Statistics

A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: www.tandfonline.com/journals/gsta20

On classes of consistent tests for the Type I Pareto distribution based on a characterization involving order statistics

Joseph Ngatchou–Wandji, Thobeka Nombebe, Leonard Santana & James Allison

To cite this article: Joseph Ngatchou–Wandji, Thobeka Nombebe, Leonard Santana & James Allison (2024) On classes of consistent tests for the Type I Pareto distribution based on a characterization involving order statistics, Statistics, 58:3, 521-551, DOI: 10.1080/02331888.2024.2347342

To link to this article: https://doi.org/10.1080/02331888.2024.2347342

Published online: 02 May 2024.

STATISTICS
2024, VOL. 58, NO. 3, 521–551
https://doi.org/10.1080/02331888.2024.2347342

Taylor & Francis
Taylor & Francis Group

On classes of consistent tests for the Type I Pareto distribution based on a characterization involving order statistics

Joseph Ngatchou–Wandji[a], Thobeka Nombebe[b], Leonard Santana[b] and James Allison[b]

[a] Université de Renne, École des hautes études en santé publique (EHESP) & Institut ´Elie Cartan de Lorraine, Nancy, France; [b]Pure and Applied Analytics, North–West University, Potchefstroom, South Africa

ABSTRACT

We propose new classes of goodness-of-fit tests for the Pareto Type I distribution. These tests are based on a characterization of the Pareto distribution involving order statistics. We derive the limiting null distribution of the tests and also show that the tests are consistent against fixed alternatives. The finite-sample performance of the newly proposed tests are evaluated and compared to some of the existing tests, where it is found that the new tests are competitive in terms of powers. The paper concludes with an application to a real world data set, namely the earnings of the 22 highest paid participants in the inaugural season of LIV golf.

ARTICLE HISTORY
Received 15 March 2023
Accepted 14 April 2024

KEYWORDS
Goodness-of-fit; consistency; Pareto distribution; characteristic function; characterization

2020 MATHEMATICS SUBJECT CLASSIFICATIONS
62F03; 62F05

1. Introduction

Many real-world phenomena exhibit measurements with heavy-tailed behaviour and, as such, lend themselves to be modelled using the Pareto distribution. First developed by economist and socialist, Vilfredo Pareto [1], the heavy-tailed Pareto distribution was initially used to model the unequal distribution of wealth in a population, but has found application in a number of other scenarios. Examples of situations modelled using this distribution include studies involving insurance claim premiums [2–4], studies involving medical insurance claims [5], and studies investigating gestation duration [6,7], to name a few (see [8]).

Due to its popularity, this distribution has enjoyed the attention of numerous researchers resulting in a number of different versions of the Pareto distribution, including the Types I, II, III, IV, and Generalised Pareto distributions. However, in this paper we will focus only on the use of the Type I Pareto distribution, which has cumulative distribution function (CDF) and probability density function (PDF) respectively given by

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

CONTACT James Allison James.Allison@nwu.ac.za 11 Hoffman street, Potchefstroom, North West Province, South Africa

where $\sigma > 0$ and $\beta > 0$ denote respectively the scale and the shape parameters. The Type I version of the Pareto distribution with $\sigma = 1$ has a number of practical applications as shown in a variety of old and new research works. In the earlier works, for example, Fisk [9] and Steindl [10] cited several examples of economic data which follow the Type I Pareto distribution, whereas Berger and Mandelbrot [11] proposed using the Type I Pareto distribution in studies of error clusters in communication circuits. It has also been shown to be useful in applications where service times and queuing systems are modelled, as discussed in Harris [12]. More recent applications include using the Type I Pareto distribution to model the wealth distribution in the Forbes 400 list [13], and modelling the city size distribution in the United States [14].

Hence, under these considerations, it is important to determine whether a realized data set, $x_1, \ldots, x_n$, from a non-negative random variable X with distribution function $F(x)$, is well-described by a Type I Pareto distribution with parameters β and σ, denoted here by $\mathcal{P}(\beta, \sigma)$. In this paper we will therefore consider using a goodness-of-fit (GOF) test to evaluate the following hypotheses regarding these data:

$$\begin{aligned} &H_0 : X \text{ follows a } \mathcal{P}(\beta,\sigma) \text{ distribution; } \exists\, \beta,\ \sigma > 0 \text{ such that } F(x) = F_{\beta,\sigma}(x),\ x \in [\sigma, \infty), \\ &\text{and} \\ &H_1 : X \text{ does not follow a } \mathcal{P}(\beta,\sigma) \text{ distribution; } \nexists\, \beta, \sigma > 0 \text{ such that } F(x) = F_{\beta,\sigma}(x), \\ &\qquad x \in [\sigma, \infty). \end{aligned} \tag{1}$$

Before discussing the existing goodness-of-fit tests for the Pareto distribution we briefly introduce the Pareto Types II, III, and IV distributions for completeness. If $X \sim \mathcal{P}(\beta, \sigma)$, then the random variable $Z = X + \mu - \sigma$ has a Pareto Type II distribution, with CDF

$$G_{\beta,\sigma,\mu}(x) = 1 - \left[1 + \left(\frac{x-\mu}{\sigma}\right)\right]^{-\beta}, \quad x \geq \mu, \tag{2}$$

where $\mu \in \mathbb{R}$ is a location parameter. Setting $\mu = 0$ in (2) we have a special case of the Pareto Type II distribution, sometimes called the Lomax distribution [15], and often appears in a reparameterised form with $\beta = 1/\xi$ and $\sigma = \delta/\xi$ (see [16] as well as Remark 1 of [17]). The Pareto Type IV distribution includes a location parameter, $\mu \in \mathbb{R}$, scale parameter, $\sigma > 0$, inequality parameter, $\gamma > 0$, and shape parameter, $\beta > 0$, and has CDF

$$H_{\mu,\sigma,\gamma,\alpha}(x) = 1 - \left[1 + \left(\frac{x-\mu}{\sigma}\right)^{1/\gamma}\right]^{-\beta}, \quad x \geq \mu. \tag{3}$$

Setting $\beta = 0$ in (3) results in the CDF of the Pareto Type III distribution. Note that the CDF of the Pareto Type I distribution can also be recovered from (3) by setting $\mu = \sigma$ and $\gamma = 1$. By setting $\gamma = 1$ in (3) the CDF of the Pareto Type II is obtained. For a full discussion on interesting properties of these distributions, as well as the relationships between the Pareto distribution and other distributions, the interested reader is referred to the monograph by Arnold [8].

Several tests have been suggested to check the goodness-of-fit of Pareto distributions; the most commonly used formal goodness-of-fit tests for Pareto distributions are those

based on the empirical distribution function (EDF), such as the Kolmogorov-Smirnov (KS) test, Cramér-von Mises (CvM) test, or Anderson-Darling (AD) test. These tests compare the empirical distribution of the data with the hypothesized theoretical Pareto distribution and assess the likelihood that the data were generated by a Pareto distribution. The results of these tests can be used to determine whether the Pareto distribution is a good fit for the data, or whether another distribution may be more appropriate. Goodness-of-fit tests for the Pareto distribution have been discussed in Beirlant et al. [18], Gulati and Shapiro [19], Martynov [20], Rizzo [21], and Falk et al. [16], among others. In Chu et al. [22] a review of established tests for the Generalized Pareto, Pareto Type I and Pareto Type II distributions is provided, whereas goodness-of-fit tests based on a variety of different characterizations for the Pareto distribution can be found in Obradovíc et al. [23], Obradovíc [24], Volkova [25], and Miloševíc and Obradovíc [26]. Ndwandwe et al. [27] provides an extensive review of the existing goodness-of-fit tests for the Pareto Type I distribution, focussing on the myriad characterizations of this distribution. Although tests specifically developed for Pareto Types II, III, and IV distributions can potentially be used to test for the Type I distribution (by exploiting relationships between these distributions), these tests will not be considered in the Monte Carlo study presented in this paper. In what follows we will refer to the Pareto Type I distribution as just the Pareto distribution.

In this paper we propose new classes of tests for the Pareto distribution. These tests are based on characterization of the Pareto distribution involving order statistics. In Section 2 we consider the case of the Type I Pareto with unit scale parameter. We present the characterization, introduce the new test statistics and derive the limiting null distribution of the tests and show that they are consistent against fixed alternatives. Section 3 is devoted to a discussion on the general Type I Pareto distribution. In Section 4 we compare the powers of our newly proposed tests with some existing tests (in the case of the general Type I Pareto distribution), while Section 5 illustrates the use of the tests in order to test the hypothesis that the 2022 season's earnings of LIV golfers (exceeding some known threshold), follows a Pareto distribution. The paper concludes in Section 6.

2. The Type I Pareto distribution with unit scale parameter

In this section, we study the case of a Type I Pareto distribution with unit scale parameter, that is, $\mathcal{P}(\beta, 1)$, $\beta > 0$.

2.1. The test statistic

Consider the following characterization of the Pareto distribution denoted here by $\mathcal{P}(\beta, 1)$, discussed in Allison et al. [28].

Characterisation. Let $X_1, \ldots, X_n$ be independent copies of a non-negative random variable X with common density function f and cumulative distribution function F. Let m be an integer such that $2 \leq m \leq n$. Then the random variables $X^{\frac{1}{m}}$ and $X_{(1)} = \min\{X_1, \ldots, X_m\}$ have the same distribution if and only if $F(x) = F_\beta(x) = F_{\beta,1}(x)$, $x \in \mathbb{R}$, $\beta > 0$.

From this characterization we have the following Theorem

Theorem 2.1: *Let $X_1, \ldots, X_n$ be copies of a non-negative random variable X with common density function f and cumulative distribution function F. Let m be an integer such that* $2 \leq$

$m \leq n$. Then the random variables $X^{\frac{1}{m}}$ and $\min\{X_1, \ldots, X_m\}$ have the same distribution if and only if for any $t \in \mathbb{R}$,

$$\mathbb{E}\left\{\frac{1}{m}\exp\left(-itX^{\frac{1}{m}}\right) - [1 - F(X)]^{m-1}\exp\left(-itX\right)\right\} = 0. \tag{4}$$

Proof: Let $2 \leq m \leq n$. It is well known that $X_{(1)} = \min\{X_1, \ldots, X_m\}$ has density function

$$\widetilde{f}(x) = m\,[1 - F(x)]^{m-1} f(x), \quad x \in \mathbb{R}.$$

It is then clear that the random variables $X^{\frac{1}{m}}$ and $X_{(1)}$ have the same distribution if and only if they have the same characteristic functions, that is, if and only if for any $t \in \mathbb{R}$,

$$\int_{\mathbb{R}} \exp\left(-itx^{\frac{1}{m}}\right) f(x)\,\mathrm{d}x = m\int_{\mathbb{R}} \exp\left(-itx\right)[1 - F(x)]^{m-1} f(x)\,\mathrm{d}x.$$

It is easy to see that the above equality is equivalent to

$$\int_{\mathbb{R}} \left\{\frac{1}{m}\exp\left(-itx^{\frac{1}{m}}\right) - \exp\left(-itx\right)[1 - F(x)]^{m-1}\right\} f(x)\,\mathrm{d}x = 0,$$

which can be written as

$$\mathbb{E}\left\{\frac{1}{m}\exp\left(-itX^{\frac{1}{m}}\right) - [1 - F(X)]^{m-1}\exp\left(-itX\right)\right\} = 0. \qquad \blacksquare$$

Let $w(\cdot)$ be any continuous function satisfying

$$w(t) > 0, \quad \lim_{t\to\pm\infty} w(t) = 0, \quad t \in \mathbb{R}, \quad 0 < \int_{\mathbb{R}} w(t)\,\mathrm{d}t < \infty \quad \text{and}$$
$$\int_{\mathbb{R}} \zeta(tx)w(x)\,\mathrm{d}x = 0, \quad t \in \mathbb{R}, \tag{5}$$

for any real-valued odd function ζ.

From the characterization and Theorem 2.1 we have that (4) characterizes the $\mathcal{P}(\beta, 1)$ distribution. Thus, suitable normalizations of empirical versions of the expectation in (4) can be used as basis for the construction of tests for that particular Pareto distribution. To this end, we propose the following test statistic

$$T_{m,n,w} = \int_{\mathbb{R}} \left|S_{m,n,\widehat{\beta}_n}(t)\right|^2 w(t)\,\mathrm{d}t, \tag{6}$$

where for all $t \in \mathbb{R}$

$$S_{m,n,\beta}(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\frac{1}{m}\exp\left(-itX_j^{\frac{1}{m}}\right) - X_j^{-\beta(m-1)}\exp\left(-itX_j\right)\right],$$

$\widehat{\beta}_n = n/\sum_{j=1}^{n}\log(X_j)$ is the maximum likelihood estimator for β.

