, \ldots, T_n$ are denoted by $T_{(1)} < \cdots < T_{(n)}$ and $\delta_{(j)}$ represents the indicator variable corresponding to $T_{(j)}$. Denote the Pareto Type I distribution with shape parameter $\beta > 0$ by

$$F_\beta(x) = 1 - x^{-\beta}, \quad x \geq 1.$$

Formally, we are interested in testing the composite hypothesis

$$H_0 : X \sim F_\beta, \tag{1}$$

for some $\beta > 0$ against general alternatives.

Based on the data (T_j, δ_j) we may estimate the distribution function underlying the lifetime using the Kaplan-Meier estimator, $\tilde{G}_n$;

$$1 - \tilde{G}_n(t) = \begin{cases} 1, & t \leq T_{(1)}, \\ \prod_{j=1}^{k-1} \left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}}, & T_{(k-1)} < t \leq T_{(k)}, \quad k = 2, \ldots, n, \\ \prod_{j=1}^{n} \left(\frac{n-j}{n-j+1}\right)^{\delta(j)}, & t > T_{(n)}. \end{cases}$$

The remainder of the article is organised as follows. In Section 2 the new test statistic is introduced and discussed. The results of a Monte Carlo study, where the size and power performance of the newly proposed test are compared to those of the two classical distribution function based tests (modified to account for censoring), are given in Section 3. The paper concludes in Section 4 with some suggestions for future research.

2. Test statistic

Throughout the paper, β is estimated using maximum likelihood. Simple calculations show that the maximum likelihood estimator of β based on the observed data is

$$\hat{\beta}(T_1, \ldots, T_n) = \frac{\sum_{j=1}^n \delta_j}{\sum_{j=1}^n \log T_j}.$$

Note that $X \sim F_\beta \iff X^\beta \sim F_1$. As a result, the tests used below are based on the transformed variables $Y_j = T_j^{\hat{\beta}(T_1,\ldots,T_n)}$, $j = 1, \ldots, n$. This is justified by the fact that

$$\hat{\beta}(Y_1, \ldots, Y_n) = \frac{\sum_{j=1}^n \delta_j}{\sum_{j=1}^n \log Y_j} = \frac{\sum_{j=1}^n \delta_j}{\hat{\beta}(T_1, \ldots, T_n) \sum_{j=1}^n \log T_j} = \frac{\hat{\beta}(T_1, \ldots, T_n)}{\hat{\beta}(T_1, \ldots, T_n)} = 1.$$

We propose a test statistic based on a fixed point characterisation of the Pareto distribution. Betsch and Ebner (2018) provide these characterisations (which are generalisations of the well-known Stein's method) for a large class of distributions. The fixed point characterisation of F_1 is given in Theorem 1.

Theorem 1. *Let Y be a continuous random variable with support $[1, \infty)$, distribution function G and $E[Y^{-1}] < \infty$. $Y \sim F_1$ if, and only if,*

$$E\left[\frac{2}{Y}(\min(Y, t) - 1)\right] = F_1(t), \quad \forall t \geq 1.$$

Theorem 1 implies that $Y \sim F_1$ if, and only if,

$$\varphi(t) := V_Y(t) - F_1(t) = 0, \tag{2}$$

for all $t > 1$, where $V_Y(t) := E\left[\frac{2}{Y}(\min(Y,t) - 1)\right]$.

Using the Kaplan-Meier estimator, $V_Y(t)$ can be estimated as

$$\hat{V}_Y(t) = \int_1^\infty \frac{2}{y}(\min(y,t) - 1)d\tilde{G}_n(y) = 2\sum_{j=1}^n \frac{\Lambda_j}{Y_j}(\min(Y_j,t) - 1), \tag{3}$$

where $\Lambda_j, j = 1, \ldots, n$ is the jump size in $\tilde{G}_n(Y_{(j)})$, given by

$$\Lambda_1 = \frac{\delta_{(1)}}{n}, \ \Lambda_n = \prod_{j=1}^{n-1}\left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}} \text{ and}$$

$$\Lambda_j = \frac{\delta_{(j)}}{n-j+1}\prod_{l=1}^{j-1}\left(\frac{n-l}{n-l+1}\right)^{\delta_{(l)}}, \ j = 2, \ldots, n-1.$$

