



## A new extended log-Weibull regression: Simulations and applications

Gauss M. Cordeiro<sup>1</sup> , Muhammad. H. Tahir<sup>2</sup> , Julio C.S. Vasconcelos<sup>\*3</sup> ,  
Edwin M.M. Ortega<sup>3</sup> , M. Adnan Hussain<sup>2</sup> 

<sup>1</sup>Department of Statistics, Federal University of Pernambuco, Recife, PE, Brazil

<sup>2</sup>Department of Statistics, The Islamia University of Bahawalpur, Bahawalpur-63100, Pakistan

<sup>3</sup>Department of Exact Sciences, University of São Paulo, Piracicaba, SP, Brazil

### Abstract

The induction of one or more parameter(s) in parent distributions opened new doors for flexible modeling in modern distribution theory. Among well-established generalized (G) classes for flexible modeling, the exponentiated-G, Marshall-Olkin-G and odd log-logistic-G families offer induction of one additional parameter while the beta-G and Kumaraswamy-G classes offer two extra shape parameters. The Marshall-Olkin-odd-log-logistic-G (MOOLL-G) family serves as an alternative to the beta-G and Kumaraswamy-G classes. A new motivation for the MOOLL-G family for competing risk scenarios, some useful properties, and parameter estimation are addressed. The new log-MOOLL-Weibull regression is useful for analysis of real life data. The accuracy of the estimates and the residuals is addressed via Monte Carlo simulations. The presented models outperform some other well-known models.

**Mathematics Subject Classification (2020).** 62E15, 62F10, 62J02

**Keywords.** generalized family, Marshall-Olkin method, maximum likelihood, regression model, Weibull distribution

### 1. Introduction

The induction of one or more parameter(s) to parent distributions encourage new concepts for flexible modeling in distribution theory. Among well-established generalized (G) classes, the exponentiated-G, Marshall-Olkin-G and odd log-logistic-G families offer induction of one extra parameter while the beta-G and Kumaraswamy-G classes offer two additional shape parameters. In the same direction, Nadarajah *et al.* [17] reviewed the exponentiated Weibull distribution and Tahir and Nadarajah [24] presented a survey on several extended classes.

\*Corresponding Author.

Email addresses: gausscordeiro@gmail.com (G.M. Cordeiro), mtahir.stat@gmail.com (M.H. Tahir), juliocezar@usp.br (J.C.S. Vasconcelos), edwin@usp.br (E.M.M. Ortega), adnanidc@gmail.com (M.A. Hussain)

Received: 14.08.2020; Accepted: 05.01.2021

Marshall and Olkin [15] pioneered a G-class by which flexibility is introduced through adding a positive shape parameter  $b$ . The cumulative distribution function (cdf) of the Marshall-Olkin-G (MO-G) class is

$$F_{\text{MOG}}(x) = \frac{H_G(x)}{1 - (1 - b)[1 - H_G(x)]}, \quad x \in \mathbb{R}, \quad (1.1)$$

where  $H_G(x)$  represents the cdf of the baseline G.

Let  $G(x) = G(x; \xi)$  be a baseline cdf with associated vector of parameters  $\xi$ . The cdf of the generalized log-logistic family, which is actually the odd log-logistic-G (OLL-G) class, can be expressed as [10] (for  $x \in \mathbb{R}$ )

$$H_{\text{OLLG}}(x) = \frac{G(x)^a}{G(x)^a + \overline{G}(x)^a}, \quad (1.2)$$

where  $\overline{G}(x) = 1 - G(x)$  and  $a > 0$ . If  $W$  is the logistic random variable (rv) with shape parameter  $a > 0$ , then the rv  $V = G^{-1}(e^W/[1 + e^W])$  has the OLL-G cdf (1.2). This is a very simple interpretation of the OLL-G class.

The extended generalized log-logistic family with two additional parameters is reported by [2] and [11] as the odd-log-logistic Marshall-Olkin-G (OLLMO-G) class. In fact, the OLLMO-G class is the Marshall-Olkin-odd-log-logistic-G (MOOLL-G) family, which follows by inserting (1.2) in Equation (1.1), having cdf (for  $x \in \mathbb{R}$ )

$$F_{\text{MOOLLG}}(x) = F_{\text{MOOLLG}}(x; a, b, \xi) = \frac{G(x)^a}{G(x)^a + b\overline{G}(x)^a}. \quad (1.3)$$

The probability density function (pdf) corresponding to (1.3) has the form

$$f_{\text{MOOLLG}}(x) = \frac{a b g(x) G(x)^{a-1} \overline{G}(x)^{a-1}}{\{G(x)^a + b\overline{G}(x)^a\}^2}. \quad (1.4)$$

Henceforth,  $X$  denotes a rv with density (1.4). The quantile function (qf) of  $X$ , say  $x = Q_{\text{MOOLLG}}(u)$ , for a given baseline G, follows by inverting Equation (1.3) as

$$x = G^{-1} \left\{ \frac{(bu)^{1/a}}{(bu)^{1/a} + (1-u)^{1/a}} \right\}, \quad u \in (0, 1). \quad (1.5)$$

**Remark 1.1.** It is clear that the MOOLL-G and MO-G classes are the same when  $a = 1$  and that the MOOLL-G and OLL-G classes are identical when  $b = 1$ . We emphasize that for  $b = 1$ , it follows the model proposed by [11], discussed and generalized by [4]. Further,  $f_{\text{MOOLLG}}(x)$  reduces to  $g(x)$  when  $a = b = 1$ , and then  $f_{\text{MOOLLG}}(x)$  can be much more flexible than  $g(x)$  for different values of  $a$  and  $b$ .

**Remark 1.2.** Alizadeh *et al.* [2] investigated some properties of the OLLMO-Lindley and OLLMO-power Lindley models (in fact the MOOLL-Lindley and MOOLL-power Lindley models). Recently, Lima *et al.* [14] studied a variant of the MOOLL-G class, called the OLL-Geometric-G (OLL-Geo-G) family, for a system of components arranged in a series structure, whose cdf takes the form (for  $x \in \mathbb{R}$ )

$$F_{\text{OLLGeoG}}(x) = \frac{G(x)^a}{G(x)^a + (1-p)\overline{G}(x)^a}, \quad a > 0, p \in (0, 1). \quad (1.6)$$

Lima *et al.* [14] obtained some mathematical properties for (1.6) such as a linear density expansion, qf, moments and generating function, and discussed parameter estimation. They provided a simulation study and proposed the OLL-Geometric-Weibull regression. Further, Pratavia *et al.* [20] introduced the OLL-Geometric-normal distribution and explored the OLL-Geometric-normal regression for some useful diagnostic measures.

In this work, it is defined the extended generalized log-logistic family with two additional parameters. One of the advantages of the family is the flexibility to accommodate different forms of failure rate and, also, to discriminate models, since it has as special cases, the classes introduced by [8] and [11]. It is constructed a new regression for the analysis of survival and reliability data.

This article presents some new properties of the MOOLL-G family, defines the log-MOOLL-Weibull regression and proves its usefulness. In Section 2, the MOOLL-G family is motivated and the estimation of the parameters is discussed. In Section 3, the Weibull is taken for baseline to verify the performance of the maximum likelihood estimates (MLEs). The log-MOOLL-Weibull (LMOLLW) regression is introduced and the accuracy of the MLEs and quantile residuals is investigated in Section 4. Three real lifetime applications illustrate the usefulness of the MOOLL-G family and the new regression in Section 5. The paper is concluded in Section 6.

