Testing the Equality of Two Pareto Distributions

Husam A. Bayoud, *Member, IAENG*

***Abstract*—This paper proposes an overlapping-based test for the equality of two Pareto distributions with different scale and shape parameters. The proposed test statistic is defined as the maximum likelihood estimate of Weitzman's overlapping coefficient, which estimates the agreement of two densities. Simulated critical points are provided for the proposed test for various sample sizes and significance levels. Statistical powers of the proposed test are computed via simulation studies. Type-I error robustness of the proposed test is studied also via simulation studies when the underlying distributions are not Pareto.**

***Index Terms*—MLE Monte Carlo Simulation, Overlapping coefficient, Simulated Power.**

I. Introduction

Testing the similarity of two populations is needed in various practical fields including pharmaceutical sciences, laboratory management, engineering sciences, food technology, management sciences, quality control and many other fields. Such tests are needed, for example, to assess the acceptability of a newly developed process\design\approach to a gold standard one. In this paper, we propose a test for the equality of two Pareto distributions of the first kind. Pareto distribution is important in practice because of its adequacy in modeling various distributions, for example, in the commercial sector, including city population distribution, stock price fluctuation, and oil field location. In addition, this distribution is highly applicable in the Air Force sector for modeling, for example, the failure time of equipment components, see Davis and Michael (1979), maintenance service times, see Harris (1968), nuclear fallout particles' distribution, see Freiling (1966), and error clusters in communications circuits, see Berger and Mandelbrot (1963).

Manuscript received Jan 16, 2017; revised Feb 17, 2017.
This work was financially supported by Fahad Bin Sultan University, Tabuk, Saudi Arabia.
Husam Awni Bayoud, College of Sciences and Humanities, Fahad Bin Sultan University, Tabuk, Saudi Arabia, P.O. Box 15700 Tabuk 71454. Phone: 00966 5 83000208; fax: 00966 4 42 76 919; e-mail: husam.awni@yahoo.com.

As described by Johnson and Kotz (1970), the probability density function (pdf) and the cumulative distribution function (cdf) of Pareto distribution with shape and scale parameters α and β, respectively, are:

$$f(x \mid \alpha, \beta) = \frac{\alpha \beta^{\alpha}}{x^{\alpha+1}}; \qquad x \geq \beta > 0, \alpha > 0\,, \tag{1}$$

and

$$F(x \mid \alpha, \beta) = 1 - \left(\frac{\beta}{x}\right)^{\alpha}; \qquad x \geq \beta > 0, \alpha > 0\,, \tag{2}$$

Throughout this article, we denote Pareto distribution by $\text{Pareto}\,(\alpha, \beta)$.

Consider for testing the similarity of two Pareto populations:

$$H_0 : F(x \mid \alpha_1, \beta_1) = F(x \mid \alpha_2, \beta_2) \text{ vs}$$
$$H_1 : F(x \mid \alpha_1, \beta_1) \neq F(x \mid \alpha_2, \beta_2)\,, \tag{3}$$

which is equivalent to test:

$$H_0 : \alpha_1 = \alpha_2 \text{ and } \beta_1 = \beta_2 \qquad \text{vs} \quad H_1 : \alpha_1 \neq \alpha_2 \text{ or } \beta_1 \neq \beta_2\,. \tag{4}$$

In general, testing the similarity of two distributions can be statistically linked to one-sided testing problem. For instance, testing the hypothesis: $H_o : F = G$ at a significance level α can be done by testing the hypothesis: $H_o : S(F,G) \geq (\leq)\ c_{\alpha}$, where $S(F,G)$ is a suitable measure of similarity (dissimilarity) of F and G, and c_{α} is the cut-off point of the measure S at a significance level α.

