Cumhuriyet Science Journal

Cumhuriyet Sci. J., 44(1) (2023) 224-228
DOI: https://doi.org/10.17776/csj.1171453

| csj.cumhuriyet.edu.tr | Founded: 2002 ISSN: 2587-2680 e-ISSN: 2587-246X Publisher: Sivas Cumhuriyet University

Parameters Estimation for the Unit log-log Distribution

Mustafa Çağatay Korkmaz [1,a], Kadir Karakaya [2,b,*], Yunus Akdoğan [2,c], Yener Ünal [3,d]

[1] *Department of Measurement and Evaluation, Faculty of Education, Artvin Çoruh University, Artvin, Türkiye*
[2] *Department of Statistics, Faculty of Science, Selçuk University, Konya, Türkiye*
[3] *Department of Statistics and Computer Science, Faculty of Science, Sivas Cumhuriyet University, Sivas, Türkiye*
**Corresponding author*

Research Article

History
Received: 06/09/2022
Accepted: 11/01/2023

ABSTRACT

In this paper, point estimation problem of two unknown parameters of the unit log-log distribution is examined. For point estimation, six methods of estimate such as maximum likelihood, maximum product spacing, Anderson-Darling, least squares, weighted least squares, and Cramer-von Mises are examined in detail. Extensive simulation experiments are conducted to compare the effectiveness of these estimators based on bias and mean squared errors. According to the simulation results, it is seen that all estimators performed well in terms of two criteria and take close values in case of large sample. Moreover, practical data applications are performed for all estimators. Results of the Kolmogorov-Smirnov statistics are reported for all estimators in practical applications.

Keywords: Unit log-log distribution, Estimation, Monte Carlo simuation, Kolmogrov-Smirnov statistic.

[a] *mcagatay@artvin.edu.tr* *https://orcid.org/0000-0003-3302-0705*
[b] *kkarakaya@selcuk.edu.tr* *https://orcid.org/0000-0002-0781-3587*
[c] *yakdogan@selcuk.edu.tr* *https://orcid.org/0000-0003-3520-7493*
[d] *uyener@cumhuriyet.edu.tr* *https://orcid.org/0000-0002-4796-8276*

Introduction

In recent years, the need to model proportional data has been increasing. Beta and Kumaraswamy [1] distributions are the best-known distributions for modeling these data. However, in recent years, many new distributions have been proposed as alternatives to these distributions. Some of these can be presented as [2-4]. One of the unit distributions proposed in recent years is the unit log-log (ULL) distribution introduced by Korkmaz and Korkmaz [5]. The ULL distribution was obtained by converting the log-log distribution [6] to the range of (0,1). The probability density function (pdf), cumulative distribution function of the ULL distribution are given, respectively, by

$$f(x,\alpha,\beta)=\frac{\alpha\log\beta(-\log x)^{\alpha-1}\beta^{(-\log x)^{\alpha}}}{x}e^{1-\beta^{(-\log x)^{\alpha}}},\ x\in(0,1) \qquad (1)$$

and

$$F(x,\alpha,\beta)=e^{1-\beta^{(-\log x)^{\alpha}}},x\in(0,1), \qquad (2)$$

where $\alpha>0$ and $\beta>1$ are the distribution parameters. The following is the quantile function of the ULL model, which we also use to generate numbers for the simulation study:

$$x_u(\alpha,\beta)=\exp\left[-\left(\frac{\log(1-\log u)}{\log\beta}\right)^{1/\alpha}\right], \qquad (3)$$

where $u\in(0,1)$. The pdf plots for some choices of α and β is presented in Figure 1. The ULL distribution has U-shaped, N-shaped, decreasing unimodal shapes and increasing, as shown in Figure 1.