## 2. Motivation

An interpretation of the new family cdf (1.3) can be encountered by combining the log-logistic distribution with the discrete geometric distribution of  $N$  independent risk components generated by the baseline odds ratio. For describing a stochastic system, let  $T$  be a rv with cdf  $G(t; \xi)$  and  $Z$  be a rv with a log-logistic density with parameter  $a > 0$  representing the *odds ratio*  $G(z; \xi)/\bar{G}(z; \xi)$  that this system will be not working at time  $z$ . So, the cdf of  $Z$  is

$$P(Z \leq z) = \int_0^{\frac{G(z; \xi)}{\bar{G}(z; \xi)}} \frac{a t^{a-1}}{(1+t^a)^2} dt = H(z; a, \xi).$$

Consider independent and identically distributed (iid) odds ratios  $Z_1, \dots, Z_N$  with common risk  $H(z; a, \xi)$ , where  $N$  is a geometric rv (with support in  $\{1, 2, \dots\}$  and probability parameter  $b$ ) independent of the  $Z_i$ 's and probability generating function (pgf)  $E(s^N) = \tau(s; b) = bs[1 - (1-b)s]^{-1}$ . For  $0 < b < 1$ , the cdf of  $X = \min\{Z_1, \dots, Z_N\}$  is just Equation (1.3) defined by

$$F_{\text{MOOLLG}}(x) = 1 - \tau(1 - H(x; a, \xi); b).$$

For the case  $b > 1$ , Equation (1.3) can also be defined by considering the geometric rv with probability  $1/b$ , i.e.,

$$F_{\text{MOOLLG}}(x) = 1 - \tau\left(1 - H(x; a, \xi); b^{-1}\right).$$

For both cases,  $F_{\text{MOOLLG}}(x)$  is the cdf of  $X = \min\{Z_1, \dots, Z_N\}$  assuming that  $Z$  has cdf  $H(z; a, \xi)$ , where  $N$  follows the geometric pgf  $\tau(s; b)$  with probability parameter  $b$  or  $b^{-1}$ .

Based on the previous construction, the MOOLL-G family can be widely used in oncology treatments, where  $N$  represents the amount of cancerous cells with potential to metastasis and the rvs  $Z_i$ 's denote the risks for these cells to metastasize. So, this family can model the recurrence time of the cancer. It also provides a strong physical motivation for explaining the failure time of a series system formed by  $N$  unknown iid components, each one having a common odds ratio to failure. Then, the MOOLL-G family explains the failure time of the whole system.

### 2.1. Estimation

The log-likelihood function  $\ell(\boldsymbol{\theta})$  for the vector of parameters  $\boldsymbol{\theta} = (a, b, \xi^\top)^\top$  from  $n$  observations  $x_1, \dots, x_n$  can be expressed as

$$\begin{aligned} \ell = \ell(\boldsymbol{\theta}) = & n \log(ab) + \sum_{i=1}^n \log g(x_i; \xi) + (a - 1) \sum_{i=1}^n \log G(x_i; \xi) \\ & + (a - 1) \sum_{i=1}^n \log \{1 - G(x_i; \xi)\} \\ & - 2 \sum_{i=1}^n \log [b \{1 - G(x_i; \xi)\}^a + G(x_i; \xi)^a]. \end{aligned} \tag{2.1}$$

Equation (2.1) can be maximized using the `AdequacyModel` program of the R software to find the MLE  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  or via other routines available for numerical maximization. The estimate  $\hat{\boldsymbol{\theta}}$  enjoys desirable properties that can be used when constructing confidence intervals for the parameters. The multivariate normal approximation for  $\hat{\boldsymbol{\theta}}$  can be easily handled numerically from the results of this program for constructing confidence intervals for the parameters.

### 3. The MOOLL-Weibull distribution

Consider the Weibull cdf  $G(x) = 1 - \exp[-(\alpha x)^\beta]$  with scale  $\alpha > 0$  and shape  $\beta > 0$ . The MOOLLW density follows from (1.4)

$$f(x) = \frac{a b \alpha^\beta \beta x^{\beta-1} \exp[-a(\alpha x)^\beta] \{1 - \exp[-(\alpha x)^\beta]\}^{a-1}}{\{[1 - \exp\{-(\alpha x)^\beta\}]^a + b \exp[-a(\alpha x)^\beta]\}^2}. \tag{3.1}$$

From now on,  $X$  denotes a rv with density (3.1). The hazard rate function (hrf) of  $X$  has the form

$$h(x) = \frac{a \alpha^\beta \beta x^{\beta-1} \{1 - \exp[-(\alpha x)^\beta]\}^{a-1}}{\{1 - \exp[-(\alpha x)^\beta]\}^a + b \exp[-a(\alpha x)^\beta]}.$$

Figures 1 and 2 display plots of the pdf and hrf of  $X$  for selected parameters, respectively. The MOOLLW density can take several forms and the MOOLLW hrf can have unimodal and bathtub shapes.

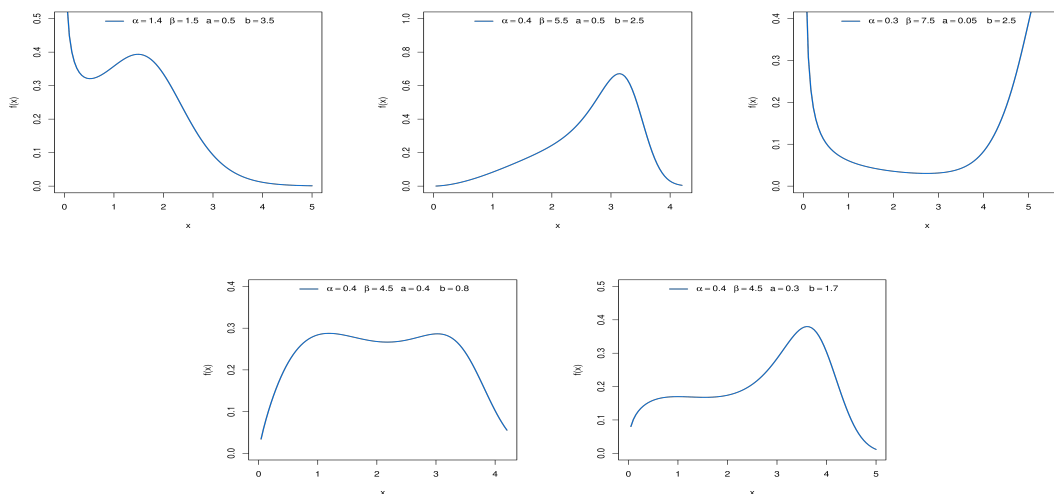


Figure 1. Plots of the MOOLLW density

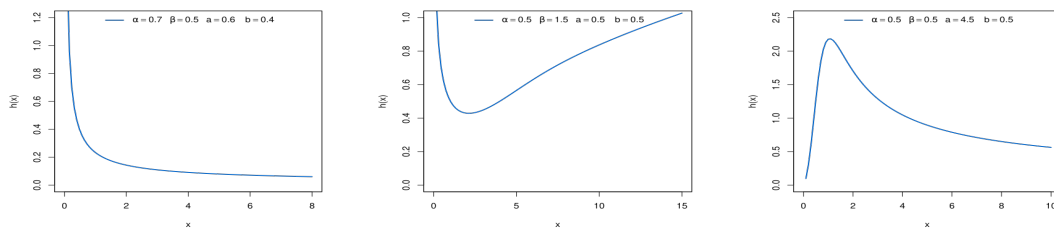


Figure 2. Plots of the MOOLLW hazard rate

Structural properties of the MOOLLW distribution can be determined from the linear representation (A.7) (derived in the Appendix) and those corresponding properties of the exponentiated Weibull with power parameter  $k \geq 0$

$$\pi_{k+1}(x) = (k + 1) \alpha^\beta \beta x^{\beta-1} \exp[-(\alpha x)^\beta] \{1 - \exp[-(\alpha x)^\beta]\}^k,$$

which are fully investigated in the literature; see, for example, [17] and references therein.

