Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.15, No.1, 2025

Improving Statistical Model through the Development of Exponentiated Generalized Exponential Pareto Distribution.

[1]Jimoh M. Afolabi. [2] H. A. Bello [3] Bukoye A. [4]Margret Igiozee
[1]follybee@auchipoly.edu.ng
[2] habello@futa.edu.ng
[3] wasiubukoye@yahoo.com
[4]meggaiu@yahoo.com
[1,3,4]Department of Statistics, Auchi Polytechnic, Auchi.
[2]Department of Statistics, Federal University of Technology, Akure.

Abstract

Most importantly, the derivation of novel class of probability distributions plays a vital role in improving the underlying structure to model complex real-world data. In this study, we propose a new distribution, Exponentiated Generalized Exponential Pareto Distribution (EGEPD), which is a compounding distribution of the Exponentiated Generalized (EG) class of distribution and Exponential Pareto distribution. The valuation of EGEPD is contemplated by colossal comparisons with other distributions such as Exponential Pareto Distribution (EPD), Exponentiated Exponential Pareto Distribution (EEPD) and Exponential Distribution (ED). EGEPD significantly outperforms these distributions according to both visual and statistical analyses. The EGEPD outperforms all alternatives in terms of Log-Likelihood values, as well as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores, confirming its goodness-of-fit and optimal complexity. Goodness-of-fit tests based on Kolmogorov-Smirnov (KS), Anderson-Darling (AD), and Cramér-von Mises (CM) statistics, etc., confirm EGEPD's performance due to the smallest statistics and the largest p-values that contribute to strong fitting of EGEPD with observed data. Parameter analysis indicates EGEPD's location parameter (α) adjusts the distribution's central tendency, while the scale parameter β controls spread and variability. The shape parameters such as a, b, and θ are able to control skewness, tail behavior, and spread of the distribution, allowing the EGEPD to be flexible and fit diverse data characteristics. The ability to adjust this prior flexibility allows for EGEPD to be an effective spatio-temporal model for complex datasets across many domains that require varying spatial and temporal scales of location, variability, or shape. The EGEPD is the best and most robust model with respect to the fitted data and the structure of the data. It proves to be a great addition among other statistical modeling approaches such as the case with Exponential Distribution and also with the medium generators such as EPD and EEPD.

Keywords: Heavy-tailed data, Goodness-of-fit tests, Log-Likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov Statistic, Anderson-Darling Statistic, Cramér-von Mises Statistic, Flexible probability distributions.

1.1 Introduction

Diaconis (1988) explained this process through a statistical science lens: "New probability distributions is a rapidly growing area of research in statistics, and there is certainly no shortage of need for better models." We introduce a new probability distribution, the Exponentiated Generalized Exponential Pareto Distribution (EGEPD) based on Karima and Boshi (2013)'s

Exponential Pareto distribution as the base and Cordeiro (2013)'s Exponentiated Generalized family of distributions. Recent work in statistical inference has emphasized the development of more sophisticated probability distributions. Cordeiro et al. (2013) first proposed the Exponentiated Generalized (EG) family of distributions, a class of flexible distributions that extends standard models by introducing intercalated parameters that better represent the behaviour of the data. This family has been critical in designing more adaptable distributions that correct some flaws of the current models. The Exponential Pareto distribution is proposed by Karima and Boshi (2013) as a combination of the exponential and Pareto distributions to describe datasets with exponential-like and heavy-tailed attributes. This distribution can be applied in several fields, including finance and insurance, where extreme value modeling is important. More recently, researchers have built upon established forms to adapt probability distributions for a wider variety of practical applications. For instance, Alzaghal et al. They use exponentiation and generalization to demonstrate highly flexible models on the Exponentiated Generalized Gamma Distribution (2020). In a similar vein, Nadarajah and Kotz (2006) developed the Exponentiated Pareto Distribution, which extends the Pareto distribution by including an additional exponentiation parameter to provide more flexibility when modeling heavy-tailed data.

Motivated by the above-mentioned work, we propose the Exponentiated Generalized Exponential Pareto Distribution (EGEPD) which incorporates some merits of the EG family of distributions and the Exponential Pareto distribution. The main goal of this work is to present the EGEPD, which is a flexible model providing a good fit to data showing both exponential and heavy-tailed behavior.

2.1 METHODOLOGY

Derivation and Development of Exponentiated Generalized Exponential Pareto Distribution (EGEPD)

Let X be a random variable with Exponential Pareto distribution (EPD) defined in [Karima and Boshi, 2013] by the CDF, probability density function (pdf) and quantile function respectively given by:

$$G(x) = 1 - e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}} \quad x > 0, \alpha, \beta \text{ and } \theta > 0 \tag{1}$$

$$g(x) = \frac{\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\left(1 - e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right) \qquad x > 0, \alpha, \beta \text{ and } \theta > 0 \tag{2}$$

$$q = -\frac{\alpha}{\beta} log(1-p)^{\frac{1}{\theta}} \tag{3}$$

while equation (4) and (5) are the probability density function (pdf) and cumulative distribution function (cdf) of Exponentiated Generalized family proposed by Cordeiro et al (2013)

The CDF of Exponentiated Generalized G family of distribution is as follows:

$$F(x) = [1 - \{1 - G(x)\}^{a}]^{b} \qquad x > 0, a, b > 0 \tag{4}$$

and the pdf is given as

$$f(x) = abg(x)\{1 - G(x)\}^{a-1}[1 - \{1 - G(x)\}^{a}]^{b-1} \qquad x > 0, a, b > 0 \tag{5}$$

2.2 PROPOSED EXPONENTIATED GENERALIZED TOPP LEONE FRÉCHET PROBABILITY OF DISTRIBUTION

The propose Exponentiated Generalized Exponential Pareto Distribution (EGEPD) CDF is defined as :

$$F(x) = \left(1 - \left(1 - \left(1 - e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b} \quad x > 0, a, b, \theta, \alpha \text{ and } \beta > 0 \tag{6}$$

The quantile of the proposed distribution EGEPD is given as:

$$x = \alpha\left(-\frac{ln\left(\left(1 - p^{\frac{1}{b}}\right)^{\frac{1}{a}}\right)}{\beta}\right)^{\frac{1}{\theta}} \tag{7}$$

The median of the proposed distribution EGEPD is given as:

$$x = \alpha\left(-\frac{ln\left(\left(1 - (0.5)^{\frac{1}{b}}\right)^{\frac{1}{a}}\right)}{\beta}\right)^{\frac{1}{\theta}} \tag{8}$$

