The Odd Log-Logistic Geometric Family with Applications in Regression Models with Varying Dispersion

Maria do Carmo S. Lima[1], Fábio Prataviera[2], Edwin M. M. Ortega[2],*, Gauss M. Cordeiro[1]

[1] *Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brazil*
[2] *Departamento de Ciências Exatas, Universidade de São Paulo, Piracicaba, Brazil*

ARTICLE INFO

ABSTRACT

We obtain some mathematical properties of a new generator of continuous distributions with two additional shape parameters called the odd log-logistic geometric family. We present some special models and investigate the asymptotes and shapes. The family density function can be expressed as a linear combination of exponentiated densities based on the same baseline distribution. We derive a power series for its quantile function. We provide explicit expressions for the ordinary and incomplete moments and generating function. We estimate the model parameters by maximum likelihood. We propose a useful regression model by varying the dispersion parameter to fit real data. We illustrate the potentiality of the proposed models by means of three real data sets.

1. INTRODUCTION

Recent developments on univariate continuous distributions have been focused to define new families by adding shape parameters to control skewness, kurtosis and tail weights, providing great flexibility in modeling skewed data in practice, including the two-piece approach introduced by Hansen [1] and the generators pioneered by Eugene *et al.* [2], Cordeiro and de Castro [3], Alexander *et al.* [4] and Cordeiro *et al.* [5]. Many subsequent articles apply these techniques to induce skewness into well-known symmetric distributions such as the symmetric Student *t*. For a review, see Aas and Haff [6].

In the last few years, new classes of distributions were proposed, for example, the generalized Weibull family by Cordeiro *et al.* [7], the exponentiated G Poisson model (Gomes *et al.*, [8]), the Poisson-X family presented by Tahir *et al.* [9], the extended-G geometric family by Cordeiro *et al.* [10], the generalized odd half-Cauchy class by Cordeiro *et al.* [11] and the generalized odd log-logistic (LL) family studied by Cordeiro *et al.* [12].

In this work, we introduce a new class of distributions called the odd log-logistic geometric-G ("OLLG-G" for short) family with two additional shape parameters in order to attract wider applications in reliability, biology and other research areas, and study some of its mathematical properties. The new family can extend several common models such as the normal and Weibull distributions by adding two extra parameters to a parent G. The proposed family is an extension of that one introduced recently by Gleaton and Lynch [13].

One way to study the effects of the explanatory variables on the lifetime or survival time is through a regression location-scale model, also known as the accelerated lifetime model. Regression models can be proposed in different forms in survival analysis, for example, Hashimoto *et al.* [14] defined the long-term survival model with interval-censored data. Ortega *et al.* [15] defined a power series beta Weibull regression model for predicting breast carcinoma, Lanjoni *et al.* [16] proposed the extended Burr XII regression model, Ortega *et al.* [17] introduced the odd Birnbaum-Saunders regression model and, more recently, Ramires *et al.* [18] predicted the cure rate of breast cancer using a regression model with four regression structures. In this paper, we develop a log-linear model using a new distribution called the *log-odd log-logistic geometric Weibull* ("LOLLG-W" for short). The new regression model by varying the dispersion parameter can be applied to censored data since it represents a parametric family of models that includes as sub-models several widely-known regression models and therefore can be used more effectively in the analysis of survival data.

---

*Corresponding author. Email: edwin@usp.br

This paper is organized as follows: In Section 2, we define the OLLG-G family and propose some new distributions. In Section 3, we derive useful expansions for the new family. In Section 4, we present moments and quantile and generating functions. Maximum likelihood estimation of the model parameters and various simulations for different parameter settings and sample sizes are performed in Section 5. In Section 6, we define a regression model of location-scale form with varying dispersion. In Section 7, we provide some applications. Section 8 offers some concluding remarks.

2. MODEL DEFINITION

Gleaton and Lynch [13], da Cruz *et al.* [19], Braga *et al.* [20] and Cordeiro *et al.* [12] studied a class of distributions called the *odd log-logistic* ("OLLG-G") family with applications in different areas. Given a continuous baseline cumulative distribution function (cdf) $G(y; \tau)$ with a parameter vector τ of dimension k, the cdf of the OLL-G class with an additional shape parameter $\alpha > 0$ is defined by

$$F(y) = \int_0^{\frac{G(y;\tau)}{\overline{G}(y;\tau)}} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2} dt = \frac{G(y;\tau)^\alpha}{G(y;\tau)^\alpha + \overline{G}(y;\tau)^\alpha}, \tag{1}$$

where $\overline{G}(y; \tau) = 1 - G(y; \tau)$. We can write

$$\alpha = \frac{\log\left[\frac{F(y)}{\overline{F}(y)}\right]}{\log\left[\frac{G(y)}{\overline{G}(y)}\right]}.$$

Thus, the parameter α represents the quotient of the log odds ratio for the generated and baseline distributions. We note that there is no complicated function in (1) in contrast with the beta generalized family Eugene *et al.* [2], which includes two extra parameters and also involves the incomplete beta function. The baseline cdf $G(y; \tau)$ is clearly a special case of (1) when $\alpha = 1$. If $G(y; \tau) = y/(1+y)$, becomes the LL distribution. Some distributions can be generated from (1).

Suppose that $\{Y_i\}_{i=1}^M$ are independent and identically distributed (iid) random variables having the OLL-G cdf (1). Let M be a geometric random variable with probability mass function given by $P(M = m) = (1-p)p^{m-1}$ for $m \in \mathbb{N}$ and $p \in (0, 1)$. Let $X = min\left(\{Y_i\}_{i=1}^M\right)$. The conditional density function of X given $M = m$ is

$$f(x|m; \tau) = \frac{m\alpha g(x; \tau) G^{\alpha-1}(x; \tau)[1 - G(x; \tau)]^{m\alpha-1}}{\left\{G^\alpha(x; \tau) + [1 - G(x; \tau)]^\alpha\right\}^{m+1}}. \tag{2}$$

Then, the probability density function (pdf) of the $(k + 2)$-parameter OLLG-G family reduces to

$$f(x; \alpha, p, \tau) = \frac{(1-p)\alpha g(x; \tau) G^{\alpha-1}(x; \tau)[1 - G(x; \tau)]^{\alpha-1}}{\left\{G^\alpha(x; \tau) + (1-p)[1 - G(x; \tau)]^\alpha\right\}^2}. \tag{3}$$

The cdf corresponding to (3) is given by

$$F(x; \alpha, p, \tau) = \frac{G^\alpha(x; \tau)}{G(x; \tau)^\alpha + (1-p)[1 - G(x; \tau)]^\alpha}. \tag{4}$$

Henceforth, a random variable X having density function (3) is denoted by $X \sim \text{OLLG-G}(\alpha, p, \tau)$. For $\alpha = 1$, we obtain the geometric-G family with $(k+1)$ parameters (τ, p). We should emphasize that (4) is the extended generalized log-logistic distribution defined by Gleaton and Lynch [21]. However, the mathematical properties of the OLLG-G family and the associated regression derived in this paper represent new results. The OLL-G class is the limiting family (the limit is defined in terms of the convergence in distribution) of the OLLG-G class when $p \to 0^+$. Setting $p = 1 - p^*$, we obtain the *Marshall-Olkin* family Marshall and Olkin [22]. On the other hand, if $p \to 1^-$, we have the distribution of a random variable Y such that $P(Y = 0) = 1$. Hence, the parameter p can be interpreted as a degeneration parameter. The OLLG-normal and OLLG-Weibull distributions are obtained by taking $G(x; \tau)$ to be the normal and Weibull cumulative distributions, respectively.

2.1. Special Models

We present two special cases of the new family: the OLLG-Weibull (OLLG-W) and OLLG-normal (OLLG-N) distributions. First, the OLLG-W distribution is defined from (3) by taking $G(x;\tau)$ and $g(x;\tau)$ to be the cdf and pdf of the Weibull, $W(a,b)$, distribution, respectively, where $\tau = (a,b)^T$. Its pdf is given by

$$f_{OLLG-W} = \frac{a\alpha\,(1-p)\,x^{a-1}\exp\left[-(\alpha-1)\left(\frac{x}{b}\right)^a\right]\left\{1-\exp\left[-\left(\frac{x}{b}\right)^a\right]\right\}^{\alpha-1}}{b^a\left\{\left\{1-\exp\left[-\left(\frac{x}{b}\right)^a\right]\right\}^{\alpha}+(1-p)\exp\left[-\alpha\left(\frac{x}{b}\right)^a\right]\right\}^2}, \tag{5}$$

where the scale parameter $b > 0$, $a > 0$, $\alpha > 0$ and $0 < p < 1$ are shape parameters. A random variable with pdf (5) is denoted by $X \sim$ OLLG-W (α, p, a, b).

For $p \to 0^+$, we obtain the odd log-logistic Weibull (OLL-W) distribution da Cruz *et al.* [19] and for $\alpha = 1$ we have the geometric Weibull (Geo-W) distribution. We display plots of the OLLG-W densities and hazards in Figures 1 and 2, respectively. These plots reveal the bimodality of the density function and increasing, decreasing, unimodal, bathtub and increasing-decreasing-increasing of the hazard rate function (hrf)