### 3.1. Simulation study

A simulation study (repeated  $N = 1,000$  times in each case) is carried out to assess the accuracy of the estimates in the MOOLLW distribution for sample sizes  $n = 100, 200, 350, 500, 1000$  under three scenarios: I:  $a = 0.5, b = 0.8, \alpha = 0.3$  and  $\beta = 1.5$ ; II:  $a = 0.5, b = 1.6, \alpha = 0.05$  and  $\beta = 0.5$  and III:  $a = 0.5, b = 1.1, \alpha = 0.3$  and  $\beta = 1.5$ . For each parameter, the average estimate (AE), the average bias (Bias) and mean squared error (MSE) are calculated from

$$Bias(\hat{\theta}) = \sum_{i=1}^N \frac{\hat{\theta}_i}{N} - \theta \quad \text{and} \quad MSE(\hat{\theta}) = \sum_{i=1}^N \frac{(\hat{\theta}_i - \theta)^2}{N}.$$

The coverage probability (CP) is also determined via simulation by calculating for each simulation the classical confidence interval (CI) and then score one for the specific simulation when the estimated CI includes the true parameter value or zero when the true parameter value falls outside the estimated CI.

Plots of the biases and MSEs are reported in Figures 3 and 4, respectively. They show that the MLEs are consistent since these averages decrease when  $n$  increases.

### 4. The LMOOLLW regression

If  $X$  has the density function (3.1), the density function of  $Y = \log(X)$  parameterized in terms of the location  $\mu \in \mathbb{R}$  and the dispersion  $\sigma > 0$  (using the transformations  $\beta = \sigma^{-1}$  and  $\alpha = e^{-\mu}$ ) can be expressed as (for  $y \in \mathbb{R}$ )

$$f(y; a, b, \sigma, \mu) = \frac{ab \exp \left[ \left( \frac{y-\mu}{\sigma} \right) - a \exp \left( \frac{y-\mu}{\sigma} \right) \right] \left\{ 1 - \exp \left[ - \exp \left( \frac{y-\mu}{\sigma} \right) \right] \right\}^{a-1}}{\sigma \left\{ \left\{ 1 - \exp \left[ - \exp \left( \frac{y-\mu}{\sigma} \right) \right] \right\}^a + b \exp \left[ -a \exp \left( \frac{y-\mu}{\sigma} \right) \right] \right\}^2}. \quad (4.1)$$

Equation (4.1) represents the (new) LMOOLLW density, i.e.

$$\text{if } X \sim \text{MOOLLW}(a, b, \alpha, \beta) \quad \text{then} \quad Y = \log(X) \sim \text{LMOOLLW}(a, b, \mu, \sigma).$$

Plots of (4.1) in Figure 5 show the bimodality and great flexibility of the density of  $Y$ .

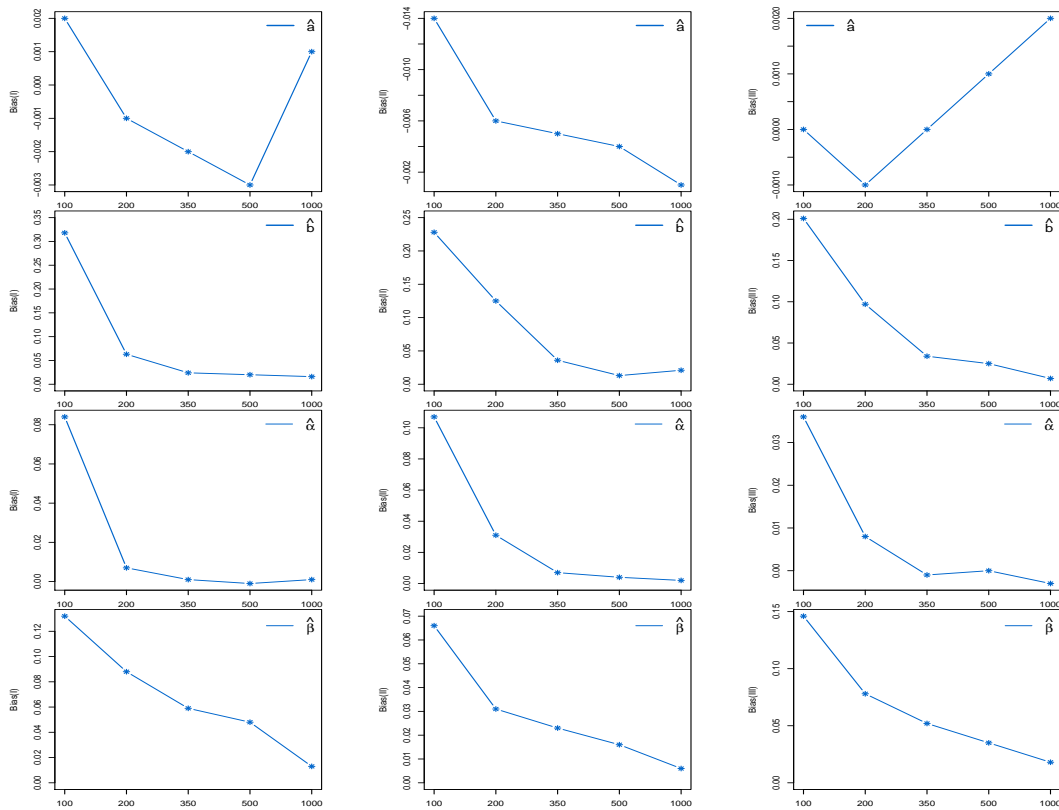


Figure 3. Biases for some parameters

The survival function (sf) of  $Y$  is

$$S(y) = \frac{b \exp \left[ -a \exp \left( \frac{y-\mu}{\sigma} \right) \right]}{\left\{ 1 - \exp \left[ - \exp \left( \frac{y-\mu}{\sigma} \right) \right] \right\}^a + b \exp \left[ -a \exp \left( \frac{y-\mu}{\sigma} \right) \right]} \tag{4.2}$$

Recently, some extended regressions can be found in [19, 21–23]. In a similar manner, we propose the LMOOLLW regression as an alternative for modeling four types of failure rate functions. First, the density function of  $Z = (Y - \mu)/\sigma$  is (for  $z \in \mathbb{R}$ )

$$\pi(z; a, b) = \frac{a b \exp[z - a \exp(z)] \{1 - \exp[-\exp(z)]\}^{a-1}}{\{ \{1 - \exp[-\exp(z)]\}^a + b \exp[-a \exp(z)] \}^2} \tag{4.3}$$

The standard extreme-value distribution follows when  $a = b = 1$ .

The LMOOLLW regression is defined by

$$y_i = \mathbf{v}_i^\top \boldsymbol{\gamma} + \sigma z_i, \quad i = 1, \dots, n, \tag{4.4}$$

where the random error  $z_i$  has density (4.3),  $\mathbf{v}_i^\top = (v_{i1}, \dots, v_{ip})$  is the explanatory variable vector,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^\top$ , and  $\sigma$ ,  $a$  and  $b$  are other unknown parameters. The location parameter vector  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$  is  $\boldsymbol{\mu} = \mathbf{V} \boldsymbol{\gamma}$ , where  $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n)^\top$  is a known model matrix.

Equation (4.4) can be useful in several applications to real data since it includes as special cases the following well-known regressions:

- Log odd log-logistic Weibull (LOLLW) regression for  $b = 1$ , which coincides with the LOLW regression [8] corresponding to the location-scale regression based on the odd log-logistic Weibull for modelling censored data.