And the pdf is given as:

$$f(x) = \frac{ab\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}} \left(1 - \left(1 - e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a-1} \left(1 - \left(1 - \left(1 - e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b-1} x > 0, a, b, \theta, \alpha \text{ and } \beta > 0 \tag{9}$$

Using binomial expansion series, we have:

$$f(x) = \frac{\mathrm{ab\beta\theta}}{\alpha}\left(\frac{\mathrm{x}}{\alpha}\right)^{\theta-1} \sum_{i}^{\infty}\sum_{j}^{\infty}\sum_{k=0}^{\infty} -1^{i+j+k}\binom{b-1}{i}\binom{\mathrm{a}-1+\mathrm{ai}}{j}\binom{j}{k}\mathrm{e}^{-\beta\left(\frac{\mathrm{x}}{\alpha}\right)^{\theta}(k+1)} \tag{10}$$

$$if\ w_{i} = \sum_{i}^{\infty}\sum_{j}^{\infty}\sum_{k=0}^{\infty} -1^{i+j+k}\binom{b-1}{i}\binom{\mathrm{a}-1+\mathrm{ai}}{j}\binom{j}{k} \tag{11}$$

Therefor the probability density function (pdf) of the Exponentiated Generalized Exponential Pareto Distribution is given as:

$$f(x) = \frac{\mathrm{ab\beta\theta}}{\alpha}\left(\frac{\mathrm{x}}{\alpha}\right)^{\theta-1} w_{i}\mathrm{e}^{-\beta\left(\frac{\mathrm{x}}{\alpha}\right)^{\theta}(k+1)} \tag{12}$$

2.3 Expansion of Properties of the Propose Distribution Function

$$E[X^{r}] = \frac{ab\alpha^{r}w_{i}\Gamma\left(\frac{r+\theta}{\theta}+1\right)}{(k+1)^{\frac{r+\theta}{\theta}}} \tag{13}$$

$$E[X] = \frac{ab\alpha w_{i}\Gamma\left(\frac{1+\theta}{\theta}+1\right)}{(k+1)^{\frac{1+\theta}{\theta}}} \tag{14}$$

$$E[X^{2}] = \frac{ab\alpha^{2}w_{i}\Gamma\left(\frac{2+\theta}{\theta}+1\right)}{(k+1)^{\frac{2+\theta}{\theta}}} \tag{15}$$

$$E[X^3] = \frac{ab\alpha^3 w_i \Gamma\left(\frac{3+\theta}{\theta}+1\right)}{(k+1)^{\frac{3+\theta}{\theta}}} \tag{16}$$

$$E[X^4] = \frac{ab\alpha^4 w_i \Gamma\left(\frac{4+\theta}{\theta}+1\right)}{(k+1)^{\frac{4+\theta}{\theta}}} \tag{17}$$

2.4 Maximum Likelihood Function

Given a sample of size n with data point $x_1, x_2, \ldots, x_n$, the likelihood function is the product of the individual PDF's evaluated at each observation:

$$L(a,b,\beta,\theta,\alpha) = \prod_{i=1}^{n} f(x_i; a,b,\beta,\theta,\alpha)$$

Taking the natural logarithm of the likelihood function (log-likelihood) simplifies the differentiation process for parameter estimation:

$$logL(a,b,\beta,\theta,\alpha) = \sum_{i=1}^{n} logf(x_i; a,b,\beta,\theta,\alpha)$$

The log-likelihood function becomes:

$$logL(a,b,\beta,\theta,\alpha) = \sum_{i=1}^{n}\left(log(ab\beta\theta) - \log(\alpha) + (\theta-1)log\left(\frac{x_i}{\alpha}\right) - \beta\left(\frac{x_i}{\alpha}\right)^{\theta}\right) + (a-1)log\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right) + (b-1)log\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)$$

The maximum likelihood estimates of the parameters a, b, β, θ and α are obtained by taking the partial derivative of the log-likelihood function with respect to each parameter, setting them equal to zero and solving.

$$\frac{\partial logL}{\partial a} = \sum_{i=1}^{n}\left(\frac{1}{a} + log\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right)\right) + \sum_{i=1}^{n}(b-1)\frac{log\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right)\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)}{1-\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right)^{a}} \tag{18}$$

$$\frac{\partial logL}{\partial b} = \sum_{i=1}^{n}\frac{log\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)}{b-1} \tag{19}$$

$$\frac{\partial logL}{\partial \beta} = \sum_{i=1}^{n}\left[\frac{1}{\beta} - \left(\frac{x_i}{\alpha}\right)^{\theta} + \frac{(a-1)\left(\frac{x_i}{\alpha}\right)^{\theta} e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}}{1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)}\right] + \sum_{i=1}^{n}\frac{(b-1)a\left(\frac{x_i}{\alpha}\right)^{\theta}\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)^{a-1}}{1-\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right)^{a}} \tag{20}$$

$$\frac{\partial logL}{\partial \theta} = \sum_{i=1}^{n}\left[\frac{1}{\theta} + log\left(\frac{x_i}{\alpha}\right) - \beta\left(\frac{x_i}{\alpha}\right)^{\theta} log\left(\frac{x_i}{\alpha}\right)\right] + \sum_{i=1}^{n}\frac{(a-1)\beta\left(\frac{x_i}{\alpha}\right)^{\theta} log\left(\frac{x_i}{\alpha}\right) e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}}{1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)} + \sum_{i=1}^{n}\frac{(b-1)a\beta\left(\frac{x_i}{\alpha}\right)^{\theta} log\left(\frac{x_i}{\alpha}\right) e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)^{a-1}}{1-\left(1-\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)\right)^{a}} \tag{21}$$

$$\frac{\partial logL}{\partial \alpha} = \sum_{i=1}^{n}\left[-\frac{\theta}{\alpha} + \left(\frac{\beta\theta x_i^{\theta}}{\alpha^{\theta+1}}\right)^{\theta}\right] + \sum_{i=1}^{n}\frac{(a-1)\beta\theta x_i^{\theta} e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}}{\left(1-e^{-\beta\left(\frac{x_i}{\alpha}\right)^{\theta}}\right)} \qquad (22)$$