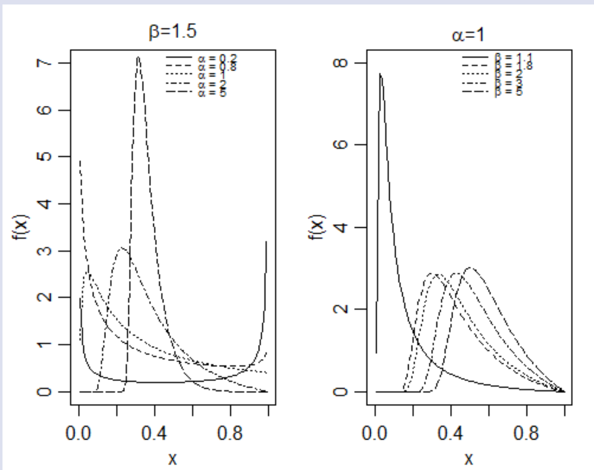

Figure 1. The pdf plots for ULL model

Some mathematical properties of the ULL model, including moments, stochastic ordering, order statistics, etc., were studied in detail by Korkmaz and Korkmaz [5]. Korkmaz and Korkmaz [5] proposed a new quantile regression model based on the ULL distribution as an alternative to beta regression [7], Kumaraswamy regression [8], and unit-Weibull regression [9]. Both parameters of ULL distribution and regression parameters were estimated by the maximum likelihood methodology in [5]. Korkmaz and Korkmaz [5] conducted comprehensive simulation experiments for the estimate of both the model parameters”” and the regression parameters under various scenarios. Korkmaz and Korkmaz [5] used only the maximum likelihood technique to estimate unknown parameters of the ULL model. The aim of the current research is to evaluate some estimators for the parameters of the ULL distribution in detail. The rest of the paper is organized as follows: Section 2 presents six estimators. In Section 3, a simulation experiment is conducted to assess the performance of these estimators based on bias and mean squared errord (MSE). Two real data applications are examined in Section 4. The study is completed with the concluding remarks in Section 5.

Estimation of Parameters Using Different Methods

In this section, six estimators are examined for the estimation of two parameters of the ULL model. These estimators are: maximum likelihood (MLE), least squares (LSE), maximum product spacing (MPS), weighted least squares (WLSE), Anderson-Darling (AD), and Cramer-von Mises (CVM). These estimators are often preferred for estimating unknown parameter of distributions. Some studies using this estimator can be given as [10-13].

Maximum Likelihood Estimation

In this subsection, we derive estimations of the parameters α and β via method of the MLE. Let $X_1, X_2, \ldots, X_n$ be a random sample from the ULL distribution with observed values $x_1, x_2, \ldots, x_n$, and Ξ $= (\alpha, \beta)^T$ be the vector of the model parameters. Then, the log-likelihood function is given by

$$\ell(\Xi) = n\log\alpha + n\log(\log\beta) + (\alpha-1)\sum_{i=1}^{n} 1'' + \log\beta\sum_{i=1}^{n}(-\log x_i)^{\alpha} + n - \sum_{i=1}^{n}\beta^{(-\log x_i)^{\alpha}}. \tag{4}$$

The MLE of α and β, say $\hat{\alpha}$ and $\hat{\beta}$, are derived by

$$\frac{\partial}{\partial\alpha}\ell(\Xi) = \frac{n}{\alpha} + \sum_{i=1}^{n}\log(-\log x_i) + \log\beta\sum_{i=1}^{n}(-\log x_i)^{\alpha}\log(-\log x_i)\left[1-\beta^{(-\log x_i)^{\alpha}}\right] = 0 \tag{5}$$

and $$\frac{\partial}{\partial\beta}\ell(\Xi) = \frac{n}{\beta\log\beta} + \frac{1}{\beta}\sum_{i=1}^{n}(-\log x_i)^{\alpha} - \sum_{i=1}^{n}(-\log x_i)^{\alpha}\beta^{(-\log x_i)^{\alpha}-1} = 0. \tag{6}$$

The $\hat{\alpha}$ and $\hat{\beta}$ are not obtained analytically from above equations. A software can be used in this regard, like R, Matlab, or Mathematica to obtain estimations via numerical methods.

