

DOUBLE LENGTH-BIASED WEIBULL DISTRIBUTION AND ITS APPLICATION

S. M. Aboukhamseen ^{1,§}, M. E. Ghitany ¹, D. K. Almutairi¹

¹ Department of Statistics and Operations Research
Faculty of Science, Kuwait University, Kuwait

Abstract

In this paper, we introduce a new double distribution, called double length-biased Weibull distribution. The distribution is generated using random sign mixture transform and has support on the whole real line. The main features of the distribution which include shape properties of the density and distribution functions as well as moments and related measures are presented. Entropy and extropy measures of the distribution are derived. Parameter estimates are derived using the maximum likelihood estimation technique and the efficiency of these estimates is assessed through a simulation study. An application to a standardized car prices data is presented to highlight the usefulness of the model.

MSC 2020: 62E15, 62F12

Key Words and Phrases: Random sign mixture transform; Bimodality; Maximum likelihood estimation; Simulations

1 Introduction

The Weibull distribution introduced in [12] has probability density function (PDF)

$$f_W(x; \alpha, \theta) = \frac{\alpha}{\theta^\alpha} x^{\alpha-1} e^{-(x/\theta)^\alpha}, \quad x > 0,$$

where $\alpha, \theta > 0$ are the shape and scale parameters, respectively. Its mean is $E(W) = \Gamma(1 + 1/\alpha) \theta$, where $\Gamma(\cdot)$ is the gamma function. The Weibull distribution may be regarded as a key probability distribution in statistics due to its extensive use in social sciences, reliability studies, and survival analysis. On the other hand, the length-biased Weibull (LBW) distribution has PDF

$$f_X(x; \alpha, \theta) = \frac{x f_W(x; \alpha, \theta)}{E(W)} = \frac{\alpha x^\alpha e^{-(x/\theta)^\alpha}}{\Gamma(1 + 1/\alpha) \theta^{\alpha+1}}, \quad x > 0. \quad (1)$$

For more details about LBW distribution, see [11] and [5].

To extended its applications for data on the whole real line, we follow the procedure presented by [4] to construct a double length-biased Weibull (DLBW) distribution using the random sign mixture transform (RSMT) given by

$$Z = YX_1 - (1 - Y)X_2, \quad (2)$$

where Y is a Bernoulli random variable (RV) with parameter $\beta \in (0, 1)$, X_1 and X_2 are independent LBW RVs independent of Y .

Using the RSMT, several double distributions appeared in the literature. For example, [3], [1], and [2] introduced double inverse-Gaussian, double log-normal, and double Chaudhry-Ahmad distributions, respectively.

The contents of this article are arranged as follows. The main statistical features of the DLBW distribution are covered in Section 2. The maximum likelihood estimates (MLEs) and their asymptotic distributions are provided in Section 3. A simulation study is conducted in Section 4 to examine the performance of the MLEs. The proposed DLBW distribution is fitted to a real dataset in Section 5. Lastly, in Section 6, the conclusions and remarks are discussed.

2 Statistical properties

2.1 Density and distribution functions

The PDF of DLBW distribution is

$$f_Z(z) = \begin{cases} \bar{\beta} f_{X_2}(|z|; \alpha_2, \theta_2), & z < 0, \\ \beta f_{X_1}(z; \alpha_1, \theta_1), & z \geq 0, \end{cases} \quad (3)$$

where $\bar{\beta} = 1 - \beta$ and, for $\alpha_j, \theta_j > 0$, $j = 1, 2$,

$$f_{X_j}(x; \alpha_j, \theta_j) = \frac{\alpha_j}{\Gamma(1 + 1/\alpha_j) \theta_j^{\alpha_j+1}} x^{\alpha_j} e^{-(x/\theta_j)^{\alpha_j}}, \quad x > 0, \quad (4)$$

are the PDFs of the LBW distributions.

The DLBW distribution has two modes given by

$$\text{Mode}(Z) = -\text{Mode}(X_2) \text{ and } \text{Mode}(X_1), \quad (5)$$

where, for $j = 1, 2$,

$$\text{Mode}(X_j) = \theta_j, \quad (6)$$

are the modes of the LBW distributions.

