

## Research Article

# The New Odd Log-Logistic Generalized Inverse Gaussian Regression Model

Julio Cezar Souza Vasconcelos,<sup>1</sup> Gauss M. Cordeiro,<sup>2</sup>  
Edwin M. M. Ortega<sup>1</sup>, and Elton G. Araújo<sup>3</sup>

<sup>1</sup>ESALQ, Universidade de São Paulo, Piracicaba, Brazil

<sup>2</sup>DEINFO, Universidade Federal de Pernambuco, Recife, Brazil

<sup>3</sup>UFMS, Universidade Federal de Mato Grosso do Sul, Paranaíba, Brazil

Correspondence should be addressed to Edwin M. M. Ortega; [edwin@usp.br](mailto:edwin@usp.br)

Received 7 August 2018; Accepted 10 December 2018; Published 10 January 2019

Academic Editor: Aera Thavaneswaran

Copyright © 2019 Julio Cezar Souza Vasconcelos et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We define a new four-parameter model called the odd log-logistic generalized inverse Gaussian distribution which extends the generalized inverse Gaussian and inverse Gaussian distributions. We obtain some structural properties of the new distribution. We construct an extended regression model based on this distribution with two systematic structures, which can provide more realistic fits to real data than other special regression models. We adopt the method of maximum likelihood to estimate the model parameters. In addition, various simulations are performed for different parameter settings and sample sizes to check the accuracy of the maximum likelihood estimators. We provide a diagnostics analysis based on case-deletion and quantile residuals. Finally, the potentiality of the new regression model to predict price of urban property is illustrated by means of real data.

## 1. Introduction

The inverse Gaussian (IG) distribution is widely used in several research areas, such as life-time analysis, reliability, meteorology and hydrology, engineering, and medicine. Some extensions of the IG distribution have appeared in the literature. For example, the *generalized inverse Gaussian* (GIG) distribution with positive support is introduced by Good [1] in a study of population frequencies. Several papers have investigated the structural properties of the GIG distribution. Sichel [2] used this distribution to construct mixtures of Poisson distributions. Statistical properties and distributional behavior of the GIG distribution were discussed by Jørgensen [3] and Atkinson [4]. Dagpunar [5] provided algorithms for simulating this distribution. Nguyen et al. [6] showed that it has positive skewness. More recently, Madan et al. [7] proved that the Black-Scholes formula in finance can be expressed in terms of the GIG distribution function. Koudou [8] presented a survey about its characterizations and Lemonte and Cordeiro [9] obtained some mathematical

properties of the *exponentiated generalized inverse Gaussian* (EGIG) distribution.

In this paper, we study a new four-parameter model named the *odd log-logistic generalized inverse Gaussian* (OLLGIG) distribution which contains as special cases the GIG and IG distributions, among others. Its major advantage is the flexibility in accommodating several forms of the density function, for instance, bimodal and unimodal shapes. It is also suitable for testing goodness-of-fit of some submodels.

Our main objective is to study a new regression model with two systematic structures based on the OLLGIG distribution. We obtain some mathematical properties and discuss maximum likelihood estimation of the parameters. For these models, we presented some ways to perform global influence (case-deletion) and, additionally, we developed residual analysis based on the quantile residual. For different parameter settings and sample sizes, various simulation studies were performed and the empirical distribution of quantile residual was displayed and compared with the standard normal distribution. These studies suggest that the empirical distribution

of the quantile residual for the OLLGIG regression model with two regression structures a high agreement with the standard normal distribution.

This paper is organized as follows. In Section 2, we define the OLLGIG distribution. In Section 3, we obtain some of its structural properties. We define the OLLGIG regression model in Section 4 and evaluate the performance of the maximum likelihood estimators (MLEs) of the model parameters by means of a simulation study. In Section 5, we adopt the case-deletion diagnostic measure and define quantile residuals for the fitted model. Further, we perform various simulations for these residuals. In Section 6, we provide two applications to real data to illustrate the flexibility of the OLLGIG regression model. Finally, some concluding remarks are offered in Section 7.

## 2. The OLLGIG Distribution

The GIG distribution [3] has been applied in several areas of statistical research. The cumulative distribution function (cdf) and probability density function (pdf) of the GIG distribution are given by (for  $y > 0$ )

$$G_{\mu, \sigma, \nu}(y) = \int_0^y \left( \frac{b}{\mu} \right)^\nu \frac{t^{\nu-1}}{2K_\nu(\sigma^{-2})} \exp \left[ -\frac{1}{2\sigma^2} \left( \frac{bt}{\mu} + \frac{\mu}{bt} \right) \right] dt \quad (1)$$

and

$$g_{\mu, \sigma, \nu}(y) = Cy^{\nu-1} \exp \left[ -\frac{1}{2\sigma^2} \left( \frac{by}{\mu} + \frac{\mu}{by} \right) \right], \quad (2)$$

where  $\mu > 0$  is the location parameter,  $\sigma > 0$  is the scale parameter,  $\nu \in \mathbb{R}$  is the shape parameter,  $K_\nu(t) = (1/2) \int_0^\infty y^{\nu-1} \exp[-(1/2)t(u + u^{-1})] du$  is the modified Bessel function of the third kind and index  $\nu$ ,  $b = K_{\nu+1}(\sigma^{-2})/K_\nu(\sigma^{-2})$ , and  $C = C(\mu, \sigma, \nu) = (b/\mu)^\nu/2K_\nu(\sigma^{-2})$ .

We denote by  $W \sim \text{GIG}(\mu, \sigma, \nu)$  a random variable having density function (2). The mean and variance of  $W$  are

$$E(W) = \mu, \quad (3)$$

$$V(W) = \mu^2 \left[ \frac{2\sigma^2}{b} (\nu + 1) + \frac{1}{b^2} - 1 \right],$$

respectively.

The moment generating function (mgf) of  $W$  reduces to

$$M(t) = \left( 1 - \frac{2\mu\sigma^2 t}{b} \right)^{-\nu/2} K_\nu(\sigma^{-2})^{-1} \cdot K_\nu \left[ \frac{1}{\sigma^2} \left( 1 - \frac{2\mu\sigma^2 t}{b} \right)^{1/2} \right]. \quad (4)$$

We use the reparameterized GIG distribution according to GAMLSS in software R. For example, we have  $\text{GIG}(\mu, \sigma\mu^{1/2}, -0.5) = \text{IG}(\mu, \sigma)$ . Other properties of the GIG distribution are investigated by Jørgensen [3].

The statistical literature is filled with hundreds of continuous univariate distributions. Recently, several methods of introducing one or more parameters to generate new distributions have been proposed. Based on the *odd log-logistic generator* (OLL-G) [10], we define the OLLGIG cdf, say  $F(y) = F(y; \mu, \sigma, \nu, \tau)$ , by integrating the log-logistic density function as follows:

$$F(y) = \int_0^{G_{\mu, \sigma, \nu}(y)/\bar{G}_{\mu, \sigma, \nu}(y)} \frac{\tau x^{\tau-1}}{(1+x^\tau)^2} dx = \frac{G_{\mu, \sigma, \nu}(y)^\tau}{G_{\mu, \sigma, \nu}(y)^\tau + \bar{G}_{\mu, \sigma, \nu}(y)^\tau}, \quad (5)$$

where  $\bar{G}_{\mu, \sigma, \nu}(y) = 1 - G_{\mu, \sigma, \nu}(y)$ ,  $\mu > 0$  is a position parameter,  $\sigma > 0$  is a scale parameter, and  $\nu \in \mathbb{R}$  and  $\tau > 0$  are shape parameters. Clearly,  $G_{\mu, \sigma, \nu}(y)$  is a special case of (5) when  $\tau = 1$ .

Henceforth, we write  $\eta(y) = G_{\mu, \sigma, \nu}(y)$  to simplify the notation. The OLLGIG density function can be expressed as

$$f(y) = f(y; \mu, \sigma, \nu, \tau) = \left( \frac{b}{\mu} \right)^\nu \frac{\tau y^{\nu-1}}{2K_\nu(\sigma^{-2})} \exp \left[ -\frac{1}{2\sigma^2} \left( \frac{by}{\mu} + \frac{\mu}{by} \right) \right] \times \{\eta(y)[1 - \eta(y)]\}^{\tau-1} \{\eta(y)^\tau + [1 - \eta(y)]^\tau\}^{-2}. \quad (6)$$

The main motivations for the OLLGIG distribution are to make its skewness and kurtosis more flexible (compared to the GIG model) and also allow bi-modality. We have  $\tau = \log[F(y)/\bar{F}(y)]/\log[\eta(y)/\bar{\eta}(y)]$ , where  $\bar{F}(y) = 1 - F(y)$  and  $\bar{\eta}(y) = 1 - \eta(y)$ . Thus, the parameter  $\tau$  represents the quotient of the log odds ratio for the new and baseline distributions. Note that the pdf and cdf of the OLLGIG distribution depend on integrals, which are calculated numerically in the same way as those of the Birnbaum-Saunders distribution.