In this paper we propose an overlapping (OVL) coefficient-based test for the similarity of two Pareto distributions. The proposed test is defined as the Maximum Likelihood Estimate (MLE) of the Weitzman's OVL coefficient; this coefficient was first proposed by Weitzman (1970) to study the agreement of two income distributions. This OVL coefficient has received the attention of different authors in the literature for its importance in measuring the agreement of two statistical distributions. Helu and Samawi (2011) derived three Weitzman's coefficient (1970), Matusita's Coefficient (1955) and Morisita's Coefficient (1959), for two Pareto distributions of kind II, and proposed point and confidence interval estimates for these coefficients based on Simple Random and Ranked Set Samples, separately. Inman and Bradley (1989) used the OVL coefficient to measure the agreement between two normal distributions assuming equal variances. Mulekar and Mishra (1994) studied various OVL coefficients for two normal populations with the same mean. Reiser and Faraggi (1999) derived the confidence interval of the OVL coefficient of two normal distributions assuming equal variances in terms of non central t and non central F distributions separately. Chaubey et. al. (2008) studied the OVL coefficient for two inverse Gaussian populations. Al-Saleh and Samawi (2007) studied OVL coefficients for two exponential populations. Recently, OVL-based tests have been proposed for testing the equality of two exponential distributions and two normal distributions separately by Bayoud (2015) and Bayoud and Kittaneh (2016), respectively. This paper proposes the same technique for testing the equality of two Pareto distributions.

The rest of this paper is organized as follows: In section 2, the proposed test statistic is derived in analytical form. Simulated critical points for the proposed test statistic are provided in Section 3 for various sample sizes and levels of significance. An approximation to the null distribution of the proposed test statistic is proposed in Section 4. In section 5, statistical power of the proposed test is studied under various scenarios. Various real datasets are analyzed in Section 6. The paper is concluded in Section 7.

II. PROPOSED TEST STATISTIC

The proposed test statistic is defined as the estimated OVL coefficient. The OVL coefficient of $f(x|\alpha_1,\beta_1)$ and $f(x|\alpha_2,\beta_2)$ is defined as:

$$\Delta = \int_x \min(f(x|\alpha_1,\beta_1), f(x|\alpha_2,\beta_2))\,dx \qquad (5)$$

The coefficient Δ ranges from zero to one. A value close to zero indicates no common area between the two density functions. A value close to one indicates that the two densities are identical.

It can be easily proved that two Pareto densities $f(x|\alpha_1,\beta_1)$ and $f(x|\alpha_2,\beta_2)$ with different parameters either intersect at only one point in the support (β^*,∞), where $\beta^* = Max(\beta_1,\beta_2)$, or they do not intersect at any point in (β^*,∞). Note that, if $\alpha_1=\alpha_2$ and $\beta_1=\beta_2$, then the two densities are clearly identical, and then $\Delta=1$. If $\alpha_1=\alpha_2$ and $\beta_1\neq\beta_2$, then the two densities do not intersect. If $\alpha_1\neq\alpha_2$ then, regardless the values of β_1 and β_2, the two densities intersect only at x_0 given by:

$$x_0 = \left(\frac{\alpha_2\,\beta_2^{\alpha_2}}{\alpha_1\,\beta_1^{\alpha_1}}\right)^{\frac{1}{\alpha_2-\alpha_1}}, \qquad (6)$$

provided that this value belongs to (β^*,∞).

Algebraically, the value of x_0 in Eq(8) might be outside the interval (β^*,∞) at some values of α_1,α_2, β_1 and β_2, For example, when $\beta_1=1.2$, $\beta_2=1$, $\alpha_1=5$ and $\alpha_2=1$, the value of $x_0=0.603$, which does not belong to $(1.2,\infty)$, this means that the two densities do not intersect in their support; and hence, there is no intersection point. In such a case, we conclude that one of the densities lies above the other at all values in (β^*,∞).

Therefore, two different Pareto densities either intersect at only one point x_0 in the interval (β^*,∞), or they do not intersect at all values $x\in(\beta^*,\infty)$.