2.2 Cumulative distribution function

The cumulative distribution function (CDF) of DLBW distribution is given by

$$F_Z(z) = P(Z \leq z) = \begin{cases} \bar{\beta} [1 - F_{X_2}(|z|; \alpha_2, \theta_2)], & z < 0, \\ \bar{\beta} + \beta F_{X_1}(z; \alpha_1, \theta_1), & z \geq 0, \end{cases} \quad (7)$$

where, for $j = 1, 2$,

$$\begin{aligned} F_{X_j}(x; \alpha_j, \lambda_j) &= P(X_j \leq x) \\ &= 1 - \frac{\Gamma(1 + 1/\alpha_j; (x/\theta_j)^{\alpha_j})}{\Gamma(1 + 1/\alpha_j)}, \quad x > 0, \end{aligned} \quad (8)$$

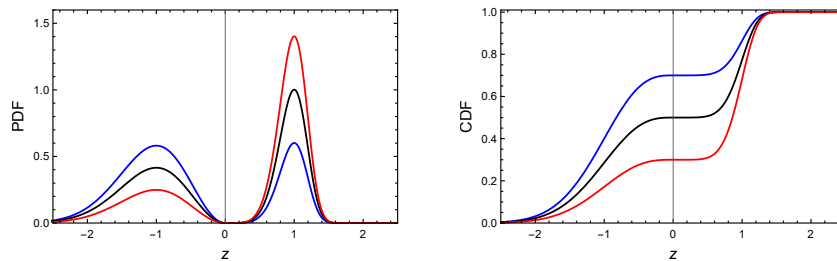


Figure 1: PDF and CDF of DLBW distribution: $(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2)$: $(0.3, 5, 1, 2, 1)$ (—), $(0.5, 5, 1, 2, 1)$ (—), $(0.7, 5, 1, 2, 1)$ (—).

are the CDFs of the LBW distributions and

$$\Gamma(a; t) = \int_t^\infty y^{a-1} e^{-y} dy, \quad a, t > 0, \quad (9)$$

is the upper incomplete gamma function.

Figure 1 shows the the PDF and CDF of the DLBW distribution for selected values of the parameters. This figure also shows that the PDF curve has two modes -1 and 1. The PDF curve on the negative (positive) side gets lower (higher) as β increases. The CDF curve gets lower as β increases.

2.3 Moments

The r th moment of the DLBW distribution is given by

$$E(Z^r) = \beta E(X_1^r) + (-1)^r \bar{\beta} E(X_2^r), \quad r \geq 1, \quad (10)$$

where, for $j = 1, 2$,

$$E(X_j^r) = \frac{(r+1)\Gamma((r+1)/\alpha_j)}{\Gamma(1/\alpha_j)} \theta_j^r, \quad (11)$$

are the r th moments of LBW distributions.

In particular, the first two raw moments of DLBW distribution, respectively, are

$$E(Z) = \beta \frac{2\Gamma(2/\alpha_1)}{\Gamma(1/\alpha_1)} \theta_1 - \bar{\beta} \frac{2\Gamma(2/\alpha_2)}{\Gamma(1/\alpha_2)} \theta_2,$$

$$E(Z^2) = \beta \frac{3\Gamma(3/\alpha_1)}{\Gamma(1/\alpha_1)} \theta_1^2 + \bar{\beta} \frac{3\Gamma(3/\alpha_2)}{\Gamma(1/\alpha_2)} \theta_2^2.$$

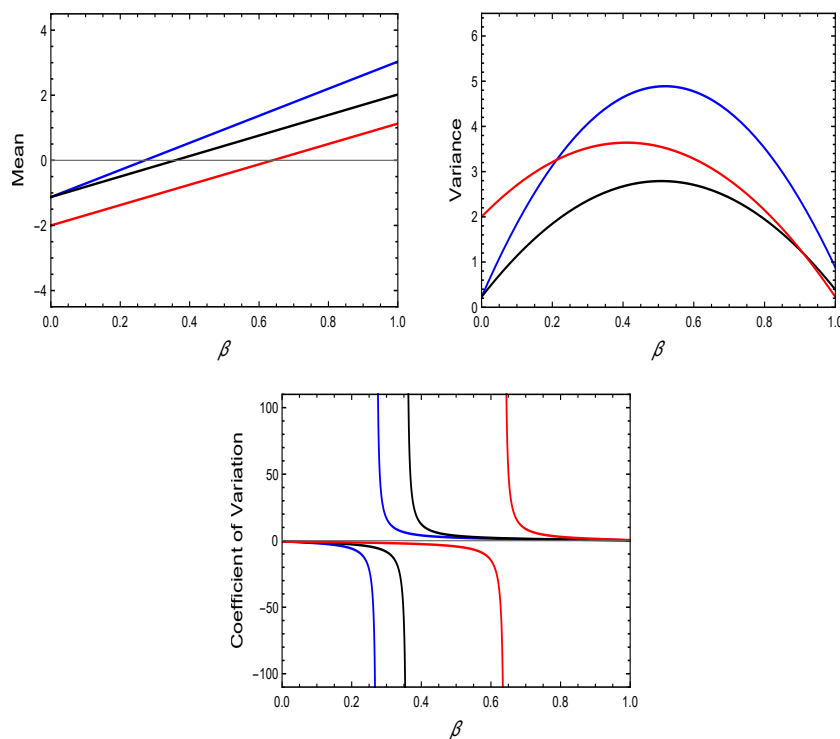


Figure 2: Mean, variance and coefficient of variation of the DLBW distribution as a function in β : $(\alpha_1, \theta_1, \alpha_2, \theta_2) : (3, 3, 2, 1)$ (—), $(3, 2, 2, 1)$ (—), $(2, 1, 1, 1)$ (—).