Second, the OLLG-N distribution is defined from (3) by taking $G(x;\tau)$ and $g(x;\tau)$ to be the cdf and pdf of the normal $N(\mu, \sigma^2)$ distribution, respectively, where $\tau = (\mu, \sigma)^T$.
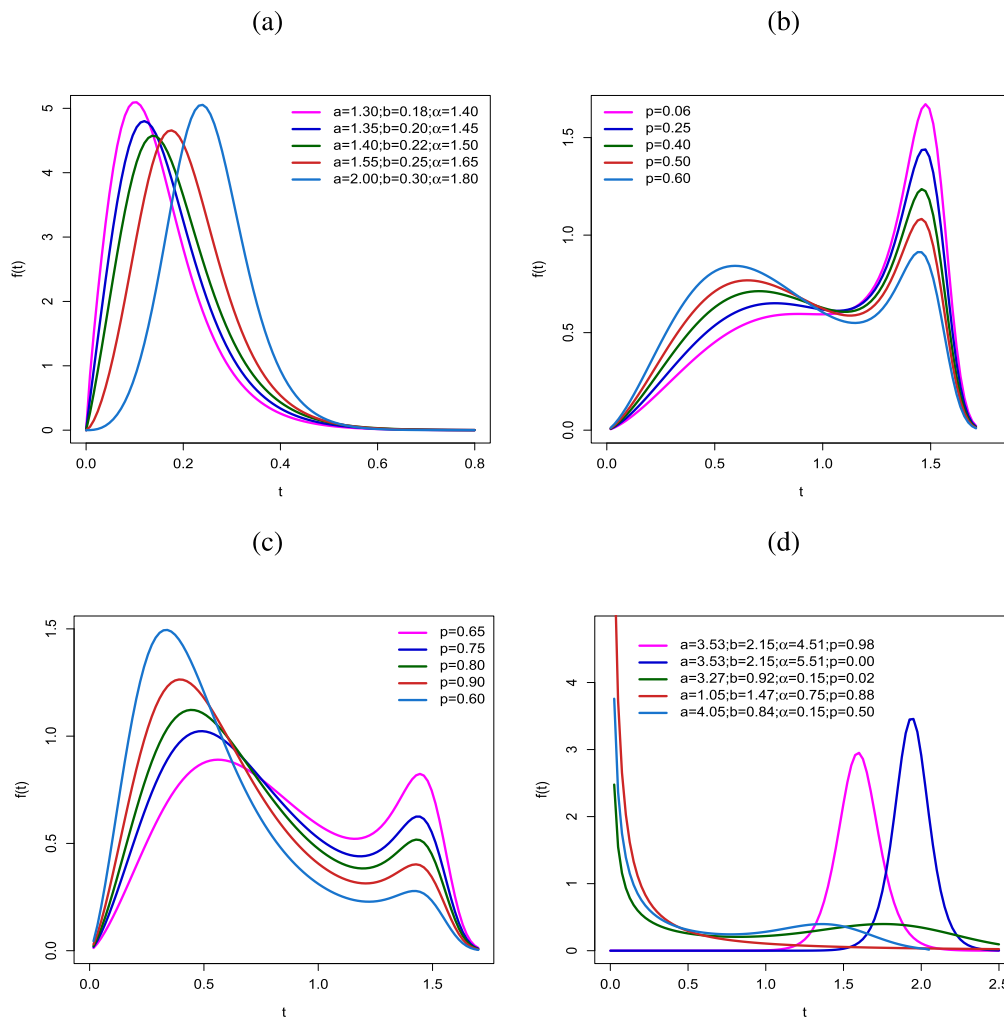

Figure 1 | Plots of the odd log-logistic geometric (OLLG)-Weibull density function for some parameter values.
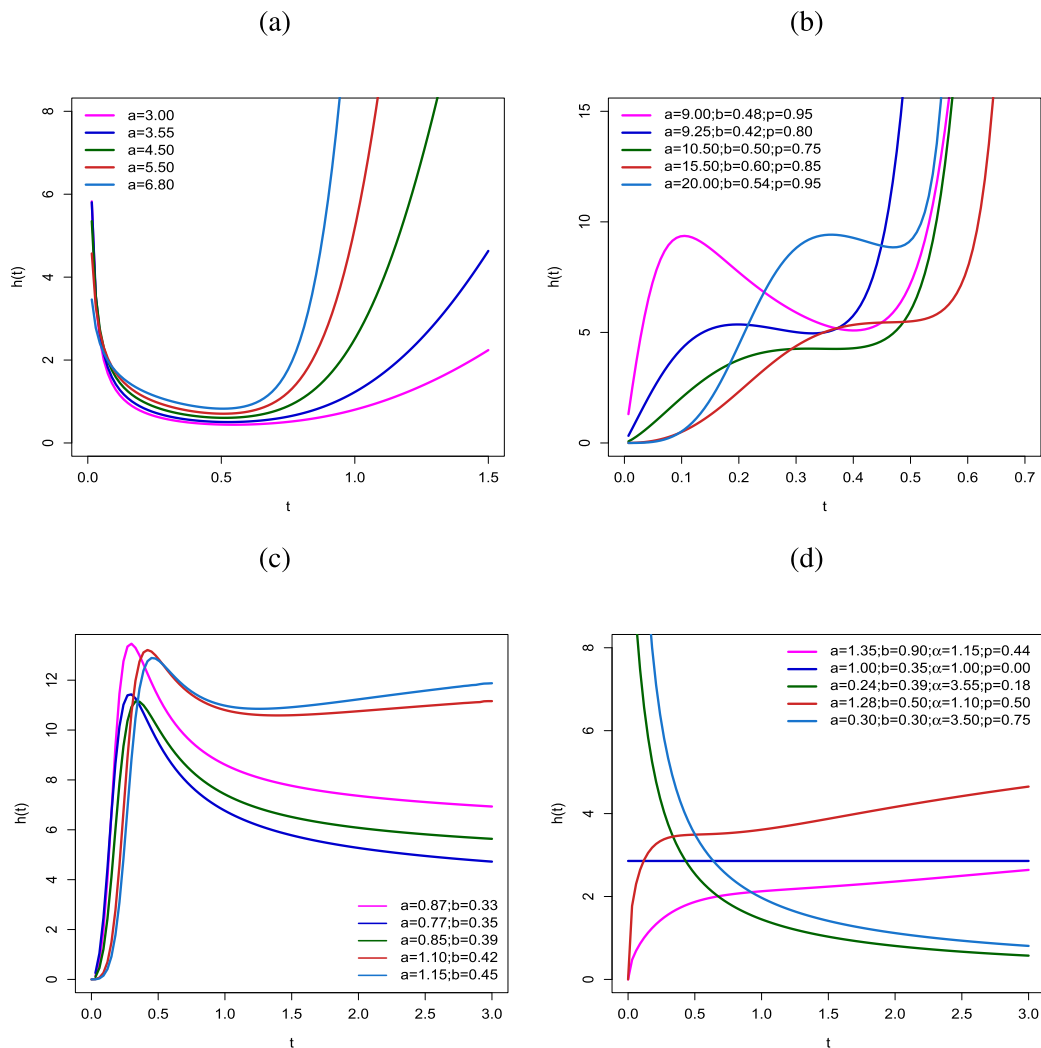(a) Fixed.

(a)

(b)

(c)

(d)

Figure 2 | Plots of the odd log-logistic geometric (OLLG)-Weibull hazard function for some parameter values. (a) Fixed.

Its density function is given by

$$f_{OLLG-N} = \frac{\alpha (1-p) \phi \left(\frac{x-\mu}{\sigma}\right) \left\{\Phi \left(\frac{x-\mu}{\sigma}\right) \left[1 - \Phi \left(\frac{x-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\left\{\Phi^\alpha \left(\frac{x-\mu}{\sigma}\right) + (1-p) \left[1 - \Phi^\alpha \left(\frac{x-\mu}{\sigma}\right)\right]\right\}^2},$$ (6)

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, $\alpha > 0$ and $0 < p < 1$ are shape parameters, and $\phi(.)$ and $\Phi(.)$ are the pdf and cdf of the standard normal distribution, respectively. A random variable with pdf (6) is denoted by $X \sim$OLLG-N $\left(\alpha, p, \mu, \sigma^2\right)$.

For $p \to 0^+$, we obtain the odd log-logistic normal (OLL-N) distribution Braga *et al.* [20] and when $\alpha = 1$ we have the geometric normal (Geo-N) distribution. For $\mu = 0$ and $\sigma = 1$, we have the standard OLLG-N distribution. Plots of the OLLG-N densities are displayed in Figure 3. These plots reveal that this density provides great flexibility in its forms, such as bimodality, asymmetry with bimodality and symmetry.

3. USEFUL EXPANSION

Theorem 3.1. *Let X be a random variable having the OLLG-G family. The cdf of X can be expressed as*

$$F(x) = \sum_{r=0}^{\infty} d_r H_r(x).$$ (7)

(a)   (b)

(c)   (d)
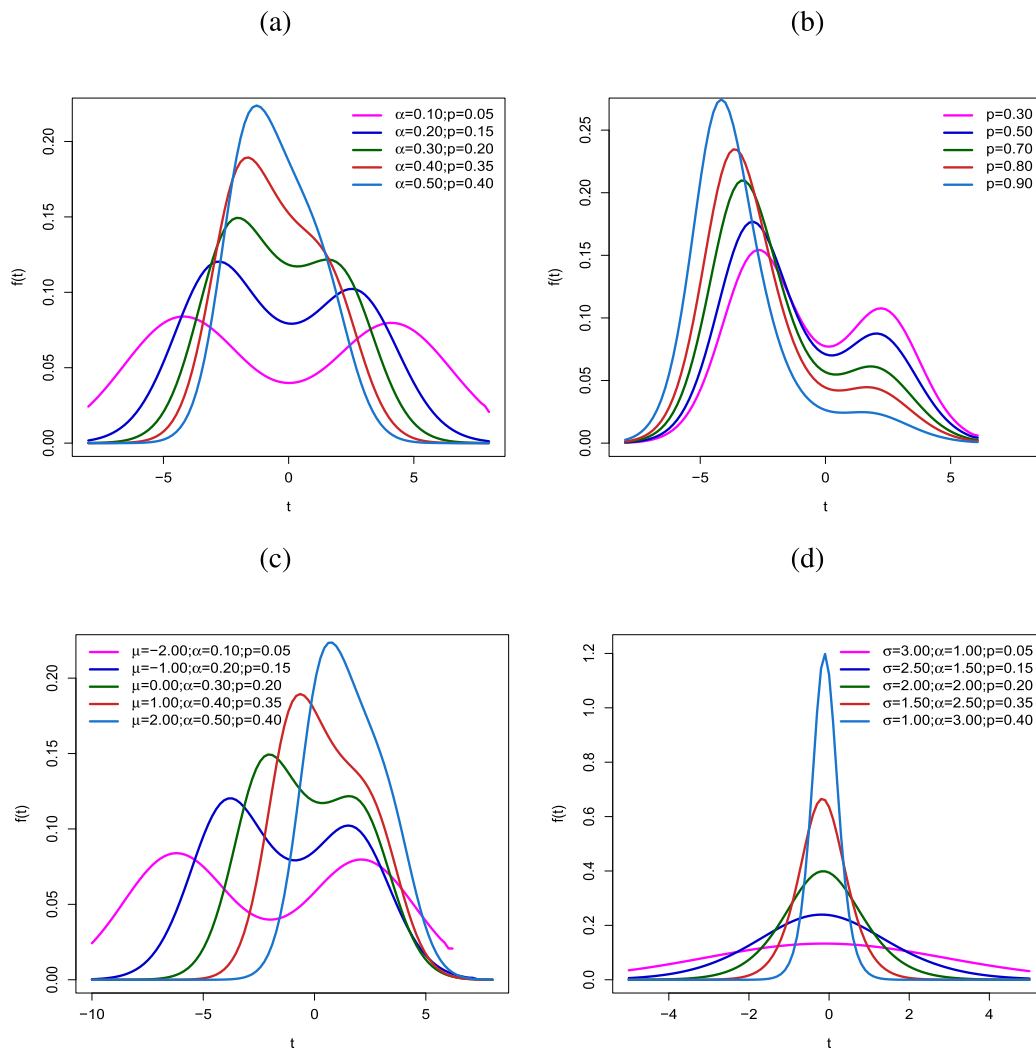
Figure 3 | Plots of the odd log-logistic geometric (OLLG)-normal density function for some parameter values.
(a) Fixed.