- Log Marshall-Olkin Weibull (LMOW) regression for  $a = 1$ , which is identical to the LMOW regression.
- Log-Weibull (LW) or extreme value regression for  $a = b = 1$ , which coincides with the log-Weibull regression [13].

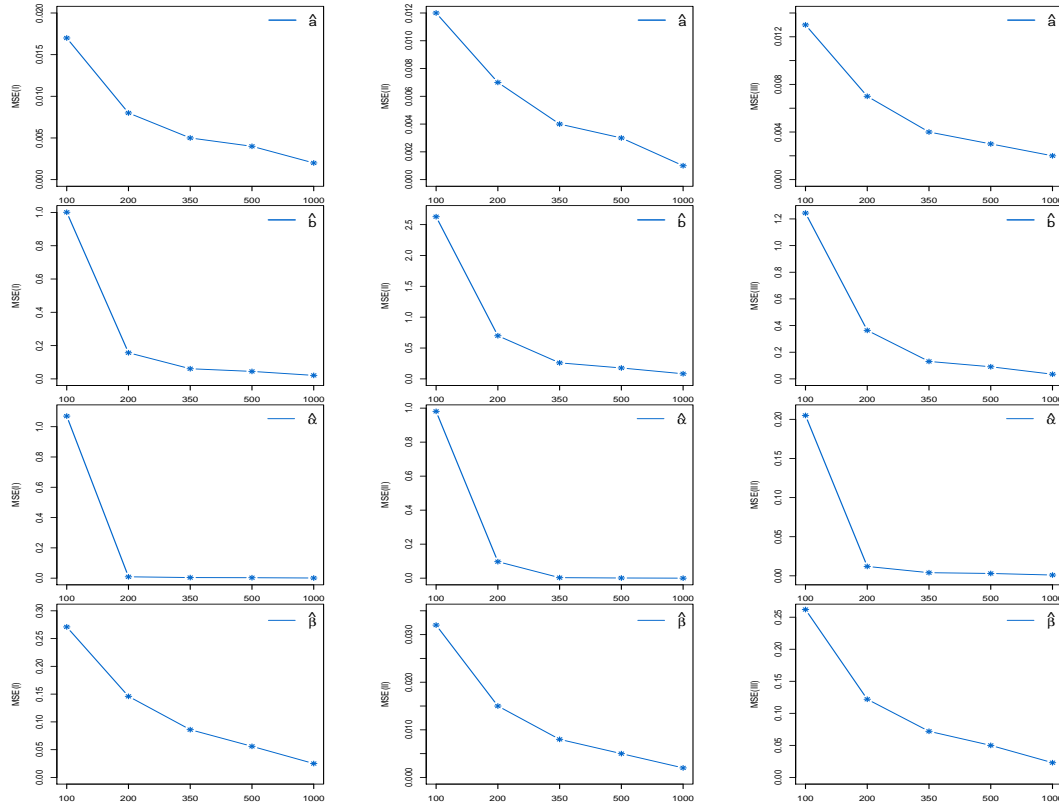


Figure 4. MSEs for some parameters

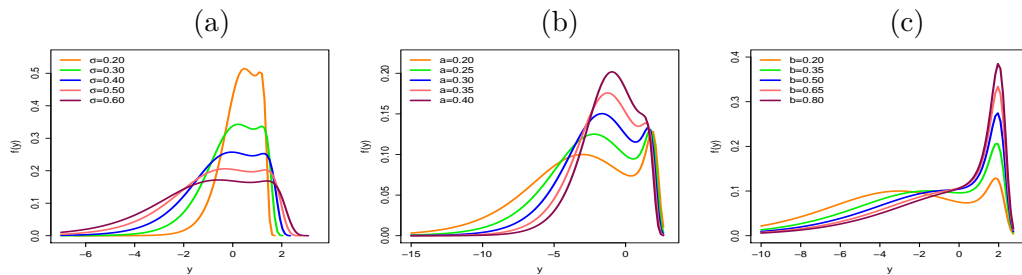


Figure 5. Plots of the LMOOLLW density (a) Varying  $\sigma$  with  $\mu = 1$ ,  $a = 0.4$  and  $b = 0.3$  (b) Varying  $a$  with  $\mu = 1$ ,  $\sigma = 0.5$  and  $b = 0.2$  (c) Varying  $b$  with  $\mu = 1$ ,  $\sigma = 0.5$  and  $a = 0.2$

The log-likelihood function for the vector  $\theta = (a, b, \sigma, \gamma^\top)^\top$  from regression (4.4) given a sample  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$  is given by

$$\begin{aligned}
 l(\theta) &= n \log \left( \frac{ab}{\sigma} \right) + \sum_{i=1}^n z_i - a \sum_{i=1}^n \exp(z_i) + (a - 1) \sum_{i=1}^n \log \{1 - \exp[-\exp(z_i)]\} \\
 &\quad - \sum_{i=1}^n \log \{ \{1 - \exp[-\exp(z_i)]\}^a + b \exp[-a \exp(z_i)] \}, \tag{4.5}
 \end{aligned}$$

where  $z_i = (y_i - \mathbf{v}_i^\top \gamma) / \sigma$ .

The MLE  $\hat{\theta}$  can be calculated by maximizing the log-likelihood (4.5) using the R software. Comparing special models with the LMOOLLW regression can be based on likelihood ratio (LR) statistics.

The quantile residuals ( $qr$ 's) can be expressed as

$$\hat{q}r_i = \Phi^{-1} \left\{ \frac{\left\{ 1 - \exp \left[ - \exp \left( \frac{y_i - \mu_i}{\sigma} \right) \right] \right\}^a}{\left\{ 1 - \exp \left[ - \exp \left( \frac{y_i - \mu_i}{\sigma} \right) \right] \right\}^a + b \exp \left[ -a \exp \left( \frac{y_i - \mu_i}{\sigma} \right) \right]} \right\}, \quad (4.6)$$

where  $\Phi^{-1}(\cdot)$  is the inverse cumulative standard normal distribution.

#### 4.1. Simulation study for the estimates

A Monte Carlo study with 1,000 replications is conducted to examine the accuracy of the MLEs in the LMOOLLW( $\mu_i, \sigma, a, b$ ) distribution using the statistical software R. The systematic structure is  $\mu_i = \gamma_0 + \gamma_1 v_{i1} + \gamma_2 v_{i2}$ , where  $a = 0.8$ ,  $b = 0.7$ ,  $\gamma_0 = 2$ ,  $\gamma_1 = 0.4$ ,  $\gamma_2 = 0.6$  and  $\sigma = 0.3$ . The response variable  $y_i$  and the covariates  $v_{i1}$  and  $v_{i2}$  are generated as follows:  $y_i \sim \text{LMOOLLW}(\mu_i, \sigma, a, b)$ ,  $v_{i1} \sim \text{Bernoulli}(0.5)$  and  $v_{i2} \sim \text{Uniform}(0, 1)$ .

The AEs, biases and MSEs for  $n = 100, 250, 500$  and  $1,000$  are reported in Table 1. For all cases, the biases and MSEs of the estimates of  $\gamma_0, \gamma_1, \gamma_2, \sigma, a$  and  $b$  decay toward zero when  $n$  increases.