Under the hypothesis stated in (1), the distribution function underlying $Y_1, \ldots, Y_n$ is approximately F_1, at least for large n. As a result,

$$\hat{\varphi}_n(t) := \hat{V}_Y(t) - F_1(t)$$

should be close to 0 for all $t > 1$. Therefore, we propose the test statistic

$$S_{n,a} = \int_1^\infty \hat{\varphi}_n^2(t)t^{-a}d\tilde{G}_n(t) = \sum_{k=1}^n \Lambda_k\left(2\sum_{j=1}^n \frac{\Lambda_j}{Y_j}[\min(Y_j,Y_k) - 1] + Y_k^{-1} - 1\right)^2 Y_k^{-a},$$

where t^{-a} is a weight function ensuring the existence of the integral and $a > 2$ is a tuning parameter. The test rejects the null hypothesis for large values of $S_{n,a}$. Since the null distribution of the test statistic is a function of the unknown censoring distribution H, we propose the following bootstrap algorithm to estimate the critical value.

1. Based on the pairs (T_j, δ_j),$j = 1, \ldots, n$, estimate β by $\hat{\beta} = \sum_{j=1}^n \delta_j / (\sum_{j=1}^n \log T_j)$.
2. Obtain a parametric bootstrap sample $X_1^*, \ldots, X_n^*$ by sampling from $F_{\hat{\beta}}$.
3. Obtain a non-parametric bootstrap sample $C_1^*, \ldots, C_n^*$ by sampling from the Kaplan-Meier estimate of the distribution of the censoring times.
4. Let
$$T_j^* = \min(X_j^*, C_j^*) \text{ and } \delta_j^* = \begin{cases} 1, & \text{if } X_j^* \leq C_j^* \\ 0, & \text{if } X_j^* > C_j^*. \end{cases}$$
5. Calculate $\hat{\beta}^* = \sum_{j=1}^n \delta_j^* / (\sum_{j=1}^n \log T_j^*)$ and obtain the transformed bootstrap values $Y_j^* = T_j^{*\hat{\beta}^*}$.
6. Calculate the test statistic, say $S^* = S((Y_1^*, \delta_1^*), (Y_2^*, \delta_2^*), \ldots, (Y_n^*, \delta_n^*))$, based on the data $(Y_j^*, \delta_j^*), j = 1, \ldots, n$.
7. Repeat steps 2–6 B times to obtain $S_1^*, \ldots, S_B^*$. Use the $(1-\alpha)$th quantile of $S_1^*, \ldots, S_B^*$ as the estimated critical value for the test.

3. Simulation study

In this section the empirical power performance of the newly proposed test $S_{n,a}$ is compared to that of the modified Kolmogorov-Smirnov (KS_n) and Cramér-von-Mises (CV_n) tests. For a discussion on these two modified tests, see D'Agostino and Stephens (1986) as well as Koziol and Green (1976). A significance level of $\alpha = 0.05$ is used throughout and critical values of all the tests are obtained using the bootstrap algorithm from Section 2. Estimated powers are shown for samples of size $n = 50$ and $n = 100$. The reported empirical powers are calculated in the case of 10% and 20% censoring for various alternative lifetime distributions. The alternative distributions used are the gamma and lognormal distributions, both shifted by 1 to ensure that these distributions have the same support as that of the Pareto distribution. The gamma distribution is denoted $\Gamma(\theta)$ and has density

$$h(x) = \frac{1}{\Gamma(\theta)}(x-1)^{\theta-1}e^{-(x-1)}, \quad x \geq 1.$$

The lognormal distribution is denoted $LN(\theta)$ and has density

$$h(x) = \exp\left(-\frac{1}{2}\left(\log(x-1)/\theta\right)^2\right)\left(\theta(x-1)\sqrt{2\pi}\right)^{-1}, \quad x \geq 1.$$

Empirical sizes are presented for the Pareto distributions with parameters 2 and 3, denoted by F_2 and F_3 in Tables 1 to 4. The censoring distribution used is the uniform distribution on the interval $(1, c)$, where $c > 1$ is chosen to produce the desired censoring proportion. For computational efficiency, power calculations are done using the warp-speed bootstrap proposed by Giacomini et al. (2013). For another example of the warp-speed bootstrap methodology used to calculate empirical powers, see Allison et al. (2019). All calculations are performed in R (R Core Team, 2020).

