International Journal of
Mathematical Modelling & Computations
Vol. 09, No. 03, Summer 2019, 201- 212

A New Five-Parameter Distribution: Properties and Applications

A. Abdollahi Nanvapisheh*

Department of Mathematics, Tehran North Branch, Islamic Azad University, Tehran, Iran.

Abstract. In this paper a new five-parameter lifetime and reliability distribution named "the exponentiated Uniform-Pareto distribution (EU-PD)," has been suggested that it has a bathtub-shaped and inverse bathtub-shape for modeling lifetime data. This distribution has applications in economics, actuarial modelling, reliability modeling, lifetime and biological sciences. Firstly, the mathematical and statistical characteristics of the proposed distribution are presented, then the applications of the new distribution are studied using the real data set. Its first moment about origin and moments about mean have been obtained and expressions for skewness, kurtosis have been given. Various mathematical and statistical properties of the proposed distribution have been discussed. Estimation of its parameter has been discussed using the method of maximum likelihood. A simulation study is given. Finally, two applications of the new distribution have been discussed with two real income and lifetime data setsThe results also confirmed the suitability of the presented models for real data collection.

Received: 05 March 2019, Revised: 26 June 2019, Accepted: 27 July 2019.

Keywords: Uniform-Pareto distribution; Exponentiated Uniform-Pareto distribution; Moments; Lifetime data; Parameter estimation; Goodness of fit.

Index to information contained in this paper

1. Introduction

Since that statistics is one of the most important applied sciences related to other fields of science, including economics, in recent years, there has been a great interest in research on parametric models of income distribution (see [3-6]). Proportional models for income distribution are presented to assess the living standards of the entire population of a country and also to compare the standard of living standards of the social classes or different regions of a country. Therefore, in order to create a probability model, a theoretical distribution function with an empirical frequency distribution characteristic is necessary to select the appropriate method for estimating the model parameters. Therefore, statistical analysis of the distribution of population income reflects the context for decision making about budget and social policies. In this paper, a new distribution is presented which is highly flexible and is a suitable distribution for modeling income data.

Identifying an appropriate distribution for data modeling is very important in statistical analysis. Presentation of a proper distribution for a set of specific data will help us to arguing about the data. For this reason, the use of methods to modify the available statistical distribution has been used to create more flexibility in the data set in various fields.

*Corresponding author. Email: anita.abdollahi@yahoo.com

http://ijm2c.iauctb.ac.ir

Recently, uniform Pareto distribution (U-PD) (see [1,2,12]), similar to the other statistical distributions, has attracted so attention because of its application in modeling the distribution of personal income and reliability, and other areas.

To check the details see resources in (see [7,8,11]). With the aim of increasing the flexibility of Pareto distribution in modeling income data, there are various data distribution (see [14-16]). In this paper, a new extension of the Pareto distribution, called the exponentiated Pareto Uniform Distribution, is presented with the main motive for modeling income data with both uniform and non-uniform risk rates. The article is organized as follows:

This study is organized in eight sections starting with this introduction. In sections 2 and 3, we present the new model and derive some of its properties. Mean deviations and Bonferroni and Lorenz curves and Maximum Likelihood Estimation is given in Section 4. In Section 5, a simulation study is given and we provide two real data applications and the paper ends with some concluding remarks.

2. The new model

A random variable X has the Uniform-Pareto distribution (U-PD) with three parameters if its cumulative distribution function (cdf) is given by:

$$G(x)=1-\frac{b(m/x)^{\theta}}{b-a},\quad m\left(\frac{b-a}{b}\right)^{\frac{-1}{\theta}}<x<\infty, \tag{1}$$

with shape parameter $\theta>0$ and scale parameter $-\infty<a\le b<\infty$. The corresponding probability density function (pdf) of the Uniform-Pareto distribution (U-PD) is as follows:

$$g(x)=\frac{b}{b-a}\frac{\theta m^{\theta}}{x^{\theta+1}},\quad m\left(\frac{b-a}{b}\right)^{\frac{-1}{\theta}}<x<\infty. \tag{2}$$

Now suppose that the random variable X has the cdf (1), then its exponentiated distribution function, which is the cdf of (EU-PD) distribution, is:

$$G(x)=\left[1-\frac{b(m/x)^{\theta}}{b-a}\right]^{\gamma},\quad m\left(\frac{b-a}{b}\right)^{\frac{-1}{\theta}}<x<\infty. \tag{3}$$

The pdf of (EU-PD) distribution is obtained by differentiating (3) and therefore we have:

$$g(x)=\left[\frac{b\theta\gamma}{b-a}\frac{m^{\theta}}{x^{\theta+1}}\right]\left[1-\frac{b(m/x)^{\theta}}{b-a}\right]^{\gamma-1},\quad m\left(\frac{b-a}{b}\right)^{\frac{-1}{\theta}}<x<\infty. \tag{4}$$

The hrf of the new distribution is given by

$$h(x)=\frac{\left[\frac{b\theta\gamma}{b-a}\frac{m^{\theta}}{x^{\theta+1}}\right]\left[1-\frac{b(m/x)^{\theta}}{b-a}\right]^{\gamma-1}}{1-\left[1-\frac{b(m/x)^{\theta}}{b-a}\right]^{\gamma}}. \tag{5}$$

The pdfs of the new distribution are plotted in Figure 1 for some selected values of parameters. It can be seen that the pdf of this distribution is inverse bathtub and increasing depending on the parameter values. Figure 2 contains the plots of the hrfs of the EU-PD distribution for different values of parameters. From Figure 2, we can observe that the hrf is bathtub-shaped and inverse bathtub-shape, depending on the parameter values.