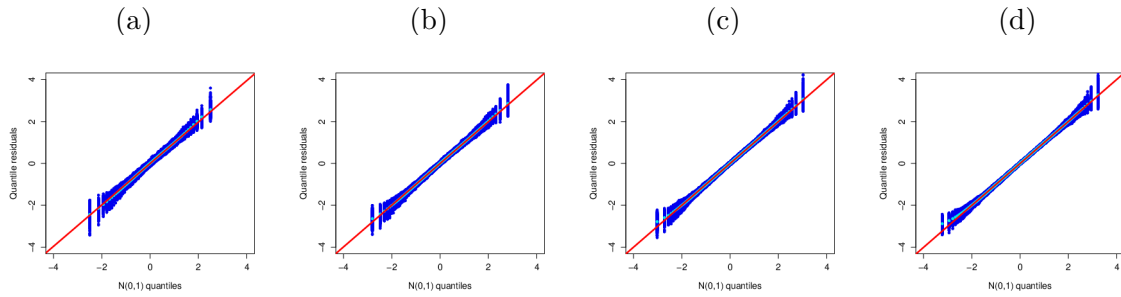
**Table 1.** Simulation results from the LMOOLLW regression

Parameter	$n = 100$			$n = 250$		
	AE	Bias	MSE	AE	Bias	MSE
$\gamma_0$	2.045	0.045	0.034	2.028	0.028	0.012
$\gamma_1$	0.399	-0.001	0.007	0.400	0.000	0.002
$\gamma_2$	0.595	-0.005	0.021	0.594	-0.006	0.007
$\sigma$	0.238	-0.062	0.018	0.272	-0.028	0.006
$a$	0.640	-0.160	0.101	0.743	-0.057	0.030
$b$	0.721	0.021	0.583	0.675	-0.025	0.132
Parameter	$n = 500$			$n = 1,000$		
	AE	Bias	MSE	AE	Bias	MSE
$\gamma_0$	2.031	0.031	0.007	2.030	0.030	0.004
$\gamma_1$	0.401	0.001	0.001	0.399	-0.001	0.001
$\gamma_2$	0.594	-0.006	0.003	0.596	-0.004	0.002
$\sigma$	0.278	-0.022	0.003	0.283	-0.017	0.002
$a$	0.763	-0.037	0.015	0.779	-0.021	0.008
$b$	0.639	-0.061	0.056	0.633	-0.067	0.034

#### 4.2. Simulation study for the residuals

The accuracy of the normal approximation for the empirical distribution of the residuals  $qr$ 's is addressed by generating 1,000 observations from the fitted LMOOLLW regression following the same procedure described in Section 4.1. So, the systematic component is  $\mu_i = \gamma_0 + \gamma_1 v_{i1} + \gamma_2 v_{i2}$  for the same parameters in this section. The plots in Figure 6 indicate that the standard normal distribution is closer to the empirical distribution of these residuals for  $n = 100, 250, 500$  and  $1,000$  corresponding to the scenarios (a), (b), (c) and (d), respectively (for moderate and large  $n$ ).





**Figure 6.** Normal probability plots for  $qr'_i$ s in the LMOOLLW regression ( $a = 0.8$  and  $b = 0.7$ )

## 5. Applications

### 5.1. Two applications of the MOOLLW distribution

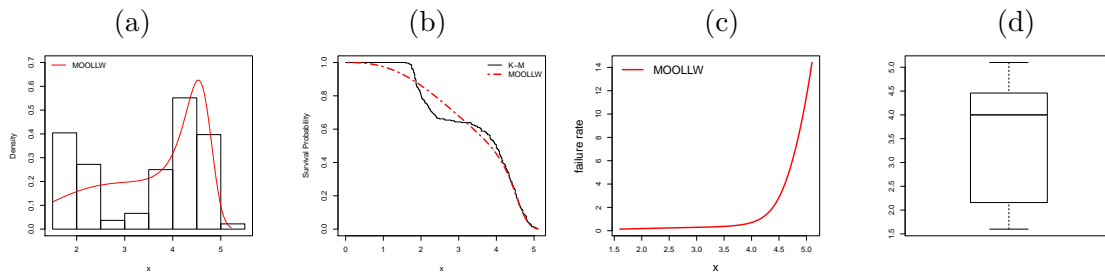
We compare the MOOLLW distribution with some well-established Weibull extensions: the Kumaraswamy-Weibull (KwW) [6], beta-Weibull (BW) [7], McDonald-Weibull (McW) [5], exponentiated-generalized Weibull (EGW) [18], exponentiated-Weibull (EW) [16] and Weibull (W) distributions. The potentiality of the proposed distribution is shown for two real data sets representing different events:

*Data set 1. (Geyser data).* The first data set comes from R built-in data (`datasets::faithful`) which describes the old Faithful Geyser data, in particular the 272 durations of the eruptions (in mins).

*Data set 2. (Aarset data).* The second data set from Aarset [1] describes the failure times of 50 industrial devices.

All computations are performed using the `AdequacyModel` program in R-language. The unknown parameters of the models are estimated by maximum likelihood. The well-known AIC (Akaike information criterion), BIC (Bayesian information criterion), HQIC (Hannan-Quinn information criterion),  $A^*$  (Anderson-Darling),  $W^*$  (Cramér-von Mises) and K-S (Kolmogorov-Smirnov) statistics are adopted for model comparisons.

Tables 2 and 4 give the MLEs and their standard errors (SEs) from seven distributions fitted to these data sets. The numbers in Tables 3 and 5 indicate that the MOOLLW distribution provides the best fit compared to the other fitted models. The plots in Figures 7 and 8 confirm this fact.



**Figure 7.** Estimated functions of the MOOLLW model for data set 1 (a) Density (b) Survival function (c) Hazard rate (d) Box-plot

**Table 2.** Estimated quantities for data set 1

Distribution	$a$	$b$	$\alpha$	$\beta$	$\nu$
MOOLLW	0.2203 (0.0175)	1.0899 (0.1486)	0.2586 (0.0029)	12.0277 (0.2425)	-
KwW	0.0823 (0.0091)	0.6255 (0.0932)	0.2138 (0.0026)	24.3823 (0.5347)	-
BW	0.0854 (0.0063)	0.3241 (0.1281)	0.2176 (0.0046)	27.7661 (0.0607)	-
EGW	0.1662 (0.0442)	0.0940 (0.0060)	4.5513 (0.0482)	26.5748 (0.0578)	-
McW	0.0676 (0.0268)	2.0811 (1.0048)	0.1998 (0.0050)	23.8858 (0.2938)	1.9476 (0.9584)
EW	0.0878 (0.0055)	-	0.2050 (0.0014)	28.3951 (0.0206)	-
W	-	-	0.2571 (0.0044)	3.6733 (0.1920)	-

**Table 3.** Adequacy measures for data set 1

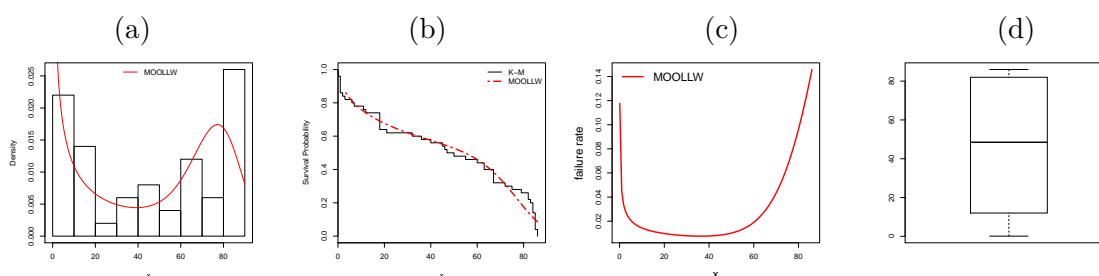
Distribution	$\hat{\ell}$	AIC	BIC	HQIC	$A^*$	$W^*$	K-S	K-S p-value
MOOLLW	346.4653	700.9307	715.3539	706.7211	4.9448	0.6701	0.1255	0.000379
KwW	363.4927	734.9854	749.4086	740.7759	6.4396	0.9351	0.1587	2.25E-06
BW	365.9784	739.9568	754.3800	745.7472	7.3288	1.1349	0.1558	3.67E-06
EGW	367.5039	743.0077	757.4310	748.7982	8.1949	1.2871	0.1606	1.61E-06
McW	367.5248	745.0497	763.0787	752.2877	7.8523	1.2287	0.1608	1.548E-06
EW	367.1236	740.2473	751.0647	744.5901	8.1046	1.2718	0.1608	1.55E-06
W	413.3641	830.7281	837.9397	833.6233	16.5785	2.8221	0.1773	7.45E-08