Tables 1 to 4 contain the percentage (rounded to the nearest integer) of 10 000 independent Monte Carlo samples for which the hypothesis stated in (1) was rejected. Each of these tables show the empirical powers associated with a given combination of sample size and censoring proportion. We display the highest power against each alternative in bold.

The power tables indicate that each of the tests maintain the specified significance level of 5% closely. The powers achieved by the tests increase with an increase in sample size and decrease as the censoring proportion increases. When considering the powers achieved against the various alternatives we see that the CV_n test is less powerful than the other tests considered for the alternatives above. However, this may be a result of the specific censoring distribution used. A more extensive Monte Carlo study will be required if more general conclusions are to be drawn regarding the power of the CV_n test.

Based on Tables 1 to 4, it seems that the KS_n test is quite powerful against gamma alternatives while $S_{n,a}$ achieves the highest powers against lognormal alternatives. Furthermore, it seems that smaller values of the tuning parameter in $S_{n,a}$ generally lead to higher powers.

4. Concluding remarks

To use our newly proposed test in a real world setting, a choice of the tuning parameter is necessary. Based on the Monte Carlo study above we suggest choosing $a = 2.1$ as this choice generally produces high powers for the alternatives considered. Tenreiro (2019) and Allison and Santana (2015) proposed

Table 1. Empirical powers for n = 50 with 10% censoring.

Distribution	KS_n	CV_n	$S_{n,2.1}$	$S_{n,2.2}$	$S_{n,2.5}$
F_2	5	4	5	5	5
F_3	5	4	5	5	5
$\Gamma(0.9)$	**29**	13	14	12	8
$\Gamma(1)$	**47**	15	28	26	17
$LN(1.2)$	51	4	**66**	64	60
$LN(1.5)$	8	4	**12**	**12**	11

Table 2. Empirical powers for n = 100 with 10% censoring.

Distribution	KS_n	CV_n	$S_{n,2.1}$	$S_{n,2.2}$	$S_{n,2.5}$
F_2	5	4	5	5	5
F_3	5	5	5	5	5
$\Gamma(0.9)$	**54**	16	23	20	12
$\Gamma(1)$	**78**	23	49	44	29
$LN(1.2)$	86	12	**93**	**93**	90
$LN(1.5)$	12	5	**21**	20	19

Table 3. Empirical powers for n = 50 with 20% censoring.

Distribution	KS_n	CV_n	$S_{n,2.1}$	$S_{n,2.2}$	$S_{n,2.5}$
F_2	5	4	5	5	4
F_3	4	4	5	5	4
$\Gamma(0.9)$	**22**	10	14	13	10
$\Gamma(1)$	**37**	10	27	25	21
$LN(1.2)$	42	8	**58**	**58**	**58**
$LN(1.5)$	7	5	**10**	**10**	**10**

Table 4. Empirical powers for n = 100 with 20% censoring.

Distribution	KS_n	CV_n	$S_{n,2.1}$	$S_{n,2.2}$	$S_{n,2.5}$
F_2	5	5	5	5	5
F_3	5	5	5	5	5
$\Gamma(0.9)$	**41**	10	25	23	17
$\Gamma(1)$	**68**	11	52	49	39
$LN(1.2)$	77	14	**92**	**92**	91
$LN(1.5)$	10	5	19	19	**20**

methods to choose tuning parameters in certain settings data-dependently. Unfortunately, neither of these methods are applicable in our situation as our critical values must be obtained using the bootstrap. A data-dependent choice for this type of scenario is still an open problem in the goodness-of-fit literature.

Another challenging possibility for future research is to develop the asymptotic theory relating to the newly proposed test. The dependence introduced by the presence of censoring complicates the derivation of the asymptotic results. Recently Fernández and Rivera (2020) studied Kaplan-Meier U- and V-statistics in the presence of random censoring. The results found in the mentioned paper may prove useful as tools to derive the asymptotic null distribution of the newly proposed test statistic.

References

ALLISON, J. AND SANTANA, L. (2015). On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, **85**, 3276–3288.