2.5 Order of Statistics

The density $f_{i:n}(x)$of the ith order statistic for i=1,…n, from independent identically distributed random variable $Y_1,\dots,\ Y_n$ is given by

$$f_{i:n}(x) = \frac{f(x)}{B(i,n-i+1)}F(x)^{i-1}\{1-F(x)\}^{n-i}$$

From equation 6 and 12, the CDF and pdf of the Exponentiated Generalized Exponential Pareto Distribution is given as:

$$f(x) = \frac{ab\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} w_i e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}(k+1)} \quad and\ F(x) = \left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b}$$

distribution is given as

$$f_{i:n}(x) = \frac{\frac{ab\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} w_i e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}(k+1)}}{B(i,n-i+1)}\left(\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b}\right)^{i-1}$$

$$*\left(1-\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b}\right)^{n-i}$$

$$let\ (1-z)^{k} = \left(1-\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b}\right)^{n-i}$$

$$where\ z = \left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b}$$

$$\sum_{l=0}^{\infty}\binom{n-i}{k}(-1)^{k}z^{k}$$

$$f_{i:n}(x) = \frac{\frac{ab\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} w_i e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}(k+1)}}{B(i,n-i+1)}\left(\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{b}\right)^{i-1}$$

$$*\sum_{j=0}^{n-i}\binom{n-i}{j}(-1)^{j}\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{bj}$$

$$f_{i:n}(x) = \frac{\frac{ab\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} w_i e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}(k+1)}}{B(i,n-i+1)} * \sum_{j=0}^{\infty}\binom{n-i}{j}(-1)^{j}\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{i+bj-1}$$

$$\left(1-\left(1-\left(1-e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)\right)^{a}\right)^{i+bj-1} = \sum_{k}^{i+bj-1}\sum_{l}^{k}\sum_{m}^{a}(-1)^{k+l+m}\binom{i+bj-1}{k}\binom{k}{l}\binom{a}{m}\left(e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)^{ml}$$

Therefore,

$$f_{i:n}(x) = \frac{\frac{ab\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} w_i e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}(k+1)}}{B(i, n-i+1)} * \sum_{j=0}^{\infty}\sum_{k}^{i+bj-1}\sum_{l}^{k}\sum_{m}^{a}(-1)^{j+k+l+m}\binom{n-i}{j}\binom{i+bj-1}{k}\binom{k}{l}\binom{a}{m}\left(e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)^{ml}$$

$$by\ expansion\ B(i, n-i+1) = \frac{n!}{\left((i-1)!\ (n-i)!\right)}$$

The order of statistic for EGEPD is given as:

$$f_{i:n}(x) = \frac{\left((i-1)!\ (n-i)!\right)\frac{ab\beta\theta}{\alpha}\left(\frac{x}{\alpha}\right)^{\theta-1} w_i e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}(k+1)}}{n!} * \sum_{j=0}^{\infty}\sum_{k}^{i+bj-1}\sum_{l}^{k}\sum_{m}^{a}(-1)^{j+k+l+m}\binom{n-i}{j}\binom{i+bj-1}{k}\binom{k}{l}\binom{a}{m}\left(e^{-\beta\left(\frac{x}{\alpha}\right)^{\theta}}\right)^{ml}$$

3.1 Data Presentation and Analysis of Data

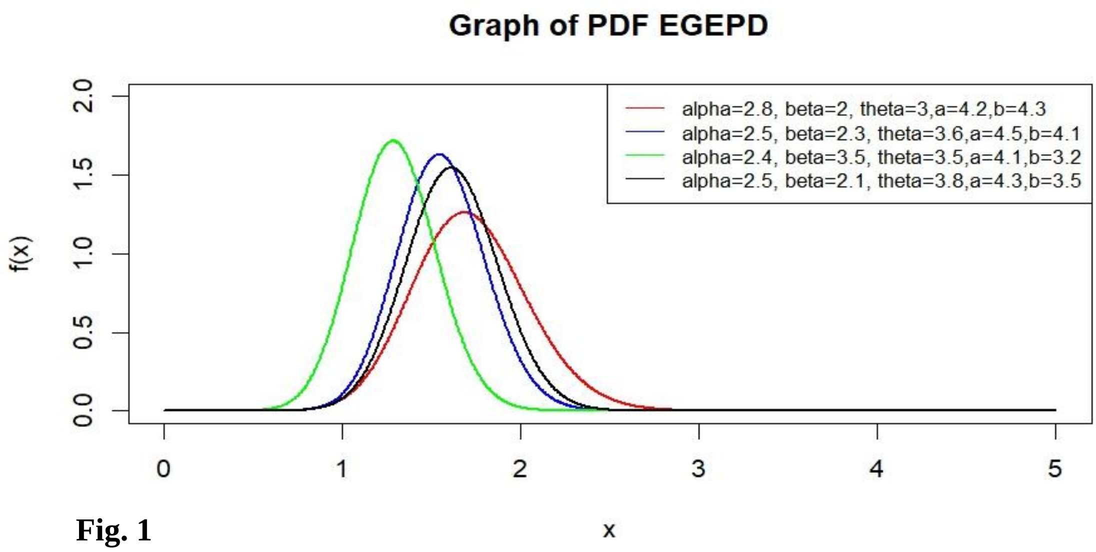

Fig. 1

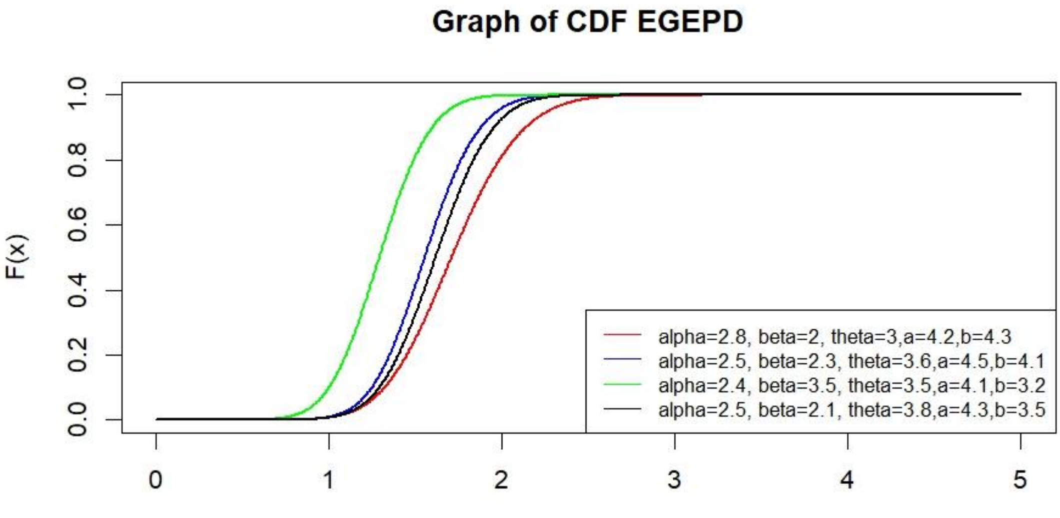
Graph of CDF EGEPD
F(x)
x
alpha=2.8, beta=2, theta=3,a=4.2,b=4.3
alpha=2.5, beta=2.3, theta=3.6,a=4.5,b=4.1
alpha=2.4, beta=3.5, theta=3.5,a=4.1,b=3.2
alpha=2.5, beta=2.1, theta=3.8,a=4.3,b=3.5