Hereafter, we assume that the random variable  $Y$  follows the OLLGIG cdf (5) with parameters  $(\mu, \sigma, \nu, \tau)^T$ , say  $Y \sim \text{OLLGIG}(\mu, \sigma, \nu, \tau)$ . The OLLGIG distribution contains as special cases the GIG distribution when  $\tau = 1$  and the IG distribution when  $\tau = 1$  and  $\nu = -0.5$ .

Some plots of the OLLGIG density for selected parameter values are displayed in Figure 1. It is evident that the proposed distribution is much more flexible, especially in relation to bi-modality (for  $0 < \tau < 1$ ), than the GIG and IG distributions.

Equation (5) has tractable properties especially for simulations, since its quantile function (qf) takes the simple form

$$y = Q_{\text{GIG}} \left( \frac{u^{1/\tau}}{u^{1/\tau} + [1 - u]^{1/\tau}} \right), \quad (7)$$

where  $Q_{\text{GIG}}(u) = G_{\mu, \sigma, \nu}^{-1}(u)$  is the qf of the GIG distribution. This scheme is useful because of the existence of fast generators for GIG random variables in some statistical packages. For example, we can fit the generalized additive models for the location, scale, and shape (GAMLSS) in R.

We use the GAMLSS package to simulate data from this nonlinear equation. The plots comparing the exact OLLGIG

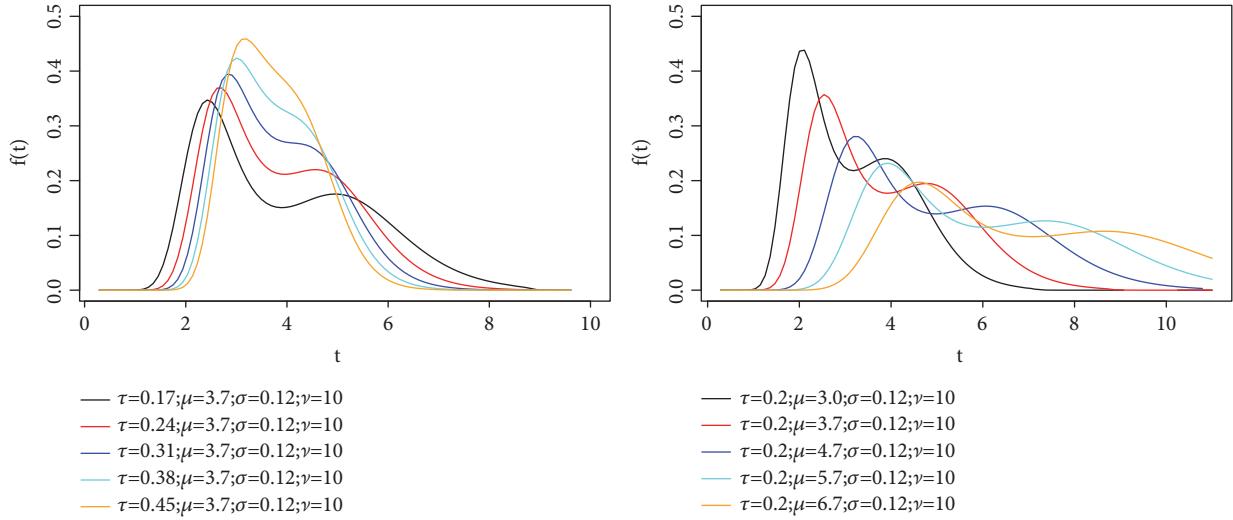


FIGURE 1: Plots of the OLLGIG density for some parameter values.

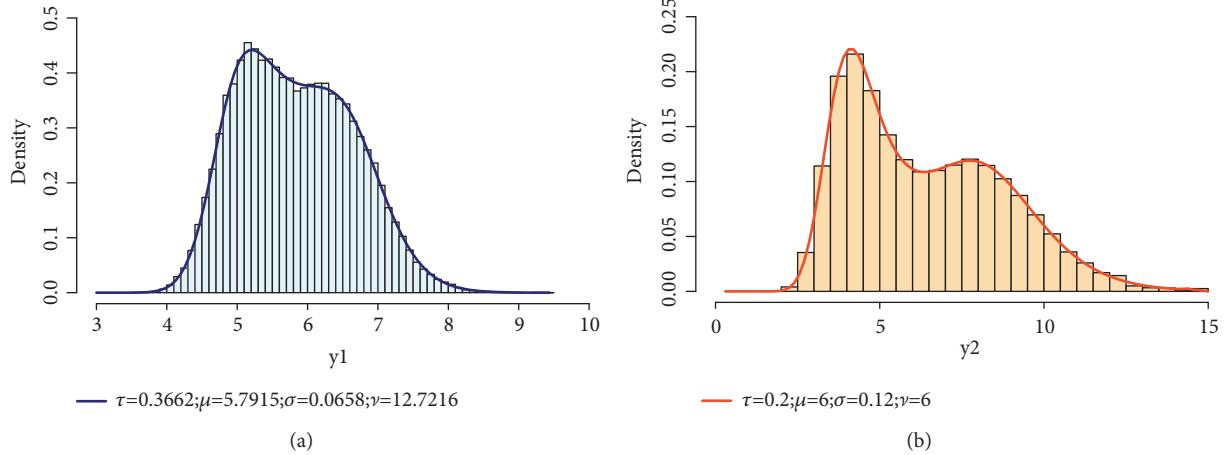


FIGURE 2: Histograms and plots of the OLLGIG density.

densities and the histograms from two simulated data sets with 100,000 replications for selected parameter values are displayed in Figure 2. These plots (and several others not shown here) indicate that the simulated values are consistent with the OLLGIG distribution.

### 3. Properties of the OLLGIG Model

**3.1. Linear Representation.** By defining the sets  $I_i = \{(k, j); k-j = i\}$  for  $i = 0, 1, \dots$ , and following the results of Lemonte and Cordeiro [9, Section 3], we can expand  $\eta(y) = G_{\mu, \sigma, \nu}(y)$  as

$$\eta(y) = 1 - \rho - y^\nu \sum_{i=0}^{\infty} d_i y^i, \quad (8)$$

where  $\rho = \rho(\mu, \sigma, \nu) = C(b/2\mu\sigma^2)^{-\nu} \sum_{j=0}^{\infty} \Gamma(\nu-j)[-(4\sigma^2)^{-1}]^j / j!$ ,  $d_i = \sum_{(k,j) \in I_i} a_{j,k}$ , and

$$a_{j,k} = a_{j,k}(\mu, \sigma, \nu) = \frac{(-1)^{k+j+1} C}{(k-j+\nu) j! k!} \frac{b^{k-j} \mu^{j-k}}{2^{k+j} \sigma^{2(k+j)}}. \quad (9)$$

To calculate  $\rho$ , the index  $j$  can stop after a large number of summands.

Further, we can rewrite  $\eta(y)$  after some algebra as

$$\eta(y) = 1 - \rho - c_0 - \sum_{i=1}^{\infty} c_i y^i, \quad (10)$$

where  $c_i = \sum_{k=0}^i f_k d_{i-k}$  (for  $i = 0, 1, \dots$ ) and  $(\nu)_r = \nu(\nu-1) \dots (\nu-r+1)$  is the descending factorial and  $f_j = \sum_{r=j}^{\infty} (-1)^{r-j} \binom{r}{j} (\nu)_r / r!$ .