Proposition 2.1: *Let* $2 \leq m \leq n$. *Then*

$$\int_{\mathbb{R}} |S_{m,n,\widehat{\beta}_n}(t)|^2 w(t)\,\mathrm{d}t = \int_{\mathbb{R}} |S^{\dagger}_{m,n,\widehat{\beta}_n}(t)|^2 w(t)\,\mathrm{d}t,$$

where for all $t \in \mathbb{R}$,

$$S^{\dagger}_{m,n,\beta}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{m} \nu\left(X_j^{\frac{1}{m}}; t \right) - X_j^{-\beta(m-1)} \nu(X_j; t) \right], \tag{7}$$

and ν *is the function defined on* $\mathbb{R} \times \mathbb{R}$ *by*

$$\nu(x; t) = \cos(tx) + \sin(tx). \tag{8}$$

Proof: Denote by $\bar{z}$ the conjugate of any complex number z. One has the following equalities:

$$\begin{aligned}
&\int_{\mathbb{R}} |S_{m,n,\widehat{\beta}_n}(t)|^2 w(t)\,\mathrm{d}t = \int_{\mathbb{R}} S_{m,n,\widehat{\beta}_n}(t)\overline{S_{m,n,\widehat{\beta}_n}(t)} w(t)\,\mathrm{d}t \\
&= \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \int_{\mathbb{R}} \left\{ \frac{1}{m^2} \exp\left[-it \left(X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}} \right) \right] - \frac{1}{m} X_k^{-\widehat{\beta}_n(m-1)} \exp\left[-it \left(X_j^{\frac{1}{m}} - X_k \right) \right] \right. \\
&\left. - \frac{1}{m} X_j^{-\widehat{\beta}_n(m-1)} \exp\left[-it \left(X_k^{\frac{1}{m}} - X_j \right) \right] + (X_j X_k)^{-\widehat{\beta}_n(m-1)} \exp\left[-it \left(X_j - X_k \right) \right] \right\} w(t)\,\mathrm{d}t.
\end{aligned}$$

By (5), since $x \mapsto \sin(x)$ is an odd function, one has that

$$\begin{aligned}
&\int_{\mathbb{R}} |S_{m,n,\widehat{\beta}_n}(t)|^2 w(t)\,\mathrm{d}t \\
&= \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \int_{\mathbb{R}} \left\{ \frac{1}{m^2} \cos\left[t \left(X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}} \right) \right] - \frac{1}{m} X_k^{-\widehat{\beta}_n(m-1)} \cos\left[t \left(X_j^{\frac{1}{m}} - X_k \right) \right] \right. \\
&\left. - \frac{1}{m} X_j^{-\widehat{\beta}_n(m-1)} \cos\left[t \left(X_k^{\frac{1}{m}} - X_j \right) \right] + (X_j X_k)^{-\widehat{\beta}_n(m-1)} \cos\left[t \left(X_j - X_k \right) \right] \right\} w(t)\,\mathrm{d}t.
\end{aligned}$$

Using the identity $\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b)$ and the fact that the function $x \mapsto \cos(x)\sin(x)$ is odd, one finally has:

$$\int_{\mathbb{R}} |S_{m,n,\widehat{\beta}_n}(t)|^2 w(t)\,\mathrm{d}t = \int_{\mathbb{R}} |S^{\dagger}_{m,n,\widehat{\beta}_n}(t)|^2 w(t)\,\mathrm{d}t. \qquad \blacksquare$$

For practical applications (and for the Monte-Carlo study in Section 4) we will choose $w(t) = e^{-a|t|}$ and $w(t) = e^{-at^2}$, the choices of which lead to the following calculable forms of the test statistic:

$$T^{(1)}_{n,m,a} = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \left[\frac{1}{m^2} \frac{2a}{a^2 + (X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}})^2} - \frac{1}{m} X_k^{-\widehat{\beta}_n(m-1)} \frac{2a}{a^2 + (X_j^{\frac{1}{m}} - X_k)^2} \right.$$

$$- \frac{1}{m} X_j^{-\widehat{\beta}_n(m-1)} \frac{2a}{a^2 + (X_k^{\frac{1}{m}} - X_j)^2} + X_j^{-\widehat{\beta}_n(m-1)} X_k^{-\widehat{\beta}_n(m-1)} \frac{2a}{a^2 + (X_j - X_k)^2} \Bigg].$$

and

$$\begin{aligned} T^{(2)}_{n,m,a} = \frac{1}{n}\sqrt{\frac{\pi}{a}} \sum_{j=1}^{n} \sum_{k=1}^{n} \Bigg[& \frac{1}{m^2} \exp\left(- \frac{\left(X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}} \right)^2}{4a} \right) \\ & - \frac{1}{m} X_k^{-\widehat{\beta}_n(m-1)} \exp\left(\frac{-\left(X_j^{\frac{1}{m}} - X_k \right)^2}{4a} \right) \\ & - \frac{1}{m} X_j^{-\widehat{\beta}_n(m-1)} \exp\left(\frac{-\left(X_k^{\frac{1}{m}} - X_j \right)^2}{4a} \right) \\ & + X_j^{-\widehat{\beta}_n(m-1)} X_k^{-\widehat{\beta}_n(m-1)} \exp\left(\frac{-\left(X_j - X_k \right)^2}{4a} \right) \Bigg] \end{aligned}$$

respectively.

2.2. Large sample properties

In this section we study the asymptotic properties of the newly proposed tests under the null hypothesis as well as under fixed alternatives. It is well known that under H_0, the maximum likelihood estimator of β is $\widehat{\beta}_n = n/\sum_{i=1}^{n} \log(X_i)$ and $\mathbb{E}(\widehat{\beta}_n) = n\beta/(n-1)$.

From this, under H_0, one can deduce the following equalities:

$$\begin{aligned} \sqrt{n}(\widehat{\beta}_n - \beta) &= \sqrt{n}\left(\frac{n}{\sum_{j=1}^{n} \log(X_j)} - \beta \right) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{\beta} - \log(X_j) \right] \left(\beta^2 + \frac{\beta n}{\sum_{j=1}^{n} \log(X_j)} - \beta^2 \right) \\ &= \frac{\beta^2}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{\beta} - \log(X_j) \right] + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{\beta} - \log(X_j) \right] \left(\frac{\beta n}{\sum_{j=1}^{n} \log(X_j)} - \beta^2 \right) \end{aligned} \tag{9}$$

Now, recall that under H_0, for any $j = 1, \ldots, n$, $\log(X_j)$ follows a gamma distribution with parameters 1 and β, $\sum_{j=1}^{n} \log(X_j)$ follows a gamma distribution with parameters n and

β and $1/\sum_{j=1}^{n}\log(X_j)$ follows an inverse gamma distribution with parameters n and β. From this, by the Strong Law of Large Numbers (SLLN) and Slutsky theorem, one has the following almost surely convergence:

$$\frac{\beta n}{\sum_{j=1}^{n}\log(X_j)} \longrightarrow \beta^2 \Longleftrightarrow \frac{\beta n}{\sum_{j=1}^{n}\log(X_j)} - \beta^2 \longrightarrow 0.$$

Next, by the Central Limit Theorem, the following convergence in distribution holds

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\frac{1}{\beta} - \log(X_j)\right] \Longrightarrow N,$$

where N is a zero-mean Gaussian random variable with variance $1/\beta^2$.

Collecting these two convergence results, one sees that the second term in (9) is $o_P(1)$.

Denote by $\mathcal{C} = C(\mathbb{R},\mathbb{R})$, the set of $\mathbb{R}$-valued continuous functions defined on $\mathbb{R}$. Define on $\mathcal{C}$ the metric

$$\rho(x,y) = \sum_{j=1}^{\infty} 2^{-j}\frac{\rho_j(x,y)}{1+\rho_j(x,y)}, \quad \forall\, j \geq 1, \quad \rho_j(x,y) = \sup_{\|w\|\leq j}\left|x(w) - y(w)\right|.$$

It is well known (see, for example, [29]) that endowed with ρ, $\mathcal{C}$ is a separable Fréchet space, and that convergence in this metric corresponds to the uniform convergence on all compact sets. That is for all $x, y \in \mathcal{C}$, $\rho(x,y) = 0 \Longleftrightarrow \forall\, j \geq 1,\ \rho_j(x,y) = 0$. For random elements x_n and y_n of $\mathcal{C}$, $\rho(x_n,y_n) \xrightarrow{P} 0 \Longleftrightarrow \forall\, j \geq 1,\ \rho_j(x_n,y_n) \xrightarrow{P} 0$.

Proposition 2.2: *Let $2 \leq m \leq n$. Under H_0, in $\mathcal{C}$, as n tends to infinity, in probability,*

$$S^{\dagger}_{m,n,\widehat{\beta}_n}(\cdot) = \widetilde{S}_{m,n,\beta}(\cdot) + o_P(1),$$

and for all $t \in \mathbb{R}$,

$$\widetilde{S}_{m,n,\beta}(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{1}{m}\nu\left(X_j^{\frac{1}{m}};t\right) - X_j^{-\beta(m-1)}\nu\left(X_j;t\right) + \left[\frac{1}{\beta} - \log(X_j)\right]\varphi(t)\right\}, \quad (10)$$

with φ standing for the function defined for any $t \in \mathbb{R}$ by:

$$\varphi(t) = (m-1)\beta^4\int_1^{\infty} x^{-\beta m-2}\nu\left(x;t\right)\,\mathrm{d}x. \quad (11)$$

Proof: Write for all $t \in \mathbb{R}$,

$$S^{\dagger}_{m,n,\widehat{\beta}_n}(t) = S^{\dagger}_{m,n,\beta}(t) + \widehat{S}_{m,n}(t),$$

where

$$\widehat{S}_{m,n}(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(X_j^{-\beta(m-1)} - X_j^{-\widehat{\beta}_n(m-1)}\right)\nu\left(X_j;t\right).$$

Now, from a first-order Taylor expansion, one has

$$X_j^{-\beta(m-1)} - X_j^{-\widehat{\beta}_n(m-1)} = (\widehat{\beta}_n - \beta)(m-1)\beta X_j^{-\beta(m-1)-1} + o_P(1).$$

Then, under H_0, for all $t \in \mathbb{R}$, one has the equalities:

$$\begin{aligned}\widehat{S}_{m,n}(t) &= (\widehat{\beta}_n - \beta)(m-1)\beta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j^{-\beta(m-1)-1} \nu\left(X_j; t\right) + o_P(1),\\ &= \sqrt{n}(\widehat{\beta}_n - \beta)\left[\frac{(m-1)\beta}{n} \sum_{j=1}^{n} X_j^{-\beta(m-1)-1} \nu\left(X_j; t\right)\right] + o_P(1).\end{aligned}$$

For all $t \in \mathbb{R}$, define the term in the brackets as $\varphi_n(t)$. By the law of large numbers, it is easy to see that $\varphi_n(t)$ converges point-wise to $\varphi(t)$ and that it is equicontinuous on every compact subset of $\mathbb{R}$. Therefore, it converges uniformly to $\varphi(t)$ on any compact subset Θ of $\mathbb{R}$. This result could also be obtained by applying Proposition 1 of Csörgö [30].

As a consequence of the above convergence, uniformly in $t \in \Theta$,

$$\widehat{S}_{m,n}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{\beta} - \log(X_j)\right] \varphi(t) + o_P(1)$$

and uniformly in $t \in \Theta$,

$$\begin{aligned}S^{\dagger}_{m,n,\widehat{\beta}_n}(t) &= S^{\dagger}_{m,n,\beta}(t) + \widehat{S}_{m,n}(t)\\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{\frac{1}{m}\nu\left(X_j^{\frac{1}{m}}; t\right) - X_j^{-\beta(m-1)}\nu\left(X_j; t\right) + \left[\frac{1}{\beta} - \log(X_j)\right]\varphi(t)\right\} + o_P(1)\\ &= \widetilde{S}_{m,n,\beta} + o_P(1).\end{aligned}$$

As this holds for arbitrary compact Θ, one can conclude that as n tends to infinity, in probability,

$$\rho(\widetilde{S}_{m,n,\beta}, S^{\dagger}_{m,n,\widehat{\beta}_n}) \longrightarrow 0,$$

which establishes the proposition. ■

Now, let k be the real-valued function defined for any $(x, t) \in \mathbb{R} \times \mathbb{R}$ by

$$k(x,t) = \frac{1}{m}\nu\left(x^{\frac{1}{m}}; t\right) - x^{-\beta(m-1)}\nu\left(x; t\right) + \left[\frac{1}{\beta} - \log(x)\right]\varphi(t).$$

Also consider the function K defined for any $t \in \mathbb{R}$ by

$$K(t) = \int_{\mathbb{R}} k(x,t)\, \mathrm{d}F(x),$$

and F_n the empirical cumulative distribution function of $X_1, X_2, \ldots, X_n$. Then one sees that

$$\varpi_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \big[k(X_j, t) - K(t)\big] = \int_{\mathbb{R}} k(x, t)\, \mathrm{d}\{\sqrt{n}[F_n(x, t) - F(t)]\}.$$

Under suitable conditions (see [30]), ϖ_n converges weakly to a zero-mean Gaussian process with covariance kernel

$$\mathbb{E}\left[\varpi_n(t)\varpi_n(s)\right] = \int_{\mathbb{R}} k(x, t)k(x, s)\, \mathrm{d}F(x) - K(t)K(s).$$

Theorem 2.2: *Let $m \in \{2, \ldots, n\}$, fixed. If $\beta > 1/m$, then under H_0, $\widetilde{S}_{m,n,\beta}(\cdot)$ converges weakly in $\mathcal{C}$ to a zero-mean Gaussian process $S_m(\cdot)$ with covariance kernel Γ_m defined for any $s, t \in \mathbb{R}$ by*

$$\begin{aligned}
\Gamma_m(s, t) = \beta \int_1^{\infty} \Bigg\{ & \frac{1}{m^2} \nu\left(x^{\frac{1}{m}}; t\right) \nu\left(x^{\frac{1}{m}}; s\right) + x^{-2\beta(m-1)} \nu\,(x; t)\, \nu\,(x; s) \\
& + \left(\frac{1}{\beta} - \log(x)\right)^2 \varphi(t)\varphi(s) - \frac{1}{m} x^{-\beta(m-1)} \nu\,(t; s)\, \nu\left(x^{\frac{1}{m}}; s\right) \\
& + \frac{1}{m}\left(\frac{1}{\beta} - \log(x)\right)^2 \nu\left(x^{\frac{1}{m}}; t\right) \varphi(s) - \frac{1}{m} x^{-\beta(m-1)} \nu\,(x; t)\, \nu\left(x^{\frac{1}{m}}; s\right) \\
& - x^{-\beta(m-1)} \left(\frac{1}{\beta} - \log(x)\right) \nu\,(x; t)\, \varphi(s) + \frac{1}{m}\left(\frac{1}{\beta} - \log(x)\right) \varphi(t) \nu\left(x^{\frac{1}{m}}; s\right) \\
& - x^{-\beta(m-1)} \left(\frac{1}{\beta} - \log(x)\right) \varphi(t) \nu\,(x; s) \Bigg\} x^{-\beta-1}\, \mathrm{d}x. \qquad (12)
\end{aligned}$$

Proof: We prove this result by showing that $\widetilde{S}_{m,n,\beta}(\cdot)$ is tight and its finite-dimensional distributions converge to those of any zero-mean Gaussian process with covariance kernel Γ_m. For this, we check the conditions (i), (i)$*$ and (ii)$*$ of Csörgö [30].