Fig. 2

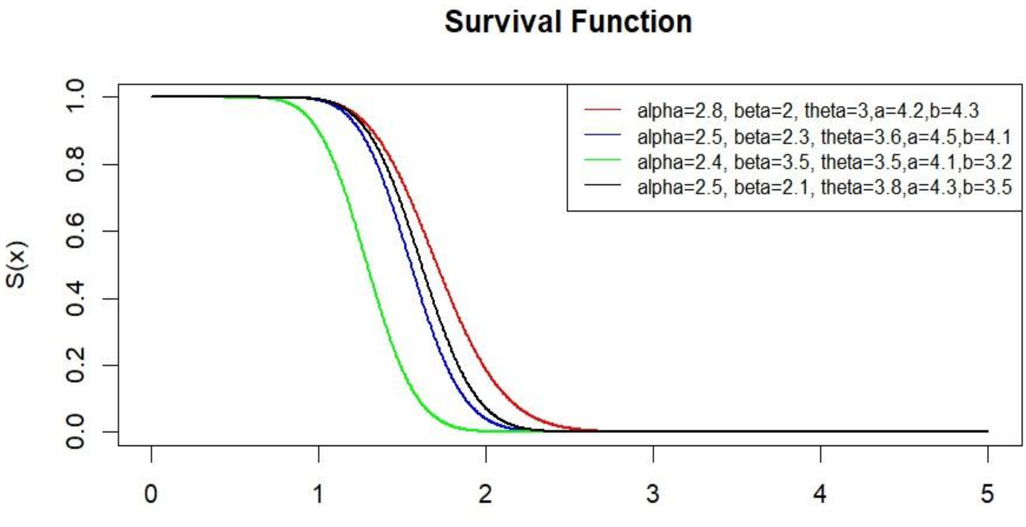
Survival Function
S(x)
x
alpha=2.8, beta=2, theta=3,a=4.2,b=4.3
alpha=2.5, beta=2.3, theta=3.6,a=4.5,b=4.1
alpha=2.4, beta=3.5, theta=3.5,a=4.1,b=3.2
alpha=2.5, beta=2.1, theta=3.8,a=4.3,b=3.5

Fig. 3

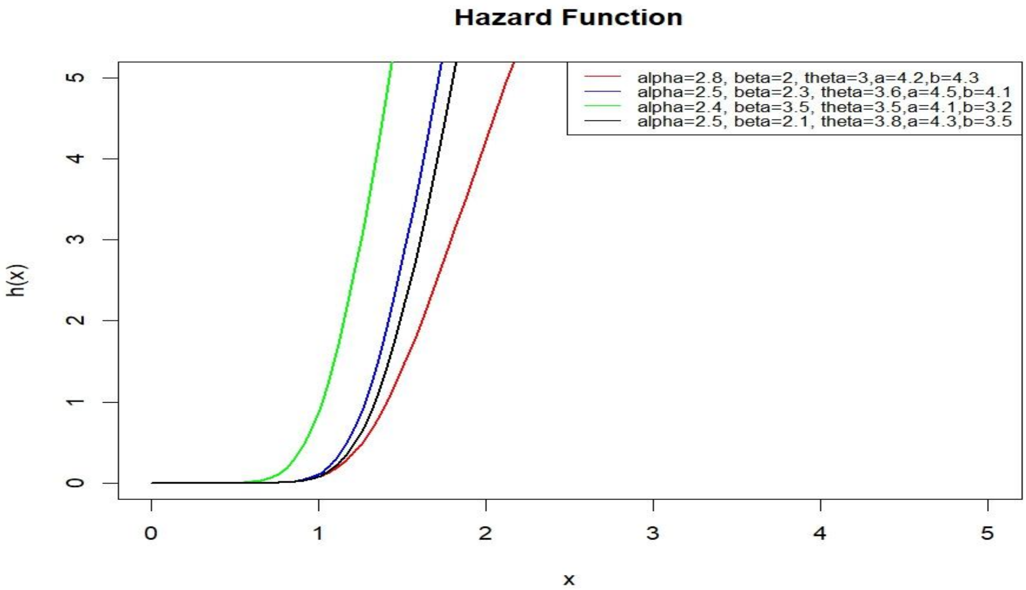

Fig. 4

3.2 REAL LIFE APPLICATION

In this section, we fit the EGEPD to real life data sets and compare it fitted values with that of its sub-models. The data comprise of Cumulative Grade Point Average (CGPA) of Auchi Polytechnic, Auchi.2022/2023 first semester result in the Department of Statistics,

Table 1: CGPA of 145 Students in the Department of statistics Auchi Polytechnic

Table: 1

3.47,3.17,2.91,2.42,3.09,2.59,3.12,2.30,2.66,2.83,3.01,3.46,2.44,3.19,2.38,2.36,2.59,2.86,2.87,2.53,2.81,3.01,
3.10,2.85,2.96,2.74,3.08,3.11,3.11,3.66,2.95,2.71,3.22,2.88,2.37,2.87,2.46,2.96,2.42,2.55,3.09,3.36,2.60,3.03,
2.86,3.00,2.99,2.66,3.18,2.75,3.04,2.64,2.59,3.49,2.26,2.74,2.55,2.79,2.43,2.67,2.83,3.08,2.84,2.40,3.14,2.50,
2.72,2.87,3.02,3.13,2.58,2.79,3.28,2.80,2.83,2.41,2.06,2.81,2.59,2.62,3.14,2.50,2.97,2.97,2.51,2.78,2.26,3.08,
2.90,2.46,3.03,2.81,2.75,2.56,2.35,2.50,2.38,3.01,3.23,2.76,3.06,2.92,2.45,2.72,2.87,2.85,2.96,2.74,2.92,2.45,
2.72,2.87,2.85,2.96,2.74,2.85,2.96,2.74,2.92,2.45,2.72,2.87,2.85,2.96,2.74,2.85,2.96,2.74,2.92,2.45,2.72,2.87,
2.85,2.96,2.74,2.85,2.96,2.74,2.92,2.45,2.72,2.87,2.85,2.96,2.74.