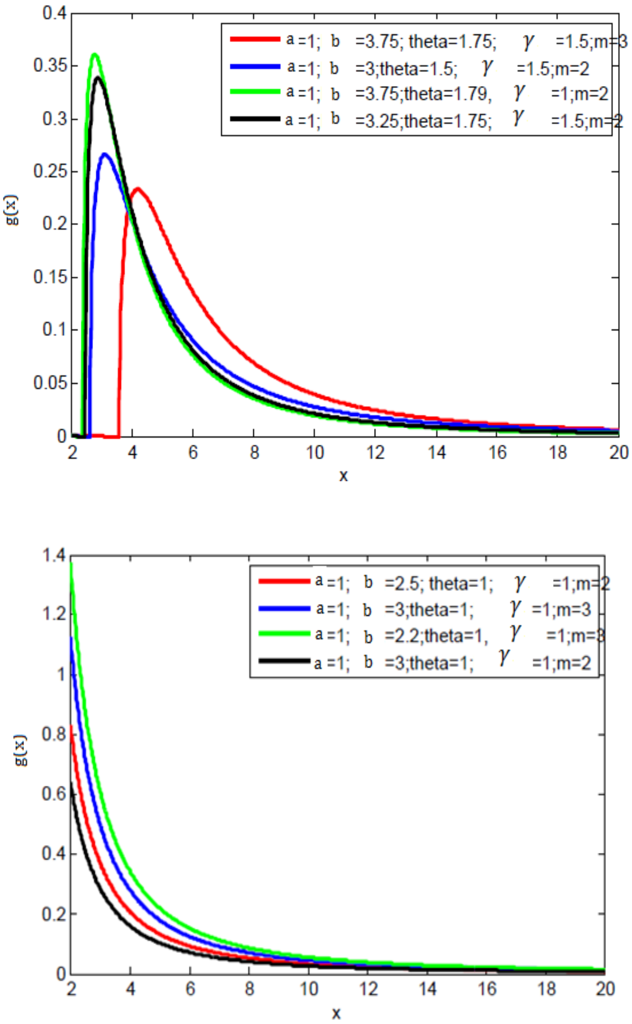

Figure 1: Pdfs of the EU-PD distribution for some selected values of α and β.

3. The statistical properties of (U-PD) and (EU-PD)

In this section, we present some of the statistical properties of (U-PD) and (EU-PD). Moments are necessary and important in any statistical analysis, especially in applications. It can be used to study the most important features and characteristics of a distribution. In this section, we present complete and incomplete moments of the EU-PD. But first, we present an expansion for $g(x)$ in order to obtain expressions for the moments. Using the binomial expansion, we have:

$$g(x)=\left[\frac{b\theta\gamma}{b-a}\frac{m^{\theta}}{x^{\theta+1}}\right]\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^{j}\left(b\left(m/x\right)^{\theta}\right)^{j}}{(b-a)^{j}}. \qquad (6)$$

Similarly, the expansion of its cumulative distribution function is as follows:

$$G(x)=\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{(-1)^{j}\left(b\left(m/x\right)^{\theta}\right)^{j}}{(b-a)^{j}}. \qquad (7)$$

Therefore, the *r*th moment of EU-PD distribution is given by:

$$E\left[x^r\right]=\int_w^\infty x^r\left[\frac{b\theta\gamma}{b-a}\frac{m^\theta}{x^{\theta+1}}\right]\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b\left(m/x\right)^\theta\right)^j}{(b-a)^j}dx$$

$$=\int_w^\infty x^{-\theta j+r-\theta-1}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^\theta\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^j\right]dx$$

$$=\frac{x^{-\theta j+r-\theta}}{-\theta j+r-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^\theta\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^j\right]_w^\infty,$$

where $W=m\left((b-a)/b\right)^{-1/\theta}$, therefore:

$$E\left[x^r\right]=-\frac{\left(m\left(\frac{b-a}{b}\right)^{-\frac{1}{\theta}}\right)^{-\theta j+r-\theta}}{-\theta j+r-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^\theta\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^\theta\right]. \tag{8}$$

