



MARSHALL-OLKIN EXTENDED POWER LINDLEY DISTRIBUTION WITH APPLICATION

RAFIF HIBATULLAH^{1*}, YEKTI WIDYANINGSIH², SARINI ABDULLAH³

^{1,2,3} Department of Mathematics, Faculty of Mathematics and Natural Sciences
University of Indonesia, West Java 16424, Indonesia

*rafif.hibatullah49@gmail.com

ABSTRAK

Lindley distribution was introduced by Lindley (1958) in the context of Bayes inference. Its density function is obtained by mixing the exponential distribution, with scale parameter β , and the gamma distribution, with shape parameter 2 and scale parameter β . Recently, a new generalization of the Lindley distribution was proposed by Ghitany et al. (2013), called power Lindley distribution. This paper will introduce an extension of the power Lindley distribution using the Marshall-Olkin method, resulting in Marshall-Olkin Extended Power Lindley (MOEPL) distribution. The MOEPL distribution offers a flexibility in representing data with various shapes. This flexibility is due to the addition of a parameter to the power Lindley distribution. Some properties of the MOEPL were explored, such as probability density function (pdf), cumulative distribution function (cdf), hazard rate, survival function, quantiles, and moments. Estimation of the MOEPL parameters was conducted using maximum likelihood method. The proposed distribution was applied to a life time data and an uncensored data corresponding to remission times (in months) of a random sample of 36 bladder cancer patients. The results were given which illustrate the MOEPL distribution and were compared to Lindley, power Lindley, gamma, and Weibull. Model comparison using the log likelihood, AIC, and BIC showed that MOEPL fit the data better than the other distributions.

Keywords: likelihood, Lindley, Marshall-Olkin, Power Lindley

1 Introduction

Lindley distribution was introduced by Dennis Victor Lindley in 1957 [14]. Lindley's distribution has been well studied by Ghitany [8] although Lindley distribution was introduced by Lindley [14] in the context of Bayes inference. According to Ghitany [8], Lindley's distribution has a more flexible characteristic than the exponential distribution. The graphic in the exponential distribution is always monotonically down but on the Lindley distribution it is not always monotonically down. Then the hazard function graph on the exponential distribution is constant while in the Lindley distribution it is not constant. The Lindley distribution can also be used in many fields, for example, for reliability analysis to model the length of waiting time of 100 bank customers before getting service (Krishna, 2011) [13] and analysis of the life time of 194 lung cancer patients with squamous cell carcinoma (Mazucheli, 2010) [16]. If $\beta > 0$ and $x > 0$, the probability density function (pdf) and the cumulative distribution function (cdf) of the Lindley distribution is given by

$$a(x) = \frac{\beta^2}{1 + \beta} (1 + x) e^{-\beta x}, \quad (1)$$

$$A(x) = 1 - \left(1 + \frac{\beta x}{1 + \beta}\right) e^{-\beta x}. \quad (2)$$

Although Lindley distribution has more flexible characteristics than exponential distributions, Lindley distribution is poor in explaining non-monotonous data compared to known distributions such as gamma or Weibull (Ghitany, 2013) [7]. Therefore, Ghitany [7] performed power transformation on Lindley distribution and produces power Lindley distribution so that the resulting distribution can explain non-monotonous data more precisely. An example of applying power Lindley distribution is to model the length of carbon fiber break time (Ghitany, 2013) [7]. The Power Lindley (PL) is constructed as follows. If a random variable X has a Lindley distribution, then the random variable $Y = X^{\frac{1}{\alpha}}$ follows an PL distribution. The cdf of PL is given by

$$G(x) = 1 - \left(\frac{\beta + \beta x^\alpha + 1}{1 + \beta}\right) e^{-\beta x^\alpha}, \quad (3)$$

where $\alpha, \beta > 0$ and $x > 0$. The pdf $g(x)$, survival function $\bar{G}(x)$, and hazard rate function $h(x)$ of the PL distribution are given by

$$g(x) = \frac{dG(x)}{dx} = \frac{\alpha\beta^2}{1 + \beta} (1 + x^\alpha) x^{\alpha-1} e^{-\beta x^\alpha}, \quad (4)$$

$$\bar{G}(x) = 1 - G(x) = \left(\frac{\beta + \beta x^\alpha + 1}{1 + \beta}\right) e^{-\beta x^\alpha}, \quad (5)$$

$$h(x) = \frac{g(x)}{\bar{G}(x)} = \frac{\alpha\beta^2(1 + x^\alpha)x^{\alpha-1}}{(\beta + \beta x^\alpha + 1)}. \quad (6)$$