Figure 2 shows the mean $\mu = E(Z)$, variance $\sigma^2 = E(Z^2) - [E(Z)]^2$ and coefficient of variation σ/μ of the DLBW distribution

as a function in β for selected values of the parameters $(\alpha_1, \theta_1, \alpha_2, \theta_2)$. This figure also shows that the mean (variance) is a linear (quadratic) function in β .

2.4 Entropy and extropy

Entropies are measures of a system's variation, instability, or uncertainty.

For a RV V with PDF $f_V(v)$, the following are two well known entropies:

1. Tsallis entropy: (see [10])

$$T_s(V) = \frac{1}{s-1} \{1 - E[f_V^{s-1}(V)]\}, \quad 0 < s \neq 1.$$

2. Shannon entropy: (see [8])

$$H(V) = E[-\ln f_V(V)] = \lim_{s \rightarrow 1} T_s(V).$$

Tsallis entropy of DLBW distribution is given by

$$T_s(Z) = T_s(Y) + \beta^s T_s(X_1) + \bar{\beta}^s T_s(X_2), \quad (12)$$

where

$$T_s(Y) = \frac{1 - \beta^s - \bar{\beta}^s}{s-1}, \quad (13)$$

is Tsallis entropy of Bernoulli distribution and, for $j = 1, 2$,

$$T_s(X_j) = \frac{1}{s-1} \left\{ 1 - \frac{\alpha_j^{2s-1} \theta_j^{1-s} \Gamma(s+1/\alpha_j)}{s^{s+1/\alpha_j} [\Gamma(1/\alpha_j)]^s} \right\}, \quad (14)$$

are Tsallis entropies of the LBW distributions.

Shannon entropy of DLBW distribution is given by

$$H(Z) = H(Y) + \beta H(X_1) + \bar{\beta} H(X_2), \quad (15)$$

where

$$H(Y) = -\beta \ln(\beta) - \bar{\beta} \ln(\bar{\beta}), \quad (16)$$

is Shannon entropy of Bernoulli distribution and, for $j = 1, 2$,

$$H(X_j) = 1 + 1/\alpha_j - \ln \left[\frac{\alpha_j}{\theta_j \Gamma(1 + 1/\alpha_j)} \right] - \psi(1 + 1/\alpha_j), \quad (17)$$

are Shannon entropies of the LBW distributions.

Alternative measure of uncertainty, called extropy, is given in [6].

For a RV V with PDF $f_V(v)$, the extropy of V is given by

$$J(V) = -\frac{1}{2} E[f_V(V)] = -\frac{1}{2} [1 - T_2(V)].$$

The extropy of DLBW distribution is given by

$$J(Z) = \beta^2 J(X_1) + \bar{\beta}^2 J(X_2), \quad (18)$$

where, for $j = 1, 2$,

$$J(X_j) = -\frac{\alpha_j(1 + \alpha_j)}{2^{3+1/\alpha_j} \Gamma(1/\alpha_j) \theta_j}, \quad (19)$$

are extropies of the LBW distributions.

Figure 3 shows that Tsallis and Shannon entropies of DLBW distribution can be negative/positive but the extropy, by definition, is always negative.

