

## Marshall-Olkin Generalized Pareto Distribution: Bayesian and Non Bayesian Estimation

Hanan A. Haj Ahmad <sup>1\*</sup>, Ehab M. Almetwally <sup>2</sup>

\* Corresponding Author



1. Basic Science Department, King Faisal University, KSA, [hananahm1@yahoo.com](mailto:hananahm1@yahoo.com)
2. Faculty of Business Administration, Delta University of Science and Technology, Egypt, [ehabxp\\_2009@hotmail.com](mailto:ehabxp_2009@hotmail.com)

### Abstract

A new generalization of generalized Pareto Distribution is obtained using the generator Marshall-Olkin distribution (1997). The new distribution MOGP is more flexible and can be used to model non-monotonic failure rate functions. MOGP includes six different sub models: Generalized Pareto, Exponential, Uniform, Pareto type I, Marshall-Olkin Pareto and Marshall-Olkin exponential distribution. We consider different estimation methods for estimating the model parameters, namely: Maximum likelihood estimator, Maximum product spacing, Least square method, weighted least square method and Bayesian Method. The Bayesian Method is considered under quadratic and Linex loss functions. Simulation analysis using MCMC technique is performed to compare between the proposed point estimation methods. The usefulness of MOGP is illustrated by means of real data set, which shows that this generalization is better fit than Pareto, Generalized Pareto and Marshall-Olkin Pareto distributions.

**Key Words:** Marshall-Olkin distribution; Generalized Pareto distribution; Maximum Product Spacing; Bayes estimation; Monte Carlo simulations.

**Mathematical Subject Classification:** 62E15, 62-07

### 1. Introduction

Statistical distributions are widely used to describe real life phenomena. For this reason, the theory of statistical distribution and generating new family of distributions are of great interest, many authors studied and generate new distributions from old ones. Many generalized classes of distributions have been developed and applied to describe various phenomena. A common advantage of these generalized distributions is that they have more parameters, and hence more flexible to model real life data. Marshal and Olkin (1997) proposed an important method of adding a new parameter to an existing distribution, thus defining a new distribution called Marshall-Olkin (MO) extended distribution. This new distribution includes the original distribution as a special case and it gives more flexibility to the model. Sankaran and Jayakumar (2008) introduced a detailed study about the physical interpretation of the MO family. Jose (2011) studied a family of Marshall-Olkin distribution and considered their application in reliability theory. Many studies have been proposed on Marshall-Olkin distribution see for example Jose and Alice (2001, 2005), Jose and Uma (2009), Haj Ahmad et al. (2017), Basheer (2019), Okasha et al. (2020) and others.

Pareto distribution is a well-known model, it was first studied by a professor of economics "Vilfredo Pareto". Many forms of Pareto distribution appeared in the literature and they were used in a wide range of scientific applications. For instance it was found that it is compliant in lifetime models such as actuarial sciences, finance, economic, life testing and climatology and it has many applications in survival analysis. Its importance also appears since it is a generalization of many other distributions such as: Exponential, Rayleigh and generalized extreme value distribution. Several generalized forms of Pareto distribution were discussed in literature. Generalized Pareto (GP) distribution was first studied by Pickands (1975), and then it was studied by many authors like Gupta et al. (1998) and Hogg et al.

(2005). The GP distribution was used as a model for excesses over thresholds. Its applications include environmental extreme events, ozone levels in the upper atmosphere, large insurance claims or large fluctuation in financial data, and reliability studies. Its areas of applications are successfully addressed in several books, such as those by Castillo et al. (2005), Kotz and Nadarajah (2000). Alice and Jose (2004) considered Marshall-Olkin Pareto and Marshall-Olkin semi-Pareto distributions; they developed time series models with modification structure. Ghitany (2005) considered MO of Pareto distribution and studied some of its statistical properties and its hazard rate. He also showed that the limiting distributions of the sample extremes were of exponential and Frichet type. Almetwally and Almongy (2019 b) used different estimation methods to estimate the parameters of a New Weibull-Pareto Distribution, Bdair and Haj Ahmad (2019), discussed different point and interval estimation for MO Pareto distribution.

In this work we introduce an extension of generalized Pareto distribution based on MO extension method, and make inferences on the parameters of MO generalized Pareto (MOGP) distribution.

We study different classical point estimation methods for the unknown parameters of MOGP distribution as well as the Bayesian method. Some properties of the density function are discussed. Numerical methods are used to solve the obtained nonlinear equations. Simulations are used to make comparison between those methods, and also to determine which method is more efficient according to the Bias and mean square error (MSE).

The rest of the paper is organized as follows: In section 2 we introduce MOGP distribution. Classical point estimation methods for the unknown parameters are discussed in section 3, while in section 4, Bayesian estimation method is considered. In section 5 simulation study and real life data analysis is presented and also comparison results among all estimation methods are provided.

## 2. Probability Density Function

Let  $F(x)=1-F(x)$  denote the survivor function of a continuous random variable X, let  $f(x)$  be the density function (pdf) associated with the cumulative distribution function (cdf)  $F(x)$ , then the MO extended distribution has survival function

$$\bar{G}(x) = \frac{\alpha \bar{F}(x)}{1 - \alpha \bar{F}(x)} \quad -\infty < x < \infty, \alpha > 0, \tag{1}$$

where  $\alpha=1-\alpha$ .

For  $\alpha=1$ ,  $G(x)=F(x)$ , hence  $F(x)$  is a special case of  $G(x)$ .

The probability density function (pdf) corresponding to Eq. (1) takes the form

$$g(x) = \frac{\alpha f(x)}{(1 - \alpha \bar{F}(x))^2}, \quad -\infty < x < \infty, \quad \alpha > 0.$$

Consider Generalized Pareto distribution (GP) with (pdf) given by

$$f(x; \theta, \lambda) = \frac{1}{\lambda} \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}-1}, \quad 0 \leq x < \infty,$$

where  $\theta$  is the shape parameter,  $\lambda>0$  is a scale parameter. For  $\theta < 0$  the range of  $x$  is  $0 < x < -\lambda/\theta$  and for  $\theta > 0$  the range is  $x > 0$ . The GP distribution becomes the uniform distribution for  $\theta=-1$  and the exponential distribution for  $\theta = 0$  (taken as the limit). The GPD has mean  $(\lambda / (1-\theta))$  and variance  $\frac{2}{(1-\theta)^2(1-2\theta)}$  provided  $\theta < 0.5$ . In this work we will consider the case when  $\theta > 0$ . The cumulative distribution function (cdf) is given by

$$F(x; \theta, \lambda) = 1 - \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}},$$

$$\bar{F}(x; \theta, \lambda) = \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}}.$$

Using Eq. (1) the survival function of Marshall-Olkin Generalized Pareto distribution (MOGP) is

$$\bar{G}(x; \alpha, \theta, \lambda) = \frac{\alpha \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \alpha \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}}}.$$

The corresponding probability density function (pdf) of MOGP is

$$g(x; \alpha, \theta, \lambda) = \frac{\frac{\alpha}{\lambda} \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}-1}}{\left(1 - \bar{\alpha} \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}}\right)^2}, \tag{2}$$

for  $\alpha > 0, \theta > 0, \lambda > 0$ , and  $x \geq 0$ .