As $k(x, t)$ is bounded with respect to t on any compact subset of $\mathbb{R}$, so is $|k(x, t)|^{2+\delta}$, for any $\delta > 0$. Thus, one can find $t_0 \in \Theta$ such that for any $x \geq 1$

$$\sup_{t \in \Theta} |k(x, t)|^{2+\delta} = |k(x, t_0)|^{2+\delta}.$$

Consequently,

$$\begin{aligned}
\int_{\mathbb{R}} \sup_{t \in \Theta} |k(x, t)|^{2+\delta}\, \mathrm{d}F(x) &= \int_{\mathbb{R}} |k(x, t_0)|^{2+\delta}\, \mathrm{d}F(x) \\
&\leq Cst \int_{\mathbb{R}} \left[1 + x^{-(2+\delta)\beta(m-1)} + \left(\frac{1}{\beta} + \log(x)\right)^{2+\delta}\right] \mathrm{d}F(x) \\
&< \infty.
\end{aligned}$$

This establishes the convergence of the finite-dimensional distributions of $\widetilde{S}_{m,n,\beta}$ and the point (i)$*$ of Csörgö [30].

It remains to show (ii)$*$. One can write:

$$|k(x,t)-k(x,s)| \le \frac{1}{m}\left|\nu\left(x^{\frac{1}{m}};t\right)-\nu\left(x^{\frac{1}{m}};s\right)\right| + x^{-\beta(m-1)}|\nu(x;t)-\nu(x;s)| + \left|\frac{1}{\beta}-\log(x)\right||\varphi(t)-\varphi(s)|.$$

By a first-order Taylor expansion of the function $t \mapsto \nu(x;t) = \cos(xt)+\sin(xt)$, one has:

$$|k(x,t)-k(x,s)| \le \frac{x^{\frac{1}{m}}}{m}|t-s| + x^{-\beta(m-1)+1}|t-s| + Cst\left|\frac{1}{\beta}-\log(x)\right||t-s|.$$

Then, one easily sees that for any $s,t \in \Theta$, one can find $\alpha \in (0,1]$ such that

$$|k(x,t)-k(x,s)| \le |t-s|^{\alpha} Cst\left(\frac{x^{\frac{1}{m}}}{m} + x^{-\beta(m-1)+1} + \left|\frac{1}{\beta}-\log(x)\right|\right) = |t-s|^{\alpha} M(x,\nu(t,s)),$$

where ν is any Θ-valued function defined on $\Theta\times\Theta$, and for any $x \ge 1$ and $t \in \Theta$, M stands for the function defined as

$$M(x,t) = Cst\left(\frac{x^{\frac{1}{m}}}{m} + x^{-\beta(m-1)+1} + \left|\frac{1}{\beta}-\log(x)\right|\right).$$

Since $M(x,t)$ does not depend on t, $\sup_{t\in\Theta} M^2(x;t) = M^2(x;t)$ and by our assumptions

$$\int_{\mathbb{R}} \sup_{t\in\Theta} M^2(x;t)\,\mathrm{d}F(x) = \int_{\mathbb{R}} M^2(x;t)\,\mathrm{d}F(x) < \infty.$$

From all these, by Csörgö [30], one can conclude that

$$\int_{\mathbb{R}} k(x,t)\,\mathrm{d}\{\sqrt{n}[F_n(x,t)-F(t)]\}$$

converges weakly to any zero-mean Gaussian process with covariance kernel

$$\int_{\mathbb{R}} k(x,t)k(x,s)\,\mathrm{d}F(x) - K(s)K(t).$$

From the following equality

$$\int_{\mathbb{R}} k(x,t)\,\mathrm{d}\{\sqrt{n}[F_n(x,t)-F(t)]\} = \widetilde{S}_{m,n,\beta}(t) - \sqrt{n}K(t), \tag{13}$$

since under H_0, $K(t)=0$, $\widetilde{S}_{m,n,\beta}(\cdot)$ converges weakly to the Gaussian process invoked in the theorem. This establishes the theorem. ■

Theorem 2.3: *Let $m \in \{2,\dots,n\}$, fixed. Assume that under H_1, $\mathbb{E}[\log(X_1)] < \infty$. Then under H_1, in probability, for any $t \in \mathbb{R}$, $S_{m,n,\widehat{\beta}_n}(t)$ has the same asymptotic behaviour as $\sqrt{n}Q(t)$, where*

$$Q(t) = \mathbb{E}\left[\frac{1}{m}\exp\left(-itX_1^{\frac{1}{m}}\right) - [1-F(X_1)]^{(m-1)}\exp\left(-itX_1\right)\right]$$

$$+\mathbb{E}\left\{\left[X_1^{-\frac{(m-1)}{\mathbb{E}[\log(X_1)]}}-[1-F(X_1)]^{(m-1)}\right]\exp\left(-itX_1\right)\right\},\quad t\in\mathbb{R}.$$

***Proof*:** Define, for any $t\in\mathbb{R}$,

$$\dot{S}_{m,n,F}=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\frac{1}{m}\exp\left(-itX_j^{\frac{1}{m}}\right)-[1-F(X_j)]^{(m-1)}\exp\left(-itX_j\right)\right].$$

Now, adding and subtracting one has

$$\frac{S_{m,n,\widehat{\beta}_n}(t)}{\sqrt{n}}=\frac{\dot{S}_{m,n,F}(t)}{\sqrt{n}}+\frac{1}{n}\sum_{j=1}^{n}\left\{X_j^{-\hat{\beta}(m-1)}-[1-F(X_j)]^{(m-1)}\right\}\exp\left(-itX_j\right)+o_P(1).$$

Then by the SLLN the first and second terms in the right-hand side of the above equation converge point-wise respectively to

$$Q_1(t)=\mathbb{E}\left[\frac{1}{m}\exp\left(-itX_1^{\frac{1}{m}}\right)-[1-F(X_1)]^{(m-1)}\exp\left(-itX_1\right)\right]$$

and

$$Q_2(t)=\mathbb{E}\left\{\left[X_1^{-\frac{(m-1)}{\mathbb{E}[\log(X_1)]}}-[1-F(X_1)]^{(m-1)}\right]\exp\left(-itX_1\right)\right\}.$$

This establishes the theorem. ■

Theorem 2.4: *Let w be any function satisfying* (5). *Let $m\in\{2,\ldots,n\}$, fixed.*

(i) *If $\beta>1/m$, then, under H_0, as n tends to infinity, in distribution*

$$T_{m,n,w}\longrightarrow\int_{\mathbb{R}}S_m^2(t)w(t)\,\mathrm{d}t,$$

where S_m is the Gaussian process invoked in Theorem 2.2.

(ii) *Under H_1, if $\mathbb{E}[\log(X_1)]<\infty$, then as n tends to infinity,*

$$T_{m,n,w}\longrightarrow\infty.$$

***Proof*:** For Part (i), first observe that since the X_i's are iid, under H_0, one has by simple computations

$$\mathbb{E}\left[\widetilde{S}^2_{m,n,\beta}(t)\right]=\mathbb{E}\left\{\frac{1}{m}\nu\left(X_1^{\frac{1}{m}};t\right)-X_1^{-\beta(m-1)}\nu\left(X_1;t\right)+\left[\frac{1}{\beta}-\log(X_1)\right]\varphi(t)\right\}^2.$$

Denote by $B_r\subset\mathbb{R}$ a ball of radius r, and $\overline{B}_r$ its complementary in $\mathbb{R}$. Integrating both sides of the above equality with respect to $w(t)\,\mathrm{d}t$ on $\overline{B}_r$, one has:

$$\int_{\overline{B}_r}\mathbb{E}\left[\widetilde{S}^2_{m,n,\beta}(t)\right]w(t)\,\mathrm{d}t$$

$$= \int_{\overline{B}_r} \mathbb{E}\left\{\frac{1}{m}\nu\left(X_1^{\frac{1}{m}};t\right) - X_1^{-\beta(m-1)}\nu\left(X_1;t\right) + \left[\frac{1}{\beta} - \log(X_1)\right]\varphi(t)\right\}^2 w(t)\,\mathrm{d}t.$$

Since the functions $t \mapsto \nu(x,t)$ and $t \mapsto \varphi(t)$ are bounded, since $w(t) \longrightarrow 0$ as t tends to infinity, it is easy to see that as r tends to infinity, the right-hand side of the last equality converges to 0. From an adaption of Theorem 2.3 of Bilodeau and Lafaye de Micheaux [29] with $f(x) = x^2$ and $\alpha = 1$, one has that, under H_0, as n tends to infinity,

$$T_{m,n,w} \longrightarrow \int_{\mathbb{R}} S_m^2(t)w(t)\,\mathrm{d}t.$$

For the proof of the second part, it follows easily from Theorem 2.3 that for larger values of n,

$$T_{m,n,w} = n\int_{\mathbb{R}} |Q(t)|^2 w(t)\,\mathrm{d}t.$$

Under H_1, there exists a $t_1 \in \mathbb{R}$, such that $Q_1(t_1) \neq 0$. The quantity $mQ_2(t)$ is the difference between the characteristic function of $X_{(1)} = \min\{X_1,\ldots,X_m\}$ under $H_0(F = F_{1/E[\log(X_1)]})$ and its characteristic function under H_1. Since $F \neq F_{1/E[\log(X_1)]}$ under H_1, $Q_2(t_2) \neq 0$ for some $t_2 \in \mathbb{R}$. This means that under H_1, there is some $t_0 \in \mathbb{R}$ for which $Q(t_0) \neq 0$. By the continuity of $|Q|$, $\int_{\mathbb{R}} |Q(t)|^2 w(t)\,\mathrm{d}t > 0$. Whence, under H_1, as n tends to infinity, in probability,

$$T_{m,n,w} \longrightarrow \infty. \qquad \blacksquare$$

Now, assume that $w(\cdot)$ is the density function (with respect to the Lebesgue's measure) of some positive measure μ with support $\mathbb{R}$. Let $\mathcal{L}_2 = L_2(\mu)$ be the collection of functions g defined on $\mathbb{R}$ such that $\int_{\mathbb{R}} g^2(t)\,\mathrm{d}\mu(t) < \infty$. For $h_1, h_2, h \in \mathcal{L}_2$, $\langle h_1, h_2\rangle = \int_{\mathbb{R}} h_1(t)h_2(t)d\mu(t)$ and $\|h\|_{\mathcal{L}_2} = \langle h,h\rangle^{\frac{1}{2}}$ respectively stand for the usual inner product and norm on $\mathcal{L}_2$.

From our assumptions, it is easy to prove that the function $\Gamma_m(s,t)$ defined by (20) is a positive semidefinite kernel. Consequently, the integral operator ∇_{Γ_m} defined on $\mathcal{L}_2$ by

$$\nabla_{\Gamma_m}h(t) = \int_{\mathbb{R}} \Gamma_m(s,t)h(s)\,\mathrm{d}\mu(s), \quad t \in \mathbb{R}. \tag{14}$$

admits eigenvalues $\xi_1, \xi_2, \ldots$ sorted so that $\xi_1 \geq \xi_2 \geq \cdots \geq 0$, and eigenfunctions $g_1, g_2, \ldots$ which form an orthonormal basis for $\mathcal{L}_2$.

Corollary 2.1: *Under the conditions of Theorem* 2.4*, under* H_0*,* $T_{m,n,w}$ *has asymptotically the same distribution as* $\sum_{j\geq 1}\xi_j\chi_j^2$*, where* ξ_j *and* χ_j^2*,* $j \geq 1$*, are respectively the eigenvalues of* ∇_{Γ_m} *and i.i.d. random variables following a chi–squared distribution with one degree of freedom.*

Proof: The Gaussian process $S_m(\cdot)$ defined in Theorem 2.2 can be viewed as a random element of $\mathcal{L}_2$. Its Karhunen-Loève representation is given by

$$S_m(t) = \sum_{j=1}^{\infty} G_j f_j(t), \quad t \in \mathbb{R},$$

where for all $j \geq 1$, $G_j = \langle S_m(\cdot), f_j \rangle$ are independent zero–mean Gaussian random variables with variances ξ_j. Thus, $\|S_m(\cdot)\|^2_{\mathcal{L}_2} = \sum_{j=1}^{\infty} G_j^2$. Recalling that $E(G_j^2) = \xi_j \geq 0$, $j \geq 1$, for nil ξ_j's, the corresponding G_j's are nil in probability. For positive ξ_j's, one can observe that $Z_j = G_j/\sqrt{\xi_j}$, $j \geq 1$, are iid standard Gaussian random variables. Thus,

$$\int_{\mathbb{R}} S_m^2(t)\, \mathrm{d}\mu(t) = \|S_m(\cdot)\|^2_{\mathcal{L}_2} = \sum_{j=1}^{\infty} \xi_j Z_j^2. \qquad \blacksquare$$

One can approximate the distribution of $\sum_{j=1}^{\infty} \xi_j \chi_j^2$ by that of $\sum_{j=1}^{J} \xi_j \chi_j^2$ for any integer J large enough. Since the ξ_j's are unknown, they can be estimated by the eigenvalues of the operator $\nabla_{\widehat{\Gamma}_{m,n}}(t)$, where $\widehat{\Gamma}_m(t)$ is any consistent estimator of $\Gamma_m(t)$. A possible way is to estimate them by the $\widehat{\xi}_j$'s from the integral equations

$$\nabla_{\widehat{\Gamma}_{m,n}} \widehat{f}_j = \widehat{\xi}_j \widehat{f}_j, \quad j \geq 1.$$

A natural estimator of $\Gamma_m(s,t)$ can be obtained by taking the empirical counterpart in the expression given in (20), in which β is replaced by $\widehat{\beta}_n$. Some indications on the computation of the cumulative distribution function of $\sum_{j=1}^{J} \widehat{\xi}_j \chi_j^2$ can be found in Ngatchou-Wandji [31] or in Fan et al. [32]. We will not pursue this further in this paper.