PL distribution is also not good to express data with a thick enough tail. On the other hand, Marshall and Olkin [15] proposed a transformation based on the survival function of a distribution by adding a new parameter to obtain a family of distributions that can make a distribution more flexible and able to reach wider data compared to the old distribution. Provided a random variable X with a survival function $\bar{G}(x)$, the Marshall-Olkin transformation is given by

$$S(x) = \frac{\theta \bar{G}(x)}{G(x) + \theta \bar{G}(x)}, \quad (7)$$

where $\theta > 0$ and $-\infty < x < \infty$. Its corresponding cdf $G(x)$, pdf $f(x)$, and hazard rate function $h(x)$ are given by

$$F(x) = 1 - s(x) = \frac{G(x)}{G(x)(1 - \theta) + \theta}, \quad (8)$$

$$f(x) = \frac{dG(x)}{dx} = \frac{\theta g(x)}{[G(x)(1 - \theta) + \theta]^2}, \quad (9)$$

$$h(x) = \frac{f(x)}{S(x)} = \frac{h(x)}{[G(x)(1 - \theta) + \theta]}. \quad (10)$$

Marshall and Olkin have noted that their transformation has a stability property, if it is applied twice, nothing new is obtained. The Marshall-Olkin extended distributions has a wider range of behavior of the probability density function (pdf) and hazard rate function than those of the baseline distributions from which they are derived [11]. Marshall-Olkin method is further used to obtain new distributions and their properties. Al-Saiari [4] introduced Marshall-Olkin extended Burr type XII distribution and studied its statistical properties. Alizadeh studied some mathematical properties of Beta Marshall-Olkin family of distribution [2] and Marshall-Olkin extended Kumaraswamy distribution [3]. Ghitany [8] studied some mathematical properties of Marshall-Olkin extended Lindley distribution. Krishna [12] studied the Marshall-Olkin generalized asymmetric Laplace distributions. Marshall-Olkin Extended Kappa distribution

was introduced by Javed [9] and Marshall-Olkin extended Zipf distribution was introduced by Perez [17].

In this paper, we will discuss Marshall-Olkin Extended Power Lindley (MOEPL) distribution. This paper would develop the theory of the power Lindley distribution that has been introduced and discussed by Ghitany and by adding new parameter using Marshall-Olkin method. To the best of our knowledge, statistical properties of the MOEPL distribution have not been known. The aim of this paper is to set the records straight about the MOEPL distribution.

2 MOEPL Distribution

2.1 Probability Density Function (pdf) and Cumulative Distribution Function (cdf)

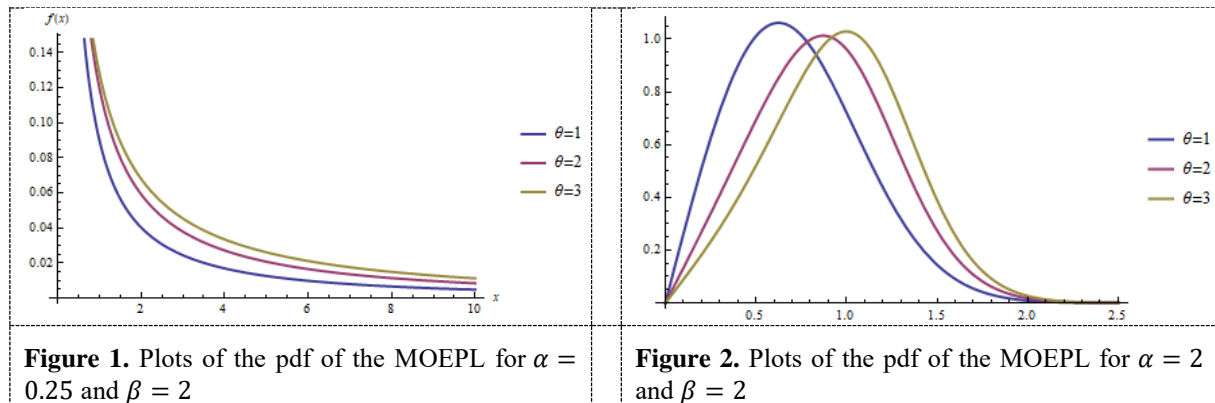
In this section, the pdf and cdf of MOEPL distribution are discussed. Let X be the random variable that have power Lindley distribution with parameters α and β , and θ is a real constant with $\theta > 0$. By substituting the cdf of PL at equation (3) to equation (9), the following cdf will be obtained