### 2.1 Monotonicity of MOGP distribution

In this subsection we study the monotonicity of MOGP distribution, for that purpose let  $V(x) = \left(1 + \frac{\theta x}{\lambda}\right)$ . Then the logarithm of pdf of MOGP in Eq. (2) can be rewritten as:

$$\log g(x; \alpha, \theta, \lambda) = \log \frac{\alpha}{\lambda} - \left(\frac{1}{\theta} + 1\right) \log V(x) - 2 \log \left(1 - \bar{\alpha} V(x)^{-\frac{1}{\theta}}\right), \tag{3}$$

where  $x \geq 0, \alpha > 0, \theta > 0, \lambda > 0$ .

The monotonicity of Eq. (3) will give an idea about the monotonicity of MOGP distribution, hence taking the derivative with respect to x we obtain

$$\frac{\partial \log g(x; \alpha, \theta, \lambda)}{\partial x} = \frac{V'(x)}{V(x)} \left[ -\left(\frac{1}{\theta} + 1\right) - 2 \frac{\bar{\alpha}}{\theta} \frac{1}{\left(V(x)^{\frac{1}{\theta}} - \bar{\alpha}\right)} \right].$$

Since  $V(x)$  and  $V'(x)$  are both positive, then if  $0 < \alpha < 1$ , the above derivative is negative, hence the MOGP distribution is decreasing for  $x > 0$ . While if  $\alpha > 1$  the derivative has a single root  $r_0 = \frac{\lambda}{\theta} \left[ \left(\frac{-2\bar{\alpha}}{\theta+1} + \bar{\alpha}\right)^\theta - 1 \right]$ , so MOGP is increasing when  $x \in [0, r_0)$ , but it is decreasing when  $x \in (r_0, \infty)$ . Figure (1) illustrates the shapes of MOGP density function. Since the GP distribution is a decreasing function, using MOGP distribution is more appropriate for non-monotone models.

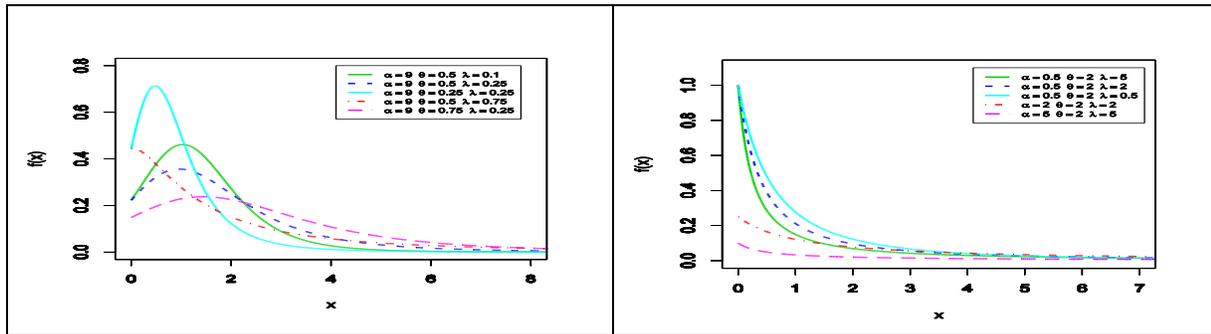


Figure 1: Plots of density of MOGP

### 2.2 Hazard Rate Function

In this subsection, we provide the hazard rate (failure rate) function of MOGP distribution. We also present some graphs of these functions for some values of parameters that illustrates their properties. Hazard rate for GP is a decreasing function for all  $x > 0$ , which is given by:

$$h(x; \theta, \lambda) = \frac{f(x)}{1 - F(x)} = \frac{1}{\lambda \left(1 + \frac{\theta x}{\lambda}\right)},$$

while MOGP distribution has a hazard rate function which takes different shapes. The hazard rate for MOGP is given by the following equation:

$$h^*(x; \alpha, \theta, \lambda) = \frac{g(x)}{1 - G(x)} = \frac{1}{\lambda \left(1 + \frac{\theta x}{\lambda}\right) \left(1 - \bar{\alpha} \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}}\right)}.$$

In order to determine the shape of the hazard rate function  $h^*(x; \alpha, \theta, \lambda)$  it is enough to determine the shape of  $\log h^*(x; \alpha, \theta, \lambda)$ . Hence,

$$\log h^*(x; \alpha, \theta, \lambda) = -\log \lambda - \log \left(1 + \frac{\theta x}{\lambda}\right) - \log \left(1 - \bar{\alpha} \left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}}\right),$$

and

$$\frac{\partial \log h^*(x; \alpha, \theta, \lambda)}{\partial x} = \frac{S(x)}{\lambda \left(1 + \frac{\theta x}{\lambda}\right) \left(\left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}} - \bar{\alpha}\right)},$$

where  $S(x) = -\theta \left(\left(1 + \frac{\theta x}{\lambda}\right)^{-\frac{1}{\theta}} - \bar{\alpha}\right)$ . Hazard rate may take several shapes according to the following cases:

1. If  $0 < \alpha < 1$ , then the derivative of  $\log h^*(x; \alpha, \theta, \lambda)$  is negative and hence the hazard rate function is decreasing.
2. If  $\alpha > 1$ , and  $\theta < \frac{-\alpha}{\bar{\alpha}}$ , then  $S(x)$  has one root  $x_0 = \frac{\lambda}{\theta} \left[\left(\bar{\alpha} \left(1 - \frac{1}{\theta}\right)^\theta - 1\right)\right]$ , so the hazard rate function is increasing on  $(0, x_0)$  and decreasing on  $(x_0, \infty)$ , with  $h(0)=1/(\alpha\lambda)$ , and  $h(\infty)=0$ .
3. If  $\alpha > 1$ , and  $\theta > \frac{-\alpha}{\bar{\alpha}}$ , we obtain a decreasing hazard rate function again.

Figure (2) gives the graphs of hazard function for different parameter values that include the above three cases.

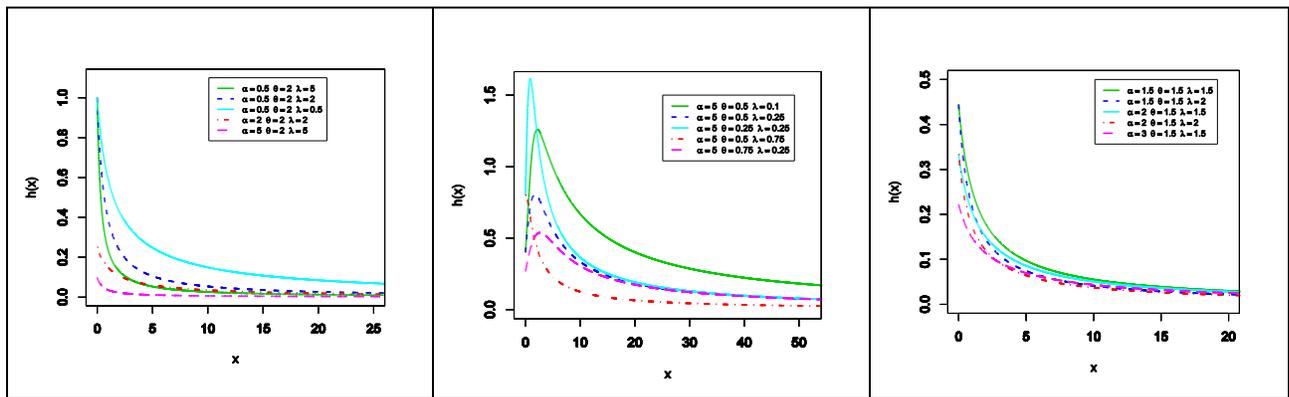


Figure 2: Plots of hazard rate function for different values of  $\alpha$ ,  $\theta$  and  $\lambda$ .