ALLISON, J. S., BETSCH, S., EBNER, B., AND VISAGIE, I. J. H. (2019). New weighted L^2-type tests for the inverse Gaussian distribution. *arXiv:1910.14119*.

AMIN, Z. H. (2007). Tests for the validity of the assumption that the underlying distribution of life is Pareto. *Journal of Applied Statistics*, **34**, 195–201.

BETSCH, S. AND EBNER, B. (2018). Characterizations of continuous univariate probability distributions with applications to goodness-of-fit testing. *arXiv:1810.06226*.

CHU, J., DICKIN, O., AND NADARAJAH, S. (2019). A review of goodness of fit tests for Pareto distributions. *Journal of Computational and Applied Mathematics*, **361**, 13–41.

D'AGOSTINO, R. B. AND STEPHENS, M. A. (1986). *Goodness-of-fit Techniques*. Marcel Dekker, New York.

DAVIS, H. T. AND FELDSTEIN, M. L. (1979). The generalized Pareto law as a model for progressively censored survival data. *Biometrika*, **66**, 299–306.

DYER, D. (1981). Structural probability bounds for the strong Pareto law. *Canadian Journal of Statistics*, **9**, 71–77.

FERNÁNDEZ, T. AND RIVERA, N. (2020). Kaplan-Meier V- and U-statistics. *Electronic Journal of Statistics*, **14**, 1872–1916.

FISK, P. R. (1961). The graduation of income distributions. *Econometrica*, 171–185.

GIACOMINI, R., POLITIS, D. N., AND WHITE, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, **29**, 567–589.

ISMAÏL, S. (2004). A simple estimator for the shape parameter of the Pareto distribution with economics and medical applications. *Journal of Applied Statistics*, **31**, 3–13.

KOZIOL, J. A. AND GREEN, S. B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika*, **63**, 465–474.

MALIK, H. J. (1970). A characterization of the Pareto distribution. *Scandinavian Actuarial Journal*, **1970**, 115–117.

NOFAL, Z. M. AND EL GEBALY, Y. M. (2017). New characterizations of the Pareto distribution. *Pakistan Journal of Statistics and Operation Research*, **13**, 63–74.

Ouyang, L. Y. and Wu, S. J. (1994). Prediction intervals for an ordered observation from a Pareto distribution. *IEEE Transactions on Reliability*, **43**, 264–269.

Pareto, V. (1897). *Cours d'Économie Politique*. F. Rouge, Lausanne.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

Tenreiro, C. (2019). On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, **89**, 1780–1797.

Chapter 6

Conclusion

In this chapter we make a few concluding remarks and suggest some avenues for future research.

6.1 Overview of the goals of the papers presented

This thesis gives a thorough examination of the existing literature on goodness-of-fit testing and proposes new classes of tests for the Pareto distribution. These tests are based on specific characterisations, one of which is related to the distribution of order statistics, while the other is based on a multiplicative memoryless property. Additionally, we have also proposed a test for evaluating whether the Pareto distribution is appropriate as a lifetime distribution when random right censoring is present. This test uses a fixed point characterisation that is based on the well-known Stein method used for distribution estimation. In the Monte Carlo simulations, we obtain and compare empirical powers against various alternative distributions. Some of the tests require parameter estimation, so we use both the maximum likelihood estimators and the method of moments estimators for this purpose. Afterwards, we assess the effectiveness of each test and determine which ones perform better than the others.

The key goals of this thesis and the results achieved can be briefly summarised as follows:

- Conduct an extensive literature review to examine previous studies and unresolved issues in the field of goodness-of-fit testing for the Pareto distribution.
  Chapter 2 of the thesis contain an extensive presentation of the literature review on the subject at hand.

- Propose new classes of goodness-of-fit tests for the Pareto Type I distribution based on a couple of characterisations.
  Chapter 3 presents two classes of tests that are based on the distribution of its order statistics, while Chapter 4 introduces two classes of tests that are based on the Pareto distribution's memoryless property.

- Investigate the large sample properties of the tests based on the order statistics characterisation. Chapter 3 contains a discussion on the asymptotic null distribution of the tests. We also prove the consistency of the tests against fixed alternative distributions.
- Propose a new class of tests for the Pareto type I distribution in the presence of random right censoring, and assess how well these new tests work in comparison to classical tests across a wide range of alternative distributions.
  Chapter 5 present a new test for the Pareto Type I lifetime distribution in the presence of random right censoring. This new test is shown to compare favourably to existing tests.
- Critically evaluate the performance of the newly proposed tests using Monte Carlo simulation. This has been done in Chapters 3, 4 and 5.