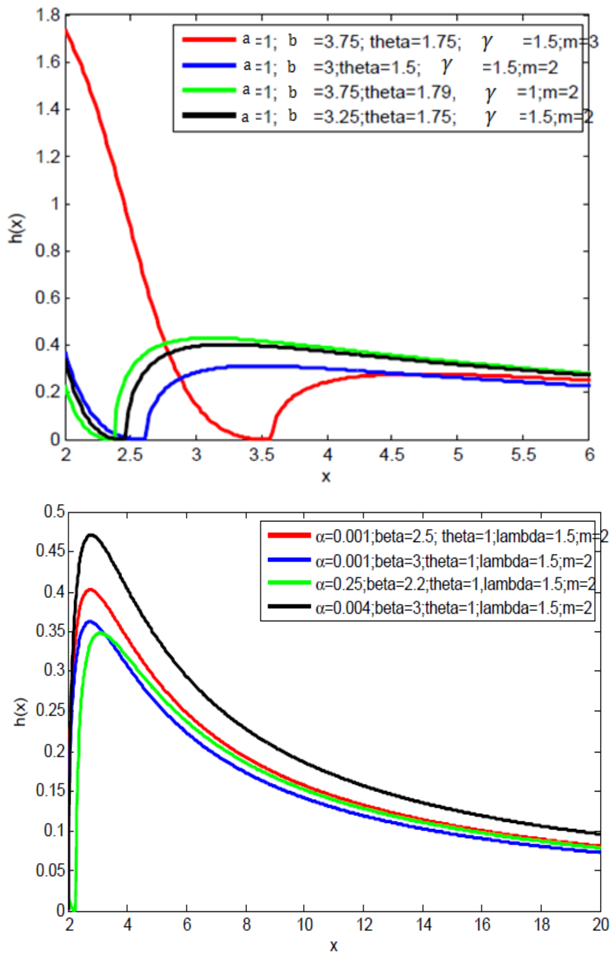

Figure 2: hrfs of the EU-PD distribution for some selected values of α and β.

Therefore, the first moment or average of EU-PD distribution is as follows:

$$E(X)=-\frac{\left(m\left(\frac{b-a}{b}\right)^{-\frac{1}{\theta}}\right)^{-\theta j+1-\theta}}{-\theta j+1-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^{\theta}\right)^j}{(b-a)^i}\left[\frac{b\theta\gamma}{b-a}m^{\theta}\right].$$

In the same way, other moments can be obtained by giving values to r in equation (8). Based on the first four moments, the Skewness and Kurtosis of this distribution are calculated from the following equations, respectively:

$$S(x)=(E(X-E(X))^3)/((E(X-E(X))^2)^{(3/2)})=(\mu_3-3\mu_1\mu_2+2\mu_1^3)/[\mu_2-\mu_1^2]^{(3/2)},$$

$$K(x)=(E(X-E(X))^4)/((E(X-E(X))^2)^2)=(\mu_4-4\mu_1\mu_3+6\mu_1^2\mu_2-3\mu_1^4)/[\mu_2-\mu_1^2]^2,$$

where μ_r for $r=1,\ldots,4$ is the rth moment of NM distribution that can be obtained by giving values to r in equation (8).

In the following, we find the terms for the incomplete moments of new model that are used to find expressions for the Lorenz curve, Bonferroni curve equations and also mean deviation from the mean and mean deviation from the middle.

Suppose X has the probability density function given in relation (3). In this case, the rth incomplete moment is:

$$\begin{aligned}\int_0^w x^r g(x)dx &= \int_0^w x^r\left[\frac{b\theta\gamma}{b-a}\frac{m^{\theta}}{x^{\theta+1}}\right]\sum_{j}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m/x)^{\theta}\right)^j}{(b-a)^j}dx\\ &=\int_0^w x^{-\theta j+r-\theta-1}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^{\theta}\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^{\theta}\right]dx\\ &=\frac{x^{-\theta j+r-\theta}}{-\theta j+r-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^{\theta}\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^{\theta}\right]_0^w,\end{aligned}$$

where $W=m\left((b-a)/b\right)^{-1/\theta}$. So we have

$$\int_0^{\infty} x^r g(x)dx=\frac{\left(m\left(\frac{b-a}{b}\right)^{-1/\theta}\right)^{-\theta j+r-\theta}}{-\theta j+r-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^{\theta}\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^{\theta}\right]. \tag{9}$$

4. Mean deviations and Bonferroni and Lorenz curves and Maximum Likelihood Estimation

Suppose X has the probability density function given in relation (4). In this case, the mean deviation from the mean is obtained from the following relation by using the relations (7) and (9):

$$\begin{aligned}\delta_1(x)&=\int_0^{\infty}|x-\mu|f(x)dx=2\mu F(\mu)-2I(\mu)\\ &=2\mu\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{(-1)^j\left(b\left(m/\mu\right)^{\theta}\right)^j}{(b-a)^j}\\ &\quad-2\left(\left(m\left(\frac{b-a}{b}\right)^{-\frac{1}{\theta}}\right)^{-\theta j+1-\theta}\Big/(-\theta j+1-\theta)\right)\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^{\theta}\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^{\theta}\right],\end{aligned} \tag{10}$$

where $I(b)=\int_0^b xf(x)dx.$ Also, if M represents the median of the new distribution, then, the mean deviation from the median is similarly as follows:

$$\delta_2(x)=\int_0^\infty |x-M|f(x)dx=2MF(M)-M+\mu-2I(M)$$
$$=2M\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{(-1)^j\left(b\left(m/M\right)^\theta\right)^j}{(b-a)^j}-M+\mu \tag{11}$$
$$-2\left(\frac{\left(m\left(\frac{b-a}{b}\right)^{-\frac{1}{\theta}}\right)^{-\theta j+1-\theta}}{-\theta j+1-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^\theta\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^\theta\right]\right).$$

In the following, we will present the Bonferroni and Lorenzo curves that have many applications in economy to examine income and poverty, reliability, medicine and insurance.