Here, $d_0 = s_0/c_0$ and, for $r \geq 1$,

$$d_r = c_0^{-1} \left(s_r - \sum_{m=1}^{r} c_m \, s_{r-m} \right),$$

where the quantities involved are defined below and $H_r(x) = G(x)^r$ is the exponentiated-G (exp-G) cdf with power parameter r.

Proof: We consider the power series

$$(1-z)^b = \sum_{k=0}^{\infty} (-1)^k \binom{b}{k} z^k. \tag{8}$$

which holds for any real non-integer $b > 0$ and $|z| < 1$.

Thus, we can rewrite (4) using (8) as

$$F(x) = \frac{G^\alpha(x)}{G^\alpha(x) + (1-p) \sum_{k=0}^{\infty} b_k \, G^k(x)}, \tag{9}$$

where $b_k = b_k(\alpha, p) = (-1)^k (1-p) \binom{\alpha}{k}$.

Further, for $z \in (0, 1)$, and real non-integer $\alpha > 0$, we have

$$z^\alpha = \sum_{r=0}^{\infty} s_r(\alpha) z^r. \tag{10}$$

where

$$s_r = s_r(\alpha) = \sum_{l=r}^{\infty} (-1)^{l+r} \binom{\alpha}{l} \binom{l}{r}.$$

Combining (9) and (10), we obtain

$$F(x) = \frac{\sum_{r=0}^{\infty} s_r(\alpha) G^r(x)}{\sum_{j=0}^{\infty} s_j(\alpha) G^j(x) + \sum_{i=0}^{\infty} J_i G^i(x)} = \frac{\sum_{r=0}^{\infty} s_r(\alpha) G^r(x)}{\sum_{t=0}^{\infty} c_t(\alpha, p) G^t(x)},$$

where $c_t = c_t(\alpha, p) = s_t(\alpha) + J_t$ and $J_t = J_t(\alpha, p) = \sum_{k=0}^{\infty} b_k s_t(k)$.

(7) follows from the ratio of two power series in the last equation.

Corollary 3.1. *The pdf of X is given by*

$$f(x) = \sum_{r=0}^{\infty} d_{r+1} h_{r+1}(x), \tag{11}$$

where d_{r+1} is given in (7) and $h_{r+1}(x)$ is the exp-G pdf with power parameter $r + 1$.

Theorem 3.1 and Corollary 3.1 are the main results of this section.

4. MATHEMATICAL PROPERTIES

4.1. Quantile Function

(4) has tractable properties especially for simulations, since the quantile function (qf) of X is given by

$$x_u = Q_G \left(\frac{u^{1/\alpha}(1-p)^{1/\alpha}}{(1-u)^{1/\alpha} + u^{1/\alpha}(1-p)^{1/\alpha}} \right), \tag{12}$$

where $Q_G(u) = G^{-1}(u)$ is the baseline qf.

Figures 4(a) and 4(b) display some plots of the OLLG-W kurtosis and skewness for some parameter values, respectively. We have positive kurtosis and positive and negative skewness. These plots reveal the flexibility of the proposed family with respect to the skewness.

4.2. Moments

In this section, we obtain the ordinary and incomplete moments of X from the moments of Y_{r+1} having the exp-G $(r + 1, \tau)$ distribution. The nth ordinary moment of X is given by

$$\mu'_n = E(X^n) = \sum_{r=0}^{\infty} d_{r+1} E\left(Y_{r+1}^n\right) = \sum_{r=0}^{\infty} (r+1) d_{r+1} \tau(n, r),$$

where d_{r+1} is given in Theorem 3.1 and $\tau(n, r) = \int_{-\infty}^{\infty} x^n g(x) G^r(x) dx = \int_0^1 Q_G^n(u) u^a du$.

Expressions for moments of several exp-G distributions are given in Nadarajah and Kotz [23].

The nth incomplete moment of X can be obtained from (11) as

$$m_n(y) = \sum_{r=0}^{\infty} d_{r+1} \int_0^y x^n h_{r+1}(x) = \sum_{r=0}^{\infty} (r+1) d_{r+1} \int_0^{G(y)} Q_G^n(x) u^r du.$$

The large number of applications of the ordinary and incomplete moments are well-known in the literature.
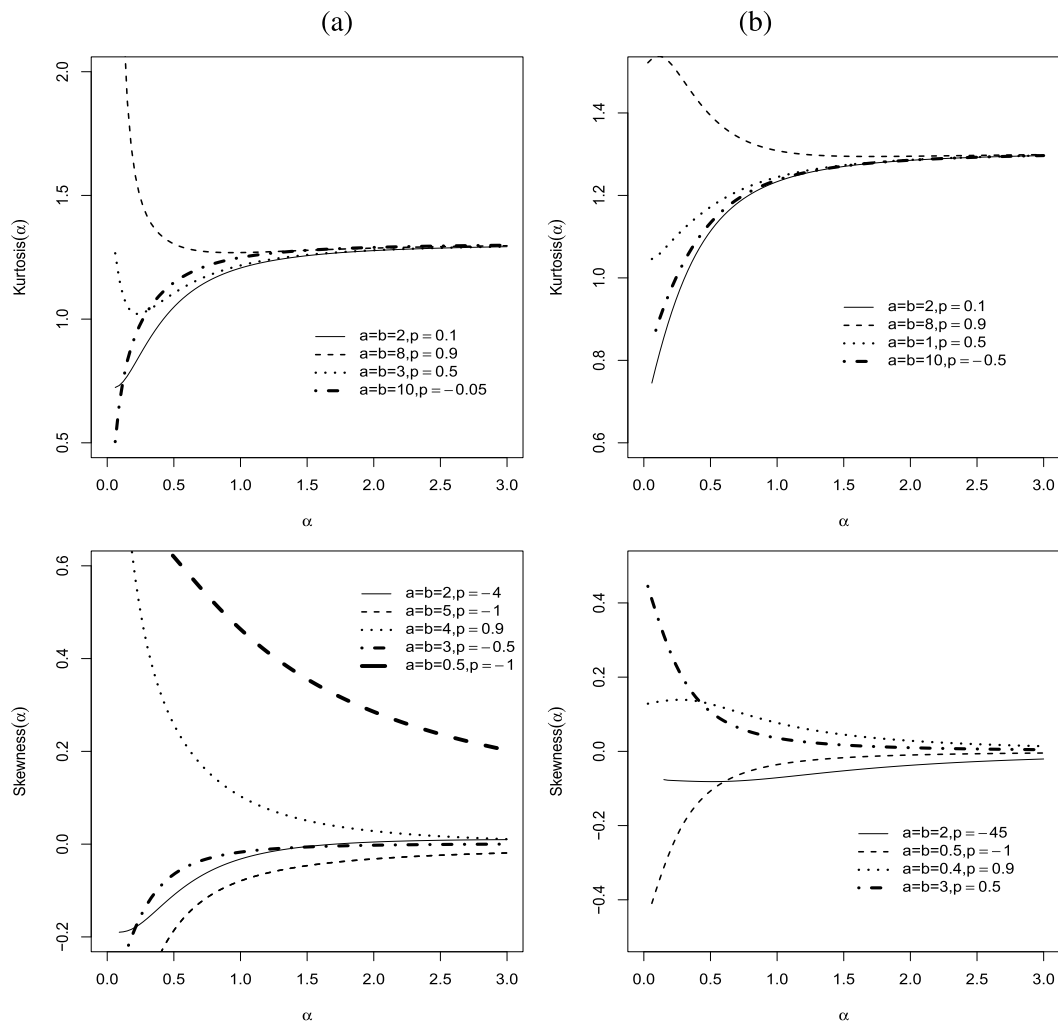
Figure 4 | Plots of the kutosis and skewness of the odd log-logistic geometric-Weibull (OLLG-W) and odd log-logistic geometric-normal (OLLG-N) models for some parameter values.

4.3. Generating Function

The moment generating function (mgf) $M(t)$ of X follows from (11) as

$$M(t) = \sum_{r=0}^{\infty} d_{r+1} M_{r+1}(t) = \sum_{r=0}^{\infty} (r+1) d_{r+1} \rho(t,r),$$

where $M_{r+1}(t)$ is the mgf of Y_{r+1} and

$$\rho(t,r) = \int_{-\infty}^{\infty} \exp(tx) g(x) G^a(x) dx = \int_0^1 \exp\{tQ_G(u)\} u^r du.$$

Thus, $M(t)$ can be determined from the exp-G generating function or from the baseline qf.

