

Robust Estimators for Marshal-Olkin Extended Linear Exponential Distribution

Ayman Orabi¹, Dalia Ziedan^{2,*}

¹Department of Management Information Systems Higher Institute for Specific Studies, Egypt

²Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt

Abstract In this paper, the estimation of the parameters for the Marshal-Olkin extended linear exponential (MOLELE) distribution is discussed in the presence of outliers or extreme observations. Three methods are used to estimate the parameters, maximum likelihood, percentile, and M methods. A simulation study is conducted in the presence of outliers to examine the performance of the estimation methods. The results confirmed that, the M-estimation method is a suitable estimation method than the other methods when there are outliers in the data. Also, a real dataset application is carried out to confirm these results.

Keywords Marshall–Olkin Extended Linear Exponential, Maximum Likelihood Estimator, Percentile Estimator, M-estimator, Robust estimator, Outliers

1. Introduction

The linear exponential (LE) distribution is a two-parameter distribution was introduced by Kodlin (1967) as a possible model for response time. The linear exponential distribution has many applications. It has been used by many biometricians, statisticians, mathematicians, medical scientists and others, for example, Broadbent (1958) used it to describe the service of milk bottles that are filled in a dairy, circulated to customers, and returned empty to the dairy. Also, Carbone *et al* (1967) used it to study the survival pattern of patients with plasmacytic myeloma.

In the literature, various methods have been used to generalize linear exponential distribution. Sarhan and Kundu (2009) introduced a generalization of the linear exponential distribution, named as the generalized linear failure rate distribution. Also, Mervoci and Elbatal (2015) introduced a four-parameter generalized version of the linear exponential distribution which is called Kumaraswamy linear exponential distribution.

Marshall and Olkin (1997) introduced a new family of the distributions based on adding a new parameter α , this distribution called Marshall-Olkin extended distribution. Suppose that $\bar{F}(x) = 1 - F(x)$ be the survival function of the baseline distribution. Then, the survival function of the Marshall-Olkin extended distribution can be defined as following

$$\bar{G}(x) = \frac{\alpha \bar{F}(x)}{1 - \bar{\alpha} \bar{F}(x)}, \quad -\infty \leq x \leq \infty \quad (1)$$

Where $\alpha > 0$ is an additional parameter and $\bar{\alpha} = 1 - \alpha$. Hence, the probability distribution (pdf) and the cumulative distribution for the new distribution are given by, respectively

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

$$G(x) = \frac{F(x)}{1 - \bar{\alpha} \bar{F}(x)} = 1 - \frac{\alpha \bar{F}(x)}{1 - \bar{\alpha} \bar{F}(x)} \quad (3)$$

Okasha and Kayid (2016) introduced a new family of Marshall-Olkin extended generalized linear exponential distribution, the unknown parameters are estimated by the maximum likelihood method.

In the presence of outliers in the data, the traditional methods of estimation do not give good results. So, the robust method of estimation can be used to estimate the unknown parameters. Kantar and Yildirim (2015) considered various robust estimators for the extended Burr Type III distribution for complete data with outliers by using different methods of robust estimation. Mousa (2017) used M-estimation as a robust method to estimate the parameters of the Marshall-Olkin extended burr III distribution for complete data with outliers. Almongy and Almetwally (2020) discussed robust estimation for point estimation of the shape and scale parameters for generalized exponential (GE) distribution using a complete dataset in the presence of various percentages of outliers.

* Corresponding author:

dalia_dalia444@yahoo.com (Dalia Ziedan)

Received: Mar. 2, 2021; Accepted: May 7, 2021; Published: May 26, 2021

Published online at <http://journal.sapub.org/statistics>

The rest of this paper is organized as follows: Section (2) is concerned with describing the MOELE distribution. The maximum likelihood, Percentile and M estimators are given in Section (3). A simulation study is carried out in Section (4). Finally, a real data example is given in Section (5).