We obtain an expansion for  $F(y)$  in (5). First, we use a power series for  $\eta(y)^\tau$  ( $\tau$  real)

$$\eta(y)^\tau = \sum_{k=0}^{\infty} p_k \eta(y)^k, \quad (11)$$

where

$$p_k = p_k(\tau) = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{\alpha}{j} \binom{j}{k}. \quad (12)$$

For any real  $\tau$ , we consider the generalized binomial expansion

$$[1 - \eta(y)]^\tau = \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} \eta(y)^k. \quad (13)$$

Inserting (11) and (13) in (5) gives

$$F(y) = \frac{\sum_{k=0}^{\infty} p_k \eta(y)^k}{\sum_{k=0}^{\infty} q_k \eta(y)^k}, \quad (14)$$

where  $q_k = q_k(\tau) = p_k(\tau) + (-1)^k \binom{\tau}{k}$  (for  $k \geq 0$ ). The ratio of the two power series in the last equation can be reduced to

$$F(y) = \sum_{k=0}^{\infty} w_k \eta(y)^k, \quad (15)$$

where the coefficients  $w_k$ 's (for  $k \geq 0$ ) are determined from the recurrence equation

$$w_k = w_k(\tau) = q_0^{-1} \left( p_k - \sum_{r=1}^k q_r w_{k-r} \right). \quad (16)$$

By differentiating (15), the pdf  $f(y)$  reduces to

$$f(y) = \sum_{k=0}^{\infty} w_{k+1} h_{k+1}(y), \quad (17)$$

where  $h_{k+1}(y) = (k+1)\eta(y)^k g_{\mu, \sigma, \nu}(y)$  is the *exponentiated generalized inverse Gaussian* (EGIG) density function with power parameter  $k+1$  (for  $k \geq 0$ ).

We can derive a linear representation for  $f(y)$  in terms of GIG densities based on the previous results and following the expansions of Lemonte and Cordeiro [9] that lead to their (24). First, we can express  $h_{k+1}(y)$  as

$$h_{k+1}(y) = \sum_{j=0}^k m_j^{(k)} \pi_j(y). \quad (18)$$

Here,  $\pi_j(y)$  represents the  $\text{GIG}(\mu, \sigma, \nu + j)$  density function and the coefficients are given by  $m_j^{(k)} = (k+1)\nu_{j,k} C(\mu, \sigma, \nu) / C(\mu, \sigma, \nu + j)$ , where  $\nu_{j,k} = \sum_{i=0}^k (-1)^i \binom{k}{i} \sum_{r=0}^i \binom{i}{r} \rho^{i-r} t_{j,r}$  and the quantities  $t_{j,r}$  are determined from the recurrence relation  $t_{j,r} = j^{-1} \sum_{m=1}^j [(r+1)m - j] c_m t_{j-m,r}$  (for  $j \geq 1$ ) and  $t_{0,r} = 1$  with  $c_m$ 's given in (10).

By combining (17) and (18) and changing  $\sum_{k=0}^{\infty} \sum_{j=0}^k$  by  $\sum_{j=0}^{\infty} \sum_{k=j}^{\infty}$ , we obtain

$$f(y) = \sum_{j=0}^{\infty} s_j \pi_j(y), \quad (19)$$

where  $s_j = \sum_{k=j}^{\infty} w_{k+1} m_j^{(k)}$ .

Equation (19) reveals that the OLLGIG density function is an infinite linear combination of GIG densities.

**3.2. Two Properties.** Equation (19) becomes useful in deriving several mathematical properties of the proposed distribution using well-known properties of the GIG distribution. We provide only two examples. The  $r$ th moment about zero of the  $\text{GIG}(\mu, \sigma, \nu)$  random variable defined by (2) is

$$E(W^r) = \left( \frac{\mu}{b} \right)^r \frac{K_{\nu+r}(\sigma^{-2})}{K_{\nu}(\sigma^{-2})}. \quad (20)$$

Then, the ordinary moments of the OLLGIG random variable  $Y$  follow from (19) as

$$E(Y^r) = \frac{\mu^r}{K_{\nu}(\sigma^{-2})} \sum_{j=0}^{\infty} s_j \frac{K_{\nu+j+r}(\sigma^{-2})}{b_j^r}, \quad (21)$$

where  $b_j = K_{\nu+j+1}(\sigma^{-2})/K_{\nu}(\sigma^{-2})$ .

By combining (19) and (4), the generating function of  $Y$  takes the form

$$M_Y(t) = \frac{1}{K_{\nu}(\sigma^{-2})} \sum_{j=0}^{\infty} s_j \left( 1 - \frac{2\mu\sigma^2 t}{b_j} \right)^{-\nu/2} \cdot K_{\nu+j} \left[ \frac{1}{\sigma^2} \left( 1 - \frac{2\mu\sigma^2 t}{b_j} \right)^{1/2} \right]. \quad (22)$$

#### 4. The OLLGIG Regression Model

In many practical applications, the lifetimes are affected by explanatory variables such as sex, smoking, diet, blood pressure, cholesterol level and several others. So, it is important to explore the relationship between the response variable and the explanatory variables. Regression models can be proposed in different forms in statistical analysis. In this section, we define the OLLGIG regression model with two systematic structures based on the new distribution. It is a feasible alternative to the GIG and IG regression models for data analysis.

Regression analysis involves specifications of the distribution of  $Y$  given a vector  $\mathbf{x} = (x_1, \dots, x_p)^T$  of covariates. We relate the parameters  $\mu$  and  $\sigma$  to the covariates by the logarithm link functions

$$\begin{aligned} \mu_i &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}_1), \\ \sigma_i &= \exp(\mathbf{x}_i^T \boldsymbol{\beta}_2), \\ i &= 1, \dots, n, \end{aligned} \quad (23)$$

respectively, where  $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{1p})^T$  and  $\boldsymbol{\beta}_2 = (\beta_{21}, \dots, \beta_{2p})^T$  denote the vectors of regression coefficients and  $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ . The most important of the parametric regression models defines the covariates in  $\mathbf{x}$  which model both  $\mu$  and  $\sigma$ .

Consider a sample  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$  of  $n$  independent observations. Conventional likelihood estimation techniques can be applied here. The total log-likelihood function for the

vector of parameters  $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \nu, \tau)^T$  from model (23) is given by

$$\begin{aligned} l(\boldsymbol{\theta}) = & n \log(\tau) + \nu \sum_{i=1}^n \log \left( \frac{b}{\mu_i} \right) + (\nu - 1) \sum_{i=1}^n \log(y_i) \\ & - \sum_{i=1}^n \log \left[ 2K_\nu \left( \frac{1}{\sigma_i^2} \right) \right] \\ & - \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left( \frac{by_i}{\mu_i} + \frac{\mu_i}{by_i} \right) \\ & + (\tau - 1) \sum_{i=1}^n \log \{ \eta(y_i) [1 - \eta(y_i)] \} \\ & - 2 \sum_{i=1}^n \log \{ \eta(y_i)^\tau + [1 - \eta(y_i)]^\tau \}, \end{aligned} \quad (24)$$

where  $K_\nu(\cdot)$  and  $\eta(\cdot)$  are defined in Section 2. The MLE  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  can be calculated by maximizing the log-likelihood (24) numerically in the GAMLSS package of the R software. The advantage of this package is that we can adopt many maximization methods, which will depend only on the current fitted model. Initial values for  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  are taken from the fit of the GIG regression model with  $\tau = 1$ . We do not have problems of maximizing this log-likelihood function. This fact is shown in Section 4.1, where some simulations of the proposed regression model are given under different scenarios.

Under general regularity conditions, the asymptotic distribution of  $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  is multivariate normal  $N_{2p+2}(0, K(\boldsymbol{\theta})^{-1})$ , where  $K(\boldsymbol{\theta})$  is the expected information matrix. The asymptotic covariance matrix  $K(\boldsymbol{\theta})^{-1}$  of  $\hat{\boldsymbol{\theta}}$  can be approximated by the inverse of the  $(2p+2) \times (2p+2)$  observed information matrix  $-\tilde{\mathbf{L}}(\boldsymbol{\theta})$ . The elements of this matrix are calculated numerically. The approximate multivariate normal distribution  $N_{2p+2}(0, -\tilde{\mathbf{L}}(\hat{\boldsymbol{\theta}})^{-1})$  for  $\hat{\boldsymbol{\theta}}$  can be used in the classical way to construct approximate confidence for the parameters in  $\boldsymbol{\theta}$ .

We can use the likelihood ratio (LR) statistic for comparing some special sub-models with the OLLGIG regression model. We consider the partition  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ , where  $\boldsymbol{\theta}_1$  is a subset of parameters of interest and  $\boldsymbol{\theta}_2$  is a subset of remaining parameters. The LR statistic for testing the null hypothesis  $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$  versus the alternative hypothesis  $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^{(0)}$  is given by  $w = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\}$ , where  $\tilde{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\theta}}$  are the estimates under the null and alternative hypotheses, respectively. The statistic  $w$  is asymptotically (as  $n \rightarrow \infty$ ) distributed as  $\chi_k^2$ , where  $k$  is the dimension of the subset of parameters  $\boldsymbol{\theta}_1$  of interest. For example, the test of  $H_0 : \tau = 1$  versus  $H : \tau \neq 1$  is equivalent to compare the OLLGIG regression model with the GIG regression model and the LR statistic reduces to  $w = 2\{l(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\nu}, \hat{\tau}) - l(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \tilde{\nu}, 1)\}$ , where  $\hat{\boldsymbol{\beta}}_1$ ,  $\hat{\boldsymbol{\beta}}_2$ ,  $\hat{\nu}$ , and  $\hat{\tau}$  are the MLEs under  $H$  and  $\tilde{\boldsymbol{\beta}}_1$ ,  $\tilde{\boldsymbol{\beta}}_2$ , and  $\tilde{\nu}$  are the estimates under  $H_0$ .