5. MAXIMUM LIKELIHOOD ESTIMATION

The parameters of the OLLG-G family are estimated by maximum likelihood from complete samples only. Let $x_1, \cdots, x_n$ be a random sample of size n from the OLLG-G (α, p, τ) distribution. The log-likelihood function for the vector of parameters $\theta = (\alpha, p, \tau^T)^T$ follows from (3) as

$$l(\theta) = n\left[\log\left(1-p\right) + \log\left(\alpha\right)\right] + \sum_{i=1}^{n} \log\left[g\left(x_i; \tau\right)\right] + \left(\alpha - 1\right)\sum_{i=1}^{n} \log\left[G\left(x_i; \tau\right)\right]$$
$$+ \left(\alpha - 1\right)\sum_{i=1}^{n} \log\left[1 - G\left(x_i; \tau\right)\right] - \sum_{i=1}^{n} \log\left\{G^{\alpha}\left(x_i; \tau\right) + \left(1-p\right)\left[1 - G(x_i; \tau)\right]^{\alpha}\right\}. \tag{13}$$

Let $\hat{\theta}$ be the maximum likelihood estimate (MLE) of θ. The log-likelihood can be maximized either directly by using the NLMixed script of SAS, the MaxBFGS sub-routine of the Ox program R software or by solving the nonlinear likelihood equations simultaneously. By using any optimization code to find $\hat{\theta}$, we face to the problem if (13) has more than one local maximum since in this situation the optimizer function with different initial values can lead to different local maximums. So, a simple procedure for finding initial parameter values is: first, plot the density function with some parameter values on the histogram of the data; second, find a good vector of parameters with good fit to the histogram and choose this vector as the initial parameter vector to the optimization problem.

The asymptotic distribution of $\left(\hat{\theta} - \theta\right)$ is $N_{k+2}\left(0, J\left(\hat{\theta}\right)^{-1}\right)$ under standard regularity conditions, where $J(\theta)$ is total observed information matrix. Based on this distribution, we can construct approximate confidence intervals for the individual parameters.

We can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain likelihood ratio (LR) statistics for testing some sub-models of the OLLG-G distribution. To ensure that $p \in (0, 1)$, we adopt the following re-parametrization for $p = \dfrac{\exp\left(p^*\right)}{1 + \exp\left(p^*\right)}$, where $p^* \in \mathbb{R}$. This procedure is used in the simulations and applications.

So, we can construct LR statistics to check if the fit using the OLLG-G distribution is statistically "superior" to a fit using the OLL-G, and geometric-G family for a given data set. In any case, hypothesis tests of the type $H_0 : \psi = \psi_0$ versus $H_1 : \psi \neq \psi_0$, where ψ is a vector formed with some components of θ and θ_0 is a specified vector, can be performed using LR statistics. For example, the test of $H_0 : \alpha = 1$ versus $H_1 : H_0$ *is not true* is equivalent to compare the OLLG-G and geometric family and the LR statistic reduces to

$$w = 2\left\{\ell\left(\hat{\alpha}, \hat{p}, \hat{\tau}\right) - \ell\left(\widetilde{\alpha}, 1, \widetilde{\tau}\right)\right\},$$

where $\hat{\alpha}$, $\hat{p}$, and $\hat{\tau}$ are the MLEs under H and $\widetilde{\alpha}$ and $\widetilde{\tau}$ are the estimates under H_0.

5.1. Simulation Study

In this section, we assess the finite sample performance of the MLEs in the OLLG-W and OLLG-N distributions, respectively, by varying the true parameter and the sample size n. We perform a Monte Carlo simulation (with $1,000$ replications) to quantify some asymptotic properties of the MLEs of the model parameters using the R software.

First, we consider the OLLG-W model under some parametric variation structure and measure the effects of the MLEs $\hat{a}$, $\hat{b}$, $\hat{\alpha}$ and $\hat{p}$. To that end, we consider $a \in \{3.53, 11.5, 5.50\}$, $b \in \{2.15, 1.25\}$, $\alpha \in \{0.35, 0.21\}$ and $p \in \{0.25, 0.60, 0.90\}$ for the sample sizes $n = 80, 150$ and 300. For each parametric point, the Monte Carlo averages of the MLEs and the mean square errors (MSEs) are determined. The figures in Table 1 reveal that the average estimates (AEs) tend to the parameter values when n increases. Thus, the first-order asymptotic properties of the MLEs hold for the OLLG-W distribution.

Next, we consider the OLLG-N model to measure both the effects of the MLEs $\hat{\mu}$, $\hat{\sigma}$, $\hat{p}$ and $\hat{\alpha}$. To that end, we consider $\mu \in \{2.00, 0.00, -1.00\}$, $\sigma \in \{1.00, 0.75, 1.00\}$, $\alpha \in \{0.50, 0.15, 0.20\}$ and $p \in \{0.40, 0.90, 0.15\}$ for the sample sizes $n = 80, 150$ and 300. For each parametric point, the Monte Carlo averages of the MLEs and MSEs are determined. The figures in Table 2 reveal that the AEs become closer to the true parameter values when the sample size increases. This fact proves empirically that the first-order asymptotic properties of the MLEs are satisfied for the OLLG-N distribution.

Table 1 | Simulations for the OLLG-W model.

(a, b, α, p)	n	AEs				MSEs			
		a	b	α	p	a	b	α	p
(3.53,2.15,0.35,0.25)	80	4.3287	2.2555	0.3277	0.2963	4.0096	0.1164	0.0088	0.0538
	150	3.8360	2.2012	0.3421	0.2667	0.8711	0.0542	0.0046	0.0336
	300	3.6832	2.1767	0.3464	0.2566	0.3894	0.0320	0.0023	0.0236
(11.5,1.25,0.21,0.60)	80	13.5820	1.2582	0.1993	0.5932	28.4536	0.0049	0.0043	0.0193
	150	12.4670	1.2547	0.2041	0.5964	8.8640	0.0023	0.0018	0.0095
	300	11.9975	1.2537	0.2068	0.6020	3.4259	0.0010	0.0009	0.0043
(5.50,1.25,0.21,0.90)	80	9.4602	1.2924	0.1806	0.8895	51.0649	0.0649	0.0183	0.0075
	150	7.1223	1.2703	0.2020	0.8929	13.8234	0.0379	0.0124	0.0025
	300	6.2040	1.2623	0.2042	0.8988	4.1319	0.0159	0.0041	0.0007

OLLG-W, odd log-logistic geometric-Weibull; AE, average estimate; MSE, mean square error.

Table 2 | Simulations for the OLLG-N model.

(μ, σ, α, p)	n	AEs				MSEs			
		μ	σ	α	p	μ	σ	α	p
(2.00,1.00,0.50,0.40)	80	2.0534	1.0035	0.5074	0.3733	0.4381	0.2232	0.0971	0.0577
	150	2.0120	0.9893	0.5022	0.3722	0.1966	0.1627	0.0779	0.0550
	300	2.0080	1.0007	0.5026	0.3885	0.0548	0.0391	0.0186	0.0217
(0.00,0.75,0.15,0.90)	80	−0.0036	0.8056	0.1945	0.8854	3.5090	0.3043	0.0405	0.0031
	150	−0.0357	0.7817	0.1752	0.8938	0.2753	0.0713	0.0137	0.0015
	300	−0.0064	0.7602	0.1606	0.8962	0.0953	0.0309	0.0055	0.0012
(−1.00,1.00,0.20,0.15)	80	−0.9114	1.0140	0.2142	0.2139	0.1436	0.0944	0.0119	0.0304
	150	−0.9345	1.0180	0.2105	0.1840	0.0823	0.0475	0.0059	0.0200
	300	−0.9462	1.0097	0.2057	0.1853	0.0405	0.0199	0.0023	0.0136

OLLG-N, odd log-logistic geometric-normal; AE, average estimate; MSE, mean square error.

6. THE LOG-OLLG-W REGRESSION MODEL WITH VARYING DISPERSION

In this section, we define a location-scale regression model with varying dispersion based on the OLLG-W distribution called the LOLLG-W regression model with varying dispersion. We consider a classic analysis for this regression model and the inferential part is carried out using the asymptotic distribution of the MLEs.

Henceforth, X is a random variable following the OLLG-W density function (5) and Y is defined by $Y = \log(X)$. The density function of Y by replacing $a = 1/\sigma$ and $b = \exp(\mu)$ reduces to

$$f(y) = \frac{(1-p)\,\alpha}{\sigma} \exp\left[\left(\frac{y-\mu}{\sigma}\right) - \alpha \exp\left(\frac{y-\mu}{\sigma}\right)\right] \left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha-1} \times$$
$$\left\{\left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha} + (1-p)\exp\left[-\alpha \exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{-2}, \tag{14}$$

where $y \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\alpha > 0$, $\sigma > 0$ and $0 < p < 1$. We refer to (14) as the LOLLG-W distribution, say $Y \sim \text{LOLLG-W}(p, \alpha, \sigma, \mu)$, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and p and α are shape parameters.

The survival function of Y reduces to

$$S(y) = \frac{(1-p)\exp\left[-\alpha \exp\left(\frac{y-\mu}{\sigma}\right)\right]}{\left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha} + (1-p)\exp\left[-\alpha \exp\left(\frac{y-\mu}{\sigma}\right)\right]}. \tag{15}$$

The random variable $Z = (Y - \mu)/\sigma$ has density function

$$f(z) = \frac{(1-p)\,\alpha\,\exp\left[z - \alpha \exp(z)\right]\left\{1 - \exp\left[-\exp(z)\right]\right\}^{\alpha-1}}{\left\{\left\{1 - \exp\left[-\exp(z)\right]\right\}^{\alpha} + (1-p)\exp\left[-\alpha \exp(z)\right]\right\}^{2}}. \tag{16}$$

A parametric model that provides a good fit to lifetime data tends to yield more precise estimates of the quantities of interest. Based on the LOLLG-W density, we propose a linear location-scale regression model with varying dispersion for the response variable y_i given by

$$y_i = \mu_i + \sigma_i z_i, \ \ i = 1, \dots, n, \tag{17}$$

where the random error z_i has density function (16), and the parameters μ_i and σ_i are parametrized as

$$\mu_i = \mu_i(\beta_1) \quad \text{and} \quad \sigma_i = \sigma_i(\beta_2),$$

where $\beta_1 = \left(\beta_{11}, \dots, \beta_{1p_1}\right)^T$ and $\beta_2 = \left(\beta_{21}, \dots, \beta_{2p_2}\right)^T$. The usual systematic component for the location parameter is $\mu_i = \mathbf{d}_i^T \beta_1$, where $\mathbf{d}_i = \left(d_{i1}, \dots, d_{ip_1}\right)^T$ is a vector of known explanatory variables, i.e., $\mu = \mathbf{D}\beta_1$, with $\mu = (\mu_1, \dots, \mu_n)^T$ and $\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_n)^T$ is a specified $n \times p_1$ matrix of full rank and $p_1 < n$. Analogously, for the dispersion parameter, we consider the systematic component $g(\sigma_i) = \eta_i = \mathbf{v}_i^T \beta_2$, where $g(\cdot)$ is a dispersion link function, and $\mathbf{v}_i = \left(v_{i1}, \dots, v_{iq}\right)^T$ is a vector of known explanatory variables, i.e., the linear structure in $g(\sigma_i)$ measures the dispersion for the ith observation. Then, we have $g(\beta) = \eta = \mathbf{V}\beta_2$, where $\beta = (\sigma_1, \dots, \sigma_n)^T$, $\eta = (\eta_1, \dots, \eta_n)^T$ and $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n)^T$ is a specified $n \times p_2$ matrix of full rank and $p_2 < n$. The dispersion covariates in $\mathbf{V}$ are commonly, but not necessary, regression covariates in $\mathbf{X}$. It is assumed that β_1 is functionally independent of β_2. Further, we consider that $g(\cdot)$ is a known one-to-one continuously twice differentiable function.