$$F(x) = \frac{1 - \left(\frac{\beta + \beta x^\alpha + 1}{1 + \beta} \right) e^{-\beta x^\alpha}}{\left(1 - \left(\frac{\beta + \beta x^\alpha + 1}{1 + \beta} \right) e^{-\beta x^\alpha} \right) (1 - \theta) + \theta} \quad (11)$$

Then, by derivative equation (11) or substituting the cdf of PL given by (3) and pdf given by (4) in Marshall-Olkin given by (9), a new distribution is yielded, denoted as MOEPL, with pdf as follows

$$\begin{aligned} f(x) = \frac{dF(x)}{dx} &= \frac{\frac{\theta \alpha \beta^2}{1 + \beta} (1 + x^\alpha) x^{\alpha-1} e^{-\beta x^\alpha}}{\left(\left(1 - \left(\frac{\beta + \beta x^\alpha + 1}{1 + \beta} \right) e^{-\beta x^\alpha} \right) (1 - \theta) + \theta \right)^2} = \\ &= \frac{\theta \alpha (1 + \beta) \beta^2 (1 + x^\alpha) x^{\alpha-1} e^{-\beta x^\alpha}}{\left((\beta + 1) e^{\beta x^\alpha} + (\theta - 1) \beta x^\alpha + (\theta - 1)(\beta + 1) \right)^2} \end{aligned} \quad (11)$$

As information, if $\theta = 1$ then we have PL distribution. If $\theta = 1$ and $\alpha = 1$ then we have Lindley distribution. Figures 1, 2, 3, and 4 show the MOEPL pdf for some α and β values; decreasing (figure 1), increasing-decreasing (unimodal) (figure 2), or decreasing-increasing-decreasing (figure 3). The influence of alpha and beta parameters already explained by Ghitany [7]. Figure 4 shows that pdf of MOEPL can be negative skew, symmetrically, or positive skew.



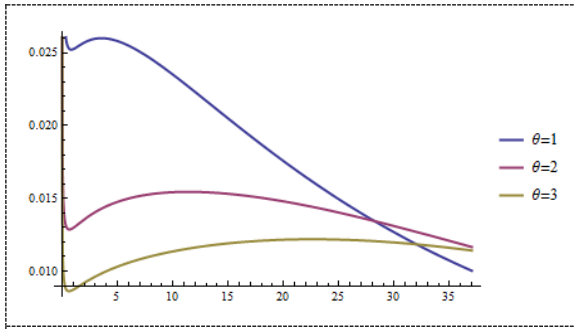


Figure 3. Plots of the pdf of the MOEPL for $\alpha = 0.75$ and $\beta = 0.85$

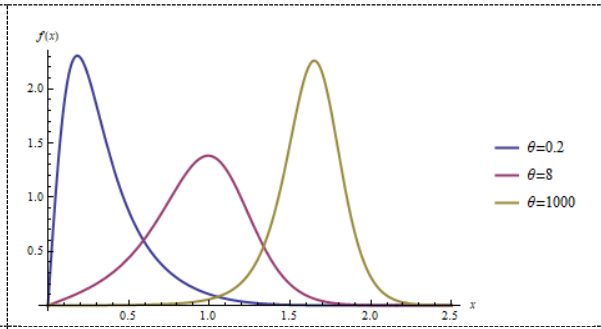


Figure 4. Plots of the pdf of the MOEPL for $\alpha = 2$ and $\beta = 3$

2.2 Survival Function and Hazard Rate Function

Let X be the random variable that have PL distribution with parameters α and β , and θ is a real constant with $\theta > 0$. By substituting equation (3) and (5) to equation (7), equation (13) the survival function of MOEPL will be obtained

$$S(x) = 1 - F(x) = \frac{\theta(\beta + \beta x^\alpha + 1)}{(1 + \beta)e^{\beta x^\alpha} + (\theta - 1)(\beta + \beta x^\alpha + 1)} \quad (13)$$