**4.1. Simulation Study.** In this part of simulation, we approach in two different ways. First, we perform a simulation to study

the behavior of the MLEs of the parameters of the OLLGIG distribution without systematic structures. Second, we evaluate the behavior of the parameter estimates considering two systematic structures.

**The OLLGIG Distribution.** Some properties of the MLEs are evaluated using a classical analysis by means of a simulation study. We simulate the OLLGIG distribution as follows:

- (i) Compute the inverse function  $F^{-1}(\cdot)$  from the cumulative distribution (1).
- (ii) Generate  $u \sim U(0, 1)$ .
- (iii) Apply  $u$  in  $F^{-1}(u) = Q(u)$  from (7).
- (iv) The values  $t = Q(u)$  are generated from the OLLGIG distribution, where  $Q(u)$  is the inverse of (1).

We take  $n = 20, 50, 150$  and  $350$  for each replication and then evaluate the estimates  $\hat{\mu}$ ,  $\hat{\sigma}$ ,  $\hat{\nu}$ , and  $\hat{\tau}$ . We repeat this process 1,000 times and then calculate the average estimates (AEs), biases, and means squared errors (MSEs). In the first scenario, we take  $\tau = 0.3662$ ,  $\mu = 5.7915$ ,  $\sigma = 0.0658$ , and  $\nu = 12.7216$ . We use the values fitted in the adjustment to the iris data set in Section 6. The estimates of the model parameters are computed using the GAMLSS package of the R software. The results of the Monte Carlo study under maximum likelihood are given in Table 1. They indicate that the MLEs are accurate. Further, the MSEs of the MLEs of the model parameters decay toward zero when  $n$  increases in agreement with first-order asymptotic theory.

**The OLLGIG Regression Model.** We examine the performance of the MLEs in the OLLGIG regression model by means of some simulations with sample sizes  $n = 100, 300$  and  $500$ . We simulate 1,000 samples from two scenarios ( $\tau = 0.5$  and  $\tau = 1.5$ ) by considering  $\mu_i = \beta_{10} + \beta_{11}x_i$  and  $\sigma_i = \beta_{20} + \beta_{21}x_i$ . For both cases, we take  $\nu = 0.53$ . The explanatory variable is generated by  $x_i \sim U(0, 1)$  and the response variable is generated by  $y_i \sim \text{OLLGIG}(\mu_i, \sigma_i, \nu, \tau)$ . For each fitted model, we compute the AEs, biases, and MSEs. Based on the results given in Table 2, we note that the MSEs of the MLEs of  $\beta_{10}$ ,  $\beta_{11}$ ,  $\beta_{20}$ ,  $\beta_{21}$ , and  $\tau$  decay toward zero when the sample size  $n$  increases, as usually expected under first-order asymptotic theory. Further, the AEs of the parameters tend to be closer to the true parameter values when  $n$  increases. These facts support that the asymptotic normal distribution provides an adequate approximation to the finite sample distribution of the estimates.

## 5. Checking Model: Diagnostic and Residual Analysis

A first tool to perform sensitivity analysis, as stated before, is by means of global influence starting from case-deletion [11, 12]. Case-deletion is a common approach to study the effect of dropping the  $i$ th observation from the data set. The

TABLE 1: AEs, biases, and MSEs for the parameters of the OLLGIG distribution.

Parameter	scenario 1			scenario 2		
	$n = 20$			$n = 50$		
	AE	Bias	MSE	Parameter	AE	Bias
$\hat{\mu}$	5.9023	0.1109	0.0895	$\hat{\mu}$	5.8618	0.0704
$\hat{\sigma}$	0.7468	0.6810	2.6764	$\hat{\sigma}$	0.2119	0.1461
$\hat{\nu}$	13.4759	0.7544	66.2484	$\hat{\nu}$	12.6530	-0.0685
$\hat{\tau}$	1.0945	0.7283	2.2952	$\hat{\tau}$	0.7115	0.3452
scenario 3				scenario 4		
Parameter	$n = 150$			$n = 350$		
	AE	Bias	MSE	Parameter	AE	Bias
	5.8354	0.0439	0.0039	$\hat{\mu}$	5.8195	0.0281
$\hat{\sigma}$	0.0822	0.0165	0.0004	$\hat{\sigma}$	0.0757	0.0099
$\hat{\nu}$	12.7241	0.0026	0.0097	$\hat{\nu}$	12.7131	-0.0085
$\hat{\tau}$	0.4822	0.1160	0.0242	$\hat{\tau}$	0.4363	0.0700

TABLE 2: AEs, biases, and MSEs for the OLLGIG regression model under scenarios 1 and 2.

Parameter	scenario 1						scenario 2		
	$n = 100$			$n = 300$					
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
$\beta_{10}$	1.5044	0.0044	0.0028	1.5011	0.0011	0.0008	1.5014	0.0014	0.0005
$\beta_{11}$	-0.6979	0.0021	0.0100	-0.6981	0.0019	0.0027	-0.6992	0.0008	0.0017
$\beta_{20}$	-1.9426	0.0574	0.1223	-1.9606	0.0394	0.0401	-1.9637	0.0363	0.0280
$\beta_{21}$	0.3636	0.0136	0.0508	0.3522	0.0022	0.0162	0.3516	0.0016	0.0093
$\tau$	0.5998	0.0998	0.0763	0.5448	0.0448	0.0228	0.5368	0.0368	0.0149
Parameter	scenario 2						$n = 500$		
	$n = 100$			$n = 300$					
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
$\beta_{10}$	1.4975	-0.0025	0.0005	1.4982	-0.0018	0.0002	1.4986	-0.0014	0.0001
$\beta_{11}$	-0.7017	-0.0017	0.0018	-0.7024	-0.0024	0.0005	-0.7025	-0.0025	0.0003
$\beta_{20}$	-2.2627	-0.2627	0.1839	-2.1766	-0.1766	0.0872	-2.1548	-0.1548	0.0659
$\beta_{21}$	0.3502	0.0002	0.0717	0.3432	-0.0068	0.0247	0.3454	-0.0046	0.0148
$\tau$	1.2052	-0.2948	0.3226	1.2720	-0.2280	0.1720	1.2932	-0.2068	0.1389

case-deletion model with systematic structures (23) is given by

$$\begin{aligned} \mu_l &= \exp(\mathbf{x}_l^T \boldsymbol{\beta}_1), \\ \sigma_l &= \exp(\mathbf{x}_l^T \boldsymbol{\beta}_2), \\ l &= 1, 2, \dots, n, \quad l \neq i. \end{aligned} \quad (25)$$

In the following, a quantity with subscript “(i)” means the original quantity with the  $i$ th observation deleted. For model (25), the log-likelihood function of  $\boldsymbol{\theta}$  is denoted by  $l_{(i)}(\boldsymbol{\theta})$ . Let  $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{\boldsymbol{\beta}}_{1(i)}^T, \hat{\boldsymbol{\beta}}_{2(i)}^T, \hat{\nu}_{(i)}, \hat{\tau}_{(i)})^T$  be the MLE of  $\boldsymbol{\theta}$  from  $l_{(i)}(\boldsymbol{\theta})$ . To assess the influence of the  $i$ th observation on the MLEs  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T, \hat{\nu}, \hat{\tau})^T$ , we can compare the difference between  $\hat{\boldsymbol{\theta}}_{(i)}$  and  $\hat{\boldsymbol{\theta}}$ . If deletion of an observation seriously influences the estimates, more attention should be paid to that observation. Hence, if  $\hat{\boldsymbol{\theta}}_{(i)}$  is far from  $\hat{\boldsymbol{\theta}}$ , then the  $i$ th observation can be regarded as influential. A first measure of the global influence

is defined as the standardized norm of  $\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}$  (generalized Cook distance) given by

$$GD_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T [\hat{\mathbf{L}}(\boldsymbol{\theta})] (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}). \quad (26)$$