**Table 4.** Estimated quantities for data set 2

Distribution	$a$	$b$	$\alpha$	$\beta$	$\nu$
MOOLLW	0.1135 (0.0200)	1.2459 (0.3550)	0.0210 (0.0010)	5.2856 (0.0165)	-
KwW	0.0667 (0.0231)	0.2680 (0.0890)	0.0167 (0.0015)	5.0653 (0.2159)	-
BW	0.1057 (0.0231)	0.0958 (0.0469)	0.0216 (0.0026)	4.7139 (0.1274)	-
EGW	0.0429 (0.0094)	0.5724 (0.0957)	5.8136 (0.0075)	1.2963 (0.0104)	-
McW	0.2377 (0.0863)	0.1796 (0.0930)	0.0190 (0.0026)	4.5870 (0.1489)	0.3328 (0.1291)
EW	0.1424 (0.0213)	-	0.0109 (0.0009)	4.8222 (0.0148)	-
W	-	-	0.0223 (0.0034)	0.9499 (0.1196)	-

**Table 5.** Adequacy measures for data set 2

Distribution	$\hat{\ell}$	AIC	BIC	HQIC	$A^*$	$W^*$	K-S	K-S p-value
MOOLLW	216.9131	441.8262	449.4743	444.7387	0.7958	0.0882	0.1177	0.4927
KwW	221.4887	450.9774	458.6254	453.8898	1.1598	0.1513	0.1310	0.3577
BW	220.0350	448.0700	455.7181	450.9824	0.9775	0.1177	0.1323	0.3453
EGW	237.8521	483.7042	491.3523	486.6167	2.7199	0.4424	0.1978	0.0400
McW	221.3386	452.6772	462.2373	456.3177	1.0747	0.1368	0.1368	0.3064
EW	228.9559	463.9117	469.6478	466.0961	1.7943	0.2710	0.2035	0.0319
W	241.0018	486.0037	489.8277	487.4599	3.0071	0.4962	0.1927	0.0488



**Figure 8.** Estimated functions of the MOOLLW model for data set 2 (a) Density (b) Survival function (c) Hazard rate (d) Box-plot

### 5.2. An application of the LMOOLLW regression

The set of 432 observations referring to a common native plant of the Peruvian Andes called *yacon* was obtained from the agricolae package [9] available in the R software with the source: CIP, Experimental eld, 2003.

The main variables involved are (for  $i = 1, \dots, 432$ ):  $y_i$ : logarithm of the weight in grams of fresh yacon roots;  $v_{i1}$ : harvest index;  $v_{i2}$ : a factor with 8 levels: AKW5075, AMM5136, AMM5150, AMM5163, ARB5125, CLLUNC118, P1385 and SAL136; and  $v_{i3}$ : number of stalks (or stems)

For constructing the covariables, we require seven dummy variables:

$$\underbrace{\text{AKW5075 AMM5136 AMM5150 AMM5163 ARB5125 CLLUNC118 P1385 SAL136}}_{\text{Processing}} \Rightarrow \underbrace{v_{21} \quad v_{22} \quad v_{23} \quad v_{24} \quad v_{25} \quad v_{26} \quad v_{27}}_{\text{7 dummy variables}}$$

First, a univariate analysis is performed considering only the response variable. Table 6 lists the MLEs and the AIC, BIC and GD (Global Deviance). In fact, the LMOOLLW distribution could be chosen as the best distribution for these data.

Some LR statistics for comparing fitted distributions to the current data are reported in Table 7. It is quite clear that the LMOLLW distribution is suitable for modeling these data comparing to the null models.

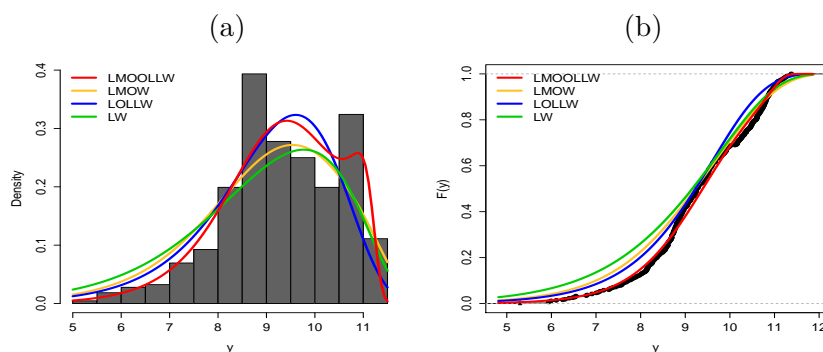
Plots of the best three fitted densities to the histogram are displayed in Figure 9(a). The plots of the estimated cdfs and the empirical distribution are given in Figure 9(b). They reveal that the new distribution provides the best fit to these data compared to the other fitted models. Note that this data set has bimodality which is captured by the new distribution.

**Table 6.** Estimated quantities for weight of fresh *yacon* roots data

Model	$\mu$	$\log(\sigma)$	$a$	$b$	AIC	BIC	GD
LMOOLLW	10.646 (0.068)	-1.273 (0.139)	0.349 (0.046)	0.209 (0.037)	1340.598	1356.872	1332.598
LMOW	10.442 (0.054)	-0.169 (0.041)	1 (-)	0.346 (0.028)	1375.490	1387.700	1369.490
LOLLW	9.882 (0.051)	0.018 (0.049)	0.977 (0.047)	1 (-)	1385.240	1397.440	1379.240
LW	9.885 (0.051)	0.033 (0.050)	(1) (-)	(1) (-)	1383.340	1391.480	1379.340

**Table 7.** LR tests for weight of fresh *yacon* roots data

Models	Hypotheses	Statistic $w$	$p$ -value
LMOOLLW vs LMOW ( $a = 1$ )	$H_0 : a = 1$ vs $H_1 : H_0$ is false	36.892	<0.001
LMOOLLW vs LOLLLW ( $b = 1$ )	$H_0 : b = 1$ vs $H_1 : H_0$ is false	46.639	<0.001
LMOOLLW vs LW	$H_0 : a = 1$ and $b = 1$ vs $H_1 : H_0$ is false	46.745	<0.001



**Figure 9.** Weight of fresh *yacon* roots data (a) Estimated densities of some fitted models (b) Estimated cumulative functions of some fitted models and the empirical cdf

Second, consider the LMOOLLW regression with systematic component ( $i = 1, \dots, 432$ )

$$\mu_i = \gamma_0 + \gamma_1 v_{i1} + \gamma_{21} v_{i21} + \gamma_{22} v_{i22} + \gamma_{23} v_{i23} + \gamma_{24} v_{i24} + \gamma_{25} v_{i25} + \gamma_{26} v_{i26} + \gamma_{27} v_{i27} + \gamma_3 v_{i3}.$$

The findings from the fitted LMOOLLW regression are reported in Table 8. The covariable  $x_1$  is significant, which means that the fresh weight of yacon root decreases when the harvest index increases. Similarly, the covariable  $x_3$  is also significant. In other words, the fresh weight of yacon root increases when the number of rods increases. Further, there is significant difference between levels (AKW5075 vs AMM5136), (AKW5075 vs ARB5125) and (AKW5075 vs CLLUNC118) in relation to fresh weight of yacon roots data.