Cdf of PL in equation (3) and hazard rate function of PL in equation (6) are substituted to equation (10) then hazard rate function of MOEPL is

$$h(x) = \frac{f(x)}{S(x)} = \frac{\alpha\beta^2 (1 + x^\alpha)x^{\alpha-1}}{(\beta + \beta x^\alpha + 1) \left(\left(1 - \left(\frac{\beta + \beta x^\alpha + 1}{1 + \beta} \right) e^{-\beta x^\alpha} \right) (1 - \theta) + \theta \right)}, \quad (14)$$

Figures 5, 6, and 7 show the graphs of hazard MOEPL distribution. The graphs can be decreasing (figure 5), increasing (figure 6), or decreasing-increasing-decreasing (figure 7). The influence of alpha and beta parameters already explained by Ghitany [7].

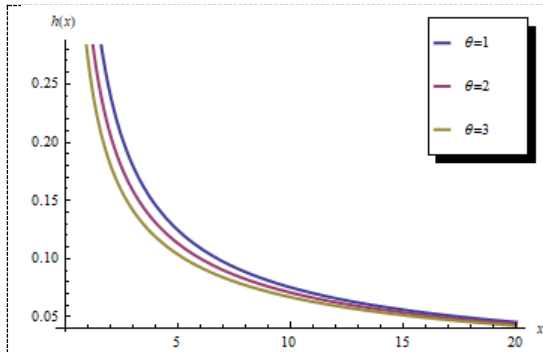


Figure 5. Plots of the hazard function of the MOEPL for $\alpha = 0.25$ and $\beta = 2$.

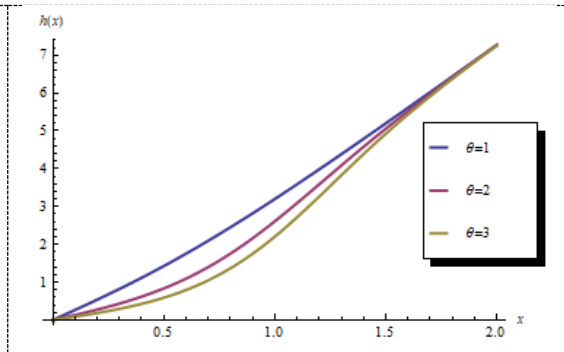


Figure 6. Plots of the hazard function of the MOEPL for $\alpha = 2$ and $\beta = 2$.

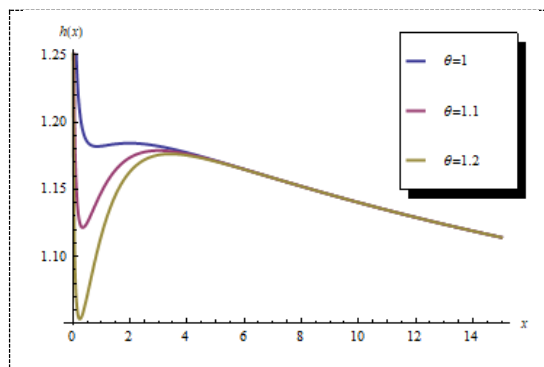


Figure 7. Plots of the hazard function of the MOEPL for $\alpha = 0.9$ and $\beta = 1.7$

2.3 Quantiles

In this section the quantiles of MOEPL distribution are discussed. For a constant $p \in (0,1)$, the quantile p of a value ω_p is

$$\omega_p = \left(\frac{-1 - \beta - W_{-1} \left(\frac{(p-1)(1+\beta)e^{-(\beta+1)}}{1-p+p\theta} \right)}{\beta} \right)^{\frac{1}{\alpha}}, \quad (15)$$

where W_{-1} denotes the negative branch of the Lambert W function that presented by Jodra (2010) [10].