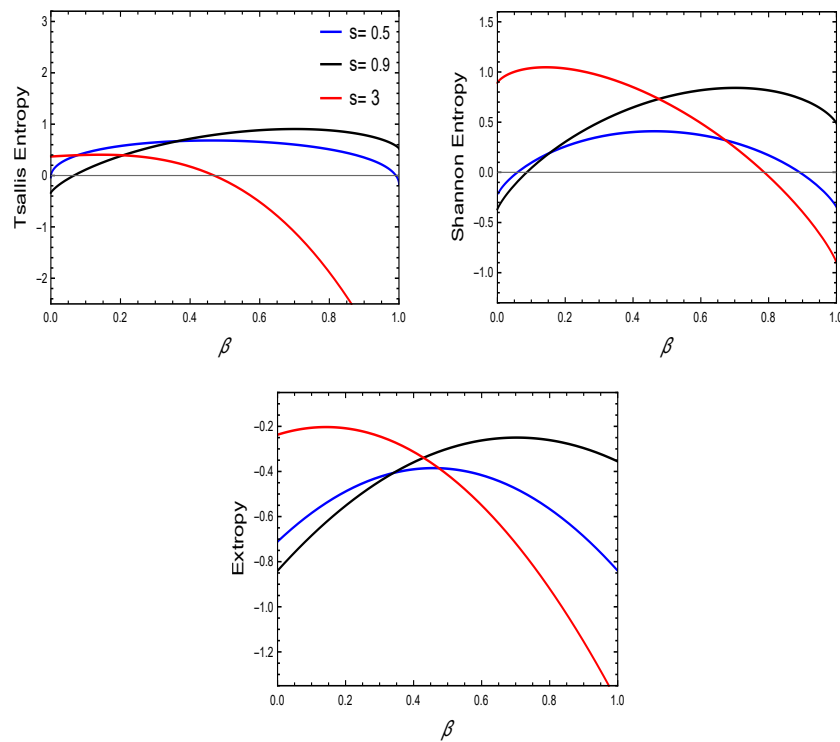


Figure 3: Tsallis, Shannon entropies and extropy of DLBW distribution as a function in β : $(\alpha_1, \theta_1, \alpha_2, \theta_2) : (6, 1, 5, 1)$ (—), $(5, 2, 6, 1)$ (—), $(5, 0.5, 5, 3)$ (—).

3 Maximum likelihood estimation

In this section, we derive the MLEs of the parameters of DLBW distribution and their asymptotic distributions.

Let $z_1, z_2, \dots, z_n$ be a random sample from $DLBW(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2)$ distribution. The log-likelihood function is given by

$$\begin{aligned} \ln L(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2) &= \sum_{z_i > 0} \ln[\beta f_{X_1}(z_i; \alpha_1, \theta_1)] \\ &+ \sum_{z_i < 0} \ln[\bar{\beta} f_{X_2}(|z_i|; \alpha_2, \theta_2)]. \end{aligned} \quad (20)$$

(i) MLE of β :

$$\hat{\beta} = n_1/n, \quad (21)$$

where n_1 is the number of positive z_i 's,

(ii) MLE of α_1 : $\hat{\alpha}_1$ is the solution, in α_1 , of the equation:

$$\frac{1}{\alpha_1} \psi(1+1/\alpha_1) + \frac{\alpha_1 C_1}{n_1} + \frac{1}{\alpha_1} \ln \left[\frac{M_1(\alpha_1)}{n_1(1+1/\alpha_1)} \right] - (\alpha_1+1) \frac{L_1(\alpha_1)}{M_1(\alpha_1)} + 1 = 0, \quad (22)$$

where

$$C_1 = \sum_{z_i > 0} \ln z_i, \quad L_1(\alpha_1) = \sum_{z_i > 0} z_i^{\alpha_1} \ln z_i, \quad M_1(\alpha_1) = \sum_{z_i > 0} z_i^{\alpha_1}. \quad (23)$$

(iii) MLE of θ_1 :

$$\hat{\theta}_1 = \left[\frac{M_1(\hat{\alpha}_1)}{n_1(1+1/\hat{\alpha}_1)} \right]^{1/\hat{\alpha}_1}. \quad (24)$$

(iv) MLE of α_2 : $\hat{\alpha}_2$ is the solution, in α_2 , of the equation:

$$\frac{1}{\alpha_2} \psi(1+1/\alpha_2) + \frac{\alpha_2 C_2}{n_2} + \frac{1}{\alpha_2} \ln \left[\frac{M_2(\alpha_2)}{n_2(1+1/\alpha_2)} \right] - (\alpha_2+1) \frac{L_2(\alpha_2)}{M_2(\alpha_2)} + 1 = 0, \quad (25)$$

where $n_2 = n - n_1$ and

$$C_2 = \sum_{z_i < 0} \ln |z_i|, \quad L_2(\alpha_1) = \sum_{z_i < 0} |z_i|^{\alpha_2} \ln |z_i|, \quad M_2(\alpha_1) = \sum_{z_i < 0} |z_i|^{\alpha_2}. \quad (26)$$

(v) MLE of θ_2 :

$$\hat{\theta}_2 = \left[\frac{M_2(\hat{\alpha}_2)}{n_2(1 + 1/\hat{\alpha}_2)} \right]^{1/\hat{\alpha}_2}. \quad (27)$$

The Fisher information matrix about $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$ is given by

$$\mathbf{I}(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2) = \text{diag} \left(I_Y(\beta), \beta \mathbf{I}_{X_1}(\alpha_1, \theta_1), \bar{\beta} \mathbf{I}_{X_2}(\alpha_2, \theta_2) \right), \quad (28)$$

where $I_Y(\beta) = \frac{1}{\beta \bar{\beta}}$ is the Fisher information about β ,

$$\mathbf{I}_{X_j}(\alpha_j, \theta_j) = \begin{bmatrix} I_{11}(\alpha_j, \theta_j) & I_{12}(\alpha_j, \theta_j) \\ I_{12}(\alpha_j, \theta_j) & I_{22}(\alpha_j, \theta_j) \end{bmatrix}, \quad (29)$$

are the Fisher information matrices about $(\alpha_j, \theta_j), j = 1, 2$, and

$$I_{11}(\alpha, \theta) = \frac{(1 + \alpha) [\psi^2(2 + 1/\alpha) + \psi'(2 + 1/\alpha)]}{\alpha^3} + \frac{2\alpha[\alpha + \psi(1/\alpha)] + \psi'(1/\alpha)}{\alpha^4}, \quad (30)$$

$$I_{12}(\alpha, \theta) = -\frac{1 + (1 + \alpha)\psi(2 + 1/\alpha)}{\alpha\theta}, \quad (31)$$

$$I_{22}(\alpha, \theta) = \frac{\alpha(1 + \alpha)}{\theta^2}. \quad (32)$$

Moreover, the asymptotic distribution of the MLEs is given by:
As $n \rightarrow \infty$,