Suppose X has the probability function in relation (4), in this case. According to (9), the Bonferroni curve equation is as follows:

$$B_F\left[F(x)\right]=\frac{1}{\mu F(x)}\int_0^x uf(u)du$$
$$=\frac{1}{\mu\left[1-\frac{b\left(\frac{m}{x}\right)^\theta}{b-a}\right]^\gamma}\left(\frac{\left(m\left(\frac{b-a}{b}\right)^{-\frac{1}{\theta}}\right)^{-\theta j+1-\theta}}{-\theta j+1-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^\theta\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^\theta\right]\right). \tag{12}$$

The Lorenz curve equation will be as follows:

$$L_F[F(x)]=B_F[F(x)]F(x)=\frac{1}{\mu}\int_0^x uf(u)du$$
$$=\frac{1}{\mu\left[1-\frac{b\left(\frac{m}{x}\right)^\theta}{b-a}\right]^\gamma}\left(\frac{\left(m\left(\frac{b-a}{b}\right)^{-\frac{1}{\theta}}\right)^{-\theta j+1-\theta}}{-\theta j+1-\theta}\sum_{j=0}^{\infty}\binom{\gamma-1}{j}\frac{(-1)^j\left(b(m)^\theta\right)^j}{(b-a)^j}\left[\frac{b\theta\gamma}{b-a}m^\theta\right]\right). \tag{13}$$

Now, let $x_1,\dots,x_n$ be random samples of size n of the EU-PD distribution. In this case, the likelihood function is as follows:

$$L=\prod_{i=1}^{n}\left[\frac{b\theta\gamma}{b-a}\frac{m^\theta}{x_i^{b+1}}\right]\prod_{i=1}^{n}\left[1-\left(b\left(m/x_i\right)^\theta/(b-a)\right)\right]^{\gamma-1}.$$

As a result, the logarithm of the likelihood function is obtained as follows:

$$l(a,b,\theta,m,\gamma)=\sum_{i=1}^{n}\log\left[\frac{b\theta\gamma}{b-a}\frac{m^\theta}{x_i^{\theta+1}}\right]+(\gamma-1)\sum_{i=1}^{n}\log\left[1-\frac{b\left(m/x_i\right)^\theta}{b-a}\right]. \tag{14}$$

Maximum likelihood estimates parameters are obtained by maximizing the likelihood logarithm function over the parameters. To this end, we derive from the logarithm of the likelihood function with respect to the parameters, then the resulting derivatives are equal to zero. As a result, maximum likelihood estimates of the parameters obtained by solving the following nonlinear equation simultaneously.

$$\frac{\partial l_n}{\partial a}=\frac{n}{b-a}+(\gamma-1)\sum_{i=1}^{n}\frac{-\frac{b\left(m/x_i\right)^{\theta}}{b-a}}{b-a-b\left(m/x_i\right)^{\theta}},$$

$$\frac{\partial l_n}{\partial b}=\sum_{i=1}^{n}\frac{\frac{\theta\gamma(b-a)-b\theta\gamma}{b-a}}{b\theta\gamma}+(\gamma-1)\sum_{i=1}^{n}\frac{-\left(m/x_i\right)^{\theta}(b-a)+b\left(m/x_i\right)^{\theta}}{b-a-b\left(m/x_i\right)^{\theta}},$$

$$\frac{\partial l_n}{\partial \theta}=\sum_{i=1}^{n}\frac{\frac{m^{\theta}\ln m-\left(x_i^{\theta+1}\right)\left(\ln x_i\right)\left(m^{\theta}\right)}{x_i^{\theta+1}}}{\theta m^{\theta}}+(\gamma-1)\sum_{i=1}^{n}\frac{-b\left(m/x_i\right)^{\theta}\ln(m/x_i)}{b-a-b\left(\ln(m/x_i)\right)^{\theta}},$$

$$\frac{\partial l_n}{\partial m}=\sum_{i=1}^{n}\frac{m/\theta}{x_i^{\theta+1}}+(\gamma-1)\sum_{i=1}^{n}\frac{\left[\frac{-b\theta m^{\theta-1}}{x_i^{\theta}}\right]}{b-a-b\left(m/x_i\right)^{\theta}},$$

There is no closed-form expression for the maximum likelihood estimator and its computation has to be performed numerically using a nonlinear optimization algorithm that can be computed in some software such as R software.