2.4 Moments

In this section the r th moments of MOEPL distribution are discussed. If X is the random variable of PL distribution with parameters α and β , and θ is a real constant with $\theta > \frac{1}{2}$ then r th moments of MOEPL is

$$E(X^r) = \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{t=0}^m (-1)^m \binom{n}{m} \binom{m}{t} \frac{\alpha \beta^{2-\frac{r}{\alpha}}}{\theta(1+\beta)^{t+1}(1+m)^{\frac{r}{\alpha}+t+1}} (n+1) \left(1 - \frac{1}{\theta}\right)^n \left[\Gamma\left(\frac{r}{\alpha} + t + 1\right) + \frac{\Gamma\left(\frac{r}{\alpha} + 2 + t\right)}{(1+m)\beta} \right] \quad (16)$$

Although r th moments of the MOEPL distribution with the parameter $\theta > \frac{1}{2}$ is not simple and for the parameter $\theta \leq \frac{1}{2}$ is difficult to obtain with ordinary integrated, but we can be found the interval value of r th moments MOEPL distribution. Let $E[X^r]^*$ is r th moments for the PL distribution [7] and $E[X^r]$ is r th moments for the MOEPL distribution. Since the r th moment of PL distribution is finite moments [7] then r th moments of MOEPL distribution is also a finite moment with the interval limit being: $E[X^r] \in \left[\frac{E[X^r]^*}{\theta}, \theta E[X^r]^* \right]$ for $\theta \geq 1$ and $E[X^r] \in \left[\theta E[X^r]^*, \frac{E[X^r]^*}{\theta} \right]$ for $\theta < 1$.

3 Maximum Likelihood Estimation

Let X_1, X_2, \dots, X_n be a random sample from MOEPL distribution, then the likelihood function is given by

$$L(x; \alpha, \beta, \theta) = \frac{(\theta\alpha(1+\beta)\beta^2)^n \prod_{i=1}^n (1+x_i^\alpha) (\prod_{i=1}^n (x_i))^{\alpha-1} e^{\beta \sum_{i=1}^n x_i^\alpha}}{\prod_{i=1}^n ((\beta+1)e^{\beta x_i^\alpha} + (\theta-1)\beta x_i^\alpha + (\theta-1)(\beta+1))}, \quad (17)$$

the logarithm of the likelihood function is then

$$\begin{aligned} \ln L(x; \alpha, \beta, \theta) &= n \ln \theta + n \ln \alpha + n \ln(1+\beta) + 2n \ln \beta + \sum_{i=1}^n \ln(1+x_i^\alpha) \\ &+ (\alpha-1) \sum_{i=1}^n \ln x_i + \beta \sum_{i=1}^n x_i^\alpha \\ &- 2 \sum_{i=1}^n \ln((\beta+1)e^{\beta x_i^\alpha} + (\theta-1)\beta x_i^\alpha + (\theta-1)(\beta+1)) \end{aligned} \quad (18)$$

The maximum of equation (18) can be obtained by taking the first partial derivatives of the log-likelihood function with respect to the three parameters, the results are:

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - 2 \sum_{i=1}^n \frac{\beta x_i^\alpha + \beta + 1}{(\beta + 1)e^{\beta x_i^\alpha} + (\theta - 1)\beta x_i^\alpha + (\theta - 1)(\beta + 1)} = 0, \quad (19)$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \alpha} = & \frac{n}{\alpha} + \sum_{i=1}^n \frac{x_i^\alpha \ln x_i}{1 + x_i^\alpha} + \sum_{i=1}^n \ln x_i + \beta \sum_{i=1}^n x_i^\alpha \ln x_i \\ & - 2 \sum_{i=1}^n \frac{\beta(\beta + 1)e^{\beta x_i^\alpha} x_i^\alpha \ln x_i + (\theta - 1)\beta x_i^\alpha \ln x_i}{(\beta + 1)e^{\beta x_i^\alpha} + (\theta - 1)\beta x_i^\alpha + (\theta - 1)(\beta + 1)} = 0, \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta} = & \frac{n}{(\beta + 1)} + \frac{2n}{\beta} + \sum_{i=1}^n x_i^\alpha \\ & - 2 \sum_{i=1}^n \frac{e^{\beta x_i^\alpha} + (\beta + 1)e^{\beta x_i^\alpha} x_i^\alpha + (\theta - 1)x_i^\alpha + \theta - 1}{(\beta + 1)e^{\beta x_i^\alpha} + (\theta - 1)\beta x_i^\alpha + (\theta - 1)(\beta + 1)} = 0, \end{aligned} \quad (21)$$

The maximum likelihood estimates of α , β , and θ are obtained by solving the nonlinear equations (19), (20), and (21). These equations are not in closed form and the values of the parameters α , β , and θ must be found by numerical methods.