Source: Department of Statistics Auchi Polytechnic, 2022/2023 first semester result

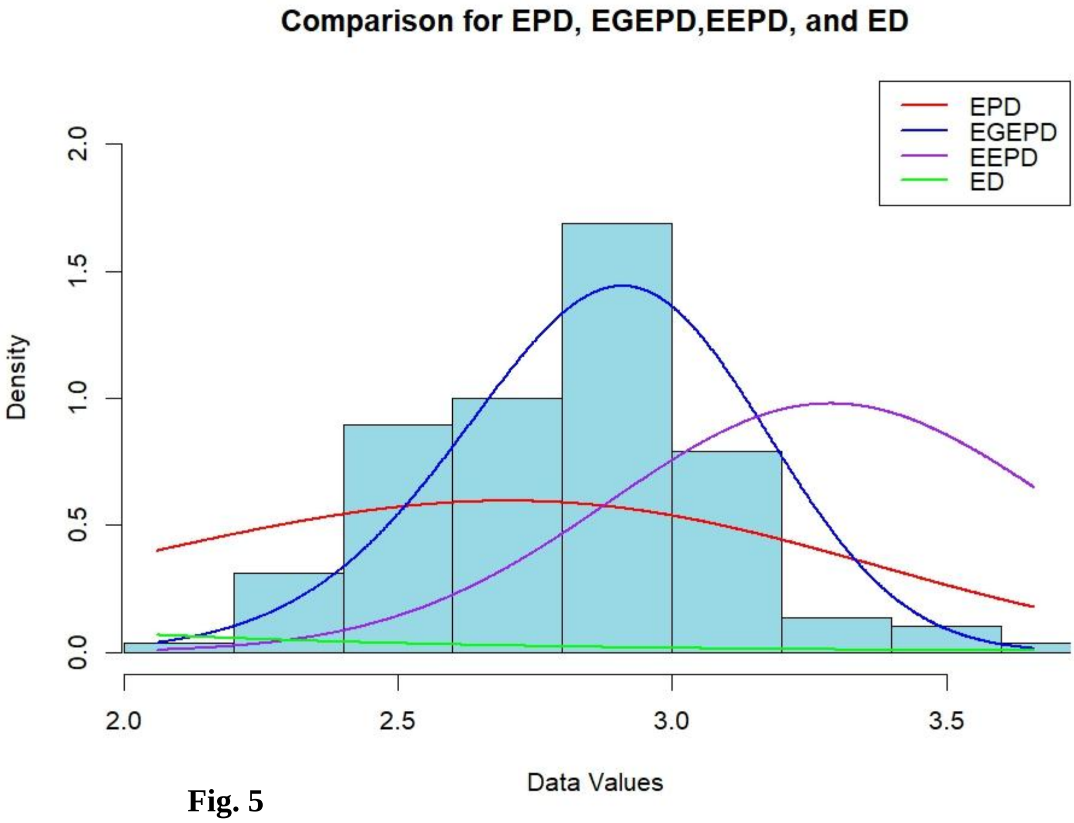

Fig. 5

This figure 5 compares the fit between 4 distributions (EPD, EGEPD, EEPD, and ED) and the data represented in the histogram. The blue bars represent the actual data Histogram The pitch of the bars indicates how common the data values are. The four colored lines in the plot are the fitted distributions so that EPD does not closely follow the form of the histogram. EGEPD fit the histogram very well, particularly around the peak. EEPD Somewhat follows the histogram but fails at parts. ED (Almost flat and clearly away from the histogram, the worst fit.) So, the EGEPD (blue line) has the best fit on the data, since it follows the histogram shape closely (especially at the peak). The Exponential Distribution (green line) has the worst fit as it is far from the data values.

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.15, No.1, 2025

Table 2: Log likelihood and Information Criterion on students CGPA

S/N	Distribution	Log Likelihood	AIC	BIC
1	**EGEPD**	**-28.99236**	**67.98473**	**81.57722**
2	EEPD	-58.64752	125.29505	136.16904
3	EP	-71.20472	148.40944	156.56493
4	ED	-683.80044	1369.60088	1372.31937

Table 2 displays the Log-Likelihood, AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values for various statistical distributions of students' CGPA data depicting the best-fit distribution for the students CGPA dataset. Log-Likelihood: A Higher value (closer to 0) means that the model fits the data better. EGEPD has the best fit: (-28.99) whereas, the worst fit is of Exponential distribution: (-683.80) AIC and BIC keep track of how well the model strikes a balance between fit and complexity. Generalization of the Model — Lower is better. In the above table EGEPD has lowest AIC (67.98) and BIC (81.57) and that has been derived only based on the fact that it performs the best across the top overall. AIC (1369.60) and BIC (1372.32) indicates that Exponential distribution is the worst fit for this data.

Table 3: Goodness of fit Criterion on students CGPA

S/N	Distribution	KS_Statistic	KS_p_value	AD_Statistic	CM_Statistic
1	**EGEP**	**0.1287149**	**4.427242e-02**	**3.60264**	**0.6172434**
2	EEP	0.5525224	6.415088e-31	16.36317	3.1746969
3	EP	0.3057557	9.180574e-10	88.21568	15.7915456
4	Exponential	0.9942006	3.702699e-99	644.98308	38.148909

Table 3: Goodness-of-Fit of Various Distributions to Students-CGPA Data (Based on 4 Tests) Kolmogorov-Smirnov Statistic (KS_Statistic) : The smaller the better. EGEPD is the best one with KS_Statistic 0.1287 and Exponential Distribution is the worst with KS_Statistic 0.9942. KS_p_value = higher p_value = better fit. If the EGEPD has the largest p-value (0.044), it indicates that it fits well. The main information that Exponential Distribution very small value (close to 0), therefore, the fitting is bad. AD_Statistic and CM_Statistic also indicate goodness of fit. Smaller values are better.EGEPD has the lowest values (3.60 for AD and 0.62 for CM),

thus it fits best. The worst fit is Exponential Distribution (note that it has the biggest values, 644.98 for AD and 38.15 for CM).