The LMOOLLW regression gives the best fit to the current data based on the statistics in Table 9.

The numbers in Table 10 indicate that the LMOOLLW regression could be chosen to explain the current data.

Figure 10(a) gives the plot of the modified martingale-type residuals ( $qr_i$ ) against the index of the observations. There is no pattern and the residuals are in the interval  $[-3, 3]$ . The normal probability plot with generated envelope displayed in Figure 10(b) show that all observations are well-fitted. Figure 10(b) indicates that the LMOOLLW regression is adequate because all observations are within the envelope.

**Table 8.** Estimated quantities from the LMOOLLW regression fitted to yacon roots data

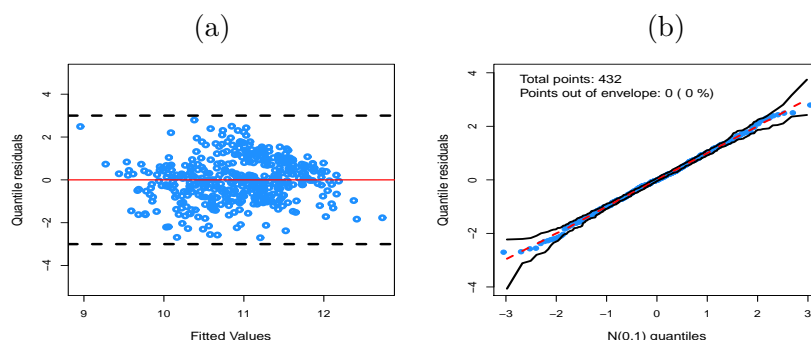
Parameter	Estimate	SE	<i>p</i> -value
$\gamma_0$	13.780	0.223	<0.001
$\gamma_1$	-12.086	0.878	<0.001
$\gamma_{21}$	-0.517	0.154	<0.001
$\gamma_{22}$	-0.043	0.201	0.831
$\gamma_{23}$	0.122	0.182	0.504
$\gamma_{24}$	-0.473	0.129	<0.001
$\gamma_{25}$	-0.358	0.144	0.013
$\gamma_{26}$	-0.210	0.204	0.304
$\gamma_{27}$	0.087	0.165	0.600
$\gamma_3$	0.102	0.030	<0.001
$\log(\sigma)$	-1.808	0.024	
<i>a</i>	0.283	0.007	
<i>b</i>	0.062	0.005	

**Table 9.** Goodness-of-fit measures for yacon roots data

Model	AIC	BIC	GD
<b>LMOOLLW</b>	<b>1219.548</b>	<b>1272.437</b>	<b>1193.55</b>
LMOW	1228.197	1277.018	1204.200
LOLLW	1234.689	1283.510	1210.690
LW	1252.922	1297.674	1230.920

**Table 10.** LR tests for yacon roots data

Models	Hypotheses	Statistic <i>w</i>	<i>p</i> -value
LMOOLLW vs LMOW	$H_0 : a = 1$ vs $H_1 : H_0$ is false	10.649	0.001
LMOOLLW vs LOLW	$H_0 : b = 1$ vs $H_1 : H_0$ is false	17.141	<0.001
LMOOLLW vs LW	$H_0 : a = 1$ and $b = 1$ vs $H_1 : H_0$ is false	37.374	<0.001



**Figure 10.** (a) Plot of the residuals ( $qr'_i$ s) versus fitted values (b) Normal probability plot with envelope for the  $qr'_i$ s from the LMOOLLW regression fitted to fresh root weight of yacon data

## 6. Conclusions

The Marshall-Olkin odd log-logistic-G (MOOLL-G) family follows by compounding of two well-known families of distributions. A motivation for the family is provided and some useful properties are presented. A special model of the MOOLL-G family is introduced

based on the Weibull and it is proposed the log-MOOLL-Weibull regression. Estimation of parameters is via maximum likelihood and some simulation studies are conducted. Three applications of the proposed models show their usefulness over well-established models. In the near future, the following works may be addressed: construct a regression for bivariate data using copulas; develop a library in R software with the proposed models; build a cured fraction model considering the distributions addressed in this work as a random effect; consider other estimation methods such as, for example, using a Bayesian approach.

## References

- [1] M.V. Aarset, *How to identify bathtub hazard rate*, IEEE Trans. Rel. **36** (1), 106-108, 1987.
- [2] M. Alizadeh, G. Ozel, E. Altun, M. Abdi and G.G. Hamedani, *The odd log-logistic Marshall-Olkin Lindley model for lifetime data*, J. Stat. Theory Appl. **16** (3), 382-400, 2017.
- [3] W. Barreto-Souza, Lemonte, A.J. and G.M. Cordeiro, *General results for the Marshall and Olkin's family of distributions*, An. Acad. Brasil. Ciênc. **85** (1), 3-21, 2013.
- [4] G.M. Cordeiro, M. Alizadeh, G. Ozel, B. Hosseinl, E.M.M. Ortega and E. Altun *The generalized odd log-logistic family of distributions: Properties, regression models and applications*, J. Stat. Comput. Simul. **87** (5), 908-932, 2017.
- [5] G.M. Cordeiro, E.M. Hashimoto and E.M.M. Ortega, *The McDonald Weibull model*, Statistics **48** (2), 256-278, 2012.
- [6] G.M. Cordeiro, E.M.M. Ortega and S. Nadarajah, *The Kumaraswamy Weibull distribution with application to failure data*, J Franklin Inst **347** (8), 1399-1429, 2010.
- [7] G.M. Cordeiro, E.M.M. Ortega and S. Nadarajah, *General results for the beta Weibull distribution*, J. Stat. Comput. Simul. **83** (6), 1082-1114, 2013.
- [8] J.N. da Cruz, E.M.M. Ortega and G.M. Cordeiro, *The log-odd log-logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis*, J. Stat. Comput. Simul. **86** (8), 1516-1538, 2016.
- [9] F. de Mendiburu and M.F. de Mendiburu, Package "agricolae", R package version: 1-2, 2020.
- [10] J.U. Gleaton and J.D. Lynch, *On the distribution of the breaking strain of a bundle of brittle elastic fibers*, Adv. in Appl. Probab. **36** (1), 98-115, 2004.
- [11] J.U. Gleaton and J.D. Lynch, *Extended generalized log-logistic families of lifetime distributions with an application*, JPSS J. Probab. Stat. Sci. **8** (1), 1-17, 2010.
- [12] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed., Academic Press, San Diego, 2000.
- [13] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 2003.
- [14] M.C.S. Lima, F. Prativiera, E.M.M. Ortega and G.M. Cordeiro, *The odd log-logistic geometric family with applications in regression models with varying dispersion*, J. Stat. Theory Appl. **18**, 278-294, 2019.
- [15] A.N. Marshall and I. Olkin, *A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families*, Biometrika **84** (3), 641-652, 1997.
- [16] G.S. Mudholkar and D.K. Srivastava, *Exponentiated Weibull family for analyzing bathtub failure-rate data*, IEEE Trans. Rel. **42** (2), 299-302, 1993.
- [17] S. Nadarajah, G.M. Cordeiro and E.M.M. Ortega, *The exponentiated Weibull distribution: a survey* Statist. Papers **54** (3), 839-877, 2013.