4 Simulation

This section presents the simulation of the data of MOEPL distribution from literature. The estimation of the MOEPL parameters was obtained using the maximum likelihood method. For the comparison MOEPL and other models, the log-likelihood values ($\log L$), the Akaike information criteria (AIC) defined by $-2\log L + 2q$, and the Bayesian information criterion (BIC) defined by $-2\log L + q \log(n)$, where q is the number of estimated parameters and n is the sample size. The best model has the highest value of $\log L$ and the lowest values of AIC and BIC. In this study, MOEPL distribution is compared to Lindley distribution, PL distribution, gamma distribution, and Weibull distribution.

4.1 Example 1

This first example deals with a life time data set taken from [1]. The data is about the lifetimes of 50 devices: 0.1, 0.2, 1, 1, 1, 1, 1, 2, 3, 6, 7, 11, 12, 18, 18, 18, 18, 18, 21, 32, 36, 40, 45, 46, 47, 50, 55, 60, 63, 63, 67, 67, 67, 67, 72, 75, 79, 82, 82, 83, 84, 84, 84, 85, 85, 85, 85, 85, 86, 86. Table 1 presents the parameters estimation, $\log L$, AIC, and BIC for the fitted MOEPL and their sub-models (Lindley, PL) distributions, gamma distribution, and Weibull distribution.

TABLE 1. Summary of fitted distributions for data set 1

Models	MLEs	log L	AIC	BIC	T
Lindley	$\beta = 0.043$	-	504.861	504.560	0.175
		251.430			
PL	$\alpha = 0.664$	-	488.175	487.573	0.176
	$\beta = 0.161$	242.087			
Gamma	$\alpha_1 = 0.799$	-	484.380	483.778	0.182
	β_1	240.190			
	$= 57.172$				
Weibull	β_1	-	484.004	483.703	0.173
	$= 44.913$	241.002			
	$\gamma_1 = 0.949$				
MOEPL	$\alpha = 0.531$	-	482.760	481.857	0.168
	$\beta = 0.493$	238.380			
	$\theta = 8.193$				

Based on Kolgomorov-Smirnov test with level of significance 0.05, the value of $w_{1-\alpha} = 0.192$. From table 1, all values of T are bigger than 0.192, so that all models fit in explaining the data. For the value of log L, MOEPL has the highest value than other models. For AIC and BIC, MOEPL has the lowest value than the other models. Therefore, MOEPL model provides the best fit to the data than other models.

4.2 Example 2

The data set represents an uncensored data corresponding to remission times (in months) of a random sample of 36 bladder cancer patients reported in Elbatal [5]: 0.08, 0.2, 0.4, 0.5, 0.51, 0.81, 0.87, 0.9, 1.05, 1.19, 1.26, 1.35, 1.4, 1.46, 1.76, 2.02, 2.02, 2.07, 2.09, 2.23, 2.26, 2.46, 2.54, 2.62, 2.64, 2.69, 2.69, 2.75, 2.83, 2.87, 3.02, 3.02, 3.25, 3.31, 3.36, 3.36. Table 2 presents the parameters estimation, log L, AIC, and BIC for the fitted MOEPL and their sub-models (Lindley, PL) distributions, gamma distribution, and Weibull distribution.

TABLE 2. Summary of fitted distributions for data set 1

Models	MLEs	log L	AIC	BIC	T
Lindley	$\beta = 0.801$	-56.698	115.398	114.954	0.168
PL	$\alpha = 1.649$	-50.841	105.682	104.795	0.107
	$\beta = 0.490$				
Gamma	$\alpha_1 = 2.309$	-54.092	112.185	111.297	0.138
	$\beta_1 = 0.840$				
Weibull	$\beta_1 = 2.164$	-51.387	106.774	105.887	0.112
	$\gamma_1 = 1.957$				
MOEPL	$\alpha = 1.282$	-49.657	105.305	103.974	0.088
	$\beta = 1.106$				
	$\theta = 5.121$				

Based on Kolgomorov-Smirnov test with level of significance 0.05, the value of $w_{1-\alpha} = 0.227$. From table 2, all values of T is bigger than 0.227, so that all models fit in explaining the data. Based on the value of log L, MOEPL has the highest value than other models. Based on

AIC and BIC, MOEPL has the lowest value than other models. Therefore, MOEPL distribution provides the best fit to this data than other models.