The LOLLG-W regression model with varying dispersion (17) opens new possibilities for fitting many different types of data. The LOLL-W regression model with varying dispersion is a special case when $p \to 0^+$. For $p \to 0^+$ and $\sigma_1 = \sigma$, we have the LOLL-W regression model defined by da Cruz *et al.* [19]. We obtain the log-Weibull (LW) regression model with varying dispersion when $p \to 0^+$ and $\alpha = 1$. For $p \to 0^+$, $\alpha = 1$ and $\sigma_i = \sigma$, we have the classical LW regression model Lawless [24].

Consider a sample $(y_1, \mathbf{d}_1, \mathbf{v}_1), \cdots, (y_n, \mathbf{d}_n, \mathbf{v}_n)$ of n independent observations, where each random response is defined by $y_i = \min\{\log(x_i), \log(c_i)\}$. We assume non-informative censoring such that the observed lifetimes and censoring times are independent. Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. Conventional likelihood estimation techniques can be applied here. The log-likelihood function for the vector of parameters $\theta = (p, \alpha, \beta_1^T, \beta_2^T)^T$ from model (17) has the form $l(\theta) = \sum_{i \in F} l_i(\theta) + \sum_{i \in C} l_i^{(c)}(\theta)$, where $l_i(\theta) = \log[f(y_i)]$, $l_i^{(c)}(\theta) = \log[S(y_i)]$, $f(y_i)$ is the density (14) and $S(y_i)$ is the survival function (15) of Y_i. The total log-likelihood function for θ reduces to

$$
\begin{aligned}
l(\theta) = {}& r \log\left[\frac{(1-p)\,\alpha}{\sigma_i}\right] + (n-r)\log(1-p) + \sum_{i \in F} z_i - \alpha \sum_{i \in F} \exp(z_i) + \\
& (\alpha - 1)\sum_{i \in F} \log\{1 - \exp[-\exp(z_i)]\} - \alpha\sum_{i \in C}\exp(z_i) - \\
& 2\sum_{i \in F}\log\{\,\{1 - \exp[-\exp(z_i)]\}^\alpha + (1-p)\exp[-\alpha\exp(z_i)]\,\} + \\
& \sum_{i \in C}\log\{\,\{1 - \exp[-\exp(z_i)]\}^\alpha + (1-p)\exp[-\alpha\exp(z_i)]\,\},
\end{aligned}
\tag{18}
$$

where $z_i = (y_i - \mu_i)/\sigma_i$ and r is the number of uncensored observations (failures). The MLE $\hat{\theta}$ of the vector of unknown parameters can be evaluated by maximizing the log-likelihood (18). We use the R software to calculate $\hat{\theta}$. Initial values for β_1 and β_2 are taken from the fit of the LW regression model with varying dispersion with $p \to 0^+$ and $\alpha = 1$. The fit of the LOLLG-W model yields the estimated survival function for y_i ($\hat{z}_i = (y_i - \hat{\mu}_i)/\hat{\sigma}_i$) given by (15) at $\hat{\theta}$, say $S(y_i; \hat{p}, \hat{\alpha}, \hat{\mu}_i, \hat{\sigma}_i)$

The asymptotic distribution of $(\hat{\theta} - \theta)$ is multivariate normal $N_{p_1+p_2+2}(0, K(\theta)^{-1})$ under standard regularity conditions, where $K(\theta)$ is the information matrix. The asymptotic covariance matrix $K(\theta)^{-1}$ of $\hat{\theta}$ can be approximated by the inverse of the $(p_1 + p_2 + 2) \times (p_1 + p_2 + 2)$ observed information matrix $-\ddot{\mathbf{L}}(\theta)$. The elements of the observed information matrix $-\ddot{\mathbf{L}}(\theta)$ can be evaluated numerically. The approximate multivariate normal distribution $N_{p_1+p_2+2}(0, -\ddot{\mathbf{L}}(\theta)^{-1})$ for $\hat{\theta}$ can be used in the classical way to construct approximate confidence regions for some parameters in θ. We can adopt LR statistics for comparing some special models with the LOLLG-W regression model in the usual way.

7. APPLICATIONS

In this section, we provide three applications to real data to prove empirically the flexibility of the OLLG-W and OLLG-N models. The computations are performed using the R software. In the first application, we compare the OLLG-N with geometric normal (Geo-N), OLL-N, Kumaraswamy normal (Kw-N), exponentiated normal (EN), gamma normal (GN) and normal distributions. In the second application, we compare the OLLG-W model with the geometric Weibull (Geo-W), OLL-W, Kumaraswamy Weibull (Kw-W), exponentiated Weibull (EW), gamma Weibull (GW), flexible Weibull (FW) and Weibull models. In the third application, a regression model is investigated considering the OLLG-W distribution.

The FW distribution Bebbington *et al.* [25], having two parameters $a_1 > 0$ and $b_1 > 0$, has pdf given by

$$
f(t) = \left(a_1 + \frac{b_1}{t^2}\right)\exp\left[\left(a_1 t - \frac{b_1}{t}\right) - \exp\left(a_1 t - \frac{b_1}{t}\right)\right], \quad t > 0.
$$

The gamma-G density function Zografos and Balakrishnan [26] is given by

$$
f(t) = \frac{g(t)}{\Gamma(k)}\{-\log[1 - G(t)]\}^{k-1}, \quad t > 0,
$$

where $k > 0$ is the shape parameter. The GW and GN are obtained by assuming the Weibull and normal baselines, respectively.

The Kumaraswamy-G (Kw-G) family was defined by Cordeiro and de Castro [3]. The associated density function, with two positive shape parameters λ and φ, is given by

$$
f(t) = \lambda\,\varphi\,g(t)\,G(t)^{\lambda-1}[1 - G(t)^\lambda]^{\varphi-1}.
$$

The exponentiated-G (Exp-G) is a special case of Kw-G when $\varphi = 1$. The Kw-W and Kw-N models are obtained by assuming the Weibull and normal baselines, respectively. The EW and EN distributions are obtained by fixing $\varphi = 1$.

We report the MLEs (and the corresponding standard errors in parentheses) of the model parameters and the following goodness-of-fit measures: Cramr-von Mises (W^*), Anderson Darling (A^*) and Kolmogarov-Smirnov (KS) statistics. The smallest values of these measures can provide the best fit among some fitted models. These measures, as well as the estimates, are obtained through the R software using the *AdequacyModel* package and the *optim* function for regression model estimates, using in both cases the BFGS method.

7.1. Uncensored Data

First, we consider two uncensored data sets described below.

Guinea pig survival times (E1): The data refer to survival times on days of 72 guinea pigs analyzed after being injected with infectious bacteria in a medical experiment.

Voltage data (E2): The data set was studied by Meeker and Escobar ([27], p. 383), which gives the times of failure and running times for a sample of devices from a field-tracking study of a larger system. At a certain point in time, 30 units were installed in normal service conditions. Two causes of failure were observed for each unit that failed: the failure caused by an accumulation of randomly occurring damage from power-line voltage spikes during electric storms and failure caused by normal product wear.

To analyze the E1 data, we consider the OLLG-N model and for the E2 data we take the OLLG-W model. Table 3 gives a descriptive summary for these data showing different degrees of skewness and kurtosis. We note that the E1 data have positive asymmetry and kurtosis and that the E2 data have negative values for both.

Tables 4 and 5 list the MLEs and standard errors (in parentheses) of the parameters of the fitted models to both data sets. The values of the statistics are presented in Table 6 to verify the goodness-of-fit of the models under study. The results in this table indicate that the OLLG-N and OLLG-W models have the lowest values of these statistics among all fitted models to both data sets.

A comparison of the proposed distributions with some of their sub-models using LR statistics is provided in Table 7. The *p*-values indicate that the proposed models yield the best fits for the two data sets. However, for the E2 set, the OLL-W model can be considered a competitive model for the OLLG-W model, because the *p*-value is close to the usual limit of 5%.

Table 3 | Descriptive statistics for the two data sets.

Data	Mean	Median	SD	Skewness	Kurtosis	Min.	Max.
E1	141.8	102.5	109.2086	2.4631	6.0742	43.0	598.0
E2	177.0	196.5	144.9922	−0.2699	−1.6416	2.0	300.0

Table 4 | MLEs of the model parameters for the E1 data set.

Model	μ	σ	α	p
OLLG-N	374.7037	30.7166	0.1058	0.9837
	(0.0852)	(0.0859)	(0.0106)	(0.0080)
OLL-N	125.6653	246.6182	3.1757	(0)
	(9.6367)	(63.4368)	(0.9672)	(−)
Geo-N	395.4403	106.1374	(1)	0.9957
	(52.3187)	(10.9766)	(−)	(0.0040)
	μ	σ	λ	φ
Kw-N	−335.3843	132.0140	961.8518	0.4019
	(0.2126)	(0.2141)	(194.1546)	(0.0539)
EN	−500.6958	209.3812	528.9657	1
	(0.0002)	(0.0244)	(0.0852)	(−)
Normal	141.8156	108.5312	1	1
	(12.7905)	(9.0547)	(−)	(−)
	μ	σ	k	
GN	−99.4173	125.2552	3.9278	
	(48.7530)	(9.6677)	(0.8960)	

OLLG-N, odd log-logistic geometric-normal; MLE, maximum likelihood estimate; OLL-N, odd log-logistic normal; Geo-N, geometric normal; Kw-N, Kumaraswamy normal; EN, exponentiated normal; GN, gamma normal; OLLG-W, odd log-logistic geometric-Weibull; OLL-W, odd log-logistic Weibull; Geo-W, geometric Weibull; Kw-W, Kumaraswamy Weibull; EW, exponentiated Weibull; GW, gamma Weibull; FW, flexible Weibull.