If the pdfs $f(x|\alpha_1,\beta_1)$ and $f(x|\alpha_2,\beta_2)$ intersect at $x_0\in(\beta^*,\infty)$, then Δ in Eq (7) can be represented as:

$$\Delta = \begin{cases} \left[\dfrac{\beta_2}{\beta^*}\right]^{\alpha_2} + \left[\dfrac{\beta_1}{x_0}\right]^{\alpha_1} - \left[\dfrac{\beta_2}{x_0}\right]^{\alpha_2}, & \text{if } f_1(\beta^*) > f_2(\beta^*) \\ \left[\dfrac{\beta_1}{\beta^*}\right]^{\alpha_1} + \left[\dfrac{\beta_2}{x_0}\right]^{\alpha_2} - \left[\dfrac{\beta_1}{x_0}\right]^{\alpha_1}, & \text{if } f_1(\beta^*) < f_2(\beta^*) \end{cases}, \qquad (7)$$

where $f_i(x)=f(x|\alpha_i,\beta_i)$ defined in Eq (1), $F_i(x)=F(x|\alpha_i,\beta_i)$ defined in Eq (2), for $i=1,2$, and x_0 is the intersection point defined in Eq(8).

On the other hand, if $f(x|\alpha_1,\beta_1)$ and $f(x|\alpha_2,\beta_2)$ do not intersect in (β^*,∞), then one of the densities is greater than the other at all values in (β^*,∞), then Δ in Eq (7) can be simplified to:

$$\Delta = \begin{cases} \int_{\beta^*}^{\infty} f(x|\alpha_2,\beta_2)\,dx, & \text{if } f_1(\beta^*) > f_2(\beta^*) \\ \int_{\beta^*}^{\infty} f(x|\alpha_1,\beta_1)\,dx, & \text{if } f_1(\beta^*) < f_2(\beta^*) \end{cases}$$

$$= \begin{cases} \left[\dfrac{\beta_2}{\beta^*}\right]^{\alpha_2}, & \text{if } f_1(\beta^*) > f_2(\beta^*) \\ \left[\dfrac{\beta_1}{\beta^*}\right]^{\alpha_1}, & \text{if } f_1(\beta^*) < f_2(\beta^*) \end{cases}. \qquad (8)$$

Clearly, Eq(10) is the limit of Eq(9) when x_0 approaches infinity.

However, the proposed test statistic for testing the null hypothesis in Eq (3) is defined as the MLE of the overlapping coefficient Δ in Eq (7) which can be simplified either to Eq (9) or Eq (10) depending on the intersection point of $f(x|\alpha_1,\beta_1)$ and $f(x|\alpha_2,\beta_2)$.

Let $x_{11},x_{12},\ldots,x_{1n_1} \overset{r.s}{\sim}$ Pareto (α_1,β_1) and $x_{21},x_{22},\ldots,x_{2n_2} \overset{r.s}{\sim}$ Pareto (α_2,β_2) be two random samples, not necessarily independent, from Pareto distributions.

The MLEs of the scale and shape parameters based on the random samples above are:

$\hat\beta_1 = x_{1(1)} = \min\{x_{11},x_{12},\ldots,x_{1n_1}\}$, $\hat\beta_2 = x_{2(1)} = \min\{x_{21},x_{22},\ldots,x_{2n_2}\}$, $\hat\alpha_1 = \dfrac{n_1}{\sum_{i=1}^{n_1}(\ln x_i - \ln x_{1(1)})}$ and $\hat\alpha_2 = \dfrac{n_2}{\sum_{i=1}^{n_2}(\ln x_i - \ln x_{2(1)})}$.

By the invariant property of the MLE, the proposed test statistic, denoted by $\hat\Delta$, can be obtained by replacing the parameters α_1,α_2, β_1, β_2, β^*, f_1, f_2 and x_0 in Eq (9) or Eq (10) by their MLEs.

The null hypothesis H_o in Eq (4) is rejected at significance level γ if $\hat\Delta < \hat\Delta_{\gamma,n_1,n_2}$, where $P(\hat\Delta<\hat\Delta_{\gamma,n_1,n_2}|H_0)=\gamma$; i.e. $\hat\Delta_{\gamma,n_1,n_2}$ is the γ^{th} percentile of the distribution of $\hat\Delta$ at n_1 and n_2 under the assumption that H_o is true. Since, deriving the null distribution of $\hat\Delta$ in analytic form is complicated. Therefore, percentiles of $\hat\Delta$ ($\hat\Delta_{\gamma,n_1,n_2}$) are simulated in Section 3 at different sample sizes and

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

significance levels. Moreover, the null distribution of the proposed test statistic is approximated by Beta distribution in Section 4.