5 Conclusion

In this paper, we introduced our proposed model, named Marshall–Olkin extended power Lindley distribution. Many properties of our proposed model were investigated, including pdf, cdf, hazard rate, survival function, quantiles and moments. The estimation of parameters was obtained using maximum likelihood method. The proposed extended model was applied on real data from [1] and [5]. The results are given which illustrate the superior performance of the MOEPL distribution compared to other models.

6 Acknowledgements

We wish to thank DRPM Universitas Indonesia through Hibah Publikasi Internasional Terindeks untuk Tugas Akhir Mahasiswa UI (PITTA) 2018 Number: 2337/UN2.R3.1/HKP.05.00/2018.

Reference

- [1] Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, 36(1), 106-108.
- [2] Alizadeh, M., Cordeiro, G. M., De Brito, E., & Demétrio, C. G. B. (2015). The beta Marshall-Olkin family of distributions. *Journal of Statistical Distributions and Applications*, 2(1), 4.
- [3] Alizadeh, M., Tahir, M. H., Cordeiro, G. M., Mansoor, M., Zubair, M., & Hamedani, G. G. (2015). The Kumaraswamy Marshal-Olkin family of distributions. *Journal of the Egyptian Mathematical Society*, 23(3), 546-557.
- [4] Al-Saiari, A. Y., Baharith, L. A., & Mousa, S. A. (2014). Marshall-Olkin extended Burr type XII distribution. *International Journal of Statistics and probability*, 3(1), 78.
- [5] Elbatal, I., Merovci, F., & Elgarhy, M. (2013). A new generalized Lindley distribution. *Mathematical theory and Modeling*, 3(13), 30-47.
- [6] Ghitany, M. E., Al-Mutairi, D. K., Al-Awadhi, F. A., & Al-Burais, M. M. (2012). Marshall-Olkin extended Lindley distribution and its application. *International Journal of Applied Mathematics*, 25(5), 709-721.
- [7] Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N., & Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference. *Computational Statistics & Data Analysis*, 64, 20-33
- [8] Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and computers in simulation*, 78(4), 493-506
- [9] Javed, M., Nawaz, T., & Irfan, M. (2018). The Marshall-Olkin Kappa Distribution: Properties and Applications. *Journal of King Saud University-Science*.
- [10] Jodrá, P. (2010). Computer generation of random variables with Lindley or Poisson–Lindley distribution via the Lambert W function. *Mathematics and Computers in Simulation*, 81(4), 851-859
- [11] Jose, K. K. (2011). Marshall-Olkin family of distributions and their applications in reliability theory, time series modeling and stress-strength analysis. *Proc. ISI 58th World Statist. Congr Int Stat Inst*, 21st-26th August, 201, 3918-392.
- [12] Krishna, E., & Jose, K. K. (2011). Marshall-Olkin generalized asymmetric Laplace distributions and processes. *Statistica*, 71(4), 453-467.
- [13] Krishna, H., & Kumar, K. (2011). Reliability estimation in Lindley distribution with progressively type II right censored sample. *Mathematics and Computers in Simulation*, 82(2), 281-294
- [14] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 102-107
- [15] 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

- [16] Mazucheli, J., & Achcar, J. A. (2011). The Lindley distribution applied to competing risks lifetime data. *Computer methods and programs in biomedicine*, 104(2), 188-1
- [17] Pérez-Casany, M., & Casellas, A. (2013). Marshall-Olkin Extended Zipf Distribution. *arXiv preprint arXiv:1304.4540*.
- [18] Rocha, R., Nadarajah, S., Tomazella, V., & Louzada, F. (2017). A new class of defective models based on the Marshall–Olkin family of distributions for cure rate modeling. *Computational Statistics & Data Analysis*, 107, 48-63