3. The case of the general Type I Pareto distribution

In this section, we indicate how to treat the case where the observations follow a more general $\mathcal{P}(\beta, \sigma)$ distribution. That is, we consider testing the following more general hypotheses:

$$H_0 : \exists\, \beta,\ \sigma > 0 \text{ such that } F(x) = F_{\beta,\sigma}(x),\ x \in \mathbb{R}$$

and

$$H_1 : \nexists\, \beta, \sigma > 0 \text{ such that } F(x) = F_{\beta,\sigma}(x), x \in \mathbb{R},$$

where we recall that

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

Remark 3.1: The above testing problem can be related to that of the preceding sections by using the fact that a non-negative random variable X follows a $\mathcal{P}(\beta, \sigma)$ distribution if and only if the scaled random variable X/σ follows a $\mathcal{P}(\beta, 1)$ distribution. This can be seen easily by observing that:

- if $X \sim \mathcal{P}(\beta, \sigma)$ then for any $x \geq 1$,

$$\mathbb{P}\left(\frac{X}{\sigma} \leq x\right) = \mathbb{P}(X \leq \sigma x) = 1 - x^{-\beta}$$

- if $X/\sigma \sim \mathcal{P}(\beta, 1)$ then for any $x \geq \sigma$,

$$\mathbb{P}\left(X \leq x\right) = \mathbb{P}\left(\frac{X}{\sigma} \leq \frac{x}{\sigma}\right) = 1 - \left(\frac{x}{\sigma}\right)^{-\beta}.$$

Now, let $Y_1, \ldots, Y_n$ be an independent and identically distributed sample following a $\mathcal{P}(\beta, \sigma)$ distribution. In the case where σ is known, by Remark 3.1, the current testing problem can be done as the one studied in Section 2, by considering the scaled observations $X_j = Y_j/\sigma, j = 1, \ldots, n$.

With this, all the results as those obtained for $\sigma = 1$ can be established.

The null distribution of the test statistic can be computed along the same lines as at the end of Section 2.

Also, the consistency of the test can be handled by establishing a result similar to Theorem 2.3 and another similar to the second part of Theorem 2.4.

The case where σ is unknown is more interesting and the one encountered in practice. Here, it is natural to consider the scaled observations $Y_j/\widehat{\sigma}_n$, $j = 1, \ldots, n$, where $\widehat{\sigma}_n$ is a consistent estimator of σ. In the sequel, we use the maximum likelihood estimators of the parameters of σ and β given by

$$\widehat{\sigma}_n = \min\{Y_1, \ldots, Y_n\} \quad \text{and} \quad \widehat{\beta}_n = \frac{n}{\sum_{i=1}^{n} \log\left(\frac{Y_j}{\widehat{\sigma}_n}\right)}.$$

Letting $X_j = Y_j/\sigma$ and observing that $F_{\beta,\sigma}(\sigma) = 0$ and $f_{\beta,\sigma}(\sigma) = \beta/\sigma$, from the Bahadur representation of sample quantiles, one can write

$$\sqrt{n}(\widehat{\sigma} - \sigma) = -\sqrt{n}\frac{F_n(\sigma)}{f_{\beta,\sigma}(\sigma)} + o_P(1) = -\frac{\sigma}{\beta}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \mathbb{I}\left(X_j \leq 1\right) + o_P(1),$$

where we recall that $\mathbb{I}(\cdot)$ is the indicator function and here $F_n(\cdot)$ is the empirical cumulative distribution function of the $Y_1, \ldots, Y_n$.

Next, by a Taylor expansion (the delta method), one has

$$\sqrt{n}[(\log(\widehat{\sigma}_n) - \log(\sigma))] = -\frac{1}{\beta}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \mathbb{I}\left(X_j \leq 1\right) + o_P(1).$$

Also, by the Bahadur representation of $\widehat{\sigma}_n$, by its consistency and the Slutsky theorem, the Bahadur representation of $\widehat{\beta}_n$ is given by:

$$\sqrt{n}(\widehat{\beta}_n - \beta) = \frac{n\beta}{\sum_{j=1}^{n} \log\left(\frac{Y_j}{\widehat{\sigma}_n}\right)} \left\{\sqrt{n}\left[\frac{1}{\beta} - \frac{\sum_{j=1}^{n} \log\left(\frac{Y_j}{\widehat{\sigma}_n}\right)}{n}\right]\right\}$$

$$= \frac{n\beta}{\log(\sigma) - \log(\widehat{\sigma}_n) + \sum_{j=1}^{n} \log\left(X_j\right)} \times \left\{ \sqrt{n} \left[\frac{1}{\beta} - \frac{\sum_{j=1}^{n} \log\left(X_j\right)}{n} - \log(\sigma) + \log(\widehat{\sigma}_n) \right] \right\}$$

$$= \frac{\beta^2}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{\beta} - \log\left(X_j\right) - \frac{1}{\beta} \mathbb{I}\left(X_j \leq 1\right) \right] + o_P(1), \tag{15}$$

Now, the test static is

$$T_{m,n,w} = \int_{\mathbb{R}} \left| W_{m,n,\widehat{\beta}_n,\widehat{\sigma}_n}(t) \right|^2 w(t)\, \mathrm{d}t, \tag{16}$$

where for all $t \in \mathbb{R}$

$$W_{m,n,\beta,\sigma}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{m} \exp\left(-itX_j^{\frac{1}{m}} \right) - X_j^{-\beta(m-1)} \exp\left(-itX_j\right) \right].$$

Then, it is easy to prove along the same lines as for Proposition 2.1 that, given an integer $m \in [2, n]$, one has

$$\int_{\mathbb{R}} |W_{m,n,\widehat{\beta}_n,\widehat{\sigma}}(t)|^2 w(t)\, \mathrm{d}t = \int_{\mathbb{R}} |W^{\dagger}_{m,n,\widehat{\beta}_n,\widehat{\sigma}}(t)|^2 w(t)\, \mathrm{d}t,$$

where for all $t \in \mathbb{R}$,

$$W^{\dagger}_{m,n,\beta,\sigma}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{1}{m} \nu\left(X_j^{\frac{1}{m}}; t \right) - X_j^{-\beta(m-1)} \nu\left(X_j; t\right) \right\}, \tag{17}$$

and $\nu(\cdot)$ is the function defined in Proposition 2.1.

For studying the asymptotic distribution of $W^{\dagger}_{m,n,\widehat{\beta}_n,\widehat{\sigma}_n}(t)$, one has to establish the corresponding versions of Proposition 2.2 and Theorem 2.2.

Proposition 3.1: *Let $2 \leq m \leq n$. Under H_0, in $\mathcal{C}$, as n tends to infinity, in probability,*

$$W^{\dagger}_{m,n,\widehat{\beta}_n,\widehat{\sigma}_n}(\cdot) = \widetilde{W}_{m,n,\beta,\sigma}(\cdot) + o_P(1),$$

where for all $t \in \mathbb{R}$,

$$\widetilde{W}_{m,n,\beta,\sigma}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{1}{m} \nu\left(X_j^{\frac{1}{m}}; t \right) - X_j^{-\beta(m-1)} \nu\left(X_j; t\right) + \left[\frac{1}{\beta} - \log\left(X_j\right) - \frac{1}{\beta} \mathbb{I}\left(X_j \leq 1\right) \right] \varphi(t) + \mathbb{I}\left(X_j \leq 1\right) \psi(t) \right\} \tag{18}$$

with φ standing for the function defined by (11) *and the function ψ is defined for any $t \in \mathbb{R}$ by*

$$\psi(t) = \frac{t}{m^2} \int_1^{\infty} x^{\frac{1}{m} - \beta - 1} \vartheta\left(x^{\frac{1}{m}}; t \right) \mathrm{d}x, \tag{19}$$

and $\vartheta(x, t)$ stands for the function $\vartheta(x, t) = \cos(xt) - \sin(xt)$.

Proof: Write for all $t \in \mathbb{R}$,

$$W_{m,n,\widehat{\beta}_n,\widehat{\sigma}_n}(t) = W_{m,n,\beta,\sigma}(t) + \widehat{W}_{m,n}(t),$$

where

$$\widehat{W}_{m,n}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\frac{1}{m} \left\{ \nu \left[\left(\frac{Y_j}{\widehat{\sigma}_n} \right)^{\frac{1}{m}}; t \right] - \nu \left[X_j^{\frac{1}{m}}; t \right] \right\} + \left(X_j^{-\beta(m-1)} - X_j^{-\widehat{\beta}_n(m-1)} \right) \nu \left(X_j; t \right) \right].$$

Now, from a first-order Taylor expansion, one has

$$\nu \left[\left(\frac{Y_j}{\widehat{\sigma}_n} \right)^{\frac{1}{m}}; t \right] - \nu \left(X_j^{\frac{1}{m}}; t \right) = (\widehat{\sigma}_n - \sigma) \frac{1}{m\sigma} X_j^{\frac{1}{m}} t\vartheta \left(X_j^{\frac{1}{m}}; t \right) + o_P(1)$$

and

$$X_j^{-\beta(m-1)} - X_j^{-\widehat{\beta}_n(m-1)} = (\widehat{\beta}_n - \beta)(m-1)\beta X_j^{-\beta(m-1)-1} + o_P(1).$$

Then, under H_0, for all $t \in \mathbb{R}$, one has the equalities:

$$\begin{aligned}
\widehat{W}_{m,n}(t) &= (\widehat{\beta}_n - \beta)(m-1)\beta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j^{-\beta(m-1)-1} \nu \left(X_j; t \right) \\
&\quad + (\widehat{\sigma}_n - \sigma) \frac{1}{m^2\sigma} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j^{\frac{1}{m}} t\vartheta \left(X_j^{\frac{1}{m}}; t \right) + o_P(1) \\
&= \sqrt{n}(\widehat{\beta}_n - \beta) \left[\frac{(m-1)\beta}{n} \sum_{j=1}^{n} X_j^{-\beta(m-1)-1} \nu \left(X_j; t \right) \right] \\
&\quad + \sqrt{n}(\widehat{\sigma}_n - \sigma) \left[\frac{t}{m^2\sigma} \frac{1}{n} \sum_{j=1}^{n} X_j^{\frac{1}{m}} \vartheta \left(X_j^{\frac{1}{m}}; t \right) \right] + o_P(1).
\end{aligned}$$

Multiply the terms in the brackets respectively by β^2 and σ/β and define the results for all $t \in \mathbb{R}$ by $\psi_{1,n}(t)$ and $\psi_{2,n}(t)$.

By the law of large numbers, it is easy to see that $\psi_{1,n}(t)$ converges point-wise to $\varphi(t)$ and that it is equicontinuous on every compact subset of $\mathbb{R}$. Therefore, it converges uniformly to $\varphi(t)$ on any compact subset Θ of $\mathbb{R}$. By the same argument, $\psi_{2,n}(t)$ converges uniformly to $\psi(t)$.