Table 5 | MLEs of the model parameters for the E2 data set.

Model	a	b	α	p
OLLG-W	18.8383	247.9113	0.0506	0.1864
	(0.0008)	(0.0008)	(0.0075)	(0.00024)
OLL-W	10.5453	218.1834	0.0772	(0)
	(0.0398)	(0.0407)	(0.0115)	(−)
Geo-W	1.2650	188.0544	(1)	2.06×10^{-09}
	(0.0068)	(0.0246)	(−)	(1.41×10^{-09})

Model	a	b	λ	φ
Kw-W	71.1698	295.9313	0.0052	0.2713
	(22.2411)	(2.8993)	(0.0024)	(0.0755)
EW	20.2227	10228.0700	0.0109	1
	(0.0046)	(109.1352)	(0.0019)	(−)
Weibull	1.2650	188.0544	1	1
	(0.2044)	(28.2172)	(−)	(−)

Model	a	b	k
GW	7.0283	351.5713	0.1282
	(0.0025)	(0.0026)	(0.0231)

Model	a_1	b_1
FW	0.0032	15.8889
	(0.0004)	(5.2696)

MLE, maximum likelihood estimate; OLLG-W, odd log-logistic geometric-Weibull; OLL-W, odd log-logistic Weibull; Geo-W, geometric Weibull; Kw-W, Kumaraswamy Weibull; EW, exponentiated Weibull, GW, gamma Weibull, FW, flexible Weibull.

Table 6 | Goodness-of-fit for the two data sets.

			E1 Data				
Model	AIC	CAIC	BIC	HQIC	W^*	A^*	KS
OLLG-N	820.3	820.9	829.4	823.9	0.3636	2.0453	0.1119
OLL-N	865.1	865.4	871.9	867.8	0.9985	5.7098	0.1568
Geo-N	832.8	833.1	839.6	835.5	0.5186	3.0225	0.1182
Kw-N	833.8	833.4	841.9	836.4	0.5292	3.1417	0.1914
EN	843.9	844.3	850.8	846.6	0.7104	4.1636	0.1830
Normal	883.1	883.3	887.7	884.9	1.3648	7.5981	0.2310
GN	879.1	879.5	886.0	881.9	1.2726	7.1386	0.2277
			E2 Data				
Model	AIC	CAIC	BIC	HQIC	W^*	A^*	KS
OLLG-W	331.4	333.0	337.0	333.2	0.0808	0.7022	0.1543
OLL-W	339.8	340.7	344.0	341.1	0.0850	0.7955	0.1688
Geo-W	374.6	375.5	378.8	375.9	0.3035	1.8208	0.2192
Kw-W	331.5	333.1	337.2	333.3	0.1129	0.9476	0.1494
EW	437.6	438.5	441.8	438.9	0.3701	2.1273	0.5391
Weibull	372.6	373.0	375.4	373.5	10.7180	56.4562	0.9987
GW	359.8	360.7	364.0	361.1	0.1963	1.2839	0.2277
FW	387.6	388.0	390.4	388.5	0.3246	2.0498	0.3942

OLLG-N, odd log-logistic geometric-normal; OLL-N, odd log-logistic normal; Geo-N, geometric normal; Kw-N, Kumaraswamy normal; EN, exponentiated normal; GN, gamma normal; OLLG-W, odd log-logistic geometric-Weibull; OLL-W, odd log-logistic Weibull; Geo-W, geometric Weibull; Kw-W, Kumaraswamy Weibull; EW, exponentiated Weibull, GW, gamma Weibull, FW, flexible Weibull.

In order to assess if the model is appropriate, plots of the fitted OLLG-N, Geo-N, Kw-N and normal density functions are displayed in Figure 5(a) and survival functions and the plots of the empirical distributions are displayed in Figure 5(b). Figure 6(a) provides plots of the fitted density functions of the OLLG-W, OLL-W, Kw-N and Weibull models and Figure 6(b) gives the survival functions and the plots to the empirical distributions. These plots indicate that the OLLG-N and OLLG-W distributions provide better fits to the current data sets.

Table 7 | LR tests for the two data sets.

	E1 Data		
Models	**Hypotheses**	**Statistic w**	***p*-value**
OLLG-N vs Normal	$H_0 : p = 0$ and $\alpha = 1$ vs $H_1 : H_0$ is false	7.0	0.0301
OLLG-N vs OLL-N	$H_0 : p = 0$ vs $H_1 : H_0$ is false	6.2	0.0121
OLLG-N vs Geo-N	$H_0 : \alpha = 1$ vs $H_1 : H_0$ is false	4.0	0.0455
	E2 Data		
OLLG-W vs Weibull	$H_0 : p = 0$ and $\alpha = 1$ vs $H_1 : H_0$ is false	6.2	0.0430
OLLG-W vs OLL-W	$H_0 : p = 0$ vs $H_1 : H_0$ is false	3.5	0.0604
OLLG-W vs Geo-W	$H_0 : \alpha = 1$ vs $H_1 : H_0$ is false	6.1	0.0131

OLLG-N, odd log-logistic geometric-normal; OLL-N, odd log-logistic normal; Geo-N, geometric normal; OLLG-W, odd log-logistic geometric-Weibull; OLL-W, odd log-logistic Weibull; Geo-W, geometric Weibull; LR, likelihood ratio.
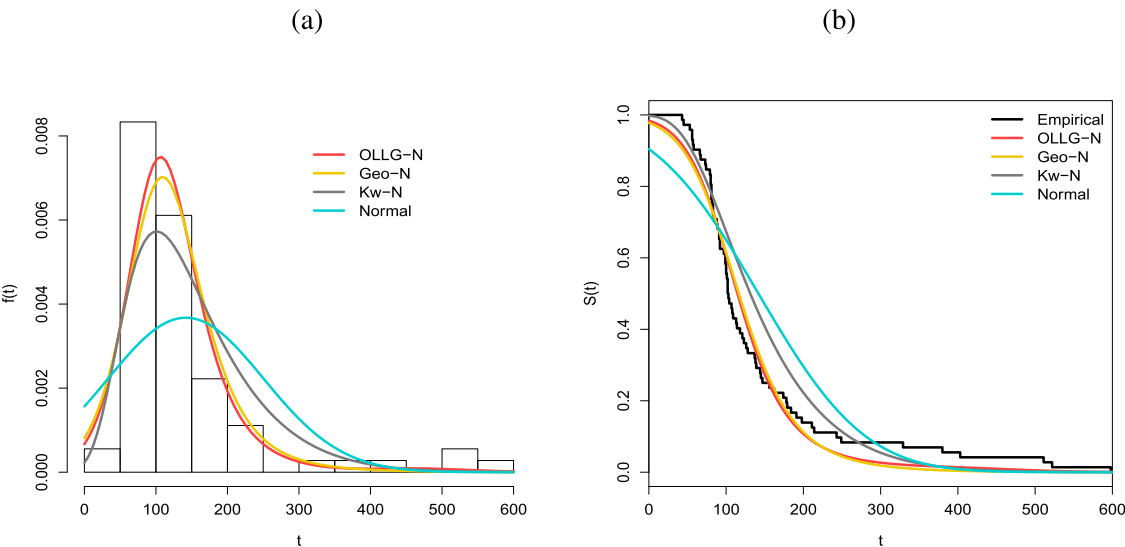

Figure 5 | (a) Estimated densities of the odd log-logistic geometric-normal (OLLG-N), geometric normal (Geo-N), Kumaraswamy normal (Kw-N) and Normal models for E1 data. (b) Estimated survival functions of the OLLG-N, Geo-N, Kw-N and Normal models for E1 data.

7.2. Regression Model with Varying Dispersion for Life Expectancy Data

The data were analyzed by Weindruch *et al.* [28] who compared the life expectancy of field mice under different diets. The authors randomly assigned 244 mice to one of four dietary treatments. The data are available in *asbio* package of the R software, which contains a data frame with 244 observations on the following two variables: t_i = lifetime in months of the ith individual in the jth treatment, $x_{ij} = i$th individual at the jth level of treatment, for $i = 1, \dots, 244$ and $j = 1, 2, 3$ and 4. The treatments correspond to a factor levels N/N85: mice were fed normally both before and after weaning (the slash distinguishes pre and post weaning). After weaning the diet consisted of 85kcal/week, a conventional total for mice rearing, N/R40: mice were fed normally before weaning, but were given a severely restricted diet of 40 kcal per week after feeding, N/R50: mice were restricted to 50kcal per week before and after weaning, R/R50: mice were fed normally before weaning, but their diet were restricted to 50 kcal per week after weaning. The analysis considering the LOLLG-W regression model is performed with the definition of dammy variables as follows: levels N/N85 ($d_{i1} = 1$, $d_{i2} = 0$ and $d_{i3} = 0$), levels N/R40 ($d_{i1} = 0$, $d_{i2} = 1$ and $d_{i3} = 0$), levels N/R50 ($d_{i1} = 0$, $d_{i2} = 0$ and $d_{i3} = 1$) and levels R/R50 ($d_{i1} = 0$, $d_{i2} = 0$ and $d_{i3} = 0$).