### 1. Classical Point Estimation Methods

In this section we consider different methods of point estimation for the MOGP parameters. Numerical techniques are helpful in obtaining the estimated values of these parameters, then simulation is used to compare between these estimation methods in order to decide which method is more efficient.

#### 3.1 Maximum Likelihood Estimation

The maximum likelihood estimation (MLE) is used in inferential statistics since it has many attractive properties, such as invariance, consistency, and normal approximation properties. It depends basically on maximizing the likelihood function of MOGP distribution. Let  $X_1, X_2, \dots, X_n$  be a random sample from MOGP distribution, then the log likelihood function for the vector of parameters  $\gamma=(\alpha, \theta, \lambda)$  can be expressed by

$$\log \ell(\gamma) = n \log \frac{\alpha}{\lambda} - \left(\frac{1}{\theta} + 1\right) \sum_{i=1}^n \log \left(1 + \frac{\theta x_i}{\lambda}\right) - 2 \sum_{i=1}^n \log \left(1 - \bar{\alpha} \left(1 + \frac{\theta x_i}{\lambda}\right)^{-\frac{1}{\theta}}\right),$$

$$\frac{\partial \log \ell(\gamma)}{\partial \alpha} = \frac{n}{\alpha} - 2 \sum_{i=1}^n \frac{u_i}{1 - \bar{\alpha} u_i},$$

$$\frac{\partial \log \ell(\gamma)}{\partial \theta} = -\frac{1}{\theta} \sum_{i=1}^n \log u_i - \frac{1}{\theta} \left(\frac{1}{\theta} + 1\right) \sum_{i=1}^n (1 - u_i^\theta) - \frac{2\bar{\alpha}}{\theta} \sum_{i=1}^n \left(\frac{\log u_i + \frac{\lambda}{\theta} (1 - u_i^\theta)}{u_i^{-1} - \bar{\alpha}}\right),$$

$$\frac{\partial \log \ell(\gamma)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda} \left(\frac{1}{\theta} + 1\right) \sum_{i=1}^n (1 - u_i^\theta) + \frac{2\bar{\alpha}}{\lambda\theta} \sum_{i=1}^n \left(\frac{u_i - u_i^{\theta+1}}{1 - \bar{\alpha}u_i}\right),$$

where  $u_i = \left(1 + \frac{\theta x_i}{\lambda}\right)^{-1/\theta}$ .

The above equations are not in closed form, therefore we must use any iterative procedure techniques such as conjugate-gradient algorithms, in order to get the needed numerical solution.

### 3.2 Maximum Product Spacing

According to Cheng and Amin (1983) Maximum Product Spacing method (MPS) was introduced as follows:

$$G = \left(\prod_{i=1}^{n+1} D_i\right)^{\frac{1}{n+1}},$$

where G is defined as the geometric mean of the product spacing function  $D_i$  such that

$$\begin{aligned} D_1 &= F(x_1), \\ D_i &= F(x_i) - F(x_{i-1}); i = 2, \dots, n, \\ D_{n+1} &= 1 - F(x_n). \end{aligned}$$

It is easy to see that  $\sum_{i=1}^{n+1} D_i = 1$ . For more applications of MPS see sing et al. (2014), Almetwally and Almongy (2019 a) and El-Sherpieny et al. (2020) . MPS method is based on the observed ordered sample  $x_1 < \dots < x_n$  from MOGP distribution, then the product spacing function is

$$G(\gamma) = \left\{ \frac{1 - \left(1 + \frac{\theta x_1}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \bar{\alpha} \left(1 + \frac{\theta x_1}{\lambda}\right)^{-\frac{1}{\theta}}} \left( \frac{\alpha \left(1 + \frac{\theta x_n}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \bar{\alpha} \left(1 + \frac{\theta x_n}{\lambda}\right)^{-\frac{1}{\theta}}} \right) \prod_{i=1}^n \left[ \frac{1 - \left(1 + \frac{\theta x_i}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \bar{\alpha} \left(1 + \frac{\theta x_i}{\lambda}\right)^{-\frac{1}{\theta}}} - \frac{1 - \left(1 + \frac{\theta x_{i-1}}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \bar{\alpha} \left(1 + \frac{\theta x_{i-1}}{\lambda}\right)^{-\frac{1}{\theta}}} \right] \right\}^{\frac{1}{n+1}}.$$

The natural logarithm of the product spacing function is

$$\begin{aligned} \ln G(\gamma) &= \frac{1}{n+1} \left\{ \ln \left( 1 - \left(1 + \frac{\theta x_1}{\lambda}\right)^{-\frac{1}{\theta}} \right) - \ln \left( 1 - \bar{\alpha} \left(1 + \frac{\theta x_1}{\lambda}\right)^{-\frac{1}{\theta}} \right) + \ln \left( \alpha \left(1 + \frac{\theta x_n}{\lambda}\right)^{-\frac{1}{\theta}} \right) - \ln \left( 1 - \right. \right. \\ &\quad \left. \left. \bar{\alpha} \left(1 + \frac{\theta x_n}{\lambda}\right)^{-\frac{1}{\theta}} \right) + \sum_{i=2}^n \ln \left[ \frac{1 - \left(1 + \frac{\theta x_i}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \bar{\alpha} \left(1 + \frac{\theta x_i}{\lambda}\right)^{-\frac{1}{\theta}}} - \frac{1 - \left(1 + \frac{\theta x_{i-1}}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \bar{\alpha} \left(1 + \frac{\theta x_{i-1}}{\lambda}\right)^{-\frac{1}{\theta}}} \right] \right\}. \end{aligned} \tag{4}$$

To obtain the normal equations for the unknown parameters, we differentiate partially Eq. (4) with respect to the vector parameter  $\gamma$  and equate them to zero, see Appendix A. The estimators of  $\gamma$  can be obtained by solving the system of nonlinear equations, so the MPS of  $\alpha$ ,  $\theta$  and  $\lambda$  can be found by using any iterative procedure technique such as conjugate-gradient algorithms solution.

### 3.3 Least Square Method

Swain et al. (1988) introduced the least square estimators method, it based on the ordered sample  $x_1 < \dots < x_n$  from MOGP distribution, the least squares method are obtained by minimizing

$$LS(\gamma) = \sum_{i=1}^n \left( \frac{1 - \left(1 + \frac{\theta x_i}{\lambda}\right)^{-\frac{1}{\theta}}}{1 - \bar{\alpha} \left(1 + \frac{\theta x_i}{\lambda}\right)^{-\frac{1}{\theta}}} - \frac{i}{n+1} \right)^2. \tag{5}$$

After differentiating equation (5) with respect to the parameters  $\alpha$ ,  $\theta$  and  $\lambda$  and then equating them to zero we get the following normal equation:

$$\begin{aligned} \frac{\partial \ln LS(\gamma)}{\partial \alpha} &= -2 \sum_{i=1}^n \left( \phi(x_i, \gamma) - \frac{i}{n+1} \right) \phi(x_i, \gamma) \phi_1(x_i, \gamma), \\ \frac{\partial \ln LS(\gamma)}{\partial \theta} &= 2 \sum_{i=1}^n \left( \phi(x_i, \gamma) - \frac{i}{n+1} \right) (d(x_i, \gamma) - c(x_i, \gamma)), \end{aligned}$$

and

$$\frac{\partial \ln LS(\gamma)}{\partial \lambda} = 2 \sum_{i=1}^n \left( \phi(x_i, \gamma) - \frac{i}{n+1} \right) \left( \frac{x_i \phi_1(x_i, \gamma) (\bar{\alpha} \phi(x_i, \gamma) - 1)}{\lambda^2 \left(1 + \frac{\theta x_i}{\lambda}\right)} \right).$$

The above nonlinear equations can't be solved analytically, so the least square estimates of  $\alpha$ ,  $\theta$  and  $\lambda$  can be used by numerical technique such as conjugate-gradient algorithms.