III. SIMULATED CRITICAL POINTS OF $\hat{\Delta}$

Since the null distribution of the proposed test statistic could not be obtained in closed form. Therefore, simulated percentiles for this distribution are generated at different significant levels, and for different sample sizes assuming, for simplicity, $n_1 = n_2$.

Table 1 shows the simulated 1st, 2.5th, 5th, and 10th percentiles of the null distribution of the proposed test statistic, these percentiles give the critical points of the proposed test statistic.

These percentiles have been computed based on 10,000 Monte Carlo Simulations generated at various significance levels under the assumption of equal sample sizes $n_1 = n_2 = n$ by using the following algorithm:

Algorithm I:

i) Arbitrary, choose $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$, based on the assumption that H_o in Eq (3) is true.

ii) Generate $x_{11}, x_{12}, \ldots, x_{1n} \overset{r.s}{\sim} \text{Pareto}\,(\alpha_1, \beta_1)$ and $x_{21}, x_{22}, \ldots, x_{2n} \overset{r.s}{\sim} \text{Pareto}\,(\alpha_2, \beta_2)$.

iii) Find the MLE $\hat{\alpha}_1, \hat{\alpha}_2$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}^* = Max(\hat{\beta}_1, \hat{\beta}_2)$ from the samples obtained in step 2. Hence, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are not necessarily equal, and $\hat{\beta}_1$ and $\hat{\beta}_2$ as well.

iv) Compute the MLE of x_0 defined in Eq (8) by replacing the parameters by their MLEs those have been computed in step (3).

v) If x_0 belongs to (β^*, ∞) then the proposed test statistic $\hat{\Delta}$ is computed using Eq (9). Otherwise, the two densities do not intersect in (β^*, ∞), and then the proposed test statistic $\hat{\Delta}$ is computed using Eq (10).

vi) Repeat steps (2-5) $m = 10{,}000$ iterations, to get $\{\hat{\Delta}_1, \hat{\Delta}_2, \ldots, \hat{\Delta}_m\}$.

vii) 1st, 2.5th, 5th, and 10th percentiles are computed for the set obtained in step (6) which are the simulated critical points of the proposed test statistic at significance level of 0.01, 0.025, 0.05 and 0.1, respectively.

IV. APPROXIMATED NULL DISTRIBUTION of $\hat{\Delta}$

In his section, we propose an approximation for the null distribution of the proposed test. Under the assumption that the null hypothesis is true, it was observed from the simulation studies in the previous section that the distribution of the test statistic $\hat{\Delta}$ is skewed to the left with values between zero and 1, far away from zero and close to 1. Eventually, Table 1 shows that $(\alpha \times 100)^{th}$ percentiles are much closer to 1 than zero for fixed n. Moreover, the percentiles become much closer to 1 when n increases. Consequently, one can approximate the null distribution of $\hat{\Delta}$ by beta distribution with parameters p and q assuming $p > q > 0$, to insure skewed to left distribution, with pdf: $f(\hat{\Delta}) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \hat{\Delta}^{p-1}(1-\hat{\Delta})^{q-1}$ provided that $p > q > 0$ and $0 < \hat{\Delta} < 1$.

TABLE I

SIMULATED $(\gamma \times 100)^{th}$ PERCENTILE $\hat{\Delta}_{\gamma,n,n}$ OF THE NULL DISTRIBUTION of $\hat{\Delta}$