The strengths of 1.5cm glass fibres were used from those obtained by workers at the UK National Physical Laboratory. Smith and Naylor, Bourguinonetal have used that data before and Merovci et al. The first run is as follows:

Table 4: The strength data concerning the 1.5cm glass fibers obtained by the workman at the UK National Physical laboratory

0.55, 0.74, 0.77, 0.81, 0.84, 1.24, 0.93, 1.04, 1.11, 1.13, 1.30, 1.25, 1.27, 1.28, 1.29, 1.48,
1.36, 1.39, 1.42, 1.48, 1.51, 1.49, 1.49, 1.50, 1.50, 1.55, 1.52, 1.53, 1.54, 1.55, 1.61, 1.58,
1.59, 1.60, 1.61, 1.63, 1.61, 1.61, 1.62, 1.62, 1.67, 1.64, 1.66, 1.66, 1.66, 1.70, 1.68, 1.68,
1.69, 1.70, 1.78, 1.73, 1.76, 1.76, 1.77, 1.89, 1.81, 1.82, 1.84, 1.84, 2.00, 2.01, 2.24.

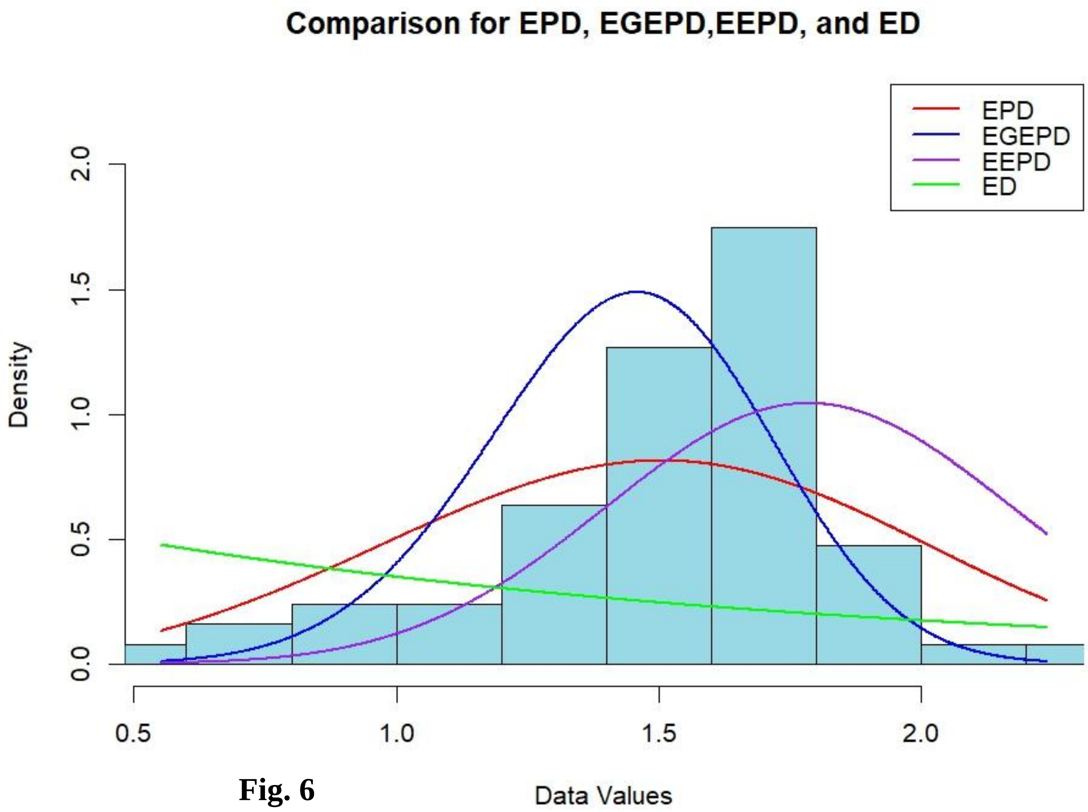

Fig. 6

Figure 6, the probability density functions, PDFs, for some four distributions: the EPD, EGEPD, EEPD, ED The histogram (light blue bars) visualizes the actual distribution of the data, that is, how often data values occur over ranges. The four lines (red, blue, purple, and green) are the theoretical density curves for the four distributions, where EPD (red line) fits moderately well, but it does not match with the peak of the histogram. EGEPD (blue line) picks out the best fit of the data its shape and height is closest to the histogram. EEPD (purple line) fits quite closely, though we see a slight deviation at the peak. ED (green line) poorly fits the data and has a visible misalignment with the histogram. Thus, from the histogram and its fit to the data, it can be concluded that the EGEPD distribution is the best fit.

Table 5: Log likelihood and Information Criterion on Glass fibre data

S/N	Distribution	LogLikelihood	AIC	BIC
1	**EGEPD**	**-30.80188**	**71.60375**	**82.31942**
2	EEPD	-42.72501	91.45002	97.87943
3	EPD	-93.40360	194.80719	203.37973
4	Exponential D.	-292.38346	586.76691	588.91005

EGEPD satisfactorily fits Glass Fibre Data best because it possesses highest Log-Likelihood along with lowest AIC and BIC values as shown in Table 5. Exponential distribution is certainly the worst fit with the least Log-Likelihood and greatest AIC and BIC values. Other models (EEPD and EPD) only achieve moderate performance, but not better than EGEPD.

Table 6 Goodness of fit Criterion on Glass fibers data

S/N	Distribution	KS_Statistic	KS_p_value	AD_Statistic	CM_Statistic
1	**EGEPD**	0.1623103	7.234397e-02	2.808005	0.3841247
2	EEPD	0.2049535	1.005585e-02	4.033532	0.7515303
3	EPD	0.3850873	1.535735e-08	17.755910	3.5482622
4	Exponential Distribution	0.9877227	8.228794e-54	610.257148	20.9588445

As seen in Table 6, EGEPD is the best-fitting for the data with the smallest KS, AD, and CM statistics, and the highest p value. The worst distribution is Exponential with the highest KS, AD and CM statistics and a near-zero p-value. EEPD and EPD moderately fit but not as EGEPD fit.

Simulated Data

This will create simulated realizations of EGEPD based on the quantile function. Here's how that simulation process works. The shape parameters are (a, b and θ), the codes simulate for different values of (a= 1.5, 2.0, 2.5), (b=1.2, 1.8, 2.0) and (θ=1.2, 1.5, 1.8) Beta (β) is the scale parameter, being held constant (1.5) in this simulation. (θ) is the location parameter which is fixed (0.8) in this simulation. Sample Size (n) — Number of random values to simulate (50 and 100). The simulation starts with drawing p uniform random samples of the size of n distributed in [0, 1]. We can create the random variables using the runif() function in R. The random variables are transformed to follow the Exponentiated Generalized Exponential Pareto (EGEPD). To obtain the desired value of x, the quantile function Q (p) is solved for x, using generated random values (p). This is a great way to convert a uniform test of random values into an EGEPD. Once the simulated data points are created, descriptive statistics (which in this case are simple averages because we are simulating independent data points) are calculated from each combination of alpha and beta values. Mean is the mean of the simulated data, Median is the middle value of the simulated data. The Standard Deviation (SD) is calculated from the spread of the simulated data. These statistics give an idea of the mean and standard deviation of the simulated data.