### 3.4 Weighted Least Squares Method

Swain et al. (1988) introduced the weighted least square estimators. We use the WLS procedure for estimating the parameters  $\alpha$ ,  $\theta$  and  $\lambda$  of the MOGP distribution. For the weighted least square estimators of the unknown parameters we need to minimize

$$WLS(\gamma) = \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left( \frac{1-(1+\frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}}}{1-\bar{\alpha}(1+\frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}}} - \frac{i}{n+1} \right)^2, \tag{6}$$

with respect to  $\alpha, \theta$  and  $\lambda$ .

After differentiating Eq. (6) with respect to parameters  $\alpha$ ,  $\theta$  and  $\lambda$  and equating them to zero we get the normal equations as follows:

$$\begin{aligned} \frac{\partial WLS(\gamma)}{\partial \alpha} &= -2 \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left( \phi(x_i, \gamma) - \frac{i}{n+1} \right) \phi(x_i, \gamma) \phi_1(x_i, \gamma), \\ \frac{\partial WLS(\gamma)}{\partial \theta} &= 2 \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left( \phi(x_i, \gamma) - \frac{i}{n+1} \right) (d(x_i, \gamma) - c(x_i, \gamma)), \end{aligned}$$

and

$$\frac{\partial WLS(\gamma)}{\partial \lambda} = 2 \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left( \phi(x_i, \gamma) - \frac{i}{n+1} \right) \left( \frac{x_i \phi_1(x_i, \gamma) (\bar{\alpha} \phi(x_i, \gamma) - 1)}{\lambda^2 \left( 1 + \frac{\theta x_i}{\lambda} \right)} \right).$$

The above nonlinear equations can't be solved analytically, so the WLS of  $\alpha$ ,  $\theta$  and  $\lambda$ , can be obtained by the use of conjugate-gradient algorithm.

## 2. Bayesian Estimation Method

Here we find Bayes estimates for the unknown parameters  $\alpha$ ,  $\theta$  and  $\lambda$ . Two different loss functions are assumed, square error (SE) and Linex loss functions. In Bayesian method all parameters are random variables with certain distribution called prior distribution. If prior information is not available which is usually the case, we need to select a prior distribution. Since the selection of prior distribution plays an important role in estimation of the parameters, our choice for the prior of  $\alpha, \theta$  and  $\lambda$  are the independent gamma distributions i.e.  $G(a_1, b_1), G(a_2, b_2)$  and  $G(a_3, b_3)$  respectively. Thus the suggested prior for  $\alpha$ ,  $\theta$  and  $\lambda$  are

$$f_1(\alpha) \propto \alpha^{a_1-1} e^{-b_1 \alpha}, \quad f_2(\theta) \propto \theta^{a_2-1} e^{-b_2 \theta}, \quad f_3(\lambda) \propto \lambda^{a_3-1} e^{-b_3 \lambda},$$

respectively, where  $a_1, a_2, a_3, b_1, b_2$  and  $b_3$  are the hyper parameters of prior distributions.

The joint prior of  $\alpha$ ,  $\theta$  and  $\lambda$  is

$$k(\alpha, \theta, \lambda) \propto \alpha^{a_1-1} \theta^{a_2-1} \lambda^{a_3-1} e^{-b_1 \alpha - b_2 \theta - b_3 \lambda}, \quad \alpha, \theta, \lambda, a_1, a_2, a_3, b_1, b_2, b_3 > 0,$$

while the joint posterior of  $\alpha$ ,  $\theta$  and  $\lambda$  is given by

$$p(\alpha, \theta, \lambda | \underline{x}) \propto L(\underline{x} | \alpha, \theta, \lambda) k(\alpha, \theta, \lambda),$$

where  $L(\underline{x} | \alpha, \theta, \lambda)$  is the likelihood function of MOGP distribution. Substituting  $L(\underline{x} | \alpha, \theta, \lambda)$  and  $k(\alpha, \theta, \lambda)$  for MOGP distribution, the joint posterior will be:

$$p(\alpha, \theta, \lambda | \underline{x}) \propto \alpha^{n+a_1-1} \theta^{a_2-1} \lambda^{a_3-n-1} e^{-b_1 \alpha - b_2 \theta - b_3 \lambda} \prod_{i=1}^n \left( 1 + \frac{\theta x_i}{\lambda} \right)^{-\frac{1}{\theta}-1} \left( 1 - \bar{\alpha} \left( 1 + \frac{\theta x_i}{\lambda} \right)^{-\frac{1}{\theta}} \right)^{-2}.$$

In the case of quadratic loss function Bayes estimate is the posterior mean, the determination of posterior mean for the purpose of obtaining Bayes estimation of the parameters  $\alpha$ ,  $\theta$  and  $\lambda$ , is not easy to obtain unless we use numerical approximation methods.

In literature there are several approximation methods available to solve this kind of problem. Here we consider Monte Carlo Markov Chain (MCMC) approximation method, see Karandikar (2006). This approximation method reduces the ratio of integrals into a whole and produces a single numerical result.

The Bayes estimates of the unknown parameters  $\alpha$ ,  $\theta$  and  $\lambda$  under Linex Loss function can be calculated through the following equation:

$$\gamma_j = \frac{-1}{v} \ln \left( \frac{\sum_{i=1}^L e^{-v\gamma^{(i)}}}{L} \right),$$

where  $v$  reflects the direction and degree of asymmetry,  $L$  is number of periods in the MCMC process, the readers may refer to Almetwally and Almongy (2018) for more details.

A wide variety of MCMC schemes are available. An important sub-class of MCMC methods are Gibbs sampling and more general Metropolis within Gibbs samplers. This is often unavailable in MLE. Indeed, the MCMC samples may

be used to completely summarize the posterior uncertainty about the parameters  $\alpha$ ,  $\theta$  and  $\lambda$ , through a kernel estimate of the posterior distribution. This is also true of any function of the parameters.

Therefore, to generate samples from MOGP distribution, we use the Metropolis-Hastings method (Metropolis et al. (1953) with normal proposal distribution). For details regarding the implementation of Metropolis-Hastings algorithm, the readers may refer to Robert and Casella (2004) and Almetwally et al. (2018).

### 3. Simulation Study and Data Analysis

In this section, we provide a complete algorithm of Monte Carlo simulation (MCS) study, in order to compare between the different estimation methods proposed in the previous sections, then we apply these computations on a real life example.

#### 5.1 Simulation Study

We explain our algorithm through an application in estimation methods framework, especially; we use Monte Carlo technique to compare between MLE, MPS, LS, WLS estimation methods and Bayesian estimation under different loss functions based on complete data for estimating MOGP distribution in life time by R language.

We start by building our model with generate all simulation controls. In this stage, we must do the following steps by order:

Step 1: Suppose the following values for the parameter vector of MOGP  $\gamma=(\alpha,\theta, \lambda)$ , case 1=(0.5, 0.5, 1), case 2=(0.5, 1.5, 1), case 3=(0.5,1.5,1.5), case 4=(0.5,1.5, 0.75), case 5=(0.9,1.5,1.5) and case 6=(1.2,1.5,1.5).