n	$\gamma = 0.01$	$\gamma = 0.05$	$\gamma = 0.1$
10	0.4614	0.5822	0.6383
11	0.5008	0.6108	0.6590
12	0.5386	0.6329	0.6808
13	0.5625	0.6492	0.6960
14	0.5829	0.6693	0.7170
15	0.5922	0.6850	0.7239
16	0.6161	0.6988	0.7341
17	0.6196	0.7038	0.7483
18	0.6423	0.7131	0.7536
19	0.6551	0.7224	0.7619
20	0.6594	0.7330	0.7734
21	0.6673	0.7443	0.7789
22	0.6749	0.7497	0.7854
23	0.6851	0.7580	0.7895
24	0.6972	0.7658	0.7971
25	0.7056	0.7692	0.7992
26	0.7116	0.7713	0.8034
27	0.7195	0.7775	0.8097
28	0.7185	0.7820	0.8148
29	0.7299	0.7862	0.8169
30	0.7357	0.7890	0.8222
31	0.7416	0.7945	0.8252
32	0.7416	0.7981	0.8278
33	0.7484	0.8012	0.8313
34	0.7510	0.8056	0.8355
35	0.7505	0.8104	0.8375
36	0.7588	0.8119	0.8393
37	0.7630	0.8134	0.8417
38	0.7653	0.8168	0.8466
39	0.7673	0.8205	0.8479
40	0.7726	0.8255	0.8489
50	0.8012	0.8456	0.8679
100	0.8600	0.8923	0.9079
200	0.9036	0.9262	0.9363

Clearly, the parameters p and q are functions in n_1 and n_2. A simulation study is performed by using Algorithm II to test whether beta distribution adequately fits the null distribution of the test statistic $\hat{\Delta}$. For simplicity, we assume $n_1 = n_2 = n$.

Algorithm II:

i) For a fixed n, Steps (i-v) in Algorithm I are performed with $m = 15000$ iterations to get $\{\hat{\Delta}_1, \hat{\Delta}_2, \ldots, \hat{\Delta}_m\}$, a set of 15000 $\hat{\Delta}$ values.

ii) The parameters p and q of beta distribution are estimated by the method of moments estimation from the set $\{\hat{\Delta}_1, \hat{\Delta}_2, \ldots, \hat{\Delta}_m\}$ obtained in Step (i).

iii) The frequency histogram for the simulated set $\{\hat{\Delta}_1, \hat{\Delta}_2, \ldots, \hat{\Delta}_m\}$ is plotted along with the fitted beta pdf using the estimated p and q in Step (ii).

Table 2 shows the frequency histogram along with the fitted beta pdf obtained from Algorithm II for various values of n. It is evident from Table 2 that beta distribution adequately fits the null distribution of $\hat{\Delta}$. Obviously, there is a positive linear relationship between n and p, and a

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

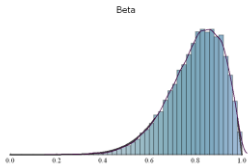
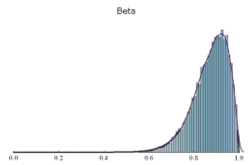
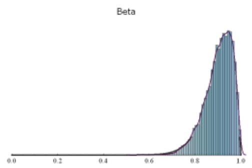
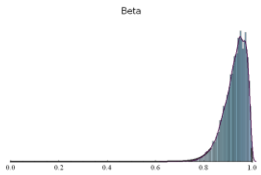
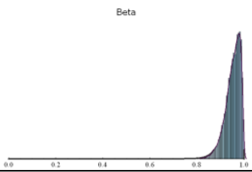
negative linear relationship between n and q. Hence, one can approximate the parameters p and q by fitting suitable functions for p and q in terms of n. For illustrative purposes, we study the relationship between n, p and q by analyzing a sample of the form $\{(n, p, q); n = 10(5)100\}$ that is generated by using Algorithm II for $n = 10(5)100$. For this sample, the estimated linear correlation coefficients between n and p and between n and q are 0.98 and -0.92, respectively. This shows that linear models could adequately fit p and q in terms of n as follows:

$$p(n) = 9.81629 + 0.321148n \text{ and } q(n) = 2.25662 - 0.0030067n$$

Therefore, the null distribution of the test statistic $\hat{\Delta}$ can be approximated by a beta distribution with parameters $p(n) = 9.81629 + 0.321148n$ and $q(n) = 2.25662 - 0.0030067n$, where n is the sample size, i.e. $Beta(9.81629 + 0.321148n, 2.25662 - 0.0030067n)$. For instance, at $n_1 = n_2 = n = 30$, the parameters p and q are estimated by 19.4507 and 2.1664, respectively. Hence, the null distribution of $\hat{\Delta}$ is approximately $Beta(19.4507, 2.1664)$.