As a consequence of the above convergence, uniformly in $t \in \Theta$,

$$\widehat{W}_{m,n}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \left[\frac{1}{\beta} - \log \left(X_j \right) - \frac{1}{\beta} \mathbb{I} \left(X_j \leq 1 \right) \right] \varphi(t) + \mathbb{I} \left(X_j \leq 1 \right) \psi(t) \right\} + o_P(1)$$

and uniformly in $t \in \Theta$,

$$
\begin{aligned}
W^{\dagger}_{m,n,\widehat{\beta}_n,\widehat{\sigma}_n}(t) &= W^{\dagger}_{m,n,\beta,\sigma}(t) + \widehat{W}_{m,n}(t) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{1}{m} \nu \left(X_j^{\frac{1}{m}}; t \right) - X_j^{-\beta(m-1)} \nu \left(X_j; t \right) \right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \left[\frac{1}{\beta} - \log\left(X_j\right) - \frac{1}{\beta} \mathbb{I}\left(X_j \leq 1\right) \right] \varphi(t) + \mathbb{I}\left(X_j \leq 1\right) \psi(t) \right\} \\
&= \widetilde{W}_{m,n,\beta,\sigma} + o_P(1).
\end{aligned}
$$

As this holds for arbitrary compact Θ, one can conclude that as n tends to infinity, in probability,

$$
\rho\left(\widetilde{W}_{m,n,\beta,\sigma}, W^{\dagger}_{m,n,\widehat{\beta}_n,\widehat{\sigma}_n}\right) \longrightarrow 0,
$$

which establishes the proposition. ■

The corresponding result to Theorem 2.2 is the following:

Theorem 3.1: *Let $m \in \{2, \ldots, n\}$, fixed. If $\beta > 1/m$, then under H_0, $\widetilde{W}_{m,n,\beta,\sigma}(\cdot)$ converges weakly in $\mathcal{C}$ to a zero-mean Gaussian process $W_m(\cdot)$ with covariance kernel Λ_m defined for any $s, t \in \mathbb{R}$ by*

$$
\begin{aligned}
\Lambda_m(s,t) = \beta \int_1^{\infty} \Bigg(& \frac{1}{m^2} \nu \left(x^{\frac{1}{m}}; t \right) \nu \left(x^{\frac{1}{m}}; s \right) + x^{-2\beta(m-1)} \nu\left(x; t\right) \nu\left(x; s\right) \\
& + \left\{ \left[\frac{1}{\beta} - \log\left(x\right) - \frac{1}{\beta} \mathbb{I}\left(x \leq 1\right) \right] \varphi(t) + \mathbb{I}\left(x \leq 1\right) \psi(t) \right\} \\
& \times \left\{ \left[\frac{1}{\beta} - \log\left(x\right) - \frac{1}{\beta} \mathbb{I}\left(x \leq 1\right) \right] \varphi(s) + \mathbb{I}\left(x \leq 1\right) \psi(s) \right\} \\
& - \frac{1}{m} x^{-\beta(m-1)} \nu\left(t; s\right) \nu \left(x^{\frac{1}{m}}; s \right) \\
& + \frac{1}{m} \left\{ \left[\frac{1}{\beta} - \log\left(x\right) - \frac{1}{\beta} \mathbb{I}\left(x \leq 1\right) \right] \varphi(s) + \mathbb{I}\left(x \leq 1\right) \psi(s) \right\} \nu \left(x^{\frac{1}{m}}; t \right) \\
& - \frac{1}{m} x^{-\beta(m-1)} \nu\left(x; t\right) \nu \left(x^{\frac{1}{m}}; s \right) \\
& - x^{-\beta(m-1)} \nu\left(x; t\right) \left\{ \left[\frac{1}{\beta} - \log\left(x\right) - \frac{1}{\beta} \mathbb{I}\left(x \leq 1\right) \right] \varphi(s) + \mathbb{I}\left(x \leq 1\right) \psi(s) \right\} \\
& + \frac{1}{m} \nu \left(x^{\frac{1}{m}}; s \right) \left\{ \left[\frac{1}{\beta} - \log\left(x\right) - \frac{1}{\beta} \mathbb{I}\left(x \leq 1\right) \right] \varphi(t) + \mathbb{I}\left(x \leq 1\right) \psi(t) \right\} \\
& - x^{-\beta(m-1)} \nu\left(x; s\right) \left\{ \left[\frac{1}{\beta} - \log\left(x\right) - \frac{1}{\beta} \mathbb{I}\left(x \leq 1\right) \right] \varphi(t) + \mathbb{I}\left(x \leq 1\right) \psi(t) \right\} \Bigg) x^{-\beta-1} \, \mathrm{d}x.
\end{aligned}
\tag{20}
$$

Proof: This result can be proved with the same techniques as in the proof of Theorem 2.2 by using, in the places of the functions $k(x,t)$ and $K(t)$, the following functions

$$h(x,t) = \frac{1}{m}\nu\left(x^{\frac{1}{m}};t\right) - x^{-\beta(m-1)}\nu\,(x;t) + \left[\frac{1}{\beta} - \log(x) - \frac{1}{\beta}\mathbb{I}\,(x \le 1)\right]\varphi(t) + \mathbb{I}\,(x \le 1)\,\psi(t)$$

and

$$H(t) = \int_{\mathbb{R}} h(x,t)dF(x).$$

With these, under conditions (i), (i)* and (ii)* of Csörgö [30] one has that

$$\pi_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[h(X_j,t) - H(t)\right]$$

converges weakly to a zero-mean Gaussian process with covariance kernel

$$\mathbb{E}\left[\pi_n(t)\pi_n(s)\right] = \int_{\mathbb{R}} h(x,t)h(x,s)\,\mathrm{d}F(x) - H(t)H(s).$$

We now check these conditions under H_0. For this, we follow the same lines as in the proof of Theorem 2.2.

It is clear that $h(x,t)$ is bounded with respect to t on any compact subset of $\mathbb{R}$. So it is a trivial matter that in any compact set there is some t_0 so that for any $\delta > 0$

$$\int_{\mathbb{R}} \sup_{t\in\Theta} |h(x,t)|^{2+\delta}\,\mathrm{d}F(x) = \int_{\mathbb{R}} |k(x,t_0)|^{2+\delta}\,\mathrm{d}F(x) < \infty$$

which shows the convergence of the finite-dimensional distributions of $\widetilde{W}_{m,n,\beta,\sigma}$, and checks the point (i)$*$ of Csörgö [30].

For checking (ii)$*$, one can write:

$$|h(x,t) - h(x,s)| \le \frac{1}{m}\left|\nu\left(x^{\frac{1}{m}};t\right) - \nu\left(x^{\frac{1}{m}};s\right)\right| + x^{-\beta(m-1)}|\nu(x;t) - \nu(x;s)| + \left|\frac{1}{\beta} - \log(x) - \frac{1}{\beta}\mathbb{I}\,(x \le 1)\right||\varphi(t) - \varphi(s)| + |\psi(t) - \psi(s)|.$$

By a first-order Taylor expansion of the functions $t \mapsto \nu(x;t) = \cos(xt) + \sin(xt)$ and $t \mapsto \vartheta(x;t) = \cos(xt) - \sin(xt)$, one has:

$$|h(x,t) - h(x,s)| \le \frac{x^{\frac{1}{m}}}{m}|t-s| + x^{-\beta(m-1)+1}|t-s| + Cst\left(\left|\frac{1}{\beta} - \log(x) - \frac{1}{\beta}\mathbb{I}\,(x \le 1)\right| + 1\right)|t-s|.$$

Then, one easily sees that for any $s,t \in \Theta$, one can find $\alpha \in (0,1]$ such that

$$|h(x,t) - h(x,s)| \le |t-s|^{\alpha} Cst\left(\frac{x^{\frac{1}{m}}}{m} + x^{-\beta(m-1)+1} + \left|\frac{1}{\beta} - \log(x) - \frac{1}{\beta}\mathbb{I}\,(x \le 1)\right| + 1\right)$$

$$= |t-s|^{\alpha} M(x, v(t,s)),$$

where v is any Θ-valued function defined on $\Theta \times \Theta$, and for any $x \geq 1$ and $t \in \Theta$, M is the function defined as

$$M(x,t) = Cst\left(\frac{x^{\frac{1}{m}}}{m} + x^{-\beta(m-1)+1} + \left|\frac{1}{\beta} - \log(x) - \frac{1}{\beta}\mathbb{I}\,(x \leq 1)\right| + 1\right).$$

As $M(x,t)$ does not depend on t, $\sup_{t\in\Theta} M^2(x;t) = M^2(x;t)$ and by our assumptions

$$\int_{\mathbb{R}} \sup_{t\in\Theta} M^2(x;t)\,\mathrm{d}F(x) = \int_{\mathbb{R}} M^2(x;t)\,\mathrm{d}F(x) < \infty.$$

From these and by Csörgö [30], it results that $\pi_n(t)$ converges weakly to any zero-mean Gaussian process with covariance kernel

$$\int_{\mathbb{R}} h(x,t)h(x,s)\,\mathrm{d}F(x) - H(s)H(t).$$

Given that under H_0, $H(t) = 0$, by the fact that

$$\pi_n(t) = \widetilde{W}_{m,n,\beta,\sigma}(t) - \sqrt{n}H(t), \tag{21}$$

one can conclude that $\widetilde{W}_{m,n,\beta,\sigma}(\cdot)$ converges weakly to the Gaussian process invoked in the theorem. This establishes Theorem 3.1.

For the consistency of the test and the approximation of the quantiles of its null distribution in this case where σ is unknown, results similar to Theorems 2.3 and 2.4 can be stated and proved with the same techniques and arguments.

In what follows, we briefly comment on the application of our method to the Pareto Type II distribution. Recall that if a random variable Z follows a Pareto Type II distribution the CDF is given by (2). It readily follows that, given the random variable Z, the random variable $Y = Z - \mu + \sigma$ follows a Type I Pareto distribution, $\mathcal{P}(\beta,\sigma)$. Given $Z_1,\ldots,Z_n$ i.i.d. observations, testing for the null hypothesis of Pareto Type II distribution is tantamount to testing if $Z_1 - \mu + \sigma,\ldots,Z_n - \mu + \sigma$ follow a $\mathcal{P}(\beta,\sigma)$. Thus, it is enough to apply the above results derived for the the Pareto Type I distribution. In the case where the parameters are unknown, which is the one encountered in practice, one must replace the unknown parameter by consistent estimators, $\widehat{\beta}_n$, $\widehat{\sigma}_n$ and $\widehat{\mu}_n$. However, in this case, the asymptotics would be difficult to treat. One of the reasons being that the Bahadur representations of the estimators may not be easy to handle. We therefore do not pursue this matter further, as the focus of the paper is on the Pareto Type I distribution. ■

4. Monte Carlo simulation study and results

This section contains the results of a Monte Carlo study where the finite-sample performance of the newly proposed tests $T^{(1)}_{n,m,a}$ and $T^{(2)}_{n,m,a}$ are compared to the following existing tests for the hypothesis in (1); i.e., the general Type I Pareto distribution:

- The traditional Kolmogorov-Smirnov (KS_n) and Cramer-von Mises (CV_n) tests.

- Two tests proposed by Zhang [33] based on the likelihood ratio, with test statistics given by

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

and

$$ZB_n = \sum_{j=1}^{n}\left[\log\left(\frac{\left(1-X_{(j)}^{-\hat{\beta}_n}\right)^{-1}-1}{\frac{(n-\frac{1}{2})}{(j-\frac{3}{4})-1}}\right)\right]^2,$$

where $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ denote the order statistics of $X_1, X_2, \ldots, X_n$.
- A test based on entropy utilizing the Kullback-Leibler divergence measure (see, e.g., [34]). The test statistic is given by

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

where

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

is an estimator for the entropy, with $X_{(j)} = X_{(1)}$ for $j < 1$, $X_{(j)} = X_{(n)}$ for $j > n$, and m is a window width subject to $m \leq \frac{n}{2}$.
We implement the test for $m = 1$ and $m = 10$.
- A test based on the empirical characteristic function proposed by Meintanis [35]. The test is a weighted $L2$ distance between the empirical characteristic function of transformed data and the characteristic function of the standard uniform distribution. Based on the transformation $\widehat{U}_j = F_{\widehat{\beta}_n}(X_j), j = 1, \ldots, n$, the test statistic is

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

The value of the tuning parameter is set to $a = 0.5$ and $a = 1$ in order to obtain the Monte Carlo results presented.
- A test based on the Mellin transform proposed by Meintanis [36]. The test statistic is given by

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

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

where

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

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

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

We present results for $a=0.5$ and $a=2$.

- A test proposed by Allison et al. [28]. The test statistic measures the difference between the empirical distribution of $\min\{X_1,\ldots,X_m\}$ and the V-empirical distribution of $\sqrt[m]{X}$, defined as

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

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

$$A1_{n,m}=\int_1^{\infty}\Delta_{n,m}(x)\,\mathrm{d}F_n(x).$$

We show results for $m=2$.

4.1. Simulation settings

A Monte Carlo study is carried out to examine the empirical power performance of the tests discussed in the previous sections against various fixed alternative distributions; this includes those listed in Table 1 along with two Pareto mixture distributions. Note that the alternatives listed in Table 1 natively have support $(0,\infty)$ and so we were required to *shift* these distributions by $\sigma=1$ unit to ensure that the simulated data has the same support as the Pareto distribution.

The first of the two mixture distributions that we use in this study places mixture probability $1-p$ on the Pareto distribution with parameter $\theta=3$ ($\mathcal{P}(3)$) and probability p on the lognormal distribution with parameters $\mu=-2.69$ and $\theta=2$ ($LN(-2.69,2)$). The second family of mixture distributions similarly places probability $1-p$ on the $\mathcal{P}(3)$ distribution and probability p on the Weibull distribution with parameter $\lambda=0.5$ and $\theta=0.25$ ($W(0.5,0.25)$). These parameter configurations are chosen to ensure that both distributions used in the mixtures share the same expected values. Random variates from many of these distributions can be obtained using, for example, the R package `PoweR` [37].