Another alternative is to assess the values of  $GD_i(\boldsymbol{\beta}_1)$ ,  $GD_i(\boldsymbol{\beta}_2)$ , and  $GD_i(\nu, \tau)$  since these values reveal the impact of the  $i$ th observation on the estimates of  $\boldsymbol{\beta}_1$ ,  $\boldsymbol{\beta}_2$ , and  $(\nu, \tau)$ , respectively. Another popular measure of the difference between  $\hat{\boldsymbol{\theta}}_{(i)}$  and  $\hat{\boldsymbol{\theta}}$  is the likelihood distance given by

$$LD_i(\boldsymbol{\theta}) = 2 \{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{(i)})\}. \quad (27)$$

Once the model is chosen and fitted, the analysis of the residuals is an efficient way to check the model adequacy. The residuals also serve to identify the relevance of an additional factor omitted from the model and verify if there are indications of serious deviance from the distribution considered for the random error. Further, since the residuals are used to identify discrepancies between the fitted model

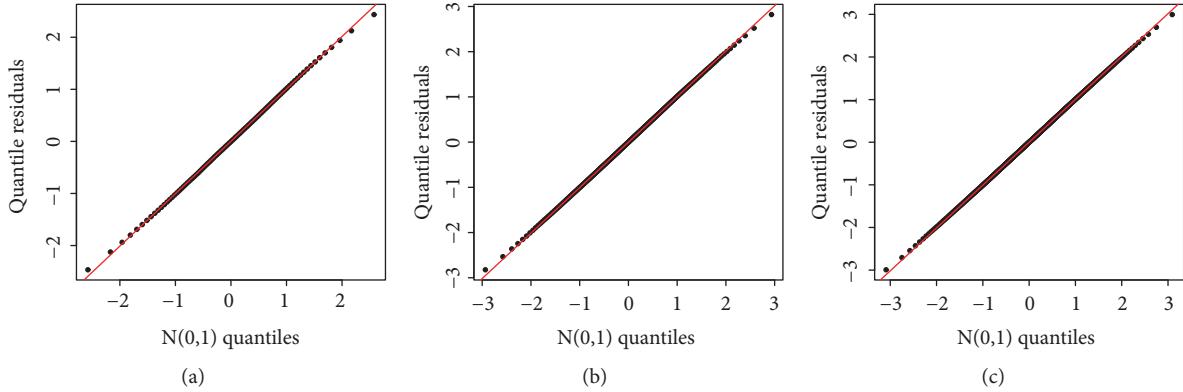


FIGURE 3: Normal probability plots for  $qr_i$  in the OLLGIG regression model under scenario 1 ( $\tau = 0.5$ ) (a)  $n = 100$ . (b)  $n = 300$ . (c)  $n = 500$ .

and the data set, it is convenient to define residuals that take into account the contribution of each observation to the goodness-of-fit measure.

In summary, the residuals allow measuring the model fit for each observation and enable studying whether the differences between the observed and fitted values are due to chance or to a systematic behavior that can be modeled. The quantile residuals (qrs) [13] for the OLLGIG regression model with two systematic structures are defined by

$$qr_i = \Phi^{-1} \left\{ \frac{\eta(y_i)^\tau}{\eta(y_i)^\tau + [1 - \eta(y_i)]^\tau} \right\}, \quad (28)$$

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

Atkinson [14] suggested the construction of an envelope to have a better interpretation of the probability normal plot of the residuals. The simulated confidence bands of the envelope should contain the residuals. If the model is well-fitted, the majority of points will be within these bands and randomly distributed. The construction of the confidence bands follows the steps:

- (i) Fit the proposed model and calculate the residuals  $qr_i$ 's;
- (ii) Simulate  $k$  samples of the response variable using the fitted model;
- (iii) Fit the model to each sample and calculate the residuals  $qr_{ij}$  ( $j = 1, \dots, k$  and  $i = 1, \dots, n$ );
- (iv) Arrange each group of  $n$  residuals in rising order to obtain  $qr_{(i)j}$  for  $j = 1, \dots, k$  and  $i = 1, \dots, n$ ;
- (v) For each  $i$ , calculate the mean, minimum and maximum  $qr_{(i)j}$ , namely,

$$qr_{(i)M} = \sum_{j=1}^k \frac{qr_{(i)j}}{k},$$

$$qr_{(i)I} = \min \{qr_{(i)j} : 1 \leq j \leq k\},$$

$$qr_{(i)S} = \max \{qr_{(i)j} : 1 \leq j \leq k\};$$
(29)

- (vi) Include the means, minimum, and maximum together with the values of  $qr_i$  against the expected percentiles of the standard normal distribution.

The minimum and maximum values of  $qr_i$ 's form the envelope. If the model under study is correct, the observed values should be inside the bands and distributed randomly.

*Simulation Study.* A simulation study is conducted to investigate the behavior of the empirical distribution of the qrs for the OLLGIG regression model. We generate 1,000 samples based on the algorithm presented in Section 4.1. We also give the normal probability plots to assess the degree of deviation from the normality assumption of the residuals. Based on the plots in Figures 3 and 4 representing the first and second scenarios, respectively, we conclude that the empirical distribution of the qrs agrees with the standard normal distribution in both scenarios. This empirical distribution becomes closer to the standard normal distribution when  $n$  increases in both scenarios.

## 6. Applications

In this section, we provide two applications to real data to prove empirically the flexibility of the OLLGIG model. The calculations are performed with the **R** software.

**6.1. Application 1: Iris Data.** In the first application, the OLLGIG distribution is compared with the nested GIG and IG distributions. The data set is *iris*, in which it provides measurements in centimeters of the variables length and width of the sepal and length and width of the petal, respectively, for 50 flowers of each of the 3 *iris* species (*setosa*, *versicolor*, and *virginica*). In this application, the variable *sepal length* (*Sepal.Length*) is used. This data set has been analyzed by several authors in multivariate analysis, for example, Anderson (1935) and Fisher [15]. We show that the distribution for these data presents bimodality.

Table 3 provides a descriptive summary for these data and indicates positively distorted distributions with varying degrees of variability, skewness, and kurtosis.

A brief descriptive analysis of the data in Table 3 reveals that the average score of the variable *sepal length* is 5.843

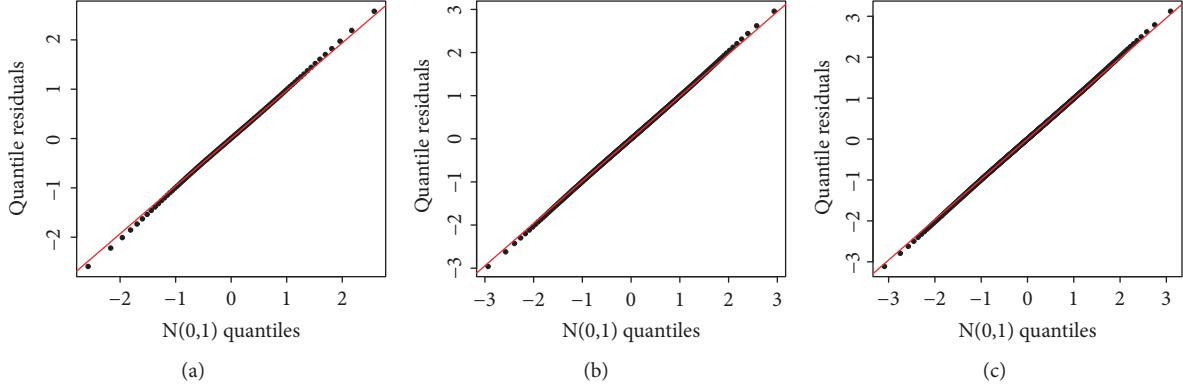
FIGURE 4: Normal probability plots for  $qr_i$  in the OLLGIG regression model for scenario 2 ( $\tau = 1.5$ ) (a)  $n = 100$ . (b)  $n = 300$ . (c)  $n = 500$ .

TABLE 3: Descriptive statistics for iris flower data.

Mean	Median	SD	Skewness	Kurtosis	Min.	Max.
5.843	5.800	0.828	0.3086	-0.6058	4.300	7.900

TABLE 4: MLEs and SEs (in parentheses) of the model parameters for the iris data.

Model	$\tau$	$\mu$	$\sigma$	$\nu$
OLLGIG	0.3662	5.7915	0.0658	12.7216
	(0.0685)	(0.0091)	(0.0079)	(0.0130)
GIG	1	5.8433	0.1413	0.1000
	(-)	(0.0674)	(0.0082)	(72.9562)
IG	1	5.8433	0.0585	-0.5
	(-)	(0.0674)	(0.0034)	(-)

TABLE 5: Goodness-of-fit measures for the iris data.