2. Marshall - Olkin Extended Linear Exponential Distribution

Let X be a random variable that have a linear exponential distribution with shape parameters $c > 0$ and $k > 0$. Then the probability density function is given by

$$f(x) = (c + kx) e^{-\left(cx + \frac{kx^2}{2}\right)}, \quad x \geq 0 \quad (4)$$

And, the cumulative density function is given by

$$F(x) = 1 - e^{-\left(cx + \frac{kx^2}{2}\right)}, \quad x \geq 0 \quad (5)$$

Substituting (4) and (5) in (2) and (3) we obtain a Marshall-Olkin Extended Linear Exponential distribution denoted by MOELE distribution with the following pdf and cdf, respectively

$$g(x) = \frac{\alpha(c + kx) e^{-\left(cx + \frac{kx^2}{2}\right)}}{\left[1 - (1 - \alpha) e^{-\left(cx + \frac{kx^2}{2}\right)}\right]^2}, \quad x \geq 0. \quad (6)$$

In particular for $\alpha = 1$ the original distribution, i.e., the linear exponential distribution is recovered. Also, for $\alpha = 1, c = 0$ the MOELE becomes the Rayleigh distribution. And, for $\alpha = 1, k = 0$, the MOELE becomes exponential distribution.

MOELE distribution is more flexible than the linear exponential distribution, because of the presence of the shape parameter. Figure (1) shows the plots of pdf for MOELE distribution for some values of the parameters.

3. Parameters Estimation

In this section, we consider three procedures to estimate the parameters of the MOELE distribution.

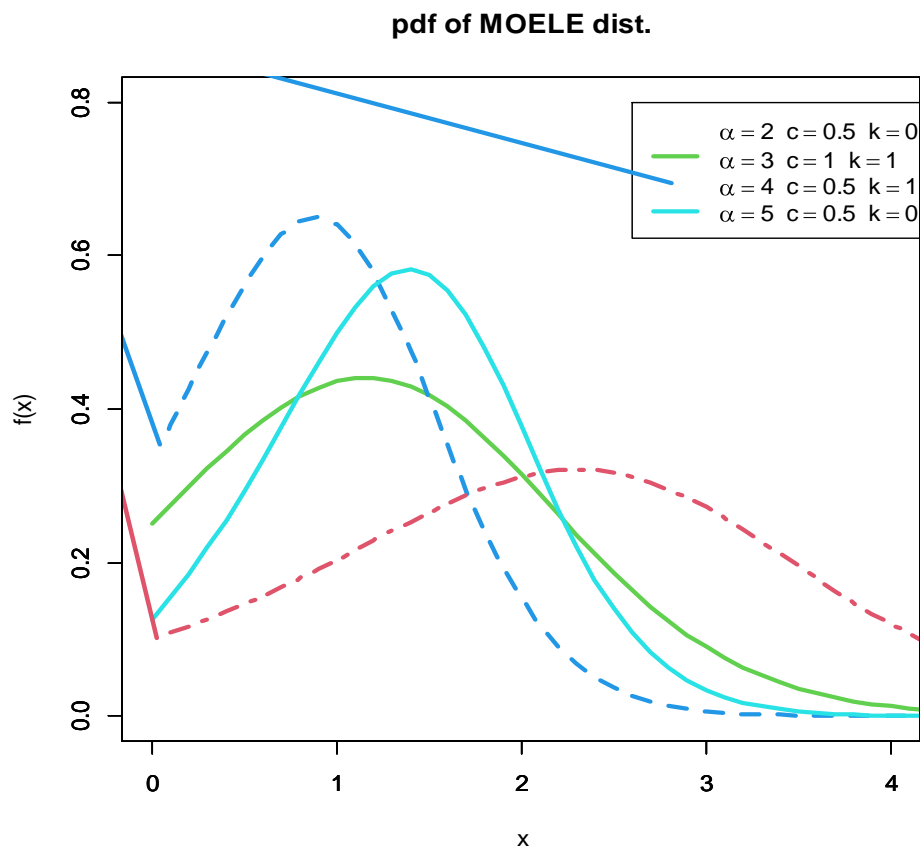


Figure (1)

3.1. The Maximum Likelihood Estimation

Suppose that x_1, x_2, \dots, x_n is a random sample from $MOELE(\alpha, c, k)$, then the log-likelihood function, $L(\alpha, c, k)$, is given by

$$L(\alpha, c, k) = n \ln \alpha + \sum_{i=1}^n \ln(c + k x_i) - \sum_{i=1}^n \left(c x_i + \frac{1}{2} k x_i^2 \right) - 2 \sum_{i=1}^n \ln \left[1 - (1 - \alpha) e^{-\left(c x_i + \frac{1}{2} k x_i^2 \right)} \right], \quad (7)$$