Step 2: Choose sample sizes  $n=50,100$  and  $150$ .

Step 3: Generate the sample random values of MOGP distribution by using quantile function  $X = \frac{\lambda}{\theta} \left( \frac{1-U}{1-U+U\alpha} \right)^{-\theta} - 1$ , where  $U$  is a uniform distribution  $(0, 1)$ .

Step 4: Solve differential equations for each estimation methods, to obtain the estimators of the parameters for MOGP distribution, so we calculate  $\alpha$ ,  $\theta$ , and  $\lambda$ .

Step 5: Repeat this experiment  $(L-1)$  times. In each experiment use the same values of the parameters. It is certain that, the values of generating random are varying from experiment to experiment even though sample size  $(n)$  does not change.

Finally, we have  $L$ -values of bias and MSE, We compute the average biases and average MSE's over 10,000 runs. This number of runs will give the accuracy in the order  $\pm 0.01$  (see Karian and Dudewicz (1999)). Therefore, we report all the results up to three decimal places. Remember that Bias estimator is  $\text{Bias}=\hat{\gamma} - \gamma$ , where  $\gamma$  is the estimated value of  $\gamma$ , and the mean squared error (MSE) of the estimator is  $\text{MSE}=\text{Mean}(\hat{\gamma} - \gamma)^2$ .

The results are presented in six different tables where the MSE is given in each cell and the corresponding bias is reported within parenthesis.

It is observed from Tables 1 to 6 that all of the estimators usually over estimate  $\alpha$  for small values of  $\alpha$ . For  $\alpha > 1$ , most of the estimators seems to be underestimates  $\alpha$ . One can also observe that for each estimation method, the MSE's decreases as the sample size increases.

It is observed from Tables 1 to 6 that all of the estimators usually overestimate  $\lambda$  for all values of  $\lambda$  for small sample size except the Bayes estimate in table 5 and 6 for Linex 0.75 and Linex 1.5 which are usually underestimate where  $\theta > 1$ . For large sample size the estimator of  $\lambda$  tends to be underestimate especially for WLS as shown in Table 5. Also in Table 6 we observe that  $\lambda$  tends to be underestimate for WLS and Bayes estimate under SE, Linex 0.75 and Linex 1.5. It is also observed that all estimates decrease as the sample size increases.

One can observe that for each estimation method, the MSE's decreases as the sample size increases and also as the value of  $\alpha$  increases, but the MSE increases as the value  $\theta$  increases.

It is observed from Tables 1 and 2 that the MLE underestimate, while in Table 4 for large value of sample size we realize that the estimators MLE, LS and WLS under estimate. Notice that in Table 6 where  $\alpha > 1$  most estimators under estimate, but the MPS method overestimates  $\theta$ .

Among all classical methods of estimation, the best method to estimate  $\alpha$  if  $\alpha < 1$  is the MPS method while for  $\alpha > 1$  the best method is the MLE with respect to the bias. Regarding  $\lambda$  the best estimation method for small sample size is the LS method, while for large samples the WLS method will serve better. For estimating  $\theta$  all classical methods are superior with different values of  $\lambda$  and  $\theta$  and for different sample sizes except the MPS which seems not appropriate to estimate  $\theta$ .

Now for non-classical Bayes point estimation method the best loss function for estimating  $\alpha$ ,  $\lambda$  and  $\theta$  is Linex 1.5 for most cases where  $\alpha < 1$ , and we realize that the square error works better for  $\alpha > 1$ .

Table 1: Bias and MSE for  $\alpha, \lambda$ , and  $\theta$

$\alpha=0.5 ; \lambda=0.5 ; \theta=1$															
		MLE		LS		MPS		WLS		SE		Linex 0.75		Linex 1.5	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	$\alpha$	0.43	0.44	0.43	0.22	0.36	0.23	0.44	0.25	0.45	0.31	0.41	0.26	0.37	0.21
	$\lambda$	0.60	0.83	0.31	0.13	0.43	0.31	0.33	0.15	0.43	0.31	0.39	0.25	0.35	0.21
	$\theta$	-0.07	0.15	0.06	0.15	0.16	0.16	0.03	0.12	0.09	0.11	0.06	0.10	0.03	0.09
100	$\alpha$	0.43	0.33	0.39	0.16	0.36	0.17	0.39	0.17	0.25	0.08	0.25	0.08	0.24	0.07
	$\lambda$	0.45	0.37	0.35	0.13	0.39	0.20	0.35	0.14	0.24	0.07	0.23	0.07	0.22	0.06
	$\theta$	-0.02	0.06	0.04	0.08	0.10	0.07	0.02	0.06	0.07	0.03	0.06	0.03	0.06	0.02
150	$\alpha$	0.41	0.24	0.39	0.16	0.36	0.15	0.40	0.17	0.21	0.05	0.20	0.05	0.20	0.05
	$\lambda$	0.42	0.26	0.36	0.13	0.38	0.16	0.36	0.14	0.20	0.05	0.20	0.05	0.20	0.04
	$\theta$	-0.02	0.03	0.02	0.04	0.05	0.03	0.00	0.03	0.07	0.01	0.06	0.01	0.06	0.01

Table 2: Bias and MSE for  $\alpha, \lambda$  and  $\theta$

$\alpha=0.5 ; \lambda=1.5 ; \theta=1$															
		MLE		LS		MPS		WLS		SE		Linex 0.75		Linex 1.5	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	$\alpha$	0.92	1.32	0.80	0.72	0.72	0.60	0.88	0.93	0.71	0.64	0.65	0.53	0.59	0.43
	$\lambda$	0.52	1.18	0.24	0.11	0.30	0.17	0.21	0.20	0.25	0.29	0.20	0.24	0.14	0.20
	$\theta$	-0.04	0.09	0.06	0.12	0.13	0.11	0.03	0.09	0.14	0.12	0.11	0.11	0.08	0.09
100	$\alpha$	0.82	0.83	0.78	0.66	0.74	0.59	0.78	0.67	0.33	0.12	0.32	0.11	0.31	0.10
	$\lambda$	0.38	0.45	0.27	0.11	0.28	0.11	0.28	0.12	0.12	0.04	0.12	0.04	0.11	0.03
	$\theta$	-0.01	0.04	0.04	0.06	0.08	0.05	0.02	0.05	0.10	0.03	0.10	0.03	0.09	0.03
150	$\alpha$	0.78	0.73	0.78	0.64	0.76	0.60	0.89	0.88	0.25	0.07	0.25	0.07	0.24	0.06
	$\lambda$	0.43	0.54	0.28	0.09	0.28	0.09	0.20	0.14	0.12	0.02	0.11	0.02	0.11	0.02
	$\theta$	-0.02	0.02	0.01	0.03	0.04	0.02	0.00	0.02	0.08	0.02	0.08	0.02	0.08	0.01