Model	AIC	CAIC	BIC	HQIC	$A^*$	$W^*$	$KS$
OLLGIG	365.0638	365.3397	377.1064	369.9563	0.3474	0.0486	0.0578
GIG	369.8170	369.9814	378.8489	373.4864	0.7242	0.1164	0.0881
IG	367.8134	367.8951	373.8347	370.2597	0.7244	0.1165	0.0881

TABLE 6: LR tests for the iris data.

Models	Hypotheses	Statistic $w$	$p$ -value
OLLGIG vs GIG	$H_0 : \tau = 1$ vs $H_1 : \tau$ is false	6.7532	0.0094
OLLGIG vs IG	$H_0 : \tau = 1$ and $\nu = -0.5$ vs $H_1 : \nu$ is false	6.7496	0.0342

and the median value is 5.800, thus indicating that the data has a symmetric distribution.

In Table 4, we report the MLEs of the model parameters and their standard errors (SEs) in parentheses. We give in Table 5 the following goodness-of-fit measures: Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), Hernnan-Quinn Information Criterion (HQIC), Cramér-von Misses ( $W^*$ ), Anderson Darling ( $A^*$ ), and Kolmogorov-Smirnov ( $KS$ ) test statistic. The small values of these measures, the better the fit. The figures in Table 5 indicate that the OLLGIG distribution has the lowest values of AIC, CAIC, BIC, HQIC,  $A^*$ ,  $W^*$ , and  $KS$  among those of the

fitted models and therefore it could be chosen as the best model.

We consider LR statistics to compare nested models. The OLLGIG distribution includes some submodels as mentioned above, thus allowing their evaluations relative to the others and to a more general model. The values of the LR statistics are listed in Table 6. It is evident from the figures in this table that the OLLGIG distribution outperforms its submodels according to the values of the LR statistics. So, it indicates that the OLLGIG model provides a better fit to these data than their sub-models.

More information is provided by a visual comparison of the histogram of the data and the fitted density functions and

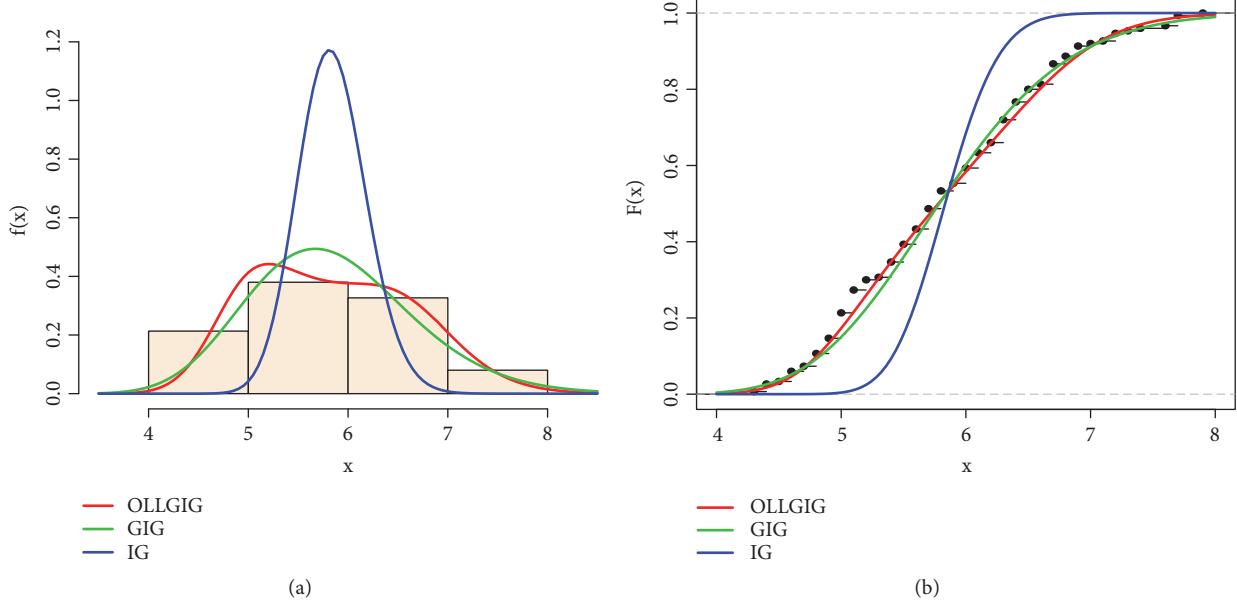


FIGURE 5: (a) Estimated densities of the OLLGIG, GIG, and IG models for iris data. (b) Estimated cumulative functions of the OLLGIG, GIG, and IG models and the empirical cdf for iris data.

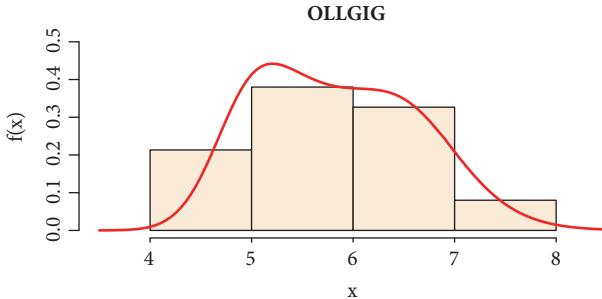


FIGURE 6: Estimated densities of the OLLGIG for iris data.

cumulative functions. The plots of the fitted OLLGIG, GIG, and IG densities are displayed in Figure 5(a). The estimated OLLGIG density provides the closest fit to the histogram of the data. In order to assess if the model is appropriate, the plots of the fitted OLLGIG, GIG, and IG cumulative distributions and the empirical cdf are displayed in Figure 5(b). They indicate that the OLLGIG distribution provides a good fit to these data.

In Figure 6, we note that the iris data has a bimodality shape, where they cannot have the GIG and IG distributions (see Figure 5(a)).

**6.2. Application 2: Price of Urban Property Data.** Here, we provide a second application of the OLLGIG regression model to evaluation the price of urban residential properties for sale in the municipality of Paranaíba in the State of Mato Grosso do Sul (MS) in Brazil. These data collected in 2017 refer to  $n = 45$  houses for sale in the municipality. In the context of real estate appraisal, it is necessary to develop statistical methodologies (characterized by the scientific accuracy) of residential property prices. Besides this aspect,

we can perceive the rare use of such methodologies by the real estate market. We construct a OLLGIG regression model with two systematic components to describe the relationship between real estate prices and other explanatory variables, thus allowing an understanding of the behavior of the price variable [16, 17]. The following explanatory variables are considered:

- (i) price of the property  $y_i$ ; this variable was divided by 10,000;
- (ii) area  $x_{i1}$  of land in square meters;
- (iii) number of parking spaces  $x_{i2}$  in the residence (0=no vacancy, 1=one vacancy, and 2=more than one vacancy); in this case, two dummy variables,  $x_{i21}$  and  $x_{i22}$ , are created;
- (iv) number of rooms with suites  $x_{i3}$  in the residence (0=no suites, 1=one suites, 2=more than one suites); in this case two dummy variables,  $x_{i31}$  and  $x_{i32}$ , are created;
- (v) if the residence has a swimming pool  $x_{i4}$  (0=no, 1=yes);
- (vi) if the residence is located in the center of the city  $x_{i5}$  (0=no, 1=yes);  $i = 1, \dots, 45$ .

In the descriptive analysis of the data from Table 7, the mean score of the variable value is 24.98, which is not close to the median value 17.00, thus indicating that the data has an asymmetric distribution.

We define the OLLGIG regression model by two systematic structures for  $\mu$  and  $\sigma$

$$\begin{aligned} \mu_i = & \exp(\beta_{10} + \beta_{11}x_{i1} + \beta_{121}x_{i21} + \beta_{122}x_{i22} + \beta_{131}x_{i31} \\ & + \beta_{132}x_{i32} + \beta_{14}x_{i4} + \beta_{15}x_{i5}) \end{aligned} \quad (30)$$

TABLE 7: Descriptive analysis of the price of urban property data\$.

Mean	Median	SD	Skewness	Kurtosis	Min.	Max.
24.98	17.00	23.9180	3.3330	14.0134	5.50	150.00

TABLE 8: MLEs, standard errors (SEs), and p-values for the OLLGIG regression model fitted for the price of urban property data.