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\alpha}_1 - \alpha_1 \\ \hat{\theta}_1 - \theta_1 \\ \hat{\alpha}_2 - \alpha_2 \\ \hat{\theta}_2 - \theta_2 \end{bmatrix} \xrightarrow{d} MVN \left(\mathbf{0}, \mathbf{I}^{-1}(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2) \right) \quad (33)$$

where $\xrightarrow{d}$ denotes *convergence in distribution* and *MVN* stands for *multivariate normal distribution*.

4 Simulation study

The generalized gamma (**GG**) distribution (Stacy, 1962) has PDF

$$f_X(x; a, b, k) = \frac{b}{\Gamma(k)} \frac{x^{b k - 1}}{a^{b k}} e^{-(x/a)^b}, \quad x > 0, \quad a, b, k > 0.$$

We understand that the $\text{GG}(a = \theta, b = \alpha, k = 1 + 1/\alpha)$ distribution is equivalent to the $\text{LBW}(\alpha, \theta)$ distribution. This link is quite beneficial for any calculations related to the LBW distribution, as the GG distribution is available in the **library(ggamma)** of the R programming language [7].

The objective of this section is to conduct a simulation study to assess the finite-sample behavior of the MLEs of the parameters of the proposed DLBW distribution. This study was conducted over 10,000 iterations. In each iteration, a random sample of sizes $n = 50, 100, \dots, 500$ is drawn from the DLBW distribution with true parameters $(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2) = (0.4, 0.5, 1, 1, 0.5), (0.5, 1, 0.5, 0.5, 1), (0.6, 2, 1, 1, 2), (0.7, 2, 1, 2, 1)$, using the following procedure:

1. Generate $Y_i \sim \text{Bernoulli}(\beta)$, $i = 1, 2, \dots, n$;
2. Generate $X_{1,i} \sim \text{GG}(\theta_1, \alpha_1, 1 + 1/\alpha_1)$, $i = 1, 2, \dots, n$;
3. Generate $X_{2,i} \sim \text{GG}(\theta_2, \alpha_2, 1 + 1/\alpha_2)$, $i = 1, 2, \dots, n$;
4. Set $Z_i = Y_i X_{1,i} - (1 - Y_i) X_{2,i}$, $i = 1, 2, \dots, n$.

The study evaluates the simulation metrics, which include average bias (AB), mean square error (MSE) of the MLEs, and the coverage probability (CP) for the 95% confidence intervals of each parameter.

Figure 4 shows that the average bias of the MLE of β can be negative/positive while the biases of the remaining MLEs are positive. All average biases are small. Figure 5 shows that the MSE of the MLEs of the parameters are small and decrease as n increases. Figure 6 shows that the coverage probability of 95% confidence interval of each parameter is close to the nominal level of 95%.

The last remarks about Figures 4 - 6 demonstrate that the MLEs of the parameters of the DLBW distribution are well behaved for point estimation and confidence intervals.