In order to present the interval estimation and a hypothesis test for the parameters, we obtain the observed Fisher information matrix. The information matrix is:

$$J(\vartheta)=-\begin{bmatrix} I_{aa} & I_{ab} & I_{a\theta} & I_{am} & I_{a\gamma}\\ I_{ab} & I_{bb} & I_{b\theta} & I_{bm} & I_{b\gamma}\\ I_{a\theta} & I_{b\theta} & I_{\theta\theta} & I_{\theta m} & I_{\theta\gamma}\\ I_{am} & I_{bm} & I_{\theta m} & I_{mm} & I_{m\gamma}\\ I_{a\gamma} & I_{b\gamma} & I_{\theta\gamma} & I_{m\gamma} & I_{\gamma\gamma}\end{bmatrix},$$

where $\vartheta=(a,b,\theta,m,\gamma)^T$ is the vector of parameters and there are clear phrases for the observed information matrix elements.

$$I_{aa}=\frac{\partial^2 l}{\partial a^2},\ l_{ab}=\frac{\partial^2 l}{\partial a\partial b},\ I_{a\theta}=\frac{\partial^2 l}{\partial a\partial\theta},\ I_{am}=\frac{\partial^2 l}{\partial a\partial m},\ I_{a\gamma}=\frac{\partial^2 l}{\partial a\partial\gamma},$$

$$I_{ba}=\frac{\partial^2 l}{\partial b\partial a},\ I_{bb}=\frac{\partial^2 l}{\partial b^2},\ l_{b\theta}=\frac{\partial^2 l}{\partial b\partial\theta},\ l_{bm}=\frac{\partial^2 l}{\partial b\partial m},\ I_{b\gamma}=\frac{\partial^2 l}{\partial b\partial\gamma},$$

$$I_{\theta a}=\frac{\partial^2 l}{\partial\theta\partial a},\ I_{\theta b}=\frac{\partial^2 l}{\partial\theta\partial b},\ I_{\theta\theta}=\frac{\partial^2 l}{\partial\theta^2},\ I_{\theta m}=\frac{\partial^2 l}{\partial\theta\partial m},\ I_{\theta\gamma}=\frac{\partial^2 l}{\partial\theta\partial\gamma},$$

$$I_{ma}=\frac{\partial^2 l}{\partial m\partial a},\ I_{mb}=\frac{\partial^2 l}{\partial m\partial b},\ I_{m\theta}=\frac{\partial^2 l}{\partial m\partial\theta},\ I_{mm}=\frac{\partial^2 l}{\partial m^2},\ I_{m\gamma}=\frac{\partial^2 l}{\partial m\partial\gamma},$$

$$I_{\gamma a}=\frac{\partial^2 l}{\partial\gamma\partial a},\ I_{\gamma b}=\frac{\partial^2 l}{\partial\gamma\partial b},\ I_{\gamma\theta}=\frac{\partial^2 l}{\partial\gamma\partial\theta},\ I_{\gamma m}=\frac{\partial^2 l}{\partial\gamma\partial m},\ I_{\gamma\gamma}=\frac{\partial^2 l}{\partial\gamma^2}.$$

Under the regularity conditions (see for example Lehmann and Casella, 1998, pp. 461-463) [9], the asymptotic inference for the vector of parameters, i.e. $\vartheta=(a,b,\theta,m,\gamma)^T$, based on normal approximation can be used. When the sample size n is large enough, then $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically a five-variate normal random vector with mean $(0,0)^T$ and the variance-covariance matrix that equates to the inverse of the expected Fisher information matrix, i.e., $J(\theta)^{-1}$. This asymmetric behavior holds if we replace $J(\theta)^{-1}$ with $\left[\frac{1}{n}I_n(\theta)\right]^{-1}$. If unknown parameters appear in the variance-covariance matrix, then they can be replaced by their respective maximum likelihood estimates. Using the normal approximation, we can obtain approximate (asymptotic) confidence intervals for the parameters.

5. A simulation study and applications and conclusion

In this section, we evaluate the performance of the MLEs of the parameters of the EU-PD model by means of a simulation study. The inverse transform algorithm is used to generate random data from the EU-PD distribution. The precision of the MLEs is discussed by means of the bias, mean squared error (MSE) and mean relative error (MRE). We generated $N=30,000$ samples of sizes $n=100,\ 150$ from the EU-PD distribution with the parameter combinations $(a,b,\theta,m,\gamma)=(2,2,2,1,1)$ and $(a,b,\theta,m,\gamma)=(2,3,2,2,1)$. We obtained the MLEs of the parameters for each generated sample and the standard errors of the MLEs were obtained by inverting the observed information matrix. Let $\hat{\theta}$ be the MLE of θ and $\hat{\theta}_i$ be the MLE of θ that is obtained in the ith iteration, then the estimated bias, MSE, and MRE of $\hat{\theta}$ can be obtained using the following equations:

$$\text{bias}_h(n)=\frac{1}{30000}\sum_{i=1}^{30000}\left(\hat{h}_i-h\right),$$

$$\text{MSE}_h(n)=\frac{1}{30000}\sum_{i=1}^{30000}\left(\hat{h}_i-h\right)^2,$$

$$\text{MRE}_h(n)=\frac{1}{30000}\sum_{i=1}^{30000}\left(\hat{h}_i/h\right).$$

Table 1. Simulation results.