TABLE II

FITTED BETA TO THE NULL DISTRIBUTION OF $\hat{\Delta}$

n	$Beta(p,q)$	Fitted Beta Curve
10	$Beta(8.42,\ 2.13)$	
20	$Beta(15.53,\ 2.24)$	
30	$Beta(20.48,\ 2.21)$	
50	$Beta(27.74,\ 2.11)$	
100	$Beta(39.61,\ 1.96)$	

Note that the 5th and 10th percentiles of $Beta(19.4507, 2.1664)$ are 0.7785 and 0.8133 which are very close to the simulated 5th and 10th percentiles $\hat{\Delta}_{0.05,30,30} = 0.7890$ and $\hat{\Delta}_{0.10,30,30} = 0.8222$, respectively.

It is worth mentioning that nonlinear models might also adequately fit p and q in terms of n. However, more studies are needed in this direction to study other models when $n_1 \neq n_2$.

V. SIMULATED POWERS

Statistical power of a test is defined as the probability of rejecting the null hypothesis H_0 when the null hypothesis is false. Unfortunately, power function of the proposed test cannot be obtained in explicit form, then simulated statistical power is computed under different alternatives, keeping Type-I error of each test at the same level. However, Table 3 shows the simulated power of the proposed test based on 5000 Monte Carlo simulations for two sample sizes $n = 10$ and 30 and various alternatives assuming the significance level $\gamma = 0.05$, and the Pareto parameters $\alpha_1 = 1$ and $\beta_1 = 2$.

TABLE III

SIMULATED POWERS OF $\hat{\Delta}$, ASSUMING $\alpha_1 = 1$, $\beta_1 = 2$ AND 0.05 LEVEL OF SIGNIFICANCE

	α_2				
	0.5	1.00	1.25	1.5	2.0
β_2					
			$n_1 = n_2 = 10$		
0.5	0.9601	0.7846	0.9275	0.9874	1.0000
1.0	0.9798	0.2618	0.6642	0.9583	0.9998
1.5	0.9980	0.0827	0.5443	0.9573	0.9998
2.0	0.9998	0.0507	0.5358	0.9711	0.9999
2.5	1.0000	0.0722	0.5863	0.9814	1.0000
3.0	1.0000	0.1106	0.6684	0.9889	0.9999
			$n_1 = n_2 = 20$		
0.5	1.0000	0.9866	0.9999	1.0000	1.0000
1.0	1.0000	0.5271	0.9759	0.9999	1.0000
1.5	1.0000	0.1285	0.9618	1.0000	1.0000
2.0	1.0000	0.0482	0.9726	1.0000	1.0000
2.5	1.0000	0.0929	0.9868	1.0000	1.0000
3.0	1.0000	0.2160	0.9942	1.0000	1.0000
			$n_1 = n_2 = 30$		
0.5	1.0000	0.9995	1.0000	1.0000	1.0000
1.0	1.0000	0.7161	0.9999	1.0000	1.0000
1.5	1.0000	0.1755	0.9988	1.0000	1.0000
2.0	1.0000	0.0460	0.9993	1.0000	1.0000
2.5	1.0000	0.1242	1.0000	1.0000	1.0000
3.0	1.0000	0.3024	1.0000	1.0000	1.0000

It is obvious from Table 3 that the statistical powers of the proposed are satisfactory. Clearly, the statistical powers of this test increases as the sample size increase. Furthermore, it appears from Table 3 that this test efficiently control the significance level, i.e. statistical powers achieve the correct significance level in simulations under the assumption that the null hypothesis is true.

However, since the proposed test is derived under the assumption that the distributions are Pareto, it is significant to study the Type-I error robustness of the proposed test for testing the similarity of non-Pareto distributions. Table 4 summarizes the significance (nominal) levels of the proposed test based on 10,000 simulated pairs of samples when the null hypothesis H0 is true for the selected population. Pareto, Normal, Gamma, Chi-square, and Exponential distributions are considered. Various assumptions of the sample sizes and the significance levels are considered. It is worth mentioning that Pareto distribution is considered again for the sake of testing the nominal levels of the test at various Type-I errors and sample sizes. The robustness of a test could be assessed by the closeness of the simulated nominal level to the correct Type-I error. It is evident from Table 4 that the proposed test is highly sensitive for violation of the Pareto assumption; which means that the test is not robust against departures from the Pareto assumption.