Table 3: Bias and MSE for  $\alpha, \lambda$  and  $\theta$

$\alpha=0.5 ; \lambda=1.5 ; \theta=1.5$															
		MLE		LS		MPS		WLS		SE		Linex 0.75		Linex 1.5	
		Bias	MSE	Bias	MSE	Bias	MSE								
50	$\alpha$	1.27	2.71	0.83	0.78	0.77	0.74	0.88	0.96	0.75	0.71	0.70	0.60	0.64	0.50
	$\lambda$	0.47	1.90	0.22	0.15	0.32	0.36	0.28	0.44	0.20	0.31	0.14	0.27	0.09	0.24
	$\theta$	0.01	0.14	0.10	0.17	0.20	0.21	0.06	0.14	0.13	0.16	0.09	0.14	0.05	0.12
100	$\alpha$	1.18	2.14	0.84	0.81	0.77	0.68	0.94	1.17	0.34	0.13	0.33	0.12	0.31	0.11
	$\lambda$	0.30	1.09	0.23	0.14	0.29	0.21	0.26	0.47	0.11	0.04	0.10	0.04	0.10	0.04
	$\theta$	0.02	0.07	0.07	0.09	0.14	0.10	0.05	0.08	0.08	0.03	0.08	0.03	0.07	0.03
150	$\alpha$	0.96	1.21	0.86	0.78	0.79	0.67	0.93	1.00	0.26	0.07	0.25	0.07	0.24	0.06
	$\lambda$	0.26	0.42	0.20	0.08	0.28	0.14	0.20	0.23	0.11	0.02	0.10	0.02	0.10	0.02
	$\theta$	0.01	0.03	0.04	0.04	0.08	0.04	0.03	0.03	0.08	0.01	0.08	0.01	0.08	0.01

Table 4: Bias and MSE for  $\alpha, \lambda$  and  $\theta$

$\alpha=0.5 ; \lambda=1.5 ; \theta=0.75$															
		MLE		LS		MPS		WLS		SE		Linex 0.75		Linex 1.5	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	$\alpha$	0.84	1.17	0.80	0.75	0.74	0.66	0.86	0.90	0.66	0.55	0.61	0.47	0.56	0.39
	$\lambda$	0.64	1.59	0.24	0.11	0.26	0.15	0.23	0.17	0.28	0.31	0.22	0.25	0.16	0.21
	$\theta$	-0.09	0.08	0.04	0.09	0.09	0.07	0.01	0.07	0.16	0.08	0.13	0.07	0.10	0.06
100	$\alpha$	0.78	0.80	0.78	0.67	0.76	0.64	0.80	0.74	0.33	0.12	0.32	0.12	0.31	0.11
	$\lambda$	0.43	0.45	0.28	0.10	0.25	0.10	0.29	0.16	0.13	0.04	0.12	0.04	0.12	0.04
	$\theta$	-0.04	0.03	0.01	0.05	0.05	0.03	0.00	0.03	0.10	0.03	0.10	0.03	0.09	0.03
150	$\alpha$	0.77	0.75	0.76	0.60	0.78	0.63	0.78	0.65	0.26	0.07	0.25	0.07	0.24	0.06
	$\lambda$	0.54	0.40	0.30	0.10	0.25	0.09	0.29	0.10	0.13	0.02	0.12	0.02	0.12	0.02
	$\theta$	-0.04	0.02	-0.01	0.02	0.02	0.01	-0.02	0.02	0.08	0.01	0.08	0.01	0.08	0.01

Table 5: Bias and MSE for  $\alpha, \lambda$  and  $\theta$

$\alpha=0.90 ; \lambda=1.5 ; \theta=1.5$															
		MLE		LS		MPS		WLS		SE		Linex 0.75		Linex 1.5	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	$\alpha$	0.59	1.22	0.20	0.10	0.14	0.15	0.28	0.25	0.28	0.23	0.23	0.18	0.19	0.14
	$\lambda$	0.23	1.41	0.02	0.11	0.16	0.31	0.03	0.25	0.04	0.24	-0.02	0.21	-0.07	0.20
	$\theta$	-0.02	0.16	0.07	0.19	0.20	0.23	0.03	0.15	0.01	0.13	-0.03	0.13	-0.06	0.12
100	$\alpha$	0.52	0.83	0.20	0.09	0.14	0.09	0.22	0.16	0.07	0.03	0.07	0.02	0.06	0.02
	$\lambda$	0.06	0.75	0.04	0.08	0.13	0.16	0.06	0.16	0.03	0.03	0.02	0.03	0.02	0.03
	$\theta$	0.00	0.08	0.05	0.10	0.13	0.11	0.03	0.09	0.00	0.02	-0.01	0.02	-0.01	0.02
150	$\alpha$	0.32	0.43	0.17	0.04	0.14	0.06	0.34	0.32	0.04	0.01	0.04	0.01	0.04	0.01
	$\lambda$	0.18	0.64	0.07	0.02	0.12	0.08	-0.01	0.25	0.02	0.01	0.02	0.01	0.01	0.01
	$\theta$	0.01	0.04	0.03	0.05	0.07	0.04	0.00	0.04	0.02	0.01	0.02	0.01	0.02	0.01

Table 6: Bias and MSE for  $\alpha, \lambda$  and  $\theta$

$\alpha=1.2 ; \lambda=1.5 ; \theta=1.5$															
		MLE		LS		MPS		WLS		SE		Linex 0.75		Linex 1.5	
		Bias	MSE	Bias	MSE	Bias	MSE								
50	$\alpha$	0.03	0.59	-0.21	0.11	-0.29	0.22	-0.19	0.15	-0.06	0.17	-0.11	0.15	-0.15	0.14
	$\lambda$	-0.04	0.88	-0.27	0.16	-0.09	0.29	-0.25	0.20	-0.13	0.28	-0.19	0.27	-0.23	0.27
	$\theta$	-0.05	0.18	0.02	0.19	0.19	0.25	-0.01	0.16	-0.05	0.12	-0.09	0.12	-0.12	0.12
100	$\alpha$	-0.03	0.42	-0.22	0.09	-0.30	0.16	-0.20	0.12	-0.11	0.04	-0.11	0.04	-0.12	0.04
	$\lambda$	-0.17	0.55	-0.25	0.12	-0.12	0.17	-0.23	0.18	-0.08	0.03	-0.09	0.04	-0.10	0.04
	$\theta$	-0.03	0.09	0.00	0.11	0.11	0.11	-0.01	0.09	-0.05	0.03	-0.06	0.03	-0.06	0.03
150	$\alpha$	-0.06	0.42	-0.25	0.09	-0.30	0.13	-0.24	0.09	-0.08	0.02	-0.08	0.02	-0.09	0.02
	$\lambda$	-0.06	0.52	-0.20	0.08	-0.13	0.09	-0.21	0.09	-0.08	0.02	-0.08	0.02	-0.08	0.02
	$\theta$	-0.02	0.05	-0.02	0.05	0.05	0.04	-0.03	0.04	-0.04	0.01	-0.04	0.01	-0.04	0.01

### 5.2 Data Analysis

We present the numerical results of the parameter estimation of MOGP distribution of a real data. This real data are mentioned by Birnbaum and Saunders (1969) which represent the data set of failure times, these refers to the fatigue times of 6061-T6 aluminum coupons. The data set consists of 101 observations with maximum stress per cycle 26,000 psi. the data are given below: Data Set: 370, 706, 716, 746, 785, 797, 844, 855, 858, 886, 886, 930, 960, 988, 990, 1000, 1010, 1016, 1018, 1020, 1055, 1085, 1102, 1102, 1108, 1115, 1120, 1134, 1140, 1199, 1200, 1200, 1203, 1222, 1235, 1238, 1252, 1258, 1262, 1269, 1270, 1290, 1293, 1300, 1310, 1313, 1315, 1330, 1355, 1390, 1416, 1419, 1420, 1420, 1450, 1452, 1475, 1478, 1481, 1485, 1502, 1505, 1513, 1522, 1522, 1530, 1540, 1560, 1567, 1578, 1594, 1602, 1604, 1608, 1630, 1642, 1674, 1730, 1750, 1750, 1763, 1768, 1781, 1782, 1792, 1820, 1868, 1881, 1890, 1893, 1895, 1910, 1923, 1940, 1945, 2023, 2100, 2130, 2215, 2268, 2440.