Parameter	Estimate	SE	p-Value
$\hat{\beta}_{10}$	7.0690	0.4428	<0.001
$\hat{\beta}_{11}$	-0.0005	0.0002	0.0679
$\hat{\beta}_{121}$	0.8069	0.2689	0.0057
$\hat{\beta}_{122}$	0.8407	0.2677	0.0041
$\hat{\beta}_{131}$	-0.8976	0.1945	<0.001
$\hat{\beta}_{132}$	0.4326	0.1872	0.0287
$\hat{\beta}_{14}$	0.5794	0.6941	0.4111
$\hat{\beta}_{15}$	-0.5323	0.1008	<0.001
$\hat{\beta}_{20}$	2.614	0.5982	<0.001
$\hat{\beta}_{21}$	0.0013	9.961e-05	<0.001
$\hat{\beta}_{221}$	-0.2054	0.1316	0.1303
$\hat{\beta}_{222}$	-0.1741	0.1223	0.1660
$\hat{\beta}_{231}$	0.5139	0.0739	<0.001
$\hat{\beta}_{232}$	0.2585	0.1077	0.0235
$\hat{\beta}_{24}$	-2.135	0.4818	<0.001
$\hat{\beta}_{25}$	0.2575	0.0481	<0.001
$\hat{\nu}$	-0.4942	0.1231	<0.001
$\hat{\tau}$	12.764	2.436	

TABLE 9: Goodness-of-fit measures for the price of urban property data.

Model	AIC	BIC	GD
OLLGIG	322.0612	354.5811	286.0612
GIG	348.8190	379.5323	314.8190
IG	333.3241	362.2307	301.3241

and

$$\sigma_i = \exp(\beta_{20} + \beta_{21}x_{i1} + \beta_{221}x_{i21} + \beta_{222}x_{i22} + \beta_{231}x_{i31} + \beta_{232}x_{i32} + \beta_{24}x_{i4} + \beta_{25}x_{i5}), \quad i = 1, \dots, 45. \quad (31)$$

We now consider the test of homogeneity of the scale parameter for the price of urban property data. The LR statistic (see Section 4) for testing the null hypothesis  $H_0 : \beta_{21} = \beta_{221} = \beta_{222} = \beta_{231} = \beta_{232} = \beta_{24} = \beta_{25} = 0$  is  $w = 31.98$  ( $p$  value  $< 0.0001$ ), which gives a favorable indication toward the dispersion not be constant.

In Table 8, we present the MLEs, SEs, and p-values. The covariates  $x_2$ ,  $x_3$ , and  $x_5$  are significant at the 5% level in the regression structure for the location parameter  $\mu$ , whereas the covariates  $x_1$ ,  $x_3$ ,  $x_4$ , and  $x_5$  are significant (at the same level) for the parameter  $\sigma$ . The figures in this table reveal that the covariate  $x_1$  is not significant with respect to the parameter  $\mu$ , but it is significant with respect to the parameter  $\sigma$ . This is due to a strong dispersion in the response variable. The covariate  $x_2$  is also significant for the number of parking spaces in the structure of  $\mu$ . The covariate  $x_3$  is significant in the location and scale structure; i.e., there is a significant

difference between the residences that do not have a suite, have a suite, or more. The covariate  $x_4$  is not significant in relation to the location, but it is significant in the structure of  $\sigma$ . There is a significant difference in the residence with or without swimming pool for the dispersion parameter. This fact can also be noted in Figure 7(a). The covariate  $x_5$  is significant in relation to both parameters  $\mu$  and  $\sigma$ ; i.e., there is a significant difference between the residences being in the center of the city and outside the center. This fact can also be noted in Figure 7(b).

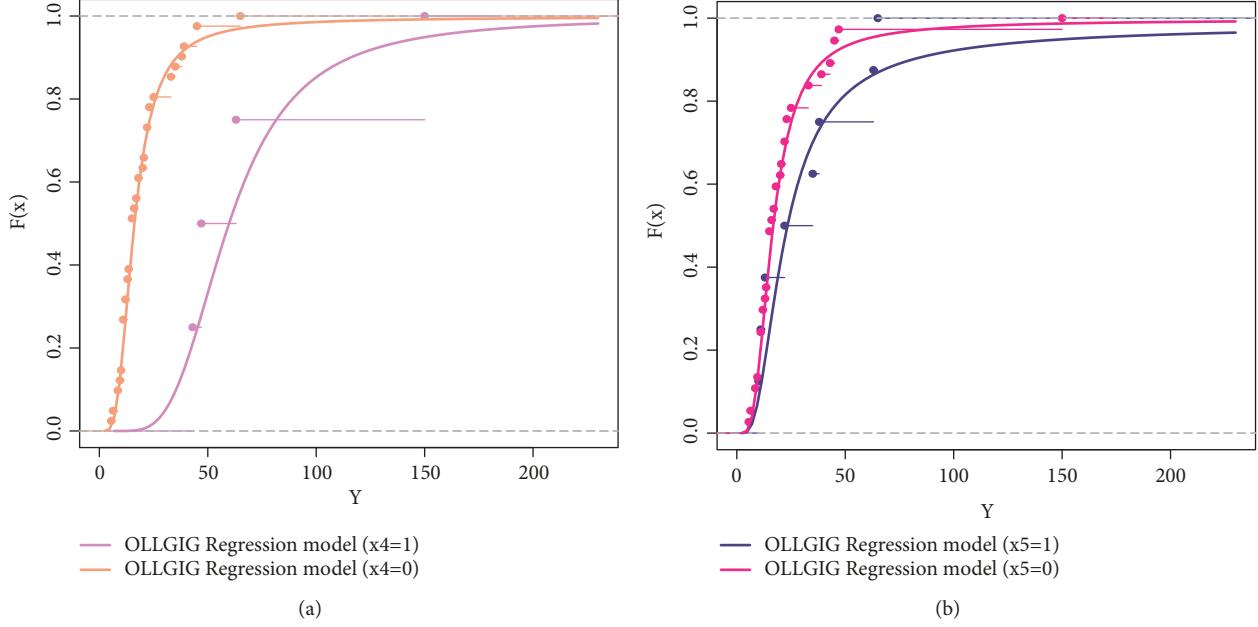
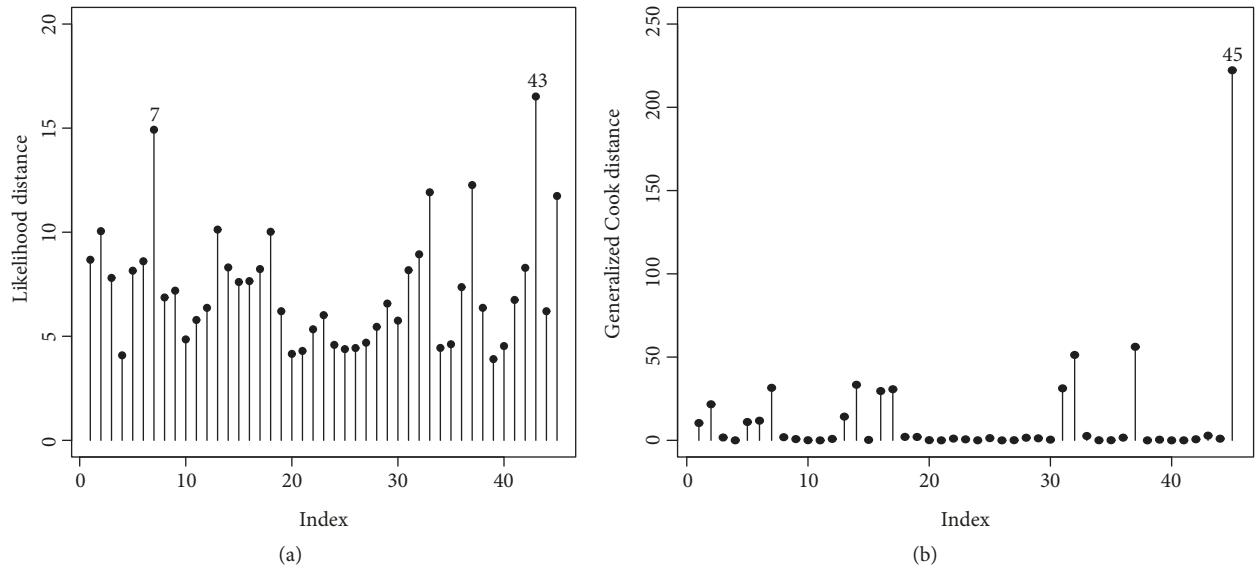
The AIC, BIC, and global deviance (GD) statistics are listed in Table 9. We note that the OLLGIG regression model presents the lowest AIC, BIC, and GD values among the other fitted models. So, there are indications that the OLLGIG model provides a better fit to these data.