- **Exploratory analysis for life expectancy data**
  Here, we perform an exploratory analysis of these data. Table 8 gives a descriptive summary of the data showing different degrees of skewness and kurtosis. In Figure 7(a) the boxplot of the response variable is shown and in Figure 7(b) the box-plots considering the treatment levels.
  Table 9 lists the MLEs and standard errors between parentheses of the parameters of the fitted models to the current data. The values of the statistics are given in Table 10 to verify the goodness-of-fit of the fitted models. The results indicate that the OLLG-W model has the lowest values of these statistics among all fitted models to the data.
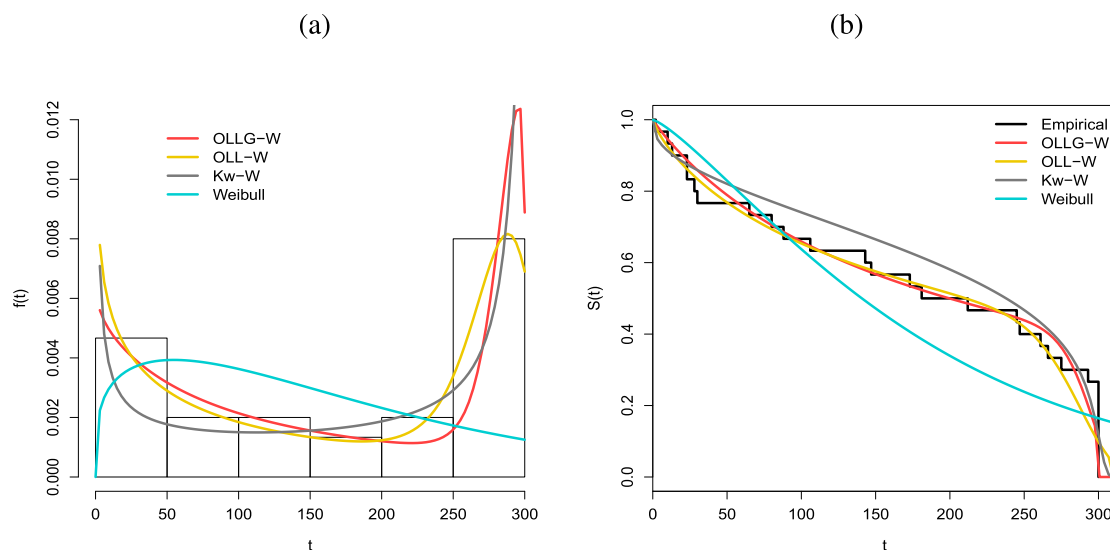
Figure 6 | (a) Estimated densities of the odd log-logistic geometric-Weibull (OLLG-W), odd log-logistic Weibull (OLL-W), Kumaraswamy Weibull (Kw-W) and Weibull models for E2 data. (b) Estimated survival functions of the OLLG-W, OLL-W, Kw-W and Weibull models for E2 data.

Table 8 | Descriptive statistics for the life expectancy data.

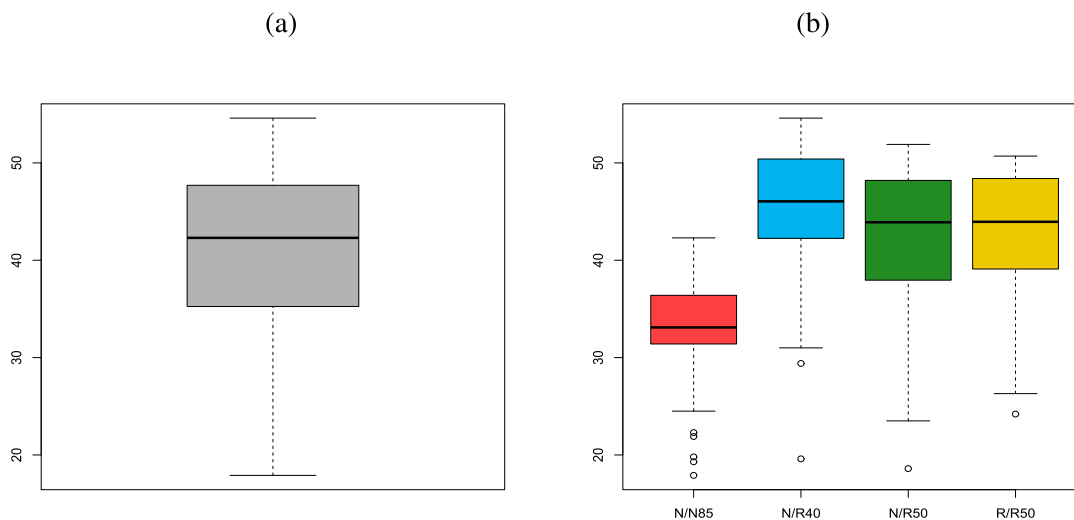|	**Mean**	**Median**	**SD**	**Skewness**	**Kurtosis**	**Min.**	**Max.**
General	40.88	42.30	8.1321	−0.5547	−0.3514	17.90	54.60
N/N85	32.69	33.10	5.1252	−1.0390	1.0485	17.90	42.30
N/R40	45.12	46.05	6.7034	−1.1755	2.0750	19.60	54.60
N/R50	42.30	43.90	7.7681	−0.9814	0.2534	18.60	51.90
R/R50	40.88	42.30	6.6831	−0.9005	0.1414	24.20	50.70

Figure 7 | Plots of the response variable. (a) General response variable. (b) Variable response per treatment level.

A comparison of the proposed distributions with some of their sub-models using LR statistics is given in Table 11. The OLL-W model can be considered a more competitive model to the OLLG-W model, because the corresponding *p*-value is greater than the usual 5%. In order to assess if the model is appropriate, plots of the fitted OLLG-W, OLL-W, Kw-W and Weibull density functions are displayed in Figure 8(a). The estimated survival functions and the plots of the empirical distributions are displayed in Figure 8(b). They indicate that the OLLG-W distribution is gives a good fit to the corresponding data set, thus capturing a slight bimodality with left asymmetry.

- **Regression model for life expectancy data**
  The LOLLG-W regression model for the life expectancy data can be expressed as follows

$$y_i = \mu_i + \sigma_i z_i,$$

Table 9 | MLEs of the model parameters for the life expectancy data.

Model	a	b	α	p
OLLG-W	16.6162	46.1443	0.3669	0.4177
	(2.8540)	(0.8573)	(0.0564)	(0.2076)
OLL-W	12.1710	44.2047	0.4414	(0)
	(1.4587)	(0.4768)	(0.0629)	(−)
Geo-W	6.1866	44.1086	(1)	0.0002
	(0.3245)	(0.4805)	(−)	(0.0091)

Model	a	b	λ	φ
Kw-W	45.2322	54.1885	0.1038	1.8377
	(13.7890)	(1.4377)	(0.0551)	(0.4821)
EW	20.1619	50.6639	0.2035	1
	(5.5796)	(0.8541)	(0.0674)	(−)
Weibull	6.1866	44.1076	1	1
	(0.3243)	(0.4791)	(−)	(−)

Model	a	b	k
GW	21.8598	51.7215	0.1875
	(2.0613)	(0.3878)	(0.0232)

Model	a_1	b_1
FW	0.0686	132.8122
	(0.0034)	(7.5067)

MLE, maximum likelihood estimate; Kw-N, Kumaraswamy normal; OLLG-W, odd log-logistic geometric-Weibull; OLL-W, odd log-logistic Weibull; Geo-W, geometric Weibull; Kw-W, Kumaraswamy Weibull; EW, exponentiated Weibull, GW, gamma Weibull, FW, flexible Weibull.

Table 10 | Goodness-of-fit measures for the life expectancy data.

Model	AIC	CAIC	BIC	HQIC	W^*	A^*	KS
OLLG-W	1676.5	1676.6	1690.5	1682.1	0.0388	0.2610	0.0365
OLL-W	1678.8	1678.8	1689.3	1683.0	0.1152	0.6285	0.0619
Geo-W	1702.1	1702.2	1712.6	1706.3	0.2701	1.7463	0.0894
Kw-W	1683.5	1683.6	1697.5	1689.1	0.1352	0.7567	0.0537
EW	1684.8	1684.9	1695.3	1689.1	0.1949	1.0666	0.0578
Weibull	1700.1	1700.2	1708.1	1702.9	25.4435	125.1121	0.9899
GW	1685.0	1685.1	1695.5	1689.3	0.1996	1.0912	0.0583
FW	1700.6	1700.6	1707.6	1703.4	0.2583	1.7219	0.0910

OLLG-W, odd log-logistic geometric-Weibull; OLL-W, odd log-logistic Weibull; Geo-W, geometric Weibull; Kw-W, Kumaraswamy Weibull; EW, exponentiated Weibull, GW, gamma Weibull, FW, flexible Weibull.

Table 11 | LR tests for the life expectancy data.

Models	Hypotheses	Statistic w	p-value
OLLG-W vs Weibull	$H_0 : p = 0$ and $\lambda = 1$ vs $H_1 : H_0$ is false	5.25	0.0720
OLLG-W vs OLL-W	$H_0 : p = 0$ vs $H_1 : H_0$ is false	1.60	0.1968
OLLG-W vs Geo-W	$H_0 : \lambda = 1$ vs $H_1 : H_0$ is false	5.26	0.0217

OLLG-W, odd log-logistic geometric-Weibull; OLL-W, odd log-logistic Weibull; Geo-W, geometric Weibull.

where $z_1, \dots, z_{244}$ are independent random variables with density function (14) and the model parameters are defined by

$$\mu_i = \beta_{10} + \beta_{11}\, d_{i1} + \beta_{12}\, d_{i2} + \beta_{13}\, d_{i3} \quad \text{and} \quad \sigma_i = \exp\left(\beta_{20} + \beta_{22}\, d_{i1} + \beta_{23}\, d_{i2}\right).$$

The MLEs for the LOLLG-W model are presented in Table 12. Thus, when establishing a significance level of 5%, we note that the regression model with varying dispersion capture significant difference between the levels of the diets.