The normal equations become:

$$\frac{\partial L}{\partial \alpha} = \frac{n}{\alpha} - 2 \sum_{i=1}^n \frac{e^{-\left(c x_i + \frac{1}{2} k x_i^2 \right)}}{1 - (1 - \alpha) e^{-\left(c x_i + \frac{1}{2} k x_i^2 \right)}} = 0 \quad (8)$$

$$\frac{\partial L}{\partial c} = \sum_{i=1}^n \frac{1}{c + k x_i} - \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \frac{(1 - \alpha) x_i e^{-\left(c x_i + \frac{1}{2} k x_i^2 \right)}}{1 - (1 - \alpha) e^{-\left(c x_i + \frac{1}{2} k x_i^2 \right)}} = 0 \quad (9)$$

$$\frac{\partial L}{\partial k} = \sum_{i=1}^n \frac{x_i}{c + k x_i} - \frac{1}{2} \sum_{i=1}^n x_i^2 - \sum_{i=1}^n \frac{(1 - \alpha) x_i^2 e^{-\left(c x_i + \frac{1}{2} k x_i^2 \right)}}{1 - (1 - \alpha) e^{-\left(c x_i + \frac{1}{2} k x_i^2 \right)}} = 0 \quad (10)$$

Since the above normal equations cannot be solved analytically, we will use some numerical methods to solve it.

3.2. Estimators Based on Percentiles

This method was originally explored by Kao (1958, 1959). In this case, the estimators are obtained by fitting a straight line to the theoretical points obtained from the distribution function and the sample percentile points. Now, we apply this approach on the MOELE distribution to obtain estimators based on percentile (PC). Since, the MOELE distribution has the form

$$G(x) = 1 - \frac{\alpha}{e^{\left(c x + \frac{1}{2} k x^2 \right) + \alpha - 1}}, \quad x \geq 0 \quad (11)$$

Now, let $X_{(i)} = x_i$ be the i th order statistics, $x_1 < x_2 < \dots < x_n$, then the quantile function is

$$m_i(\alpha, c, k) = \left[\frac{2}{k} \ln \left(\frac{1 + (\alpha - 1) p_i}{1 - p_i} \right) + \frac{c^2}{k^2} \right]^{1/2} - \frac{c}{k} \quad (12)$$

so,

$$x_i = m_i(\alpha, c, k) + u_i \quad (13)$$

where, p_i denotes some estimate of $G(x_i; \alpha, c, k)$ and u_i is the error term, then the estimate of c, k and α can be obtained by minimizing

$$Q(\alpha, c, k) = \sum_{i=1}^n \left[x_i - \left[\frac{2}{k} \ln \left(\frac{1 + (\alpha - 1) p_i}{1 - p_i} \right) + \frac{c^2}{k^2} \right]^{1/2} - \frac{c}{k} \right]^2 \quad (14)$$

In fact, there are several estimators of p_i , the unbiased estimator of them of $G(x_i; \alpha, c, k)$ is $p_i = i/(1+n)$, so we consider it. Hence,

$$\frac{\partial Q(\alpha, c, k)}{\partial \alpha} = \sum_{i=1}^n \left[x_i - \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{1/2} - \frac{c}{k} \right] \left[\frac{p_i}{1 - (1 - \alpha) p_i} \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{-1/2} \right] = 0 \quad (15)$$

$$\frac{\partial Q(\alpha, c, k)}{\partial c} = \sum_{i=1}^n \left[x_i - \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{1/2} - \frac{c}{k} \right] \left[\frac{2c}{k} \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{-1/2} - 1 \right] = 0 \quad (16)$$

$$\frac{\partial Q(\alpha, c, k)}{\partial k} = \sum_{i=1}^n \left[x_i - \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{1/2} - \frac{c}{k} \right] \left[\left(A_i(\alpha) + \frac{c^2}{k} \right) \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{-1/2} + \frac{c}{2} \right] = 0 \quad (17)$$

where $A_i(\alpha) = \ln \left(\frac{1 + (\alpha - 1) p_i}{1 - p_i} \right)$. The equations (15), (16) and (17) is a non-linear system. So, it is possible to use some numerical methods to estimate c, k and α simultaneously. These estimators we call as percentile estimators (PCE_s).

3.3. Robust Estimation for the MOELE Distribution

Robust estimation is an estimation method that is used when there are some outliers that affect the model. Robust estimation is used to detect outliers and provide results that

are resistant to the outliers. One of the robust estimation methods is M estimator. In the past three decades, there are considerable works in the literature devoted to developing statistical procedures that are resistant to outliers and stable (or robust) with respect to deviations from a given distributional model. In particular, methods for robust regression, estimation, and testing on regression models have received much attention. Among these, procedures based on M-estimators play an important and complementary role.