The MLEs estimates of the parameters with the corresponding standard errors (SE) given in parenthesis are reported in Table (7), MOGP model is compared with other special case models like GP, Pareto type 1 and MOP. This comparison was conducted using Kolmogorove-Smerнове (KS) test with the corresponding p-values. Also Akaike information criterion (AIC) such that  $AIC = -2 L(\gamma) + 2p$ , where p is the number of parameters in the model and L is the maximized value of the likelihood function for the model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Bayesian information criterion (BIC) is also used as a criteria for comparison between models where BIC can be defined as:  $BIC = -2 L(\gamma) + p \ln(n)$ , where n is the sample size. As a model selection criterion, the researcher should choose the model with the minimum BIC value. The last criterion is the consistent Akaike information criteria (CAIC), which is defined by  $CAIC = -2 L(\gamma) + p(\ln n + 1)$ . As a model selection criterion, the researcher should choose the model with the minimum CAIC value. The MLEs of  $\alpha, \lambda, and \theta$  are computed numerically using the function optimal in R statistical package. The values of the KS statistic with p-values, AIC, CAIC, BIC and HQIC are reported in Table (7). When comparing the values of KS between MOGP and other sub models like GP, Pareto type 1 and MOP we obtain the minimum KS for MOGP which equals 0.054 with related p-value of 0.927. Therefore, this indicates that the MOGP distribution fits the data set well and better than other distributions. This is also indicates the needs of new distributions in managing some sets of data. The same result is found using AIC, CAIC and HQIC criteria since the smallest of these values is noticed for MOGP amongst other distribution. While the minimum value of BIC is indicated to MOP with small difference between it and BIC of MOGP. So in general we can tell that the new distribution is superior according to other sub models.

We have estimated parameters of MOGP distribution and compute SE under different estimation methods as shown in Table (8). According to the above real data we realize that WLS estimate better for  $\alpha$  and  $\lambda$ , while Linex 0.75 is better to be used for estimating  $\theta$

Table 7: MLEs, KS p-values and different model goodness of fit criterion for failure times

	MOGP	GP	pareto type I	MOP
	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)
$\alpha$	750.8335 (1.494)			8131.9573 (0.2403)
$\theta$	194.15938 (19.428)	102.0144 (228.55)	584.269943 (264.156)	298.3765 (38.7808)
$\lambda$	0.02323 (0.025)	507.8105 (1.54665)	1.060191 (2.5656)	5.91657 (0.4880)
D	0.054	0.9816	0.63304	0.0572
P-value	0.927	2.20E-16	2.20E-16	0.896
AIC	1500.158	2720.843	1625.145	1504.308
CAIC	1500.405	2720.966	1625.267	1512.154
BIC	1508.003	2726.073	1630.375	1507.484
HQIC	1503.334	2722.961	1627.262	1504.556

Table 8: MLEs, LS, WLS, SE and Linex for real data

	MLE	LS	WLS	SE	Linex $v=1.5$	Linex $v=0.75$
	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)
$\alpha$	750.8335(1.4943)	628.541602 (8.4635)	619.26e (0.3105)	751.08 (0.675)	750.97 (0.579)	750.925 (0.539)
$\theta$	194.15938(19.4283)	186.951415 (8.1255)	195.79 (3.0832)	195.042 (0.524)	194.972 (0.4896)	194.938 (0.476)
$\lambda$	0.02323(0.0255)	0.022744 (0.0341)	0.030172 (0.0044)	0.023 (0.009)	0.0233 (0.007)	0.0232 (0.007)

The MPS method is the difference between the observations of the cumulative function, so this method cannot be used for that data because of similar values in it.

#### 4. Conclusions

In this study we considered a generalization of GP distribution which is MOGP this new distribution proved to be more flexible and more appropriate for monotone life time data. The pdf of MOGP is monotonic depending on the values of parameters. Also we realized that the hazard rate is increasing-decreasing function which can be widely used to model different real data. We studied different classical and non-classical point estimation methods and by numerical methods we conducted a comparison between these methods. Between classical methods we concluded that the best method for estimating  $\alpha$  is the MPS method and for estimating  $\lambda$  are LS and WLS depending on sample size while for estimating  $\theta$  we found that MLE and WLS are most appropriate methods. Our choice of best estimating method was with respect to minimum value of absolute bias. Regarding Bayes methods Linex loss function was better than square error loss function for estimating all parameters of MOGP distribution. The flexibility of this distribution was illustrated in an application to a real data set.

#### References

1. Alice, T. & Jose, K. K. (2004). Marshall-Olkin exponential time series modeling. STARS International Journal 5(1):12-22.
2. Almetwally, E. M., & Almongy, H. M. (2019 a). Maximum Product Spacing and Bayesian Method for Parameter Estimation for Generalized Power Weibull Distribution under Censoring Scheme. Journal of Data Science, 17(2), 407-444.
3. Almetwally, E. M., & Almongy, H. M. (2019 b). Estimation Methods for the New Weibull-Pareto Distribution: Simulation and Application. Journal of Data Science, 17(3), 610-630.
4. Almetwally, E. M., Almongy, H. M., & El sayed Mubarak, A. (2018). Bayesian and Maximum Likelihood Estimation for the Weibull Generalized Exponential Distribution Parameters Using Progressive Censoring Schemes. Pakistan Journal of Statistics and Operation Research, 14(4), 853-868.
5. Almetwally, E. M., & Almongy, H. M. (2018). Estimation of the Generalized Power Weibull Distribution Parameters Using Progressive Censoring Schemes. International Journal of Probability and Statistics, 7(2), 51-61.