In order to assess if the model is appropriate, the plots comparing the empirical survival function and estimated survival function for the LOLLG-W regression model are displayed in Figures 9(a) and 9(b). These plots indicate that the LOLLG-W regression model with varying dispersion provides a better fit to the data set in Figure 9(b) compared to the homoscedastic regression model Figure in 9(a).
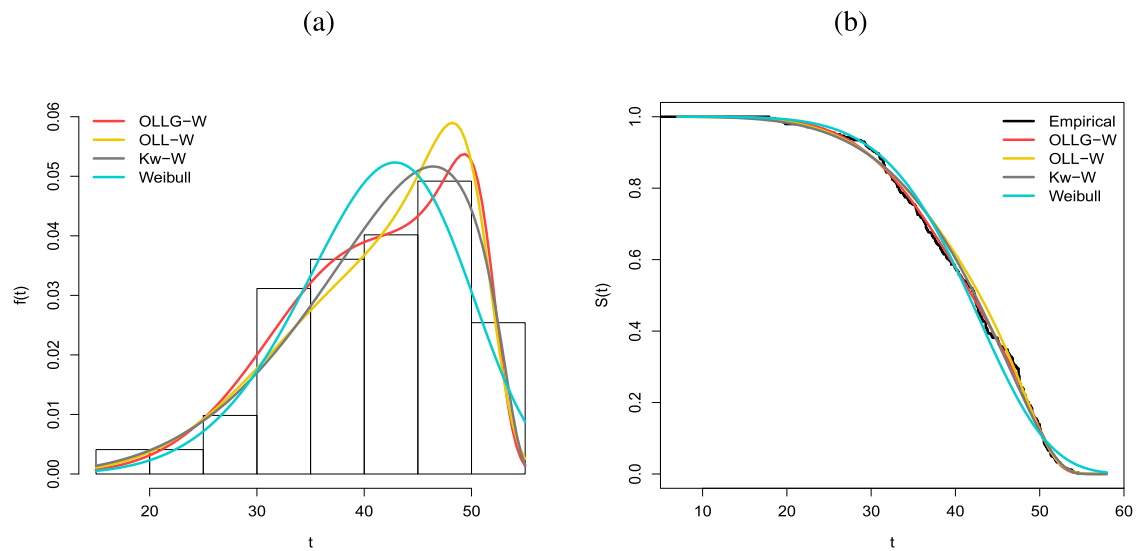
Figure 8 (a) Estimated densities of the odd log-logistic geometric-Weibull (OLLG-W), odd log-logistic Weibull (OLL-W), Kumaraswamy Weibull (Kw-W) and Weibull models for the life expectancy data. (b) Estimated survival functions of the OLLG-W, OLL-W, Kw-W and Weibull models and the empirical cdf for the life expectancy data.

Table 12 MLEs, SEs and *p*-values for the LOLLG-W regression model fitted to the life expectancy data.

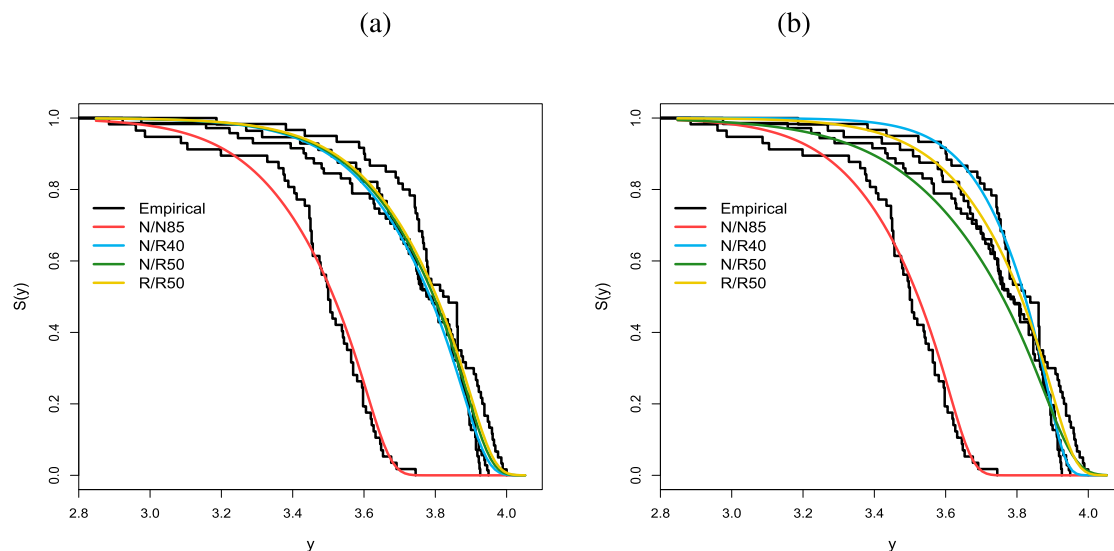| Regression model with constant dispersion				Regression model with varying dispersion			
Parameter	**Estimate**	**SE**	***p*-Value**	**Parameter**	**Estimate**	**SE**	***p*-Value**
σ	0.0738	0.0082	–	α	0.5276	0.0729	–
α	0.5194	0.0722	–	p	0.00006	0.0063	–
p	0.0285	0.0075	–	β_{10}	3.8373	0.0141	< 0.0001
β_{10}	3.8337	0.0139	< 0.0001	β_{11}	−0.2883	0.0175	< 0.0001
β_{11}	−0.2887	0.0181	< 0.0001	β_{12}	0.0046	0.0182	0.8006
β_{12}	−0.0183	0.0168	0.2771	β_{13}	−0.0431	0.0237	0.0702
β_{13}	−0.0078	0.0167	0.6408	β_{20}	−2.6363	0.1201	< 0.0001
				β_{22}	−0.2864	0.1337	0.0331
				β_{23}	0.5276	0.1272	0.0197

Figure 9 Estimated survival considering the log-odd log-logistic geometric Weibull (LOLLG-W) regression model for the life expectancy data. (a) Regression model with constant variance. (b) Regression model with varying dispersion.

8. CONCLUDING REMARKS

The odd log-logistic geometric-*G* (OLLG-*G*) family provides a rather general and flexible framework for statistical analysis of data. It unifies some previously known distributions and yields a general overview of these distributions for theoretical studies. It also represents a rather

flexible mechanism for fitting a wide spectrum of real world data sets. The OLLG-*G* family is motivated by the wide use of the odd LL and geometric distributions in practice, and also for the fact that the generalization provides more flexibility to analyze skewed data. This extension provides a continuous crossover toward cases with different shapes (e.g., a particular combination of skewness and kurtosis). We derive an expansion for the density function as a linear combination of exp-*G* density functions. We obtain explicit expressions for the moments and generating function. Also, based on the odd log-logistic geometric-Weibull (OLLG-W) distribution, we propose a very suitable regression model with varying dispersion for modeling lifetime data, which allows us to jointly model the location and dispersion parameters. The estimation of parameters is approached by the method of maximum likelihood. Applications of the OLLG-*G* family to real data reveal that it could provide a better fit than other statistical models frequently used in lifetime data analysis.

CONFLICT OF INTEREST

There are no conflicts of interest.

AUTHORS' CONTRIBUTIONS

– Maria do Carmo S. Lima and Gauss M. Cordeiro contributed to the definition and properties of the new distribution.

– Fábio Prataviera and Edwin M. M. Ortega contributed to the regression model and applications

Funding Statement

ACKNOWLEDGMENTS

REFERENCES

1. B.E. Hansen, Int. Econ. Rev. 35 (1994), 705–730.
2. N. Eugene, C. Lee, F. Famoye, Commun. Stat. Theory Methods. 31 (2002), 497–512.
3. G.M. Cordeiro, M. de Castro, J. Stat. Comput. Sim. 81 (2011), 883–898.
4. C. Alexander, G.M. Cordeiro, E.M.M. Ortega, J.M. Sarabia, Comput. Stat. Data Anal. 56 (2012), 1880–1897.
5. G.M. Cordeiro, E.M.M. Ortega, D.C. da Cunha, J. Data Sci. 11 (2013), 1–27.
6. K. Aas, I.H. Haff, J. Financ. Economet. 4 (2006), 275–309.
7. G.M. Cordeiro, E.M.M. Ortega, T.G. Ramires, J. Stat. Distrib. Appl. 2 (2015), 131–145.
8. A.E. Gomes, C.Q. Da-Silva, G.M. Cordeiro, Commun. Stat. Theory Methods. 44 (2015), 4217–4240.
9. M.H. Tahir, M. Zubair, G.M. Cordeiro, A. Alzaatreh, M. Mansoor, J. Stat. Comput. Sim. 86 (2016), 2901–2921.
10. G.M. Cordeiro, G.O. Silva, E.M.M. Ortega, J. Stat. Distrib. Appl. 3 (2016), 3.
11. G.M. Cordeiro, M. Alizadeh, T.G. Ramires, E.M.M. Ortega, Commun. Stat. Theory Methods. 46 (2017), 5685–5705.
12. G.M. Cordeiro, M. Alizadeh, G. Ozel, B. Hosseini, E.M.M. Ortega, E. Altun, J. Stat. Comput. Sim. 87 (2017), 908–932.
13. J.U. Gleaton, J.D. Lynch, J. Probab. Stat. Sci. 4 (2006), 51–64.
14. E.M. Hashimoto, E.M.M. Ortega, G.M. Cordeiro, V.G. Cancho, Sankhy B. 77 (2015), 207–239.
15. E.M.M. Ortega, G.M. Cordeiro, A.K. Campelo, M.W. Kattan, V.G. Cancho, Stat. Med. 34 (2015), 1366–1388.
16. B. Lanjoni, E.M.M. Ortega, G. Cordeiro, J. Agric. Biol. Environ. Stat. 21 (2016), 203–224.
17. E.M.M. Ortega, A.J. Lemonte, G.M. Cordeiro, N.J. da Cruz, J. Stat. Theory Pract. 10 (2016), 780–804.
18. T.G. Ramires, E.M.M. Ortega, G.M. Cordeiro, G.A. Paula, N. Hens, Commun. Stat. Simul. Comput. (2017).
19. J.N. da Cruz, E.M.M. Ortega, G.M. Cordeiro, J. Stat. Comput. Sim. 86 (2016), 1516–1538.
20. A.S. Braga, G.M. Cordeiro, E.M.M. Ortega, J.N. da Cruz, J. Stat. Theory Pract. 10 (2016), 311–335.
21. J.U. Gleaton, J.D. Lynch, J. Probab. Stat. Sci. 8 (2010), 1–17.
22. A.W. Marshall, I. Olkin, Biometrika. 84 (1997), 641–652.
23. S. Nadarajah, S. Kotz, Acta Appl. Math. 92 (2006), 97–111.
24. J.F. Lawless, Statistical Models and Methods for Lifetime Data, John Wiley, New York, 2003.
25. M. Bebbington, C.D. Lai, R. Zitikis, Reliab. Eng. Syst. Safe. 92 (2007), 719–726.
26. K. Zografos, N. Balakrishnan, Stat. Methodol. 6 (2009), 344–362.
27. W.Q. Meeker, L.A. Escobar, Statistical Methods for Reliability Data, John Wiley & Sons, New York, 2014.
28. R. Weindruch, R.L. Walford, S. Fligiel, D. Guthrie, J. Nutr. (1986), 641–654. 4