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

Alternative	Density function	Notation
Gamma (θ)	$\frac{1}{\Gamma(\theta)}(x-1)^{\theta-1}\mathrm{e}^{-(x-1)}$	$\Gamma(\theta)$
Weibull(λ,θ)	$\frac{\theta}{\lambda}(\frac{x-1}{\lambda})^{\theta-1}\exp\{-(\frac{x-1}{\lambda})^{\theta}\}$	$W(\lambda,\theta)$
Log-normal(μ,θ)	$\exp\{-\frac{1}{2}((\log(x-1)-\mu)/\theta)^2\}/\{\theta(x-1)\sqrt{2\pi}\}$	$LN(\mu,\theta)$
Linear failure rate(θ)	$(1+\theta(x-1))\exp(-(x-1)-\theta(x-1)^2/2)$	$LF(\theta)$
Beta exponential(θ)	$\theta\mathrm{e}^{-(x-1)}(1-\mathrm{e}^{-(x-1)})^{\theta-1}$	$BE(\theta)$
Dhillon(θ)	$\frac{\theta+1}{x+1}\exp\{-(\log(x+1))^{\theta+1}\}(\log(x+1))^{\theta}$	$DH(\theta)$
Log-Weibull(θ)	$\frac{(1+\theta)\log^{\theta}x}{x^{2+\theta}}$	$LW(\theta)$
Benini(θ)	$\frac{\exp^{-\theta\log^2 x}}{x^2}(1+2\theta\log x)$	$BN(\theta)$

The results are given in Tables 2–5 and display the estimated powers (the percentage of times the null hypothesis is rejected in $MC = 20{,}000$ independent Monte Carlo replications) calculated for sample sizes $n = 20$ and $n = 30$. Note that the results from the test statistic $T^{(1)}_{n,m,a}$ are omitted from the simulation results as they were found to be similar to $T^{(2)}_{n,m,a}$. The 'warp-speed' bootstrap method [38] is employed in order to simultaneously calculate the bootstrap critical values and Monte Carlo approximations of the power. In addition, all results are calculated by estimating the parameters using either maximum likelihood estimation (MLE),

$$\widehat{\sigma}_{MLE} = \min(X_1,\ldots,X_n) \quad \text{and} \quad \widehat{\beta}_{MLE} = \frac{n}{\sum_{i=1}^{n}\log\left(X_i/\widehat{\sigma}_{MLE}\right)}$$

or method of moments estimation (MME),

$$\widehat{\sigma}_{MME} = \frac{\bar{X}_n(\widehat{\beta}_{MME}-1)}{\widehat{\beta}_{MME}} \qquad \text{and} \qquad \widehat{\beta}_{MME} = \frac{n\bar{X} - \min(X_1,\ldots,X_n)}{n(\bar{X}_n - \min(X_1,\ldots,X_n))}.$$

The results obtained when MLE is employed are given in Tables 2 and 4, whereas the results associated with MME are given in Tables 3 and 5. A significance level of 5 % is used throughout the study and all calculations were executed using R v4.2.2 [39]. Note that all the code used in these simulations can be found here:

https://github.com/LSantanaZA/ParetoGOFUsingOrderStatistics-2023

The empirical power results for the two test statistics developed in this paper make use of specific configurations of the tuning parameters m and a, as these combinations were found to produce high powers in a separate exploratory preliminary simulation. Specifically, for $T^{(2)}$, the parameter settings used are the pairings of $m = 2$ & $a = 0.5$, $m = 3$ & $a = 0.5$ and $m = 4$ & $a = 0.5$.

4.2. Simulation results

We begin by discussing the results obtained using MLE estimators for the parameters for $n = 20$ and $n = 30$, respectively, and then move on to the results obtained when using MME estimation. The two highest values for each alternative distribution (rows) are highlighted to make comparison easier. From Tables 2–5 it is clear that all the tests closely maintain the nominal significance level of 5%.

Table 2. Estimated powers for sample size $n = 20$ based on MLE.

	KS_n	CV_n	ZA_n	ZB_n	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_{n,2}$	$T^{(2)}_{n,2,0.5}$	$T^{(2)}_{n,3,0.5}$	$T^{(2)}_{n,4,0.5}$
$\mathcal{P}$(5,1)	5	5	5	6	5	4	5	5	5	5	5	5	5	5
$\mathcal{P}$(10,1)	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Γ(0.5)	23	25	**38**	**37**	26	1	25	23	26	12	28	15	26	32
Γ(0.8)	7	7	6	5	**10**	**18**	7	4	1	4	3	7	3	2
Γ(1.0)	15	16	10	4	12	**39**	16	13	2	11	8	**19**	9	3
Γ(1.2)	25	29	19	6	18	**56**	29	25	3	21	16	**32**	18	7
W(0.5)	34	37	**52**	**50**	35	1	38	35	49	29	39	30	39	46
W(0.8)	6	6	6	6	**10**	**15**	6	4	2	2	3	7	3	2
W(1.2)	29	**32**	21	6	18	**63**	31	28	6	24	19	**32**	19	7
W(1.5)	47	**56**	42	14	30	**83**	55	52	21	48	38	53	38	19
LN(0,1.0)	13	13	7	3	8	**26**	13	11	1	10	8	**14**	9	3
LN(0,1.2)	7	7	4	2	6	**16**	7	6	1	4	4	**9**	4	2
LN(0,1.5)	4	4	3	3	**6**	**7**	4	3	2	2	3	**6**	3	2
LN(0,2.5)	29	32	37	33	21	0	33	31	**43**	34	32	25	38	**41**
LFR(0.2)	20	23	15	6	16	**51**	22	19	4	17	11	**24**	12	4
LFR(0.5)	24	27	18	6	18	**60**	27	23	6	22	14	**28**	14	5
LFR(0.8)	27	**30**	21	7	19	**64**	**30**	27	8	24	17	29	16	6
LFR(1.0)	28	**32**	23	7	21	**66**	31	28	9	25	18	30	17	6
BE(0.5)	19	21	**35**	**34**	26	2	22	19	22	9	22	11	21	26
BE(0.8)	8	8	7	6	**11**	**21**	8	6	1	5	4	8	3	2
BE(1.0)	15	16	10	4	13	**39**	15	12	2	11	7	**19**	8	3
BE(1.5)	34	41	28	9	20	**66**	40	37	7	33	25	**44**	29	12
DH(0.2)	6	6	3	3	6	**11**	6	4	1	3	3	**7**	4	2
DH(0.4)	11	12	6	2	7	**23**	11	10	1	7	6	**13**	8	3
DH(0.6)	21	22	12	4	11	**36**	**23**	20	1	16	13	**23**	16	7
DH(0.8)	30	**34**	22	6	16	**50**	33	31	3	27	21	**34**	25	12
LW(0.1)	4	4	3	3	**5**	**7**	4	3	2	3	3	4	3	3
LW(0.3)	**8**	**8**	4	2	6	**17**	**8**	7	1	5	4	**8**	6	5
LW(0.5)	15	15	9	3	9	**29**	**16**	14	1	11	8	**16**	12	11
LW(0.7)	25	**29**	16	5	13	**44**	28	26	2	21	17	25	22	21
BN(0.1)	4	4	3	4	**6**	**7**	4	3	2	2	3	4	3	3
BN(0.3)	6	6	4	4	**7**	**15**	6	4	1	3	3	5	4	3
BN(0.5)	**8**	**8**	5	3	7	**20**	**8**	6	1	5	4	6	5	4
BN(0.7)	**11**	**11**	6	3	8	**26**	**11**	9	1	7	5	8	6	6
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.1	7	8	**9**	**9**	6	3	8	8	**9**	**9**	8	**9**	**9**	**9**
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.3	18	21	24	21	11	1	20	23	**25**	21	21	22	**25**	**27**
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.5	34	39	43	37	20	0	38	41	**45**	37	39	40	**45**	**48**
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.7	56	62	63	55	33	0	61	64	**67**	58	57	62	**67**	**69**
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.1	6	6	**8**	**8**	6	4	7	7	7	7	7	7	7	**8**
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.3	11	13	**19**	**18**	11	2	13	14	16	14	15	14	16	17
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.5	24	26	**37**	34	20	1	25	27	32	23	30	25	32	**35**
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.7	39	43	**54**	50	33	0	42	45	49	37	47	40	49	**54**

When considering the estimated powers under MLE estimation, the test $KL_{n,10}$ does best for the majority of the alternatives given in Table 1, closely followed by $T^{(2)}_{n,2,0.5}$. However, we note that while the $KL_{n,10}$ test has good performance among these non-mixture alternatives, it has remarkably poor performance when applied to the mixture distributions; displaying almost no power for this sequence of alternatives. When considering the two mixture alternatives, the statistic $T^{(2)}_{n,4,0.5}$ produces some of the highest estimated powers across all mixing proportions used. In particular, it has the best performance for the Pareto-log-normal mixture alternative and is tied for the best performing test statistic (with ZA_n and ZB_n) for the Pareto-Weibull mixture.

Turning our attention to the performance of the tests under the MME estimation scheme, we find that the $G_{n,2}$ test statistic generally produces the best power performance among the non-mixture alternatives, followed, once again, by $T^{(2)}_{n,2,0.5}$. We note that, in stark

Table 3. Estimated powers for sample size $n = 20$ based on MME.

	KS_n	CV_n	ZA_n	ZB_n	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_{n,2}$	$T^{(2)}_{n,2,0.5}$	$T^{(2)}_{n,3,0.5}$	$T^{(2)}_{n,4,0.5}$
$\mathcal{P}$(5,1)	5	5	5	5	5	4	5	5	5	5	5	5	5	5
$\mathcal{P}$(10,1)	5	5	5	5	5	5	5	5	5	4	5	5	5	5
Γ(0.5)	10	10	10	11	**26**	1	11	8	9	4	12	6	12	**15**
Γ(0.8)	16	18	15	17	10	18	17	19	15	**22**	10	**21**	15	12
Γ(1.0)	34	41	37	40	13	37	40	43	36	**48**	22	**45**	37	31
Γ(1.2)	54	61	57	61	20	53	60	63	57	**68**	39	**64**	58	51
W(0.5)	12	10	14	16	**35**	1	12	8	12	5	16	14	15	**20**
W(0.8)	16	18	15	17	10	15	18	19	12	**22**	8	**20**	14	11
W(1.2)	53	60	57	61	20	61	59	62	58	**67**	42	**63**	57	50
W(1.5)	71	79	78	81	32	**82**	78	80	80	**82**	66	81	77	71
LN(0,1.0)	32	38	37	38	9	24	37	**40**	37	**44**	24	38	**40**	36
LN(0,1.2)	24	29	27	27	7	15	27	**30**	24	**33**	12	27	27	25
LN(0,1.5)	14	16	16	16	7	6	15	**17**	10	**18**	5	14	12	11
LN(0,2.5)	12	11	**31**	**36**	28	0	12	9	3	8	22	11	11	11
LFR(0.2)	41	49	45	50	17	50	48	50	45	**55**	30	**52**	45	38
LFR(0.5)	45	53	51	57	19	**59**	52	55	51	**61**	36	**59**	49	42
LFR(0.8)	47	55	54	59	20	**62**	54	57	55	**61**	40	59	53	45
LFR(1.0)	50	58	57	62	21	**65**	57	60	58	**63**	42	60	54	47
BE(0.5)	8	8	8	10	**26**	3	9	6	7	4	10	5	8	**11**
BE(0.8)	19	21	18	21	11	20	21	22	18	**26**	10	**23**	18	15
BE(1.0)	35	40	36	40	14	37	39	42	36	**48**	22	**44**	37	33
BE(1.5)	64	71	69	73	23	64	70	73	69	**77**	51	**76**	69	63
DH(0.2)	22	25	24	23	7	10	24	**26**	18	**28**	10	21	21	20
DH(0.4)	35	43	41	41	9	21	41	**44**	35	**47**	19	40	39	36
DH(0.6)	50	57	54	55	14	34	56	**59**	51	**62**	33	55	55	51
DH(0.8)	61	69	66	67	19	47	68	**71**	64	**74**	46	**71**	67	63
LW(0.1)	13	16	17	16	6	6	15	17	10	**18**	6	**18**	12	12
LW(0.3)	28	34	33	32	8	16	33	**36**	27	**38**	14	34	29	31
LW(0.5)	42	50	47	47	11	27	48	**52**	43	**55**	26	**52**	45	47
LW(0.7)	54	64	61	62	16	41	62	65	58	**68**	39	**66**	60	61
BN(0.1)	13	**16**	14	15	6	7	15	**16**	10	**17**	6	15	11	12
BN(0.3)	20	24	20	21	8	14	22	24	18	**27**	10	**25**	19	20
BN(0.5)	22	27	23	25	8	19	25	**28**	23	**31**	14	**28**	24	24
BN(0.7)	26	31	27	29	9	25	30	32	28	**36**	19	**33**	28	29
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.1	**6**	**6**	**6**	5	**7**	3	**6**	**6**	**6**	**6**	5	**6**	**6**	**6**
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.3	12	13	12	10	12	1	13	12	13	11	9	12	**14**	**16**
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.5	24	25	24	22	22	0	26	25	27	20	14	24	**29**	**32**
$\mathcal{P}$(3,1) & LN(-2.69,2), p =0.7	43	46	45	42	36	0	46	45	48	35	18	42	**50**	**54**
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.1	5	5	5	5	**6**	4	5	5	5	5	5	5	5	**6**
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.3	7	7	7	6	**11**	2	7	7	8	6	8	7	8	**9**
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.5	12	14	13	12	**20**	1	14	13	15	10	13	12	16	**20**
$\mathcal{P}$(3,1) & W(0.5,0.25), p =0.7	23	25	27	25	**34**	0	26	24	27	17	22	22	30	**35**

contrast to the powers obtained under MLE estimation, the $KL_{n,10}$ test is not competitive at all when using the MME estimation technique for the alternatives in Table 1. For the mixture alternatives, the test proposed in this paper produces some of the highest estimated powers for almost all of their parameter configurations. In particular, our test with setting $m = 4$ and $a = 0.5$ performs best for both of these mixture alternatives under almost all mixing proportions considered.

It is clear that the proposed test produces higher estimated powers for the alternatives in Table 1 when using MME estimation compared to the corresponding powers obtained under MLE estimation. Conversely, these tests calculated under MLE have much higher powers than their MME counterparts when considering the mixture alternatives. As is common with these kinds of analyses, we conclude that there is no single test with the best overall power performance, but we find that our proposed test is competitive against the

Table 4. Estimated powers for sample size $n = 30$ based on MLE.