n	$(a,b,\theta,m,\gamma)=(2,2,2,1,1)$				n	$(a,b,\theta,m,\gamma)=(2,3,2,2,1)$			
	Parameters	Bias	MSE	MRE		Parameters	Bias	MSE	MRE
100	a	0.18	0.95	1.09	100	a	0.40	0.96	1.09
	b	0.11	0.74	1.08		b	0.28	0.62	1.05
	θ	0.09	0.63	1.05		θ	0.10	0.39	1.04
	m	0.07	0.58	1.01		m	0.08	0.24	1.03
	γ	0.05	0.52	1.03		γ	0.07	0.18	1.01
150	a	0.07	0.61	1.06	150	a	0.08	0.09	1.05
	b	0.05	0.42	1.04		b	0.07	0.07	1.04
	θ	0.04	0.22	1.03		θ	0.06	0.05	1.02
	m	0.03	0.16	1.001		m	0.04	0.04	1.01
	γ	0.02	0.09	1.005		γ	0.01	0.01	1.002

We can obtain the estimated bias, MSE, and MRE of $\hat{a},\ \hat{b},\ \hat{m},\ \hat{\gamma}$ similarly. The numerical results of the simulation are given in Table 1. It is clear from Table 1 that the estimated biases and MSEs decrease when the sample size *n* increases. The estimated MREs of all parameters are close to one and approach this nominal value when the sample size increases. Now, we present applications of EU-PD distribution using real data. These applications demonstrate the flexibility of this distribution compared to the other models for the real data set. We compare the fit of the EU-PD distribution with those of some other lifetime distributions which are Pareto, dagum and Beta exponentiated Pareto. The Pareto distribution is

$$f(X)=\frac{2\alpha\left(\beta/x\right)^{\alpha}}{x\left[1+\left(\beta/x\right)^{\alpha}\right]^{2}}=\frac{2\alpha\beta^{\alpha}x^{\alpha-1}}{\left(x^{\alpha}+\beta^{\alpha}\right)^{2}},\quad x\geq\beta,$$

and the Beta Exponentiated Pareto Distribution is

$$f(X)=\frac{\alpha k\beta^{k}}{B(a,b)}\frac{1}{x^{k+1}}\left[1-\left(\beta/x\right)^{k}\right]^{\alpha a-1}\left\{1-\left[1-\left(\beta/x\right)^{k}\right]^{\alpha}\right\}^{b-1},$$

and the dagum distribution is

$$f(X)=\alpha\beta\gamma x^{-(1+\gamma)}\left(1+\beta x^{-\gamma}\right)^{-(1+\alpha)}.$$

All the computations presented in this section were done using the MATLAB and R software.

The first set of data, named dataset II in this article, is relates to the test of the deep groove strength of the ball bearing, which includes the number of turns (per hundred million) for each of the 23 ball bearings. These data are taken from [10].

0.1788, 0.2892, 0.3300, 0.4152, 0.4212, 0.4560, 0.4848, 0.5184, 0.5196, 0.5412, 0.5556, 0.6780, 0.6864, 0.6864, 0.6888, 0.8412, 0.9312, 0.9864, 1.0512, 1.0584, 1.2792, 1.2804, 1.734.

The second data set (Set I) is related to the **National income data (in Rls. Billion)** of Iran in the years of 1959-2014 that is taken from Central Bank of Iran.

111257.8, 124298.4, 136981.4, 153789.4, 166084.9, 167425.3, 188821.2, 202564.7, 207174.4, 234197.3, 258982.0, 284830.5, 388985.7, 455285.8, 689874.2, 1077603.7, 1069051.5, 1259465.5, 1195498.9, 977492.5, 998628.7, 674656.1, 623984.0, 759981.3, 800077.5, 718224.0, 711357.3, 545700.5, 685626.3, 427235.8, 484050.3, 589257.6, 638427.2, 651409.6, 886092.3, 878198.7, 855566.2, 900151.8, 863284.1, 779184.7, 924250.1, 955521.6, 989637.3, 1163029.1, 1243801.0, 1334424.7, 1507359.8, 1600344.8, 1789984.5, 1778177.4, 1704565.1, 1871637.6, 1990261.3, 1649248.1, 1606316.0, 1580348.0.

We use the Crame'r-VonMisses (CVM), Anderson-Darling (ADR) test statistics, Akaike Information Criterion (AIC) and the Bayesian criterion (BIC) in order to compare the fits. The computed MLEs and the values of AIC and BIC, (CVM) and (ADR) for both data sets are given in Table 2. These criteria are widely utilized to check how closely a specified cdf fits the empirical distribution of a given data set. It is well-known that the smaller values of AIC, BIC, (CVM) and (ADR), mean a better fit to the data. Here, it is observed from Table 2 that the EU-PD model outperforms all the other considered models in the sense of the considered criteria.

Table 2. The maximum likelihood estimates of the parameters and the values of AIC and BIC

for real data.