6.2 Overview of results

We briefly summarise the findings and conclusions of the three research articles and the proceeding.

Testing for the Pareto type I distribution: A comparative study

In the first article we provide an overview of the existing tests for the Pareto distribution, focussing specifically on the Pareto type I distribution. In addition, we consider two different methods of estimating the parameters of this 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. The numerical powers of the various test statistics are investigated and compared using a Monte Carlo study.

On a new class of tests for the Pareto distribution using Fourier methods

In the second article we proposed two new classes of goodness-of-fit tests for the Pareto distribution utilising a characterisation that is based on the order statistics. These tests, which are U and V statistics, are derived from a characterisation of the Pareto distribution that uses the distribution of the sample minimum. We conducted a thorough Monte Carlo study to examine the performance of the new tests in finite sample settings. The simulation study considers two sets of power results; the first one is concerned with powers against fixed alternative distributions and the second one considers some local power estimates, where data is simulated from two families of mixture distributions. The findings show that for most alternative distributions considered, the proposed V statistic based test performs better than all other tests. The U statistic based test is not as powerful as the V statistic, however it has comparable power to traditional test statistics.

Revisiting the memoryless property - testing for the Pareto distribution

In the third article we developed two new classes of goodness-of-fit tests for the Pareto type I distribution based on a memoryless property of this distribution. We conducted a Monte Carlo study to evaluate the effectiveness of the proposed tests compared to some of the most commoly used tests, which include the Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling tests, a test based on likelihood ratios, and a test based on the Mellin transform. The findings clearly indicated that in the vast majority of cases, the results obtained using the method of moments estimator is superior to those associated with maximum likelihood. In both cases, the newly proposed tests are found to be competitive and frequently perform better than existing tests against the alternative distributions under consideration.

A new fixed point characterisation based test for the Pareto distribution in the presence of random censoring

We propose a new goodness-of-fit test for the Pareto type I lifetime distribution in the presence of random right censoring. The test is based on a fixed point characterisation, which is obtained using the well-known Stein method for the approximation of distributions. A limited Monte Carlo study was conducted to compare the empirical power performance of the new test to that of the modified Cramér-von Mises and Kolmogorov-Smirnov tests, taking into account two different censoring proportions and two alternative lifetime distributions. The findings of the Monte Carlo study reveal that both the Kolmogorov-Smirnov test and the newly proposed test are particularly effective in terms of their power against various alternative distributions.

6.3 Concluding remarks and future research

The primary goal of this thesis was to develop new classes of goodness-of-fit tests for the Pareto type I distribution. This was achieved by utilising three characterisations; one based on order statistics, another based on a multiplicative memoryless property and a third based on a fixed point characterisation. Throughout the thesis we demonstrated that the newly proposed tests perform favourably, in terms of powers against existing tests. The paper in Chapter 2 of the thesis has been submitted to the journal Metron where reviewers requested minor changes, Chapter 3 has been accepted for a publication in the journal Stat, Chapter 4 is to be submitted to the journal Computational statistics, while Chapter 5 has been published in the conference proceedings of the South African Statistical Association.

We conclude the thesis by suggesting some potential areas for further research. One possible avenue is developing new goodness-of-fit tests for the Lomax distribution, which is a special case of the Pareto type II distribution and bears similarities to the Pareto type I distribution. This distribution has recently been suggested as a possible error distribution when fitting duration models. To determine the adequacy of the fit of the duration models, suitable goodness-of-fit tests need to be developed (currently only modified versions of traditional tests are used for this purpose). Secondly we suggest delving deeper into goodness-of-fit tests for the Pareto distribution when censoring is present. In this setting, it will be possible to base a test on the multiplicative memoryless property considered in Chapter 4. In this case, the Kaplan-Meier estimator can be use to obtain a non-parametrical estimate of the survival function, while a parametric estimate can be obtained by fitting the Pareto distribution. A goodness-of-fit test may be constructed by comparing these estimates.