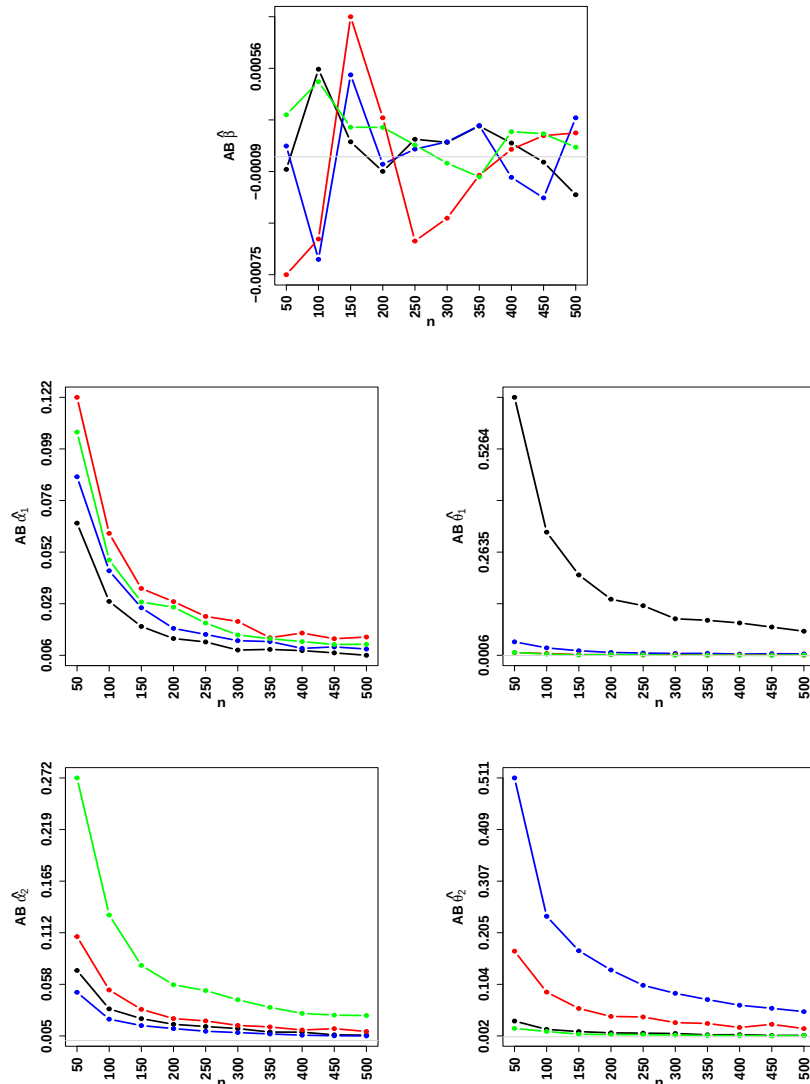


Figure 4: Average bias of the MLEs of the parameters of DLBW distribution: $(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2) : (0.4, 0.5, 1, 1, 0.5)$ (—), $(0.5, 1, 0.5, 0.5, 1)$ (—), $(0.6, 2, 1, 1, 2)$ (—), $(0.7, 2, 1, 2, 1)$ (—).

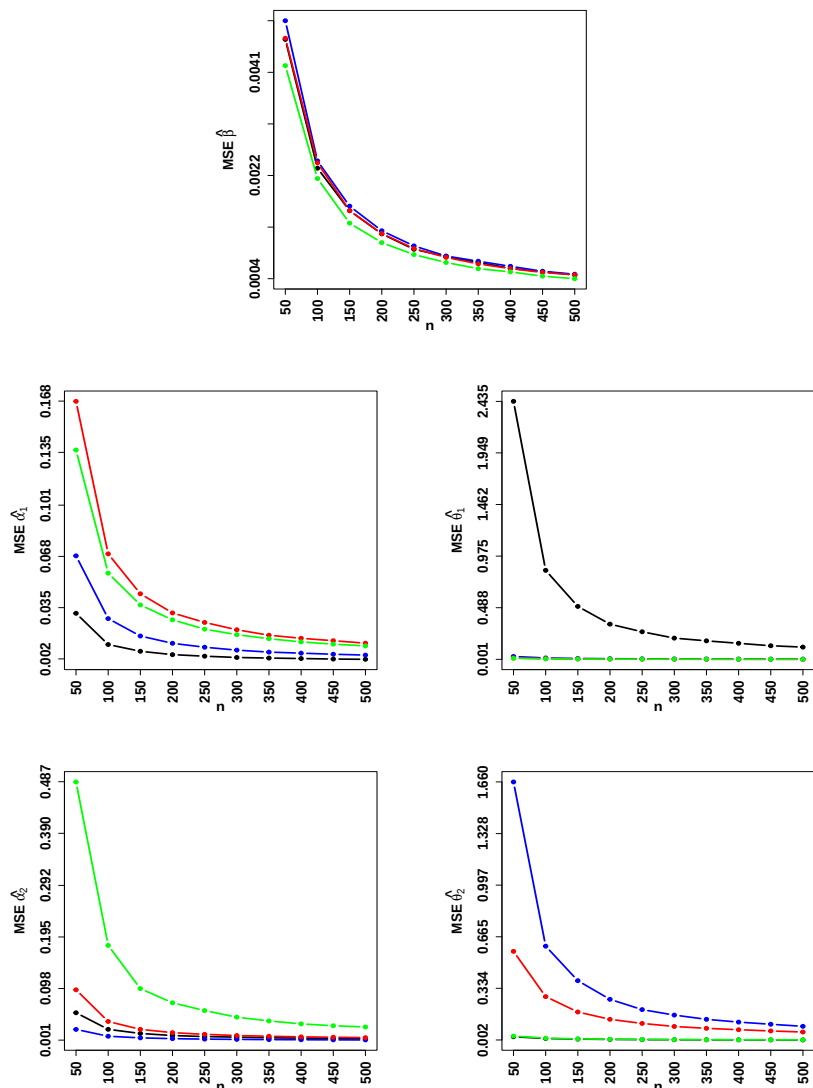


Figure 5: MSE of the MLEs of the parameters of DLBW distribution: $(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2)$: $(0.4, 0.5, 1, 1, 0.5)$ (—), $(0.5, 1, 0.5, 0.5, 1)$ (—), $(0.6, 2, 1, 1, 2)$ (—), $(0.7, 2, 1, 2, 1)$ (—).