DIST.	MLEs	CVM	ADR	AIC	BIC
New Model	$\boldsymbol{a=0.6520,\ b=1.1298,\ \theta=0.2524,\ m=0.5392,\ \gamma=0.3034}$	**0.0867**	**0.4631**	**889.7**	**895.1**
Pareto	$\alpha=2.7563,\ \beta=0.7999$	0.1428	0.7512	986.3	992.4
Dagum	$\alpha=1.4399,\ \beta=1375.63,\ \gamma=0.00237$	0.2531	0.9645	995.8	997.2
Beta Exponentiated Pareto	$\alpha=0.2775,\ \beta=0.5229,\ k=0.06234$ $a=1.3645,\ b=0.1072$	0.3652	1.2365	1356.6	1415.1
New Model	$\boldsymbol{a=0.0474,\ b=87168,\ \theta=0.00470,\ m=0.7683,\ \gamma=0.4887}$	**0.0428**	**0.3452**	**157.31**	**167.37**
Pareto	$\alpha=18.79,\ \beta=5943.59$	0.0854	0.8627	173.25	174.12
Dagum	$\alpha=5.5415,\ \beta=2.6884,\ \gamma=34.0665$	0.3676	1.4012	183.42	181.31
Beta Exponentiated Pareto	$\alpha=0.74923,\ \beta=0.0000001,\ k=1.01337$ $a=0.2719,\ b=0.715$	0.4756	2.8923	189.56	196.64

It can be seen from the Table 2, the new distribution has the smallest values of the smallest AIC, BIC, CVM and ADR values. Therefore, it can be concluded that the best fits belong to the new introduced model, i.e. the EU-PD distribution among the other considered distribution in this section.

For the sake of visual comparison, the estimated pdfs of the considered distributions as well as the empirical histograms of the data sets are given in Figure 3. Also, a normal Q-Q plot and a boxplot charts for the data set I and II are plotted in Figures 4 and 5. It is obvious from the figures, that the proposed new distribution provides the best fit for real data set in comparison with the other considered distributions.

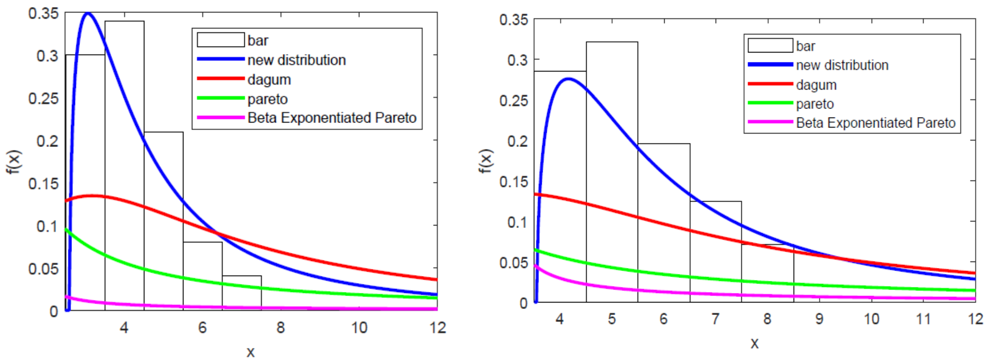

Figure 3. The plot of the fitted probability density functions of considered distributions as well as the histogram for the data set I and the dataset II.

We note that the figure plotting and computations of this paper have been performed using R [13]. Moreover, the package survival was used [15].

In this paper, a new distribution was introduced and some of its math properties were discussed. The new distribution has five parameters and its cdf and hrf have simple forms. The hrf of the new model can be bathtub and inverse bathtub shaped depending on the values of the parameters. Generally, we can say that the proposed distribution provides a more flexible model for fitting a wide range of real data sets in comparison with some other distributions and thus it can be an appropriate alternative distribution for some other existing models, in modelling real data that may appear in many areas like engineering, survival analysis, hydrology, economics and so on.

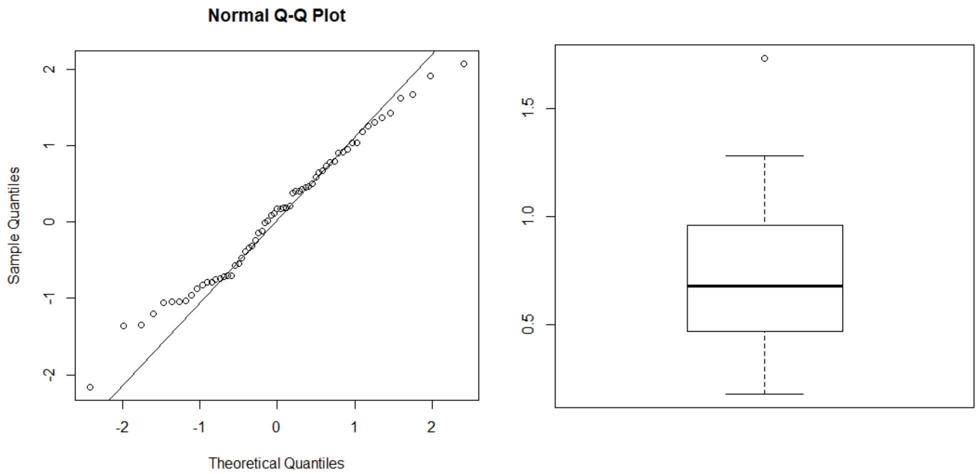

Figure 4. A normal Q-Q plot and a boxplot charts for the data set I.