In this paper we proposed a robust estimation method based on M-estimation method proposed by Huber (1964). The robust M-estimator method to estimate the parameters of the MOELE distribution is performed as following:

$$x_i = m_i(\alpha, c, k) + e_i \quad (18)$$

where $m_i(\alpha, c, k)$ is the quantile function was given in equation (12), and e_i is the error term after scaling as following

$$e_i = \frac{u_i}{s} \quad (19)$$

Where, u_i is the error term was given in equation (9), $s = \frac{MAD(u)}{0.6745}$ and $MAD(u) = MAD(u_1, u_2, \dots, u_n) = Median[u - Median(u)]$. [See Hampel *et.al.* (1986)]

Now, minimize the objective function (ρ) for all invariant errors with respect to the parameters c, k and α . There many ρ functions used in robust statistical analysis, we will use Tukey's Bisquare and Huber's weight [See Huber (1981)].

Tukey's Bisquare objective function is

$$\rho(e_i) = \begin{cases} 1 - \left(1 - \left(\frac{e_i}{a}\right)^2\right)^3 & |e_i| \leq a \\ 1 & |e_i| > a \end{cases} \quad (20)$$

with derivative is

$$\rho'(e_i) = \begin{cases} \frac{6e_i}{a^2} \left(1 - \left(\frac{e_i}{a}\right)^2\right)^2 & |e_i| \leq a \\ 0 & |e_i| > a \end{cases} \quad (21)$$

where $a = 4.685$ the tuning constant determines if an observation is an outlier or not.

Huber's weight objective function is

$$\rho(e_i) = \begin{cases} \frac{1}{2} e_i^2 & |e_i| \leq a \\ a|e_i| - \frac{1}{2} a^2 & |e_i| > a \end{cases} \quad (22)$$

with derivative is

$$\rho'(e_i) = \begin{cases} e_i & |e_i| \leq a \\ a \operatorname{sign} e_i & |e_i| > a \end{cases} \quad (23)$$

where $a = 1.345$

Since ρ is differentiable, M estimates can be obtained for the two selected objective function by minimize $\sum_{i=1}^n \rho(e_i)$ with respect to c, k and α , and equating to zero as following:

$$\sum_{i=1}^n \rho'(e_i) \frac{\partial m_i(\alpha, c, k)}{\partial \alpha} = \sum_{i=1}^n \frac{p_i}{1 - (1 - \alpha) p_i} \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{-1/2} = 0 \quad (24)$$

$$\sum_{i=1}^n \rho'(e_i) \frac{\partial m_i(\alpha, c, k)}{\partial c} = \sum_{i=1}^n \left[\frac{2c}{k} \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{-1/2} - 1 \right] = 0 \quad (25)$$

$$\sum_{i=1}^n \rho'(e_i) \frac{\partial m_i(\alpha, c, k)}{\partial k} = \sum_{i=1}^n \left[\left(A_i(\alpha) + \frac{c^2}{k} \right) \left[\frac{2}{k} A_i(\alpha) + \frac{c^2}{k^2} \right]^{-1/2} + \frac{c}{2} \right] = 0 \quad (26)$$

Since equations (24) - (26) are non-linear equations, so the numerical methods will apply to solve these equations.

4. Simulation Study

In this section, a simulation study is conducted in the presence of outliers to examine the performance of the estimation methods was given in section 3. The data were generated from the MOELE distribution by using inverse transform method, and the different values of c, k and α are used. The procedures are performed as:

Step (1): We generate random samples x_1, x_2, \dots, x_n of sizes $n = 20, 40$ and 100 from the MOELE distribution. We have taken parameter values $(\alpha, c, k) = (3, 2, 1), (2, 1, 1)$, and $(2, 1, 0.5)$.

Step (2): For each random sample, the outliers are generated from the uniform distribution as $U(\bar{x} + 4s, \bar{x} + 7s)$, where \bar{x} is the sample mean and s is the standard deviation of x_1, x_2, \dots, x_n . For the small sample size ($n = 20$) one outlier is taken, for the moderate sample size ($n = 40$) two outliers are taken and for the largest sample size ($n = 100$) five outliers are taken. [See Wei and Fung (1999)].

Step (3): Solve equations (8) - (10) simultaneous to obtain the ML estimators, and solve equations (15) - (17) simultaneous to obtain the PC estimators. Also, solve equations (24) - (26) simultaneous to obtain the M estimators.

Step (4): Calculate the bias and the mean square error (MSE) for the estimators obtained in step (3).