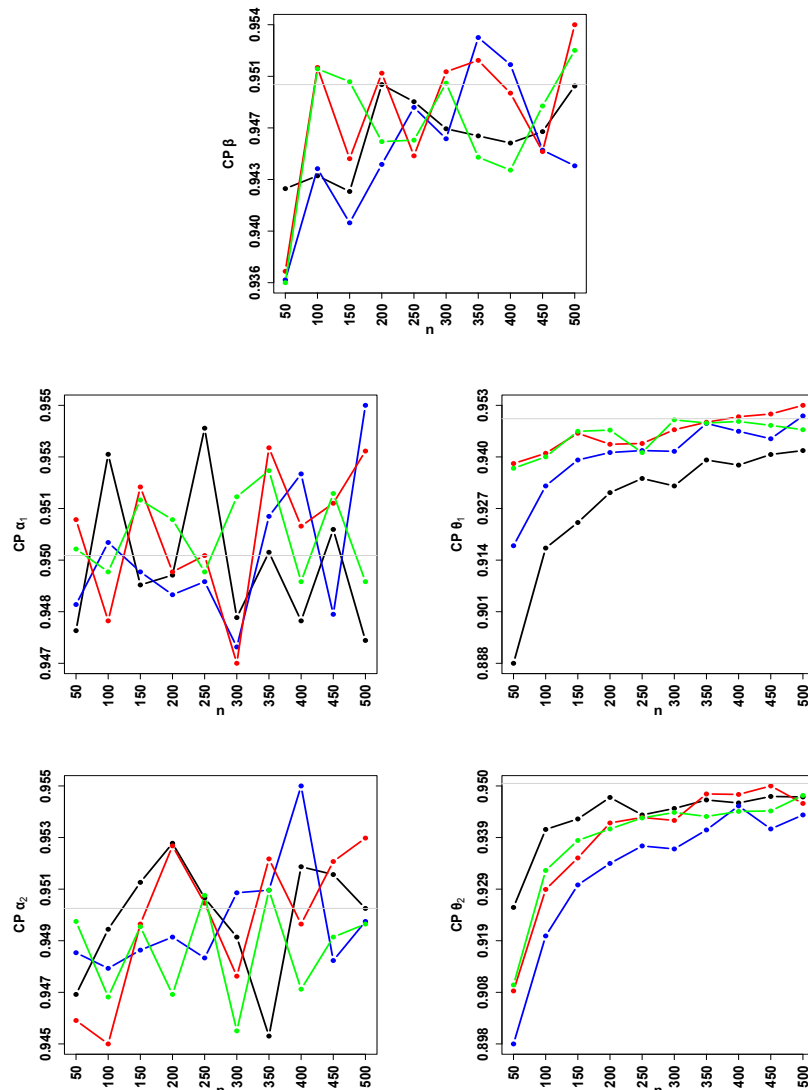


Figure 6: CP of the 95% confidence intervals of the parameters of DLBW distribution: $(\beta, \alpha_1, \theta_1, \alpha_2, \theta_2) : (0.4, 0.5, 1, 1, 0.5)$ (—), $(0.5, 1, 0.5, 0.5, 1)$ (—), $(0.6, 2, 1, 1, 2)$ (—), $(0.7, 2, 1, 2, 1)$ (—).

5 Application

Car prices data, $x_1, x_2, \dots, x_{135}$, in Polish currency (in thousands) of 135 cars of Toyota Verso produced in 2013-2016 offered for sale by Polish automotive portals are reported in[9]. We standardized this data using the transformation $z_i = (x_i - \bar{x})/s$, $i = 1, 2, \dots, 135$, where $\bar{x} = 49.065$ and $s = 10.235$ are the sample mean and standard deviation of car prices, respectively.

Figure 7 shows the density histogram of the data $z_1, z_2, \dots, z_{135}$. This figure also shows that this data is bimodal with one mode on each side of the origin.

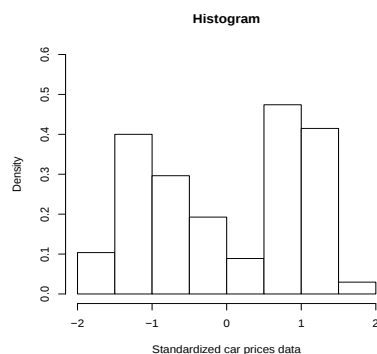


Figure 7: Density histogram of the standardized car prices data.

Table 1: Summary of the fitted DLBW model for the standardized car prices data.