Maximum Product Spacing Estimation

MPS technique was presented by [14]. Let be the $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ ordered statistics from ULL distribution with sample size *n* and $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the ordered observed values from the ULL distribution. Geometric mean (GM) of the differences is given as

$$GM = \sqrt[n+1]{\prod_{i=1}^{n+1}\left[F(x_{(i)},\alpha,\beta) - F(x_{(i-1)},\alpha,\beta)\right]}, \tag{7}$$

where $F(x_{(n+1)},\alpha,\beta) = 1$ and $F(x_{(0)},\alpha,\beta) = 0$. MPSE, $\hat{\alpha}_{MPS}$ and $\hat{\beta}_{MPS}$, of the α and β parameters are acquired by maximizing GM. If we take the logarithm of the above expression, we get the following:

$$MPS(\Xi) = \frac{1}{n+1}\sum_{i=1}^{n+1}\log\left[e^{1-\beta^{(-\log x_{(i)})^{\alpha}}} - e^{1-\beta^{(-\log x_{(i-1)})^{\alpha}}}\right]. \tag{8}$$

Least Squares Estimation

The LSE $\hat{\alpha}_{LSE}$ and $\hat{\beta}_{LSE}$, of α and β, respectively, are achieved by minimizing the following

$$LSE(\Xi) = \sum_{i=1}^{n}\left(e^{1-\beta^{(-\log x_{(i)})^{\alpha}}} - \frac{i}{n+1}\right)^2. \tag{9}$$

Weighted Least Squares Estimation

The WLSE, $\hat{\alpha}_{WLSE}$ and $\hat{\beta}_{WLSE}$, of the α and β parameters are achieved by minimizing the following

$$WLSE(\Xi)=\sum_{i=1}^{n}\frac{(n+2)(n+1)^2}{i(n-i+1)}\left(e^{1-\beta^{\left(-\log x_{(i)}\right)^{\alpha}}}-\frac{i}{n+1}\right)^2. \quad (10)$$

Anderson-Darling Estimation

The AD, $\hat{\alpha}_{AD}$, $\hat{\beta}_{AD}$, of the α, β parameters are obtained by minimizing

$$AD(\Xi)=-n-\sum_{i=1}^{n}\frac{2i-1}{n}\left[\log\left\{e^{1-\beta^{\left(-\log x_{(i)}\right)^{\alpha}}}\right\}+\log\left\{1-e^{1-\beta^{\left(-\log x_{(i)}\right)^{\alpha}}}\right\}\right]. \quad (11)$$

Cramer-von Mises Estimation

The CVM, $\hat{\alpha}_{CVM}$ and $\hat{\beta}_{CVM}$, of the α and β parameters are acquired by minimizing

$$CVM(\Xi)=\frac{1}{12n}+\sum_{i=1}^{n}\left[e^{1-\beta^{\left(-\log x_{(i)}\right)^{\alpha}}}-\frac{2i-1}{2n}\right]^2. \quad (12)$$

Because Equations (4), (8)-(12) involve non-linear functions, explicit forms of all estimators cannot be obtained directly. As a result, numerical tecniques like quasi-Newton and Newton-Raphson algorithms must be used to solve them.

Simulation Study

In this section, the success of the estimators for ULL model parameters are examined different sample size *n*. It is genetared $N=1000$ samples of size $n=20,25,\ldots,1000$ from the ULL model based on the true parameter values $\alpha=5$ and $\beta=5$. The constrOptim coomand in the R is used to acquire all estimations. In addition, for comparisons between the methods, we calculate the bias and MSE of the estimators. The bias and MSE are computed as for $h=\alpha$ or β

$$Bias_h(n)=\frac{1}{N}\sum_{i=1}^{N}\left(h_i-\hat{h}_i\right), \quad (13)$$

and

$$MSE_h(n)=\frac{1}{N}\sum_{i=1}^{N}\left(h_i-\hat{h}_i\right)^2, \quad (14)$$

respectively.