Mean, Median, and SD calculated for each combination of alpha and beta would be saved in a results data frame. I then reshaped this data frame so the statistics were in their own columns for easier viewing.

Table 7: Mean, Standard Deviation and Median from Exponentiated Generalized Exponential Pareto (EGEPD) of sample Size n=50

			Scale Parameters					
Shape Parameters			$\boldsymbol{\alpha = 2}$ $\boldsymbol{\beta = 4}$			$\boldsymbol{\alpha = 4}$ $\boldsymbol{\beta = 6}$		
$\boldsymbol{\theta}$	$\boldsymbol{a}$	$\boldsymbol{b}$	Mean	Median	SD	Mean	Median	SD
1.2	2.0	1.5	0.4412	0.3541	0.3087	0.6294	0.5051	0.4404
1.2	2.0	2.0	0.4706	0.4195	0.2334	0.6713	0.5984	0.3329
1.2	2.0	2.5	0.5326	0.4618	0.3512	0.7598	0.6587	0.5010
1.2	2.5	1.5	0.3345	0.2663	0.2498	0.4772	0.3798	0.3563
1.2	2.5	2.0	0.4167	0.4020	0.2531	0.5944	0.5735	0.3611
1.2	2.5	2.5	0.3874	0.3574	0.1948	0.5527	0.5098	0.2778
1.2	3.0	1.5	0.3088	0.2691	0.1975	0.4405	0.3838	0.2817
1.2	3.0	2.0	0.3291	0.3043	0.1977	0.4694	0.4341	0.2821
1.2	3.0	2.5	0.3785	0.3314	0.2561	0.5400	0.4728	0.3654
1.5	2.0	1.5	0.5118	0.4575	0.2674	0.7811	0.6982	0.4081
1.5	2.0	2.0	0.6282	0.5723	0.2854	0.9588	0.8735	0.4357
1.5	2.0	2.5	0.6902	0.5733	0.3458	1.0535	0.8750	0.5278
1.5	2.5	1.5	0.4675	0.3977	0.2662	0.7136	0.6070	0.4063
1.5	2.5	2.0	0.5065	0.4884	0.2322	0.7731	0.7454	0.3543
1.5	2.5	2.5	0.5504	0.5332	0.2596	0.8401	0.8138	0.3963
1.5	3.0	1.5	0.3977	0.3619	0.2147	0.6071	0.5523	0.3276
1.5	3.0	2.0	0.4721	0.4585	0.2027	0.7205	0.6998	0.3094
1.5	3.0	2.5	0.5159	0.5192	0.2184	0.7874	0.7924	0.3333

1.8	2.0	1.5	0.6880	0.6319	0.3258	1.0984	1.0088	0.5202
1.8	2.0	2.0	0.6994	0.6570	0.3239	1.1166	1.0490	0.5172
1.8	2.0	2.5	0.7975	0.7520	0.3057	1.2733	1.2007	0.4881
1.8	2.5	1.5	0.6207	0.5979	0.2780	0.9910	0.9545	0.4439
1.8	2.5	2.0	0.6713	0.6742	0.3210	1.0719	1.0764	0.5126
1.8	2.5	2.5	0.6972	0.7005	0.2644	1.1131	1.1185	0.4221
1.8	3.0	1.5	0.4959	0.4771	0.2048	0.7917	0.7618	0.3270
1.8	3.0	2.0	0.5689	0.5358	0.2377	0.9084	0.8555	0.3795
1.8	3.0	2.5	0.6130	0.6117	0.2054	0.9788	0.9766	0.3279

Source: Computation from simulated data on EGEPD distribution

Table 8: Mean, Standard Deviation and Median from Exponentiated Generalized Exponential Pareto (EGEPD) of sample Size n=100

			Scale Parameters					
Shape Parameters			$\alpha = 2$ $\beta = 4$			$\alpha = 4$ $\beta = 6$		
θ	a	b	Mean	Median	SD	Mean	Median	SD
1.2	2.0	1.5	0.4102	0.3298	0.2783	0.5851	0.4704	0.3970
1.2	2.0	2.0	0.4901	0.3925	0.3448	0.6992	0.5600	0.4918
1.2	2.0	2.5	0.5249	0.4816	0.2822	0.7487	0.6870	0.4026
1.2	2.5	1.5	0.3412	0.3026	0.2215	0.4867	0.4317	0.3159
1.2	2.5	2.0	0.3988	0.3623	0.2611	0.5688	0.5169	0.3725
1.2	2.5	2.5	0.4140	0.3531	0.2368	0.5905	0.5037	0.3378
1.2	3.0	1.5	0.3171	0.2695	0.2164	0.4523	0.3845	0.3087
1.2	3.0	2.0	0.3504	0.3085	0.2186	0.4999	0.4400	0.3119
1.2	3.0	2.5	0.3478	0.3086	0.1869	0.4962	0.4402	0.2666
1.5	2.0	1.5	0.5577	0.4980	0.2833	0.8512	0.7601	0.4325
1.5	2.0	2.0	0.6105	0.5451	0.3178	0.9318	0.8320	0.4850
1.5	2.0	2.5	0.6544	0.6235	0.3109	0.9988	0.9516	0.4746
1.5	2.5	1.5	0.4449	0.4068	0.2323	0.6790	0.6209	0.3546
1.5	2.5	2.0	0.4829	0.4606	0.2400	0.7371	0.7031	0.3662
1.5	2.5	2.5	0.5936	0.5594	0.2685	0.9060	0.8538	0.4099
1.5	3.0	1.5	0.3801	0.3605	0.2106	0.5802	0.5502	0.3215
1.5	3.0	2.0	0.4760	0.4171	0.2685	0.7266	0.6366	0.4097
1.5	3.0	2.5	0.5200	0.4979	0.1999	0.7937	0.7600	0.3050
1.8	2.0	1.5	0.6510	0.6079	0.3119	1.0394	0.9706	0.4980
1.8	2.0	2.0	0.7546	0.7280	0.3080	1.2048	1.1623	0.4917
1.8	2.0	2.5	0.7750	0.7782	0.2993	1.2374	1.2424	0.4778
1.8	2.5	1.5	0.5769	0.5686	0.2551	0.9210	0.9078	0.4072
1.8	2.5	2.0	0.7042	0.6841	0.3044	1.1243	1.0922	0.4860
1.8	2.5	2.5	0.7465	0.7279	0.2981	1.1919	1.1623	0.4760
1.8	3.0	1.5	0.5348	0.5036	0.2865	0.8538	0.8041	0.4574
1.8	3.0	2.0	0.5862	0.5571	0.2487	0.9359	0.8895	0.3970
1.8	3.0	2.5	0.6269	0.5809	0.2710	1.0010	0.9275	0.4326