Parameter	MLE	S.E.	$\ln \hat{L}$
β	0.504	0.043	-151.635
α_1	2.668	0.297	
θ_1	0.883	0.048	
α_2	1.600	0.205	
θ_2	0.732	0.079	

Table 2: Goodness-of-fit tests of the fitted DLBW model.

Goodness-of-fit test	Statistic	p -value
Kolmogrov-Smirnov	0.060	0.717
Anderson-Darling	0.358	0.889
Cramer-von Misses	0.048	0.891

Table 1 provides a summary of the fitted DLBW model while Table 2 shows the suitability of the fitted DLBW model. Figure 8 provides extra diagnostic plots which confirm the suitability of the fitted DLBW model.

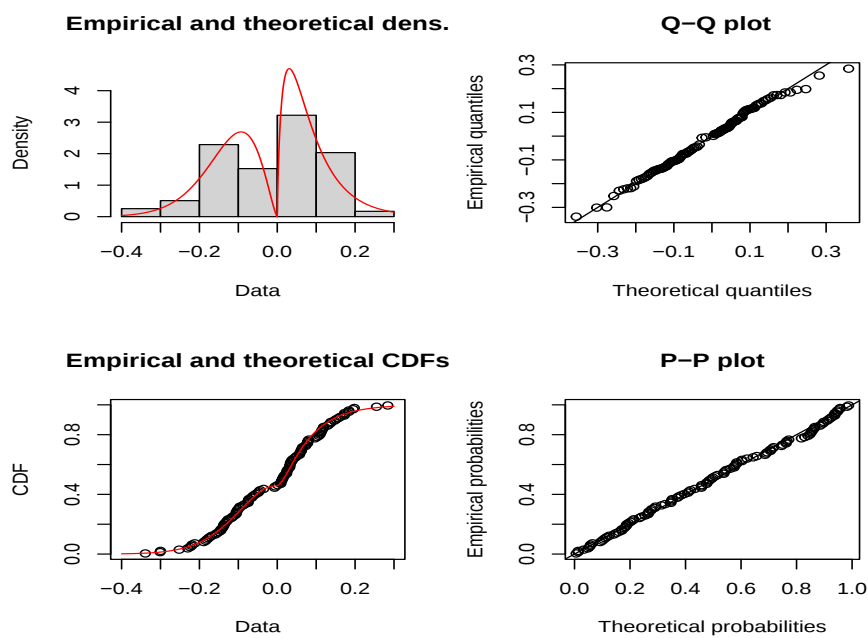


Figure 8: Diagnostic plots of the fitted DLBW model.

6 Conclusion and comments

The double length-biased Weibull model is introduced here through a random sign mixture transformation procedure. Various statistical characteristics of the proposed model are discussed. The maximum likelihood estimation technique is employed to estimate the parameters of the model. A simulation study is conducted to evaluate the accuracy of these estimates. The application of the proposed model to a real dataset consisting of standardized car prices demonstrates a very good fit. Future research may investigate Bayesian estimation method or address censored data. We believe that data analysts will find the model presented here to be beneficial.

References

- [1] M. F. Alfahad, M. E. Ghitany, A. N. Alothman, S. Nadarajah, A bimodal extension of the log-normal distribution on the real line with an application to DNA microarray data. *Mathematics* **11**(15), (2023), 3360.
- [2] H. M. Alkhezi, M.E. Ghitany, M. F. Alfahad, J. Mazucheli, A new bimodal double distribution on the real line and its application. *International Journal of Applied Mathematics* **37**(6), (2024), 627- 648.
- [3] A. Almutairi, M. E. Ghitany, A. N. Alothman, R. C. Gupta, Double inverse-Gaussian distributions and associated inference. *Journal of the Indian Society for Probability and Statistics* **24**, (2023), 157-182.
- [4] E. Aly, A unified approach for developing Laplace-type distributions. *Journal of the Indian Society for Probability and Statistics* **19**, (2018), 245-269.

- [5] D. E. Giles, Improved maximum likelihood estimation for the Weibull distribution under length-biased sampling. *Journal of Quantitative Economics* **19**, (2021), 59–77.
- [6] F. Lad, G. Sanfilippo, G. Agro, Extropy: Complementary dual of entropy. *Statistical Science* **30** (1), (2015), 40–58.
- [7] R Core Team R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2023.
- [8] C. E. Shannon, A mathematical theory of communication. *Bell System Technical Journal* **27** (3) (1948), 379-423.
- [9] P. Sulewski, Two-piece power normal distribution. *Communications in Statistics-Theory and Methods* **50**, (2021), 2619-2639.
- [10] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **52**, (1988), 479-487.
- [11] G. Tzavelas, D. Panagiotakos, Statistical inference for the size-biased Weibull distribution. *Journal of Statistical Computation and Simulation* **83**(7), (2013), 1252–1265.
- [12] W. Weibull, A statistical distribution function of wide applicability. *Journal of Applied Mechanics* **18**, (1951), 293–296.