Figures 2-7 illustrate the outcomes of this simulation experiment. Figures 2-7 indicate that six estimators are consistent, as MSE, bias decrease and the empirical means are goes true parameters values with increasing *n*. In addition, six estimators are asymptotic unbiased. Quantity of biases and MSEs in the MLE methodology are initially higher than other methods, but they become very close as the sample size increases. Although the bias criteria suggests that LSE is the best estimator, as *n* increases, the bias of all estimators approach one another and finally reach zero. According to MSE criterion, MPS method is better than other methods. As a result, it can be concluded that any estimator can be chosen for the ULL model in case of large sample. For other parameter settings, same results can be achieved.

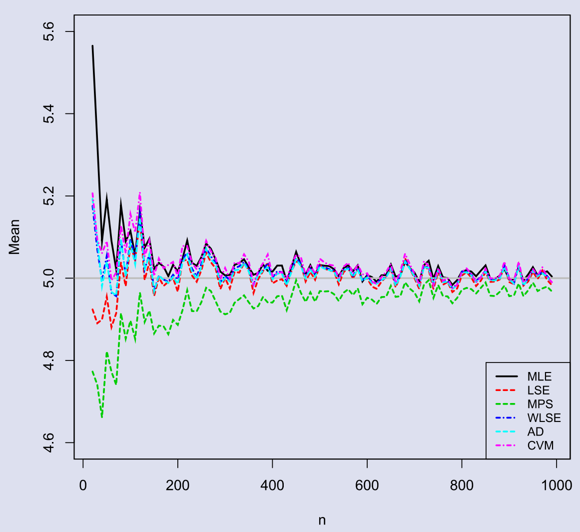

Figure 2. The empirical means of parameter for all estimators

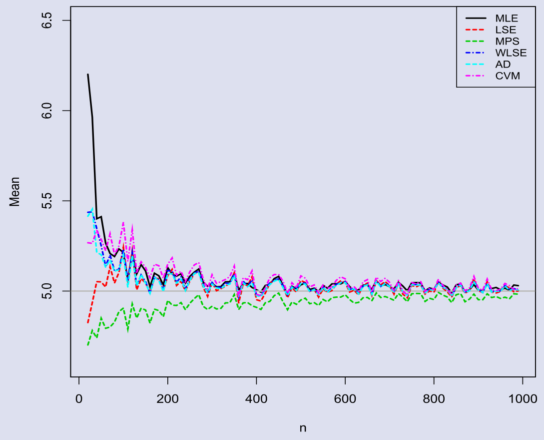

Figure 3. The empirical means of parameter for all estimators

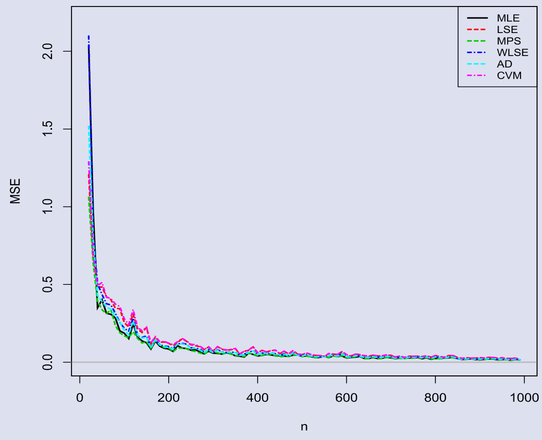

Figure 6. The MSEs of parameter α for all estimators

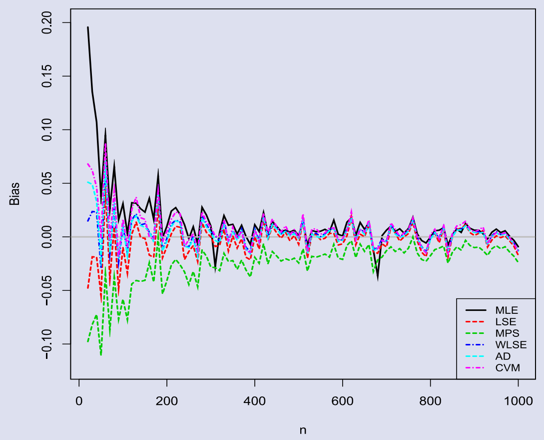

Figure 4. The biases of parameter α for all estimators

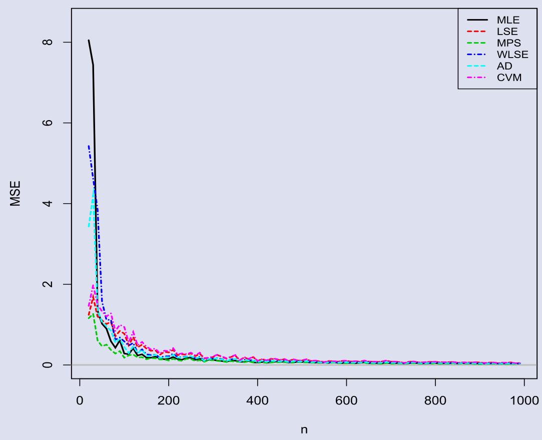

Figure 7. The MSEs of parameter β for all estimators

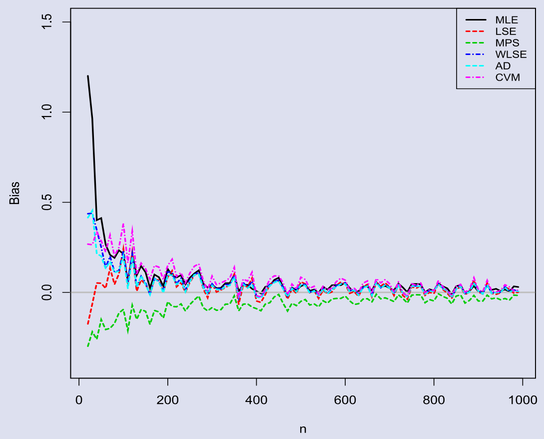

Figure 5. The biases of parameter β for all estimators

Applications Based on Real Data