Step (5): Steps from (1) to (4) will be repeated 100 times.

The tables from 1 to 3 show the values of the bias and the RMSE for the ML, PC and robust (Tukey and Huber)

estimators under different values of (α, c, k) and different values of n in the presence of outliers. It is obvious from the tabulated results that:

- (i) The robust estimator based on Tukey's Bisquare function has the smallest bias and the smallest RMSE in the most.
- (ii) The robust estimator based on Huber's weight function has bias and RMSE smaller than ML and PC estimators.

Table 1. The Bias and RMSE for $(\alpha, c, k) = (3, 2, 1)$ and different sample sizes

n=20 with one outlier						
methods	α		c		k	
	bias	RMSE	Bias	RMSE	bias	RMSE
MLE	1.08081	1.131003	1.221879	1.131851	0.927393	0.92632
PC	0.938655	0.933885	0.923202	0.789228	0.65826	0.789228
Huber	0.710563	0.723149	0.689234	0.601699	0.315546	0.601699
Tukey	0.493915	0.478535	0.47977	0.36217	0.197385	0.36217
n=40 with two outlier						
MLE	0.8877	0.91842	0.73914	0.54862	0.39185	0.5388
PC	0.565074	0.793638	0.594504	0.468144	0.34344	0.454638
Huber	0.395437	0.462812	0.457996	0.350196	0.229397	0.343581
Tukey	0.263145	0.28355	0.30689	0.117035	0.137075	0.114706
n=100 with five outlier						
MLE	0.39086	0.52333	0.42712	0.14972	0.1996	0.146741
PC	0.344664	0.408168	0.278586	0.126369	0.157257	0.123854
Huber	0.251377	0.259994	0.101612	0.058303	0.101493	0.057143
Tukey	0.15959	0.02711	0.05786	0.01467	0.032035	0.014378

Table 2. The Bias and RMSE for $(\alpha, c, k) = (2, 1, 1)$ and different sample sizes

n=20 with one outlier						
methods	α		c		k	
	bias	RMSE	Bias	RMSE	bias	RMSE
MLE	0.95668	0.85349	0.97078	0.84697	0.77054	0.85945
PC	0.829188	0.699066	0.861093	0.75744	0.553059	0.740196
Huber	0.632625	0.436191	0.640346	0.51443	0.363293	0.49525
Tukey	0.44637	0.29022	0.435215	0.21374	0.23923	0.314955
n=40 with two outlier						
MLE	0.88752	0.46909	0.82421	0.35963	0.47182	0.4721
PC	0.759879	0.389916	0.679788	0.307377	0.395181	0.401823
Huber	0.555884	0.293139	0.474551	0.217217	0.290094	0.268709
Tukey	0.36939	0.2093	0.257545	0.13614	0.166385	0.144775
n=100 with five outlier						
MLE	0.67983	0.39442	0.2715	0.18836	0.28073	0.27744
PC	0.475074	0.303669	0.198603	0.155223	0.200403	0.223758
Huber	0.340683	0.22176	0.143241	0.06608	0.154665	0.125489
Tukey	0.21664	0.09884	0.096455	0.040875	0.043675	0.04678

Table 3. The Bias and RMSE for $(\alpha, c, k) = (2, 1, 0.5)$ and different sample sizes

n=20 with one outlier						
methods	α		c		k	
	bias	RMSE	Bias	RMSE	bias	RMSE
MLE	0.866871	0.746289	0.772308	0.747522	0.506889	0.604593
PC	0.637875	0.553889	0.599746	0.482657	0.28966	0.437388
Huber	0.38088	0.353475	0.418325	0.32671	0.18707	0.2914
Tukey	0.866871	0.746289	0.772308	0.747522	0.506889	0.604593
n=40 with two outlier						
MLE	0.60812	0.52511	0.60684	0.6254	0.32069	0.49515
PC	0.512064	0.465489	0.533196	0.441666	0.248814	0.342819
Huber	0.391321	0.356139	0.39032	0.322238	0.168875	0.24472
Tukey	0.18424	0.253695	0.231925	0.20639	0.118105	0.08093
n=100 with five outlier						
MLE	0.34717	0.47392	0.4421	0.39958	0.23562	0.1264
PC	0.223938	0.370134	0.348183	0.301779	0.202581	0.091611
Huber	0.153209	0.255703	0.265335	0.229726	0.119686	0.067711
Tukey	0.105625	0.1523	0.15668	0.14498	0.08149	0.020315

5. Real Data Example

In this section, we apply on a real data set to verify how our estimators work in practice. The data given by shao (2000) is used. This data is used by Mousa (2017) to fit a Marshall –Olkin extended Burr III distribution. The data set is resulted from the study of influence of the proportion of toxicity of chromium in marine water. Table (4) contains 36 values for chromium marine water.