Source: Computation from simulated data on EGEPD distribution

Tables 7 and 8 presents the simulated outputs for EGEPD shows how the different parameters affect the statistical properties (mean, median and standard deviation) of the distribution. As the location parameter α increases the distribution shift right and both mean and median increase. This indicates that the EGEPD is well suited to datasets that have a baseline that shifts the central tendency. A big value of the scale parameter β stretches the distribution and leads to larger mean, median, and SD, which shows that β controls the "spreadness" of the distribution. Increasing α reduces the mean and variability in general when α, β and θ are fixed. An increase in the size of b results in a larger mean and standard deviation (SD) and therefore affects the tail behavior and variability of the distribution. The result is that larger θ corresponds to a wider and more spread out distribution, thus greater variability. The adjustment of the parameters (especially the simultaneous growth of α and β) will make the distribution shift to the right (larger values) and more diverse. This interaction highlights the EGEPD's versatility in producing different shapes and spreads.

The findings show that the EGEPD can be tuned to fit different modeling scenarios due to its parameterization structure: location parameter α affecting the center tendency, scale parameter β adjusting to spread and variance, and shape parameters a, b and θ determining finer details such as skewness, tail behavior, and overall spread of the distribution. The flexibility of the EGEPD provides a strong approach to modeling complex datasets, especially those data sets that need to be fitted to the location, variability, or shape.

4.0 Findings

The results show that, in all indices and figures, the best fit is provided by the EGEPD distribution. As shown in Figures 5 and 6, the EGEPD curve closely matches both the histogram peaks and the overall shape of the data, thus visually outperforming the other distributions. This result is confirmed by Tables 2 and 5, where EGEPD emerges with the highest Log-Likelihood and the lowest AIC and BIC, indicating its ability to balance goodness of fit and model complexity effectively. Moreover, these observations are also supported by Tables 3 and 6, where EGEPD has the smallest KS, AD, and CM statistics with the highest p-value, implying the closest fit among compared methods to fit the data. Conversely, the ED is the worst among the analyses. The numbers, underscore that its curve is virtually flat, and very far from the data and unable to capture the shape of the histogram. Tables 2 and 5 confirm this poor performance with its low Log-Likelihood and largest AIC and BIC values. Finally, as shown in Tables 3 and 6, ED has the largest KS, AD, and CM respectively, along with a p-value of nearly 0, further supporting this observation. Though the EEPD and EPD distributions give moderate performance, they are substantially outperformed by EGEPD, in terms of all evaluation criteria.

So, in conclusion, the analysis shows that the EGEPD distribution provides the best fit among all the fitted distributions since it achieves better performance in terms of all the goodness-of-fit measures for all the datasets. The Exponentiated Generalized Exponential Pareto Distribution (EGEPD) statistical properties are notably impacted the values of their parameters. As the location parameter α increases, the whole distribution shifts rightward with an increasing mean and median. A greater scale parameter β stretches it out: larger mean, median, and variability. Specific shape parameters can control the distribution's spread, tail behavior, and variance: a usually decreases variance, b increases it, and θ expands the entire scale. EGEPD has high flexibility for applications, where the parameters interacting simultaneously allow for various shapes of the distribution.

4.1 Conclusion

The analysis shows that the Exponentiated Generalized Exponential of Pareto Distribution (EGEPD) is the most preferable distribution tfor the given data sets. It consistently yields better performance on all evaluation metrics log-likelihood, AIC, BIC, and goodness-of-fit tests such as KS, AD and CM statistics. Its strong fit is further confirmed by the clear dyeing of the EGEPD curve with respect to the histogram. In contrast, the Exponential Distribution (ED) fares the worst in terms of its ability to represent the data, clearly not aligning in statistical or graphical terms as seen in Figure [5] and Figure [6] The EEPD and EPD models provide moderate performance compared to the benchmark EGEPD, but their result is inferior. As a result, the EGEPD is the least liable and stable selection tool to classify this data, reflecting the latent features and construction quite well. The outcomes demonstrate that the parameter structure of EGEPD allows the model to adjust to different modeling contexts. Where the location parameter α determines the central tendency, the scale parameter β adjusts spread and variability, and the shape parameters a, b and θ when appropriate define finer details like skewness, tail behavior, and overall spread of the distribution. This comprehensive flexibility endows the EGEPD as a powerful tool for modeling complex datasets, especially those that need to be adjusted in location, variability or shape.

Reference

Cordeiro, G. M., Ortega, E. M., & da Cunha, D. C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, *11*(1), 1-27.

Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, 81(7), 883 - 898.

Bourguignon, M. and Cordeiro, G.M. (2016). The exponentiated generalized extended exponential distribution. Journal of Data Science, 14, 393C414.

De Andrade, T.A.N., Rodrigues, H. and Cordeiro, G.M. (2015). The exponentiated generalized Gumbel distribution. Revista Colombiana de Estadstica, 38, 123C143. https://doi.org/10.15446/rce.v38n1.48806.

Elbatal, I. and Muhammed, H.Z. (2014). Exponentiated generalized inverse Weibull distribution. Applied Mathematical Sciences, 8, 3997-4012.

Eugene and Jones, M. C. (2008). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*. **6**:70-81.

Galton, F. (1883). Enquiries into Human Faculty and its Development, Macmillan and Company, London.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society London, 115, 513-585.

Gupta, R.C., Gupta, P.L. and Gupta, R.D. (1998). Modeling failure time data by Lehman alternatives. Communications in Statistics, Theory and Methods, 27, 887-904.

Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions. Australian & New Zealand Journal of Statistics, 41(2), 173-188. 358 On Development of Four-Parameters Exponentiated Generalized.

Johnson, N.L., Kotz, S. and Balakrishnan, N (1994). Continuous Univariate Distributions, vol. 1, 2nd Edn. Wiley, New York.

Kapur, J.N. and Sexena, H.C. (1960). Mathematical Statistics, First Edition S. Chand and Company Ltd. Ram Nagar, New Delhi.

Kareema and Boshi, Exponential Pareto Distribution, Mathematical Theory and Modeling, 3 (2013), no. 5, 135-146.