In this section, two real data applications for ULL distribution are examined. The ULL distribution is modelled to two practical datasets by estimating the two unknown parameters using the all estimators discussed in Section 2. The MLE, LSE, QE, WLSE, AD, and CVM the parameters α and β of ULL distribution are obtained by BFGS algorithm. The results and Kolmogorov-Smirnov statistics (KS) and related p values for all estimators are reported in Table 1.

The first data is originates from [15] and indicates the Susquehanna River flood levels in Harrisburg, Pennsylvania. The data are 0.654, 0.613, 0.315, 0.449, 0.297, 0.402, 0.379, 0.423, 0.379, 0.3235, 0.269, 0.740, 0.418, 0.412, 0.494, 0.416, 0.338, 0.392, 0.484 and 0.265. For the first data, some descriptive statistics are as follows: there are 20 observations, the mean is 0.4231, the standard deviation is 0.1252, the median is 0.4070, the skewness is 1.1560, and the kurtosis is 1.1530.

The second data set taken from [16] and represents the strengths of 1.5 cm glass fibers, which were first measured by researchers at the UK National Physical Laboratory. The data are 0.17, 0.13, 0.16, 0.14, 0.20, 0.15, 0.13, 0.11, 0.15, 0.12, 0.12, 0.15, 0.12, 0.16, 0.21, 0.20, 0.23, 0.16, 0.12, 0.10, 0.32, 0.33, 0.33, 0.36, 0.38, 0.20 and 0.26. For the second data, some descriptive statistics are as follows: there are 27 observations, the mean is 0.1930, the standard deviation is 0.0831, the median is 0.1600, the skewness is 1.0710, and the kurtosis is -0.056.