TABLE IV

NOMINAL LEVELS FOR $\hat{\Delta}$ WHEN H0 IS TRUE FOR THE GIVEN POPULATION

	Significance Level		
Population	0.01	0.05	0.10
	$n_1 = n_2 = 10$		
$Pareto(9,2)$	0.0110	0.0533	0.1019
$N(5,1)$	0.2237	0.4264	0.5187
$Gamma(2,3)$	0.2249	0.4293	0.5390
χ^2_{10}	0.1884	0.3952	0.4896
$Exp(9)$	0.2759	0.4641	0.5627
	$n_1 = n_2 = 20$		
$Pareto(9,2)$	0.0085	0.0485	0.1059
$N(5,1)$	0.4342	0.5741	0.6478
$Gamma(2,3)$	0.4335	0.5841	0.6507
χ^2_{10}	0.3848	0.5412	0.6169
$Exp(9)$	0.4754	0.6038	0.6842
	$n_1 = n_2 = 30$		
$Pareto(9,2)$	0.0111	0.0524	0.1007
$N(5,1)$	0.5280	0.6353	0.7090
$Gamma(2,3)$	0.5281	0.6430	0.6937
χ^2_{10}	0.4689	0.5897	0.6677
$Exp(9)$	0.5546	0.6655	0.7313

VI. REAL APPLICATION

In this section, a real data obtained from Watthanacheewakul and Suwattee (2010) is analyzed, this data summarizes the major rice crop (in Kilograms) in the crop year 2001/2002 (April 1st, 2001 to March 31st, 2002) from two Tambols, Nongyang and Nongjom, of Amphoe Sansai in the Chiang Mai province, Thailand. Samples (shown in Table 5) of sizes 28 and 30 were drawn from Nongyang and Nongjom Tambols, respectively.

First, it was checked whether Pareto distribution can be used or not to analyze these data sets.
For Nongyang's data, the MLEs of α and β are 2.75765 and 3000, respectively. The Kolmogorov-Smirnov (KS) distance between the empirical distribution function and the fitted distribution function has been used to check the goodness of fit. The KS statistic value is 0.107, and the KS critical value is 0.2250 at at $n = 28$ and $\alpha = 0.05$. Accordingly, one cannot reject the hypothesis that the data are coming from T-L distribution. For Nongjom's data, the MLEs of α and β are 3.01544 and 3600, respectively. The KS statistic value is 0.133, and the KS critical value is 0.2176 at at $n = 30$ and $\alpha = 0.05$. Accordingly, one cannot reject the hypothesis that the data are coming from T-L distribution.

TABLE V

THE MAJOR RICE CROP IN KILOGRAMS FROM THE TWO TAMBOLS FOR THE CROP YEAR 2001/2002 (APRIL 1ST , 2001 TO MARCH 31ST, 2002)

Nongyang	Nongjom
3440	3600
3200	5000
5400	7500
3800	7800
4300	3600
7000	4000
3700	4000
6000	4800
3250	4900
3500	4100
3000	4500
3400	4200
5000	6800
3600	4000
18000	7000
3150	4500
8500	8300
4500	3800
4250	24000
3500	5800
3000	3720
3600	6000
5000	5400
3000	3600
4000	4500
3600	4220
4270	3600
5800	4800
	4200
	4113

Now, we need to test the null hypothesis that the rice crop of Nongyang and Nongjom are the same by using the proposed test. The proposed test statistic is equal to $\Delta = 0.6048$ which is less than the critical value $\hat{\Delta}_{0.05,28,30} = 0.7857$, this means that the hypothesis that the two crops are the same is rejected at significance level of 0.05, which agrees with what was concluded by Watthanacheewakul and Suwattee (2010).