We adopt again the LR statistics to compare the fitted models in Table 10. We reject the null hypotheses in the two tests in favor of the wider OLLGIG regression model. Rejection is significant at the 5% level and provides clear evidence of the need of the shape parameter  $\tau$  when modeling real data.

We use the R software to compute the  $LD_i(\theta)$  and  $GD_i(\theta)$  measures in the diagnostic analysis presented in

TABLE 10: LR tests for the price of urban property data.

Models	Hypotheses	Statistic $w$	$p$ -value
OLLGIG vs GIG	$H_0 : \tau = 1$ vs $H_1 : H_0$ is false	28.7579	<0.001
OLLGIG vs IG	$H_0 : \tau = 1$ and $\nu = -0.5$ vs $H_1 : H_0$ is false	15.2629	<0.001

FIGURE 7: Estimated cdf from the fitted OLLGIG regression model and the empirical cdf for the price of urban property data. (a) For covariate  $x_4$ , and (b) for covariate  $x_5$ .FIGURE 8: Index plot for  $\theta$ : (a)  $LD_i(\theta)$  (likelihood distance) and (b)  $GD_i(\theta)$  (generalized Cook's distance).

Section 5. The results of such influence measures index plots are displayed in Figure 8. These plots indicate that the cases #7, #43, and #45 are possible influential observations.

In addition, Figure 9(a) provides plots of the qrs for the fitted model, thus showing that all observations are in the interval  $(-3, 3)$  and a random behavior of the residuals.

Hence, there is no evidence against the current suppositions of the fitted model. In order to detect possible departures from the distribution errors in model, as well as outliers, we present the normal plot for the qrs with a generated envelope in Figure 9(b). This plot reveals that the OLLGIG regression model is very suitable for these data, since there

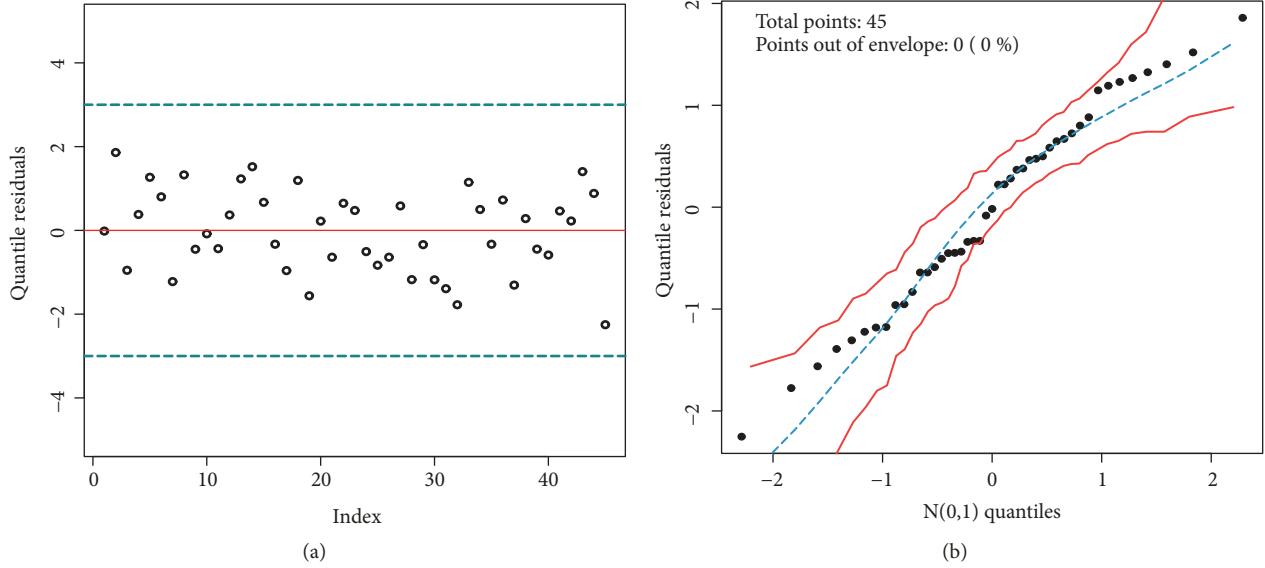


FIGURE 9: (a) Index plot of the qrs and (b) normal probability plot with envelope for the qrs from the fitted OLLGIG regression model fitted to urban property data.

are no observations falling outside the envelope. Also, no observation appears as a possible outlier.

## 7. Concluding Remarks

We present a four-parameter distribution called the *odd log-logistic generalized Gaussian inverse* (OLLGIG) distribution, which includes as special cases the generalized Gaussian inverse (GIG) and inverse Gaussian (IG). We provide some of its mathematical properties. Further, we define the OLLGIG regression model with two systematic structures based on this new distribution, which is very suitable for modeling censored and uncensored data. The proposed model serves as an important extension to several existing regression models and could be a valuable addition to the literature. Some simulations are performed for different parameter settings and sample sizes. The maximum likelihood method is described for estimating the model parameters. Diagnostic analysis is presented to assess global influences. We also discuss the sensitivity of the maximum likelihood estimates from the fitted model via quantile residuals. The utility of the proposed OLLGIG regression model is demonstrated by means of a real data set for price data of urban residential properties in the municipality of Paranaíba in the State of Mato Grosso do Sul, Brazil.

## Data Availability

The [DATA TYPE] data used to support the findings of this study were supplied by Universidade Federal de Mato Grosso do Sul under license and so cannot be made freely available.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.

## References

- [1] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, vol. 40, pp. 237–264, 1953.
- [2] H. S. Sichel, "On a Distribution Law for Word Frequencies," *Journal of the American Statistical Association*, vol. 70, pp. 542–547, 1975.
- [3] B. Jørgensen, *Statistical Properties of the Generalized Inverse Gaussian distribution*, Springer, New York, NY, USA, 1982.
- [4] A. C. Atkinson, "The simulation of generalized inverse Gaussian and hyperbolic random variables," *SIAM Journal on Scientific Computing*, vol. 3, no. 4, pp. 502–515, 1982.
- [5] J. S. Dagpumar, "An easily implemented generalised inverse Gaussian generator," *Communications in Statistics—Simulation and Computation*, vol. 18, no. 2, pp. 703–710, 1989.
- [6] T. T. Nguyen, J. T. Chen, A. K. Gupta, and K. T. Dinh, "A proof of the conjecture on positive skewness of generalised inverse Gaussian distributions," *Biometrika*, vol. 90, no. 1, pp. 245–250, 2003.
- [7] D. Madan, B. Roynette, and M. Yor, "Unifying black-scholes type formulae which involve brownian last passage times up to a finite horizon," *Asia-Pacific Financial Markets*, vol. 15, no. 2, pp. 97–115, 2008.
- [8] A. E. Koudou and C. Ley, "Efficiency combined with simplicity: new testing procedures for generalized inverse Gaussian models," *TEST*, vol. 23, no. 4, pp. 708–724, 2014.
- [9] A. J. Lemonte and G. M. Cordeiro, "The exponentiated generalized inverse Gaussian distribution," *Statistics & Probability Letters*, vol. 81, no. 4, pp. 506–517, 2011.

- [10] J. U. Gleaton and J. D. Lynch, "Properties of generalized logistic families of lifetime distributions," *Journal of Probability and Statistical Science*, vol. 4, no. 1, pp. 51–64, 2006.
- [11] R. D. Cook, "Detection of influential observation in linear regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.
- [12] R. D. Cook and S. Weisberg, *Residuals and influence in regression*, Monographs on Statistics and Applied Probability, Chapman & Hall, NY, USA, 1982.
- [13] P. K. Dunn and G. K. Smyth, "Randomized Quantile Residuals," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 236–244, 1996.
- [14] A. C. Atkinson, *Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis*, Clarendon Press Oxford, Oxford, UK, 1985.
- [15] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.
- [16] J. W. M. Bertrand and J. C. Fransoo, "Modelling and simulation: Operations management research methodologies using quantitative modeling," *International Journal of Operations and Production Management*, vol. 22, pp. 241–264, 2002.
- [17] E. G. Araújo, J. C. Pereira, F. Ximenes, C. P. Spanhol, S. Garson, and E. G. Araújo, "Proposta de uma metodologia para a avaliação do preço de venda de imóveis residenciais em Bonito/MS baseado em modelos de regressão linear múltipla," *Pesquisa e Desenvolvimento Engenharia de Produção*, vol. 10, pp. 195–207, 2012.