- [18] P.E. Oguntunde, O.A. Odetunmibi and A.O. Adejum, *On the exponentiated generalized Weibull distribution: A generalization of the Weibull distribution*, Indian J Sci Technol **8** (35), 1-7, 2015.
- [19] E.M.M. Ortega, J.N. da Cruz and G.M. Cordeiro, *The log-odd logistic-Weibull regression model under informative censoring*, Model Assist Stat Appl **14** (3), 239-254, 2019.
- [20] F. PrataViera, G.M. Cordeiro, E.M.M. Ortega and A.K. Suzuki, *The odd log-logistic geometric normal regression models with applications*, Adv. Data Sci. Adapt. Anal. **11** (01n02), 1-25, 2019.
- [21] F. PrataViera, E.M.M. Ortega, G.M. Cordeiro and A.S. Braga, *The heteroscedastic odd log-logistic generalized gamma regression model for censored data*, Comm. Statist. Simulation Comput. **48** (6), 1815-1839, 2018.
- [22] F. PrataViera, E.M.M. Ortega, G.M. Cordeiro, R.R. Pescim and B.A.W. Verssani, *A new generalized odd log-logistic flexible Weibull regression model with applications in repairable systems*, Reliab. Eng. Syst. Saf. **176**, 13-26, 2018.
- [23] J.C Souza Vasconcelos, G.M. Cordeiro, E.M.M. Ortega and E.G. Araújo, *The new odd log-logistic generalized inverse gaussian regression model*, J. Probab. Stat., 1-13, 2019.
- [24] M.H. Tahir and S. Nadarajah, *Parameter induction in continuous univariate distributions: Well-established G families*, An. Acad. Brasil. Ciênc. **87** (2), 539-568, 2015.

### Appendix

Following [3], we can write the linear representation for Equation (1.3)

$$F_{\text{MOOLLG}}(x) = \sum_{i=0}^{\infty} w_i H_{\text{OLLG}}(x)^{i+1}, \tag{A.1}$$

where the coefficients are (for  $i = 0, 1, \dots$ )

$$w_i = w_i(b) = \begin{cases} \frac{(-1)^i b}{(i+1)} \sum_{j=i}^{\infty} \binom{j}{i} (j+1)(1-b)^j, & b \in (0, 1), \\ b^{-1}(1-1/b)^i, & b > 1. \end{cases}$$

Equation (A.1) can be rewritten as a sum of exponentiated-G (exp-G) cdfs with non-negative integer powers. For doing this, four convergent power series are required. First, the generalized binomial expansion

$$(1-z)^a = \sum_{j=0}^{\infty} (-1)^j \binom{a}{j} z^j,$$

is convergent for  $|z| \leq 1$ .

Second, a power series for  $z^a = [1 - (1-z)]^a$  follows using the last expansion twice (for  $|z| \leq 1$ )

$$z^a = \sum_{k=0}^{\infty} p_k(a) z^k, \tag{A.2}$$

where

$$p_k(a) = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{a}{j} \binom{a}{k}.$$

Third, an expansion for  $z^a + (1-z)^a$  follows from the last two power series (for any real  $a > 0$  and  $|z| \leq 1$ )

$$z^a + (1-z)^a = \sum_{k=0}^{\infty} q_k(a) z^k, \tag{A.3}$$

where  $q_k(a) = p_k(a) + (-1)^k \binom{a}{k}$ .

Combining (A.2) and (A.3) and from [12] Section 0.313

$$\frac{z^a}{z^a + (1 - z)^a} = \frac{\sum_{k=0}^{\infty} p_k(a) u^k}{\sum_{k=0}^{\infty} q_k(a) z^k} = \sum_{k=0}^{\infty} s_k(a) z^k, \tag{A.4}$$

where  $s_0(a) = p_0(a)/q_0(a)$  and the quantities  $s_k(a)$ 's (for  $k \geq 1$ ) are obtained as

$$s_k(a) = \frac{1}{p_0(a)} \left[ p_k(a) - \frac{1}{q_0(a)} \sum_{r=1}^k p_r(a) q_{k-r}(a) \right].$$

Fourth, based on (A.4), we can rewrite  $H(x)^{i+1}$  in Equation (A.1) (for  $x \in \mathbb{R}$ ) as

$$H_{\text{OLLG}}(x)^{i+1} = \left\{ \frac{G(x)^a}{G(x)^a + [1 - G(x)]^a} \right\}^{i+1} = \left\{ \sum_{k=0}^{\infty} s_k(a) G(x)^k \right\}^{i+1}.$$

Next, the following expansion holds ([12], Section 0.314)

$$H_{\text{OLLG}}(x)^{i+1} = \left( \sum_{k=0}^{\infty} s_k(a) G(x)^k \right)^{i+1} = \sum_{k=0}^{\infty} t_{i+1,k}(a) G(x)^k, \tag{A.5}$$

where, for  $i = 0$  and  $k \geq 0$ ,  $t_{i+1,k}(a) = s_k(a)$  and, for  $i \geq 1$ ,  $t_{i+1,0}(a) = s_0(a)^{i+1}$  and, for  $k \geq 1$ ,

$$t_{i+1,k}(a) = \frac{1}{k s_0(a)} \sum_{r=1}^k [r(i + 2) - k] s_r(a) t_{i+1,k-r}(a).$$

By inserting (A.5) in Equation (A.1) yields

$$F_{\text{MOOLLG}}(x) = \sum_{k=0}^{\infty} v_k G(x)^k, \tag{A.6}$$

where  $v_k = v_k(a, b) = \sum_{i=0}^{\infty} w_i(b) t_{i+1,k}(a)$  for  $k \geq 0$ .

By differentiating (A.6), the pdf of  $X$  admits the linear representation

$$f_{\text{MOOLLG}}(x) = \sum_{k=0}^{\infty} v_{k+1} \pi_{k+1}(x), \tag{A.7}$$

where  $\pi_{k+1}(x) = (k + 1) G(x)^k g(x)$  denotes the exp-G density with power parameter  $k + 1$ . Thus, the MOOLL-G family density is a linear combination of exp-G densities which can be adopted to find its structural properties.

Let  $Y_{k+1}$  be a rv having exp-G density  $\pi_{k+1}(x)$ . The  $n$ th ordinary moment of  $X$  can be obtained from (A.7) as

$$\mu'_n = \mathbb{E}(X^n) = \sum_{k=0}^{\infty} v_{k+1} \mathbb{E}(Y_{k+1}^n) = \sum_{k=0}^{\infty} (k + 1) v_{k+1} \tau_{n,k}, \tag{A.8}$$

where  $\tau_{n,k} = \int_{-\infty}^{\infty} x^n G(x)^k g(x) dx = \int_0^1 Q_G(u)^n u^k du$ . Several papers dealing with exp-G moments are given by [24]. The  $n$ th incomplete moment of  $X$  can be expressed as

$$m_n(y) = \int_{-\infty}^y x^n f_{\text{MOOLLG}}(x) dx = \sum_{k=0}^{\infty} (k + 1) v_{k+1} \int_0^{G(y)} Q_G(u)^n u^k du. \tag{A.9}$$

Equation (A.9) (with  $n = 1$ ) allows to find conditional moments, mean deviations and Bonferroni and Lorentz curves of  $X$ .



The moment generating function (mgf) of  $X$  follows from (A.7) as

$$M(t) = \sum_{k=0}^{\infty} v_{k+1} M_{k+1}(t) = \sum_{i=0}^{\infty} (k+1) v_{k+1} \rho_k(t),$$

where  $M_{k+1}(t)$  is the mgf of  $Y_{k+1}$  and  $\rho_k(t) = \int_{-\infty}^{\infty} e^{tx} G(x)^k g(x) dx = \int_0^1 \exp[t Q_G(u)] u^k du$ . Several mgfs for MOOLL-G distributions can be directly obtained from this equation and the mgfs in the papers listed in Table 1 of [24].