VII. CONCLUSIONS

In this paper, a non classical test has been proposed to test the similarity of two Pareto populations. The proposed test statistic is defined as the MLE of Weitzman's OVL coefficient, which estimates the agreement of two densities. This test statistic has been derived in closed form. Since deriving the exact null distribution of the proposed test statistic is not simple, simulated percentiles have been provided at various equal sample sizes and significance levels. However, the null distribution of the proposed test statistic was approximated by beta distribution, but still further studies are needed in this direction. Simulated powers of the proposed test have been computed. It has been concluded that the proposed test almost has satisfactory power and it efficiently controls the significance level. Furthermore, we studied Type-I error robustness of the proposed test when the underlying distributions are non-

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Pareto. It has been noticed that this test the proposed test is highly sensitive for violation of the Pareto assumption; which means that the test is not robust against departures from the Pareto assumption.

REFERENCES

[1] H. T. Davis, and M. L. Feldstein, "The Generalized Pareto Law as a Model for Progressively Censored Survival Data," Biometrika, vol. 66, pp.299-306, 1979.

[2] C. M. Harris, "The Pareto Distribution as a Queue Service Discipline," Operations Research, vol. ii, pp. 307-313, January-February 1968.

[3] E. C. Freiling, “A Comparison of the Fallout Mass-Size Distributions Calculated by Lognormal and Power-Law Models,” San Francisco: U.S. Naval Radiological Defense Laboratory, 1966 (AD-646 019).

[4] J. M. Berger and B. Mandelbrot, "A New Model for Error Clustering in Telephone Circuits," IBM Journal of Research and Development, vol. 7, pp. 224-236, July 1963.

[5] N. L. Johnson and S. Kotz, “Continuous Uniwriute Distri- hutions-I,” Houghton Mifflin Co. 1970.

[6] M.S. Weitzman, “Measures of overlap of income distributions of white and Negro families in the United States,” Technical paper No. 22, Department of Commerce, Bureau of Census, Washington. U.S. (1970).

[7] A. Helu and H. Samawi, “On Inference of Overlapping Coefficients in Two Lomax Populations Using Different Sampling Methods,” Journal of Statistical Theory and Practice, vol. 5(4), pp. 683-696, 2011.

[8] K. Matusita, "Decision rules based on the distance for problem of fir, two samples, and Estimation," Ann. Math. Statist. vol. 26, pp. 631–640, 1955.

[9] M. Morisita, "Measuring interspecific association and similarity between communities. Memoirs of the faculty of Kyushu University," Series E. Biology, vol. 3, pp. 36–80, 1959.

[10] H.F. Inman and E.L. Bradley, “The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities,” Commun Stat – Theory Methods, vol. 18(10), pp. 3851–3874, 2008.

[11] M. Mulekar and S. N. Mishra, “Overlap coefficients of two normal densities: equal means case,” J. Japan. Statist. Soc., vol. 24, pp. 169-180, 1994.

[12] B. Reiser and D. Faraggi, “Confidence Intervals for the overlapping coefficient: the normal equal variance case,” The Statistician, vol. 48 (3), pp. 413-418, 1999.

[13] Y. P. Chaubey, D. Sen, and S. N. Mishra, “Inference on Overlap for Two Inverse Gaussian Populations: Equal Means Case,” Communications in Statistics-Theory and Methods, vol. 37, 12, 1880-1894, 2008.

[14] M. F. AL-Saleh and H. M. Samawi, "Inference on overlapping coefficients in two exponential populations," J. of Modern Statist. Sci., vol. 6 (2), pp. 503–516, 2007.

[15] H. A. Bayoud and O. A. Kittaneh, "Testing the equality of Two exponential Distributions," Communications in Statistics-Simulation and Computation, vol. 45(7), pp. 2249-2256, 2016.

[16] H. A. Bayoud, "Testing the Similarity of Two Normal Populations with Application to the Bioequivalence Problem," Journal of Applied Statistics, vol. 43(7), pp. 1322-1334, 2016.

[17] L. Watthanacheewakul and S. Suwattee, "Comparisons of Several Pareto Population Means," Proceeding of International MultiConference of Engineers and Computer Scientists, vol III, 2010.