	KS_n	CV_n	ZA_n	ZB_n	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_{n,2}$	$T^{(2)}_{n,2,0.5}$	$T^{(2)}_{n,3,0.5}$	$T^{(2)}_{n,4,0.5}$
$\mathcal{P}(5,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\mathcal{P}(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	33	36	**54**	**52**	35	2	37	33	36	16	41	21	36	44
$\Gamma(0.8)$	10	11	12	7	11	**26**	10	8	3	7	5	**14**	5	3
$\Gamma(1.0)$	24	29	23	8	16	**54**	28	28	9	25	16	**35**	21	10
$\Gamma(1.2)$	45	54	45	16	26	**76**	53	53	21	51	38	**59**	44	27
W(0.5)	49	54	**71**	**69**	49	1	54	50	66	41	58	43	55	63
W(0.8)	9	9	11	8	**12**	**22**	9	6	2	5	4	**12**	4	3
W(1.2)	50	59	49	16	27	**79**	58	58	31	55	42	**62**	47	31
W(1.5)	77	**86**	79	45	50	**95**	85	**86**	66	83	73	85	76	60
LN(0,1.0)	26	**30**	20	6	11	**36**	29	**30**	13	28	21	28	26	18
LN(0,1.2)	12	13	8	2	7	**21**	13	12	3	11	8	**15**	11	6
LN(0,1.5)	4	4	3	2	5	**9**	4	3	2	3	3	**9**	3	2
LN(0,2.5)	41	45	52	45	28	0	46	43	**57**	46	48	34	52	**57**
LFR(0.2)	35	42	34	10	21	**68**	41	39	16	38	25	**46**	29	15
LFR(0.5)	42	50	44	14	26	**79**	48	49	24	47	33	**54**	35	20
LFR(0.8)	47	57	51	18	30	**83**	55	55	30	54	38	**59**	40	24
LFR(1.0)	51	60	54	20	31	**85**	58	59	35	57	41	**61**	44	27
BE(0.5)	27	30	**50**	**49**	33	3	32	25	31	11	33	15	28	37
BE(0.8)	11	12	13	7	12	**30**	12	9	3	9	5	**14**	6	3
BE(1.0)	26	30	23	7	16	**53**	29	28	10	27	17	**35**	21	11
BE(1.5)	60	71	60	23	33	**85**	70	71	37	68	56	**74**	62	44
DH(0.2)	8	8	5	3	6	**14**	8	7	1	6	5	**9**	6	3
DH(0.4)	22	**24**	15	4	10	**30**	**24**	**24**	5	21	16	**24**	21	13
DH(0.6)	37	44	32	10	16	**50**	44	**45**	15	41	32	44	40	28
DH(0.8)	54	**62**	49	19	25	**68**	**62**	**62**	27	60	49	**62**	57	41
LW(0.1)	**5**	**5**	4	3	**5**	**8**	**5**	4	2	4	3	**5**	4	4
LW(0.3)	14	**15**	9	3	8	**21**	14	14	2	11	9	14	12	12
LW(0.5)	29	**34**	23	7	13	**41**	33	**34**	8	29	24	33	30	30
LW(0.7)	47	54	40	15	21	**60**	53	**55**	21	50	41	53	49	49
BN(0.1)	5	5	4	4	**6**	**8**	5	4	2	3	3	5	4	4
BN(0.3)	9	9	7	4	8	**19**	9	8	2	7	5	**11**	7	7
BN(0.5)	14	**16**	10	4	9	**28**	**16**	14	4	12	9	**16**	12	11
BN(0.7)	18	**20**	14	4	11	**35**	**20**	19	6	17	12	19	15	14
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.1	8	8	**10**	**10**	7	3	8	9	**10**	9	**10**	9	**10**	**11**
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.3	23	26	**32**	27	13	1	26	28	**32**	26	30	27	**32**	**35**
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.5	48	54	58	50	26	0	53	56	**61**	50	57	54	**61**	**64**
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.7	73	79	79	71	45	0	78	80	82	73	76	78	**83**	**85**
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.1	6	7	**9**	**9**	6	4	7	7	8	7	8	7	8	**9**
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.3	15	17	**25**	**24**	14	2	17	18	21	16	21	17	21	23
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.5	31	35	**50**	**48**	26	1	34	36	41	30	42	33	41	**48**
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.7	53	59	**72**	67	44	0	57	59	64	49	64	54	65	**69**

majority of other tests encountered in the literature when either MLE or MME estimation is used; this is true for both mixed and non-mixed alternatives.

Finally, it is clear from Tables 2–5 that, for the test statistic $T^{(2)}_{n,m,a}$, the choice of the parameters m and a potentially have a pronounced impact on the estimated powers produced. This fluctuation in power performance is briefly explored in Table 6 for the 'Benini' and 'log-Weibull' alternatives using a variety of different configurations of the parameters m and a. These particular alternatives were selected because they can be considered local alternatives; setting the parameter θ to zero the Pareto distribution is obtained, with deviations from Pareto occurring when $\theta > 0$. The powers are obtained using a sample size of $n = 20$ and the parameters σ and β are estimated using MME. From Table 6 it is clear that, in general, the highest powers are associated with smaller values of both m and a; the powers taper off as both m and a increase. This tendency is observed in all six alternatives

Table 5. Estimated powers for sample size $n = 30$ based on MME.

	KS_n	CV_n	ZA_n	ZB_n	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_{n,2}$	$T^{(2)}_{n,2,0.5}$	$T^{(2)}_{n,3,0.5}$	$T^{(2)}_{n,4,0.5}$
$\mathcal{P}(5,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\mathcal{P}(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	15	15	18	22	**35**	2	16	11	14	4	**27**	8	16	24
$\Gamma(0.8)$	21	23	19	24	12	**25**	22	23	18	**27**	12	**25**	18	14
$\Gamma(1.0)$	48	56	51	57	18	51	54	58	51	**64**	35	**61**	51	43
$\Gamma(1.2)$	74	81	78	82	30	73	80	83	78	**86**	61	**84**	78	70
W(0.5)	17	15	25	30	**49**	1	19	10	21	5	**42**	19	24	31
W(0.8)	21	24	19	23	13	20	23	24	16	**28**	9	**26**	17	13
W(1.2)	72	81	79	82	30	78	80	83	80	**86**	65	**84**	78	71
W(1.5)	91	95	95	**96**	53	95	95	**96**	95	**96**	89	**96**	94	90
LN(0,1.0)	49	57	56	57	14	34	55	**58**	57	**61**	40	56	**58**	56
LN(0,1.2)	33	39	39	**40**	9	20	37	**40**	36	**44**	20	36	37	36
LN(0,1.5)	17	**20**	**20**	**20**	7	8	19	**20**	13	**22**	6	16	15	13
LN(0,2.5)	13	11	38	**46**	35	0	13	8	5	6	**40**	13	14	16
LFR(0.2)	59	67	64	70	23	67	66	69	63	**73**	46	**71**	62	53
LFR(0.5)	64	73	72	77	27	77	72	75	70	**79**	56	**78**	68	59
LFR(0.8)	69	78	76	81	31	**82**	76	79	76	**82**	61	80	72	63
LFR(1.0)	71	79	79	82	32	**84**	78	81	78	**83**	64	**83**	74	66
BE(0.5)	11	10	14	18	**33**	3	12	7	9	3	**20**	5	10	16
BE(0.8)	25	29	24	28	13	29	28	29	23	**34**	14	**31**	22	17
BE(1.0)	49	58	52	58	18	51	56	59	52	**64**	35	**61**	54	45
BE(1.5)	84	90	88	90	38	83	89	91	88	**92**	77	**92**	88	83
DH(0.2)	27	33	32	32	8	13	32	**34**	26	**35**	13	28	28	27
DH(0.4)	51	59	56	56	14	28	57	**60**	52	**62**	33	54	55	53
DH(0.6)	69	77	74	75	22	48	76	**78**	73	**80**	54	75	76	72
DH(0.8)	81	87	85	86	32	65	86	88	85	**89**	71	**89**	88	82
LW(0.1)	17	**20**	**20**	**20**	7	7	19	**20**	14	**21**	7	**20**	14	15
LW(0.3)	40	**46**	**46**	45	11	20	45	**46**	39	**50**	22	**46**	41	43
LW(0.5)	61	69	67	67	18	38	68	71	63	**72**	43	**72**	65	66
LW(0.7)	76	83	81	82	27	56	82	**85**	80	**86**	63	84	81	82
BN(0.1)	17	**20**	18	19	7	8	19	**20**	13	**22**	7	19	14	15
BN(0.3)	26	30	27	28	9	18	29	**31**	23	**34**	14	**31**	25	26
BN(0.5)	33	39	34	37	11	26	37	40	33	**44**	21	**41**	34	35
BN(0.7)	38	44	40	43	12	33	43	46	41	**50**	26	**47**	42	41
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.1	6	6	6	5	**7**	3	6	6	6	6	6	6	**7**	**7**
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.3	16	17	17	16	14	1	17	17	19	15	17	17	**20**	**22**
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.5	37	40	39	38	28	0	40	39	43	31	35	37	**46**	**50**
$\mathcal{P}(3,1)$ & LN(-2.69,2), p =0.7	63	66	65	64	49	0	66	65	69	52	48	60	**70**	**75**
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.1	5	5	5	5	**6**	4	5	5	5	5	5	5	5	**6**
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.3	9	9	9	9	**14**	2	9	9	10	8	**13**	9	11	**13**
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.5	18	20	22	22	**27**	1	20	19	23	14	26	18	24	**29**
$\mathcal{P}(3,1)$ & W(0.5,0.25), p =0.7	36	39	42	43	45	1	39	36	42	25	45	34	**46**	**53**

considered here but was also found to be true for many of the other alternatives considered in the main simulation study. There were a few exceptions to this, but we only report these results as they are representative of the common trend observed.

5. Practical data application

To further investigate the behaviour of the test statistics studied in this paper, we now present the results obtained by applying those tests to a practical data set. The data set concerns the earnings for the 2022 inaugural season of LIV golf. LIV golf is a new golf series backed by the Saudi Ariabian sovereign wealth fund which aims to be an alternative to the PGA Tour by attracting star players and providing larger paydays for winners. The data were obtained from www.spotrac.com [40] and collects the player earnings in the

Table 6. Estimated powers of the test statistic $T^{(2)}_{n,m,a}$ for varying choices of a and m (using MME estimation and sample size $n = 20$) for 6 alternatives.

LW(0.1)	$m = 2$	$m = 3$	$m = 4$	$m = 10$	**BN(0.1)**	$m = 2$	$m = 3$	$m = 4$	$m = 10$
$a = 0.5$	18	12	12	12	$a = 0.5$	15	11	12	12
$a = 0.75$	13	13	11	11	$a = 0.75$	13	13	10	11
$a = 1$	11	12	12	9	$a = 1$	11	11	11	8
$a = 2$	9	9	9	9	$a = 2$	9	9	9	9
LW(0.5)	$m = 2$	$m = 3$	$m = 4$	$m = 10$	**BN(0.5)**	$m = 2$	$m = 3$	$m = 4$	$m = 10$
$a = 0.5$	52	45	47	48	$a = 0.5$	28	24	24	24
$a = 0.75$	48	49	42	44	$a = 0.75$	25	25	21	21
$a = 1$	44	44	44	30	$a = 1$	21	21	22	14
$a = 2$	30	30	30	30	$a = 2$	14	14	14	14
LW(0.9)	$m = 2$	$m = 3$	$m = 4$	$m = 10$	**BN(0.9)**	$m = 2$	$m = 3$	$m = 4$	$m = 10$
$a = 0.5$	76	70	71	72	$a = 0.5$	36	31	31	32
$a = 0.75$	72	72	66	67	$a = 0.75$	32	32	28	28
$a = 1$	67	67	67	47	$a = 1$	28	28	28	19
$a = 2$	47	47	47	47	$a = 2$	19	19	19	19

Table 7. Data set: LIV golf earnings data set (accessed 2023-09-11).

33,509,017	16,700,083	13,726,499	13,254,785	11,040,714	9,297,500	8,068,000	8,033,500
7,982,785	7,345,000	7,268,167	5,580,314	5,518,500	4,992,618	4,849,000	4,596,000
4,454,500	4,408,000	4,230,964	4,195,367	3,726,583	3,558,666	3,424,333	3,411,314
3,399,100	3,248,000	3,085,000	2,948,500	2,944,950	2,676,500	2,675,833	2,569,917
2,533,833	2,421,714	2,379,700	2,188,285	2,016,350	1,899,000	1,861,667	1,573,000
1,566,999	1,471,200	1,377,000	1,366,000	1,321,950	1,274,285	1,185,000	1,097,000
1,050,000	1,043,000	1,019,000	958,750	958,750	840,000	800,000	800,000
683,000	651,000	540,000	530,000	437,000	360,000	324,000	268,000
250,000	232,000	208,750	190,000	160,000	156,000	150,000	140,000
140,000	133,750	120,000					

2022 season. The data are shown in Table 7 and a box-plot of these values is provided in Figure 1.