Table 1: Results of the parameter estimations and KS for two data sets for all estimators

		Parameter		KS	
Data	**Estimators**	α	β	**Statistics**	**p-values**
Flood levels	MLE	2.9192	1.9334	0.1366	0.8493
	LSE	2.7806	1.9296	0.1360	0.8529
	MPS	2.5363	1.9388	0.1354	0.8099
	WLSE	2.6582	1.9170	0.1390	0.8339
	AD	2.8151	1.9312	0.1359	0.8535
	CVM	3.0179	1.9503	0.1317	0.8783
Glass fiber	MLE	4.3774	1.0422	0.1120	0.8868
	LSE	3.7747	1.0602	0.1063	0.9200
	MPS	3.9008	1.0574	0.1107	0.8589
	WLSE	3.8547	1.0583	0.1046	0.9291
	AD	3.9660	1.0544	0.1050	0.9271
	CVM	3.9888	1.0525	0.0990	0.9538

Concluding Remarks

The ULL distribution introduced by [5] is studied in this work with relation to six point estimation methods. For one parameter setting as $\alpha = 5$ and $\beta = 5$ and different sample sizes, simulations are performed. When the sample size is increased, the MSEs and biases are observed to decrease and approach zero. In addition, the estimations and KS outcomes for all estimators are examined for two real datasets. The simulation results show that all estimators perform well on the bias and MSE criteria and close each other in a large sample.

Conflicts of interest

The authors state that did not have a conflict of interests

References

[1] Kumaraswamy, P., A generalized probability density function for double-bounded random processes , *J. Hydrol.*, 46 (1980), 79-88.

[2] Gómez-Déniz, E., Sordo, M.A., Calderín-Ojeda, E., The log–lindley distribution as an alternative to the beta regression model with applications in insurance, *Insurance Math. Econ.*, 54 (2014), 49-57.

[3] Mazucheli, J., Menezes, A.F.B., Chakraborty, S., On the one parameter unit-Lindley distribution and its associated regression model for proportion data, *J. Appl. Stat.*, 46 (2019), 700-714.

[4] Korkmaz, M.Ç., Chesneau, C. On the unit burr-xii distribution with the quantile regression modeling and applications, *Comput. Appl. Math.*, 40 (2021), 1-26.

[5] Korkmaz, M.Ç., Korkmaz, Z.S., The unit log–log distribution: a new unit distribution with alternative quantile regression modeling and educational measurements applications*, Journal of Applied Statistics*, (2021) 1-20.

[6] Pham, H., A vtub-shaped hazard rate function with applications to system safety, *Int. J. Reliab. Appl.*, 3 (2002), 1–16.

[7] Ferrari, S., Cribari-Neto, F., Beta regression for modelling rates and proportions, *Journal of Applied Statistics*, 31(7) (2004) 799–815.

[8] Mitnik, P.A., Baek, S., The Kumaraswamy distribution: Median-dispersion re-parameterizations for regression modeling and simulation-based estimation, *Stat. Pap.*, 54 (2013) 177–192.

[9] Mazucheli, J., Menezes, A., Fernandes, L., de Oliveira, R., Ghitany, M., The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates, *Journal of Applied Statistics*, 47 (2020) 954–974.

[10] Kınacı, İ., Kuş, C., Karakaya, K., Akdoğan, Y., APT-Pareto Distribution and its Properties. *Cumhuriyet Science Journal,* (2019) 40 (2) 378-387.

[11] Karakaya, K., Tanış, C., Different methods of estimation for the one parameter Akash distribution*, Cumhuriyet Science Journal*, (2020) 41 (4) 944-950 .

[12] Tanış, C., Saraçoğlu, B., Kuş, C., Pekgör, A., Transmuted complementary exponential power distribution: properties and applications. *Cumhuriyet Science Journal*, (2020) 41 (2) 419-432.

[13] Hamedani, G.G., Korkmaz, M.Ç., Butt, N.S., Yousof, H.M., The Type I Quasi Lambert Family, *Pakistan Journal of Statistics and Operation Research*, 17(3) (2021) 545-558.

[14] Cheng, R.C.H., Amin, N.A.K., Maximum product of spacings estimation with application to the lognormal distribution, Math Report, (1979) 791.

[15] Dumonceaux, R., Antle, C.E., Discrimination between the log-normal and the Weibull distributions, *Technometrics*, 15(4) (1973) 923-926.

[16] Elgarhy, M., Exponentiated generalized Kumaraswamy distribution with applications, *Annals of Data Science*, 5(2) (2018) 273-292.