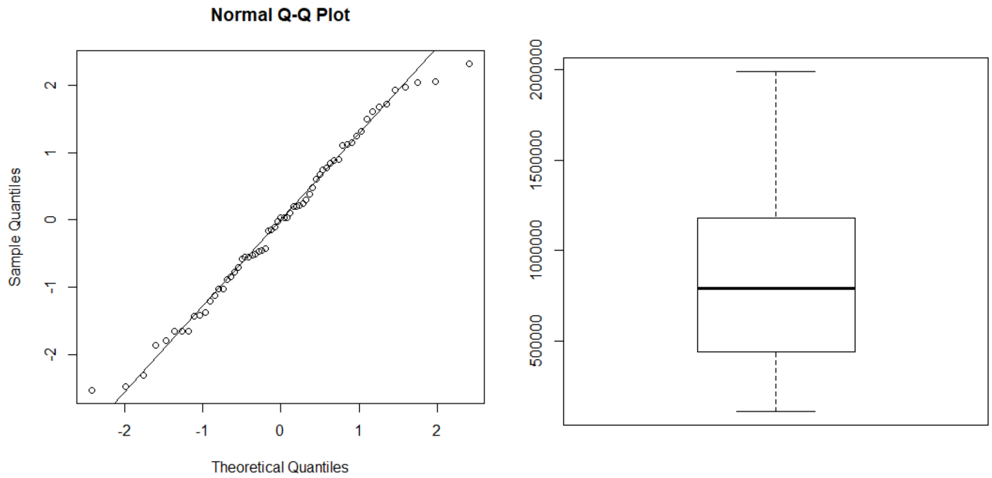

Figure 5. A normal Q-Q plot and a boxplot charts for the data set II.

Several properties of the new distribution, like sequential statistics, sequential statistical moments, the asymptotic distribution of extreme values, stochastic orderings and distribution of the ratio of two random variables have not been presented in this paper. In addition, there exist some inferential topics related to the new introduced distribution like Bayesian estimation of the parameters, estimation based on censored samples, estimation by means of other methods like the diagonally weighted least-squares method, prediction of the future observations from this distribution and so on. We hope to work on some of the mentioned topics and report our findings in future.

References

[1] Akinsete, F. Famoye and C. Lee, The Beta-Pareto distribution, Statistics, **42 (4)** (2008) 547–563.

[2] K. A. Al Kadhim and A. D. I. Al Musawy, Uniform-Pareto distribution, Journal of University of Babylon for Pure and Applied Sciences, **25 (5)** (2017) 1663-1673.

[3] M. Bahmani-Oskooee and A. Motavallizadeh-Ardakani, On the effects of income volatility on income distribution: Asymmetric evidence from state level data in the U.S., Journal of Research in Economics, **72 (2)** (2018) 224-239.

[4] F. Clementi, T. Di Matteo, M. Gallegati, and G. Kaniadakis, The k-generalized distribution: A new descriptive model for the size distribution of incomes, Physica A: Statistical Mechanics and its Applications, **387** (2008)

3201–3208.
[5] F. Clementi, M. Gallegati and G. Kaniadakis, A k-generalized statistical mechanics approach to income analysis, Journal of Statistical Mechanics: Theory and Experiment, (2009), doi: 10.1088/1742-5468/2009/02/P02037.
[6] F. Clementi, M. Gallegati and G. Kaniadakis, The k-generalised statistics in personal income distribution, The European Physical Journal B, **57** (2007) 187–193.
[7] A. Drăgulescu and V. M. Yakovenko, Evidence for the exponential distribution of income in the USA, The European Physical Journal B, **20** (2001) 585–589.
[8] F. C. Figueira, N. J. Moura Jr. and M. B. Ribeiro, The Gompertz Pareto income distribution, Physica A: Statistical Mechanics and its Applications, **390 (4)** (2011) 689–698.
[9] E. L. Lehmann and G. Casella, Theory of Point Estimation, 2nd Edition, Springer-Verlag, New York, (1998).
[10] J. Lieblein and M. Zelen, Statistical investigation of the fatigue life of deep-groove ball bearings, Journal of Research of the National Bureau of Standards, **57 (5)** (1956) 273-316.
[11] N. J. Moura and M. B. Ribeiro, Jr., Evidence for the Gompertz curve in the income distribution of Brazil 1978–2005, The European Physical Journal B, **67** (2009) 101–120.
[12] V. Pareto, Cours d'Économie Politique, Lausanne, (1987).
[13] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, (2017).
[14] C. A. Silva, Applications of physics to finance and economics: returns, trading activity and income, Ph.D. Thesis, University of Maryland, (2001).
[15] T. M. Therneau, survival: A package for survival analysis in S, R package version 2.38, (2015), URL: https://cran.r-project.org/package=survival.
[16] V. M. Yakovenko and J. B. Rosser, Colloquium: statistical mechanics of money, wealth, and income, Reviews of Modern Physics, **81** (2009) 1703–1725.