It is clear from Figure 1 that the data values have a right-skewed tendency, with some extremely large outliers. The Pareto distribution is a distribution with heavy tails and is often used to model income above a specified threshold, so it sensible to also consider this distribution as a possible model for these earnings values above a threshold. We therefore focus our attention on the LIV golf earnings above the threshold of 3,500,000, indicated in Figure 1 with a dashed grey line. The values above the threshold are extracted from the data set, scaled by dividing them by the *known* value of 3,500,000, and their empirical distribution is plotted in Figure 2. Overlaid on this empirical distribution plot are two parametric Pareto distributions: one where the parameter is estimated using MLE (producing the estimate $\widehat{\beta} = 1.781$) and the other where the parameter is estimated using MME (producing the estimate $\widehat{\beta}_n = 1.932$). From these figures it seems that the Pareto might be a good fit for the data, but to more formally test this assertion we now apply all of the tests discussed in this paper to the above-threshold, scaled LIV golf earnings data, with the results found in Table 8. The estimated p-values reported in Table 8 were obtained by making use of a parametric bootstrap employing $B = 10{,}000$ bootstrap replications from a Pareto distribution with parameter estimated using either the MLE or the MME. From these p-values it

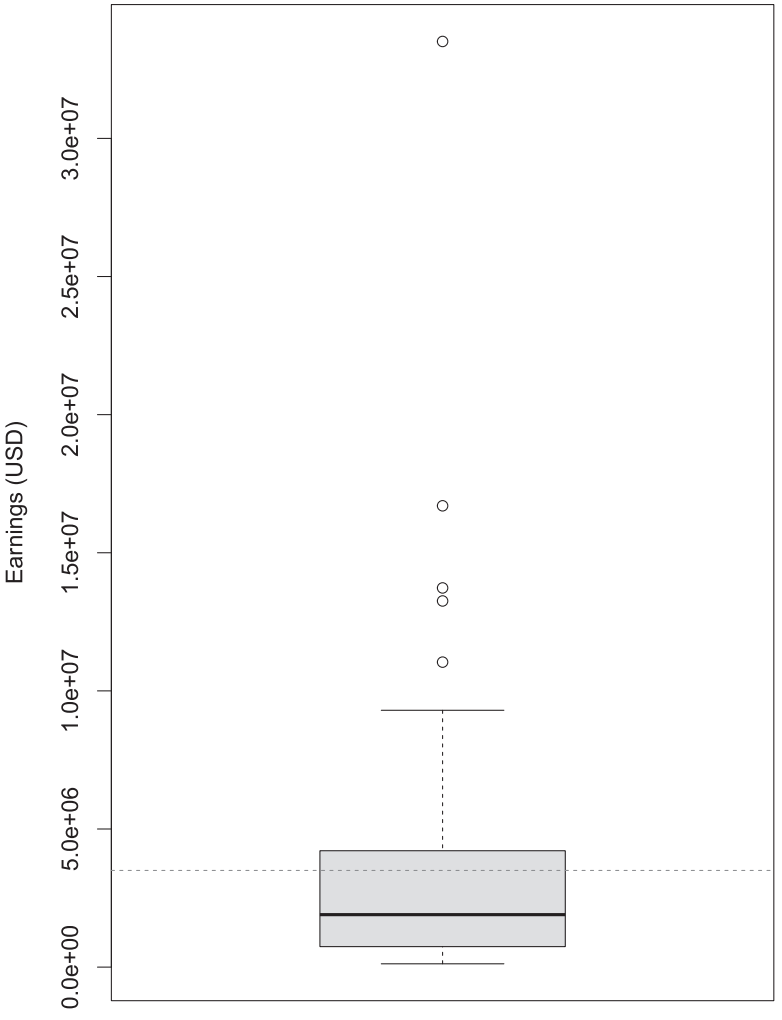

Figure 1. Box-plot of the LIV golf 2022 earnings. The dashed line represents the threshold of $3,500,000.

is clear that *all* the tests do not reject the null hypothesis that the 2022 season's earnings of LIV golfers exceeding 3,500,000, follows a Pareto distribution.

6. Concluding remarks

In this paper, we propose two new classes of goodness-of-fit tests for the Pareto Type I distribution based on a characterization involving order statistics. In addition, we also derive the null distribution of these test statistics and show that the tests are indeed consistent. A Monte Carlo simulation study is presented to demonstrate the finite-sample performance of these tests under a variety of alternative distributions. Through the inclusion of other similar tests for the Pareto distribution, the simulation study also demonstrates that our tests are competitive when compared to these other tests. Finally, the choice of the two tuning parameters appearing in these tests was also roughly explored, with the finding that,

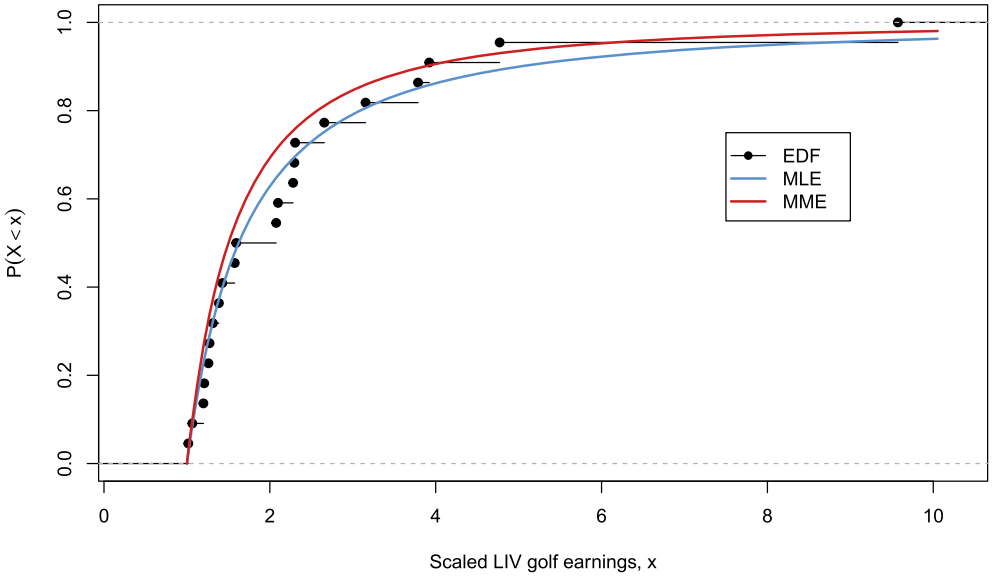

Figure 2. Empirical distribution function of the scaled above-threshold LIV golf 2022 earnings with two Pareto distribution overlays (one where the parameter is estimated via MLE and the other where it is estimated via MME).

Table 8. *p*-values for the LIV golf earnings data set (accessed 2023-09-11).

Test	MLE	MME
KS_n	0.45	0.22
CV_n	0.55	0.24
ZA_n	0.80	0.50
ZB_n	0.68	0.50
$KL_{n,1}$	0.75	0.72
$KL_{n,10}$	0.39	0.46
$M_{n,0.5}$	0.58	0.27
$M_{n,1}$	0.40	0.20
$G_{n,0.5}$	0.38	0.22
$G_{n,2}$	0.30	0.16
$A1_{n,2}$	0.61	0.61
$T^{(2)}_{n,2,0.5}$	0.47	0.22
$T^{(2)}_{n,3,0.5}$	0.36	0.22
$T^{(2)}_{n,4,0.5}$	0.37	0.24

when the tests are to be implemented in a practical setting, the choices $a = 0.5$ and $m = 2$ or $m = 3$ can be recommended.

Acknowledgments

The work of the third and fourth authors is based on research supported by the National Research Foundation (NRF). Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Pareto V. The new theories of economics. J Polit Econ. 1897;4:485–502. doi: 10.1086/250454
[2] Beirlant J, de Wet T, Goegebeur Y. Pricing risk when distributions are fat tailed. J Appl Probab. 2004;41A:157–175.
[3] Brazauskas V. Robust parametric modeling of the proportional reinsurance premium when claims are approximately Pareto-distributed. Proc Business Economic Stats Sect. 2000;A:144–149.
[4] Mert M, Saykan Y. On a bonus-malus system where the claim frequency distribution is geometric and the claim severity distribution is Pareto. Hacet J Math Stat. 2005;34:75–81.
[5] Zisheng O, Chi X. Generalized Pareto distribution fit to medical insurance claims data. Appl Math J Chinese Universities. 2006;21:21–29. doi: 10.1007/s11766-996-0018-z
[6] Keiding N, Hansen OKH, Sorensen DN, et al. The current duration approach to estimating time to pregnancy. Scand J Stat. 2012;39:185–204. doi: 10.1111/sjos.2012.39.issue-2
[7] Keiding N, Kvist K, Hartvig H, et al. Estimating time to pregnancy from current durations in a cross-sectional sample. Bio-statistics. 2002;3:565–578.
[8] Arnold B. Pareto distributions. Boca Raton, FL: CRC Press, Taylor and Francis Group; 2015.
[9] Fisk P. The graduation of income distributions. Econometrica. 1961;29:171–185. doi: 10.2307/1909287
[10] Steindl J. Random processes and the growth of firms. Madison, WI: Hafner Publishing; 1965.
[11] Berger J, Mandelbrot B. A new model for error clusterng in telephone circuits. IBM J Res Dev. 1963;7:224–236. doi: 10.1147/rd.73.0224
[12] Harris CM. The Pareto distribution as a queue service discipline. Oper Res. 1968;16:307–313. doi: 10.1287/opre.16.2.307
[13] Klass O, Biham O, Levy M, et al. The Forbes 400 and the Pareto wealth distribution. Econ Lett. 2006;90:290–295. doi: 10.1016/j.econlet.2005.08.020
[14] Ioannides Y, Skouras S. US city size distribution: Robustly Pareto, but only in the tail. J Urban Econ. 2013;73:18–29. doi: 10.1016/j.jue.2012.06.005
[15] Lomax K. Business failures: another example of the analysis of failure data. J Am Stat Assoc. 1954;49:847–852. doi: 10.1080/01621459.1954.10501239
[16] Falk M, Guillou A, Toulemonde G. A LAN based Neyman smooth test for Pareto distributions. J Stat Plan Inference. 2008;138(10):2867–2886. doi: 10.1016/j.jspi.2007.10.007
[17] Charpentier A, Flachaire E. Pareto models for risk management. Working Papers hal-02423805, HAL. 2019. Available from: https://ideas.repec.org/p/hal/wpaper/hal-02423805.html.
[18] Beirlant J, de Wet T, Goegebeur Y. A goodness-of-fit statistic for the Pareto-type behavior. J Comput Appl Math. 2006;186:99–116. doi: 10.1016/j.cam.2005.01.036
[19] Gulati S, Shapiro S. Goodness-of-fit tests for Pareto distribution. In: Statistical models and methods for biomedical and technical systems. Boston, MA: Springer; 2008. p. 259–274.
[20] Martynov G. Cramér-von Mises test for the Weibull and Pareto distributions. In: Proceedings of Dobrushin International Conference Moscow, Moscow, Russia, 2009. p. 117–122.
[21] Rizzo ML. New goodness-of-fit tests for Pareto distributions. Astin Bull. 2009;39:69–715. doi: 10.2143/AST.39.2.2044654
[22] Chu J, Dickin S, Nadarajah S. A review of goodness of fit tests for Pareto distributions. J Comput Appl Math. 2019;361:13–41. doi: 10.1016/j.cam.2019.04.018
[23] Obradovíc M, Jovanovíc M, Miloševíc B. Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics. Statistics. 2015;49:1026–1041. doi: 10.1080/02331888.2014.919297
[24] Obradovíc M. On asymptotic efficiency of goodness of fit tests for Pareto distribution based on characterizations. Filomat. 2015;29:2311–2324. doi: 10.2298/FIL1510311O

[25] Volkova K. Goodness-of-fit tests for Pareto distribution based on its characterization. Stat Methods Appl. 2016;25:351–373. doi: 10.1007/s10260-015-0330-y
[26] Miloševíc B, Obradovíc M. Two-dimensional Kolmogorov-type goodness-of-fit tests based on characterizations and their asymptotic efficiencies. J Nonparametr Stat. 2016b;28:413–427. doi: 10.1080/10485252.2016.1163358
[27] Ndwandwe L, Allison J, Santana L, et al. Testing for the Pareto type I distribution: A comparative study. 2022. arXiv preprint arXiv:2211.10088.
[28] Allison J, Milošević B, Obradović M, et al. Distribution-free goodness-of-fit tests for the Pareto distribution based on a characterization. Comput Stat. 2022;37:403–418. doi: 10.1007/s00180-021-01126-y
[29] Bilodeau M, Lafaye de Micheaux P. A multivariate empirical characteristic function test of independence with normal marginals. J Multivar Anal. 2005;9(2):345–369. doi: 10.1016/j.jmva.2004.08.011
[30] Csörgö S. Kernel-transform empirical processes. J Multivar Anal. 1983;13:517–533. doi: 10.1016/0047-259X(83)90037-4
[31] Ngatchou-Wandji J. Testing for symmetry in multivariate distributions. Stat Methodol. 2009;6(3):230–250. doi: 10.1016/j.stamet.2008.09.003
[32] Fan Y, Lafaye de Micheaux P, Penev S, et al. Multivariate nonparametric test of independence. J Multivar Anal. 2017;153:189–210. doi: 10.1016/j.jmva.2016.09.014
[33] Zhang J. Powerful goodness-of-fit tests based on the likelihood ratio. J R Stat Soc B Stat Methodol. 2002;64(2):281–294. doi: 10.1111/1467-9868.00337
[34] Ahrari V, Baratpour S, Habibirad A, et al. Goodness of fit tests for Rayleigh distribution based on quantiles. Commun Stat. 2002;51(2):341–357. doi: 10.1080/03610918.2019.1651336
[35] Meintanis S. A unified approach of testing for discrete and continuous Pareto laws. Stat Papers. 2009;50(3):569–580. doi: 10.1007/s00362-007-0103-2
[36] Meintanis S. Goodness-of-fit tests and minimum distance estimation via optimal transformation to uniformity. J Stat Plan Inference. 2009;139(2):100–108. doi: 10.1016/j.jspi.2008.03.037
[37] Lafaye de Micheaux P, Tran VA. `PoweR`: A reproducible research tool to ease Monte Carlo power simulation studies for goodness-of-fit tests in R. J Stat Softw. 2016;69(3):1–42. doi: 10.18637/jss.v069.i03
[38] Giacomini R, Politis DN, White H. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Econ Theory. 2013;29(3):567–589. doi: 10.1017/S0266466612000655
[39] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2021. Available from: https://www.R-project.org/.
[40] `www.spotrac.com`. LIV golf results by year. [accessed 2023 September 11]. Available from: https://www.spotrac.com/liv/rankings/year/2022/.