6. Basheer, A. M. (2019). Marshall-Olkin alpha power inverse exponential distribution: properties and applications, *Annals of data science*, <https://doi.org/10.1007/s40745-019-00229-0>, (2019).
7. Bdair, O. & Haj Ahmad, H. (2019). Estimation of the Marshall-Olkin Pareto Distribution Parameters: Comparative Study. *REVISTA INVESTIGACION OPERACIONAL*, **41(2)**, forthcoming.
8. Birnbaum, Z. W., & Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, *6(2)*, 328-347.
9. Castillo, E., Hadi, A. S., Balakrishnan, N. & Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Hoboken, NJ: Wiley. MR2191401
10. Cheng, R. C. H. & Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *J. Roy. Statist.Soc. Ser. B*, *45*, 394 - 403.
11. El-Sherpieny, E. S. A., Almetwally, E. M., & Muhammed, H. Z. (2020). Progressive Type-II hybrid censored schemes based on maximum product spacing with application to Power Lomax distribution. *Physica A: Statistical Mechanics and its Applications*, 124251.
12. Ghitany, M.E. (2005) Marshall-Olkin extended Pareto distribution and its application. *International Journal of Applied Mathematics*, *18*, No.1, 17-31.
13. Gupta, R.C., Gupta, R.D. & Gupta, P.L. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics: Theory and Methods*, *27*, 887-904.
14. Haj Ahmad, H., Bdair, O. & Ahsanullah, M. (2017). On Marshall-Olkin Extended Weibull Distribution. *Journal of Statistical Theory and Applications*, Vol. 16, No. 1, 1–17
15. Hogg, R.V., McKean, J.W. & Craig, A.T. (2005). *Introduction to Mathematical Statistics*, 6th ed. Pearson Prentice-Hall, New Jersey, 2005.
16. Jose, K.K. (2011). Marshall-Olkin Family of Distributions and their applications in reliability theory, time series modeling and stress-strength analysis. *int. statistical inst. proc 58th world statistical congress*, Dublin (Session CPS 005).
17. Jose, K. K. & Alice, T. (2001) Marshall-Olkin Generalized Weibull distributions and applications, *STARS: int. Journal*, *2*, 1, 1-8.
18. Jose, K. K. & Alice, T. (2005) *Marshall-Olkin Family of Distributions: Applications in Time series modelling and Reliability*, J.C Publications, Palakkad.
19. Jose, K. K. & Uma, P. (2009) On Marshall-Olkin Mittag-Leffler distributions and processes, *Far East Journal of Theoretical Statistics*, *28*, 189-199.
20. Karian, Z. A. & Dudewicz, E. J. (1999). *Modern Statistical Systems and GPSS Simulations*, 2nd edition, CRC Press, Florida.
21. KARANDIKAR, R.L. (2006). On Markov Chain Monte Carlo (MCMC) Method, *Sadhana*, *31*, 81-104.
22. Kotz, S. & Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.
23. Marshall, A. W. & Olkin, I. (1997). A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families, *Biometrika* *84(3)*, 641-652.
24. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M., Teller, A. H. & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, *21*, 1087; doi: 10.1063/1.1699114
25. Okasha, H. M., El-Baz, A. H. & Basheer A. M., (2020). On Marshall-Olkin extended inverse Weibull distribution: properties and estimation using type-II censoring data, *Journal of Statistics Applications & Probability Letters*, *7(1)*, 9-21.
26. Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, *3*, 119--131.
27. Robert, C.P. & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York. <http://dx.doi.org/10.1007/978-1-4757-4145-2>
28. Sankaran, P. G. & Jayakumar, K. (2006). On proportional odds model, *Statistical Papers* *49*, 779-789.
29. Singh, U., Singh, S. K., & Singh, R. K. (2014). A comparative study of traditional estimation methods and maximum product spacing method in generalized inverted exponential distribution. *Journal of Statistics Applications & Probability*, *3(2)*, 153.
30. Swain, J., Venkatraman, S. & Wilson, J. (1988). Least Squares Estimation of Distribution Function in Johnson's Translation System, *Journal of Statistical Computation and Simulation* *29*, 271-297.

**Appendix A**

$$\frac{\partial \ln G(\gamma)}{\partial \alpha} = -\phi_1(x_1, \gamma) + \frac{1}{\alpha} - \phi_1(x_1, \gamma) + \sum_{i=2}^n \frac{\phi_1(x_{i-1}, \gamma)\tau(x_{i-1}, \gamma) - \phi_1(x_i, \gamma)\tau(x_i, \gamma)}{\tau(x_i, \gamma) - \tau(x_{i-1}, \gamma)},$$

$$\frac{\partial \ln G(\gamma)}{\partial \theta} = a(x_1, \gamma) - b(x_n, \gamma) + \sum_{i=2}^n c(x_i, \gamma) - d(x_i, \gamma) + c(x_{i-1}, \gamma) - d(x_{i-1}, \gamma),$$

and

$$\begin{aligned} & \frac{\partial \ln G(\gamma)}{\partial \lambda} \\ &= \frac{x_1(1 + \frac{\theta x_1}{\lambda})^{-1-\frac{1}{\theta}}}{\lambda^2(1 - \zeta(x_1, \theta, \lambda))} + \frac{\bar{\alpha}x_1\tau(x_1, \gamma)}{\lambda^2(1 + \frac{\theta x_1}{\lambda})} + \frac{x_n + \bar{\alpha}x_n\tau(x_n, \gamma)}{\lambda^2(1 + \frac{\theta x_n}{\lambda})} \\ &+ \sum_{i=2}^n \frac{\frac{x_i\tau(x_i, \gamma)}{\lambda^2(1 + \frac{\theta x_i}{\lambda})} - \frac{x_{i-1}\tau(x_{i-1}, \gamma)}{\lambda^2(1 + \frac{\theta x_{i-1}}{\lambda})} + \frac{\bar{\alpha}x_i\tau(x_i, \gamma)(\zeta(x_i, \theta, \lambda) - 1)}{\lambda^2(1 + \frac{\theta x_i}{\lambda})v(x_i, \gamma)} - \frac{\bar{\alpha}x_{i-1}\tau(x_{i-1}, \gamma)(\zeta(x_{i-1}, \theta, \lambda) - 1)}{\lambda^2(1 + \frac{\theta x_{i-1}}{\lambda})v(x_{i-1}, \gamma)}}{\tau(x_i, \gamma) - \tau(x_{i-1}, \gamma)}, \end{aligned}$$

where

$$\phi_1(x_i, \gamma) = \frac{(1 + \frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}}}{1 - \bar{\alpha}(1 + \frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}}}, \zeta(x_i, \theta, \lambda) = (1 + \frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}}, v(x_i, \gamma) = 1 - \bar{\alpha}(1 + \frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}},$$

$$\tau(x_i, \gamma) = \frac{1 - (1 + \frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}}}{1 - \bar{\alpha}(1 + \frac{\theta x_i}{\lambda})^{-\frac{1}{\theta}}}, a(x_i, \gamma) = \frac{1}{\zeta(x_i, \theta, \lambda) - 1} \left( \frac{\zeta(x_i, \theta, \lambda) \ln(1 + \frac{\theta x_i}{\lambda})}{\theta^2} - \frac{x_i \zeta(x_i, \theta, \lambda)}{\theta \lambda (1 + \frac{\theta x_i}{\lambda})} \right) - \frac{\bar{\alpha} \left( \frac{\zeta(x_i, \theta, \lambda) \ln(1 + \frac{\theta x_i}{\lambda})}{\theta^2} - \frac{x_i \zeta(x_i, \theta, \lambda)}{\theta \lambda (1 + \frac{\theta x_i}{\lambda})} \right)}{v(x_i, \gamma)},$$

$$b(x_i, \gamma) = \frac{1}{\zeta(x_i, \theta, \lambda)} \left( \frac{\zeta(x_i, \theta, \lambda) \ln(1 + \frac{\theta x_i}{\lambda})}{\theta^2} - \frac{x_i \zeta(x_i, \theta, \lambda)}{\theta \lambda (1 + \frac{\theta x_i}{\lambda})} \right) - \frac{\bar{\alpha} \left( \frac{\zeta(x_i, \theta, \lambda) \ln(1 + \frac{\theta x_i}{\lambda})}{\theta^2} - \frac{x_i \zeta(x_i, \theta, \lambda)}{\theta \lambda (1 + \frac{\theta x_i}{\lambda})} \right)}{v(x_i, \gamma)},$$

$$c(x_i, \gamma) = \frac{1}{v(x_i, \gamma)} \left( \frac{-\zeta(x_i, \theta, \lambda) \ln \zeta(x_i, \theta, \lambda)}{\theta} - \frac{x_i \zeta(x_i, \theta, \lambda)}{\theta \lambda (1 + \frac{\theta x_i}{\lambda})} \right),$$

$$\text{and } d(x_i, \gamma) = \bar{\alpha} \tau(x_i, \gamma) c(x_i, \gamma).$$