To estimate the parameters, the values of data set are divided by 365 for more fitting to MOLE distribution and Kolmogorov-Smirnov (KS) test is used. The estimation of the unknown parameters (α, c, k) and P -values are obtained in table (5).

Table 4. Chromium in marine water

2000	776.25	89.13	2000	177.83	1174.89
12.59	1258.93	4.79	199.5	540	9.55
199.53	3311.31	8800	2.4	1456	3090.3
4	39.81	10000	728	2511.89	602.56
56	2630.27	1600	1122.02	1200	1995.26
264.03	187.2	602.56	140	478.63	210

Table 5. Estimation of parameters for real data set

method	α	c	k	$k S$	p-value
MLE	0.78027	0.70017	0.00156	1.321	0.061
PC	0.7692	0.6972	0.0034	1.312	0.064
Huber	0.3856	0.41898	0.0406	1.168	0.131
Tukey	0.36934	0.42023	0.03747	1.125	0.159

From table (5), we note that

- (i) The Tukey estimator is the best estimator because the

KS value is the smallest value and the P-value is the biggest value.

- (ii) the Huber estimator is better than the ML and PC estimators based on the values of KS and P-value.

6. Conclusions

In this paper, the estimation of the unknown parameters for the MOELE distribution is discussed in the presence of outliers. The traditional estimation methods such as ML and PC methods are used. Also, the M-estimation method is used based on the two objective functions, i. e. tukey's Bisquare and Huber. The simulation study and the real data are proved that, the M- estimation method is more efficient in estimating the unknown parameters of the MOELE distribution than the traditional methods when the data contains outliers or extreme observations.

REFERENCES

- [1] Almongy, H. M., and Almetwally, E. M. (2020). Robust estimation methods of generalized exponential distribution with outliers. Pak.j.stat.oper.res, 16(3), 545-559.
- [2] Broadbent, S., (1958). Simple mortality rates. Journal of Applied Statistics, 7, 86.
- [3] Carbone, P., Kellerthouse, L., and Gehan, E. (1967). Plasmacytic Myeloma: A study of the relationship of survival to various clinical manifestations and anomalous protein type in 112 patients. American Journal of Medicine, 42, 937-948.
- [4] Kodlin, D. (1967). A new response time distribution.

- Biometrics, 23, 227-239.
- [5] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). Robust statistics: The approach based on influence functions. Wiley, New York.
 - [6] Huber, P. J. (1964). Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35, 73-101.
 - [7] Huber, P.J. (1981). Robust statistics. Wiley, New York.
 - [8] Kantar, Y. M., and Yildirim, V. (2015). Robust estimation for parameters of the extended burr type III distribution. Communications in Statistics-Simulation and Computation, 44(7), 1901-1930.
 - [9] Kao, J. H. K. (1958). Computer methods for estimating weibull parameters in reliability studies. Transaction of IRE-Reliability and Quality Control, 13, 15-22.
 - [10] Kao, J. H. K. (1959). A graphical estimation of mixed weibull parameters in life testing electron tubes. Technometrics, 1, 389-407.
 - [11] Marshall, A. W., and I. Olkin. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika. 84, 641-652.
 - [12] Merovci, F., and Elbatal, I. (2015). A new generalization of linear exponential distribution. J. Stat. Appl. Pro. Lett, 2(1), 1-14.
 - [13] Mousa, S. A. (2017). Robust estimators for marshall-olkin extended burr III distribution. International Journal of Research in Applied Natural and Social Sciences. 5, 69-79.
 - [14] Okasha, H.M. and Kayid, M. (2016). A new family of Marshall-Olkin extended generalized linear exponential distribution. Journal of Computational and Applied Mathematics. 296, 576-592.
 - [15] Sarhan, A., and Kundu, D. (2009). Generalized linear failure rate distribution. Commun. Statist. Theory Methods. 38 (5), 642-660.
 - [16] Shao, Q. (2000). Estimation for hazardous Concentrations based on NOEC toxicity data: An Alternative Approach. Environmetrics. 11, 583-595.
 - [17] Wei, W.H., and Fung W.K. (1999). The mean-shift outlier model in general weighted regression and its application. Computational Statistics & Data Analysis. 30, 429-441.