ORIGINAL RESEARCH PAPER

A regression model for extreme events and the presence of bimodality with application to energy generation data

Julio Cezar Souza Vasconcelos | Gauss Moutinho Cordeiro
Edwin Moises Marcos Ortega | João Gabriel Ribeiro

1 Departamento de Ciências Exatas, Universidade de São Paulo, ESALQ, Piracicaba, Brazil
2 Departamento de Estatística, Universidade Federal de Pernambuco, UFPE, Recife, Brazil
3 Departamento de Matemática, Universidade do Estado de Mato Grosso, UNEMAT, Sinop, Brazil

Correspondence
Edwin Moises Marcos Ortega, Departamento de Ciências Exatas, Universidade de São Paulo, ESALQ, Piracicaba, Brazil. Email: edwin@usp.br

Funding information
CNPq, CAPES, Santo Antônio Hydroelectric Plant

Abstract
The application of the theory of extreme values has been growing due to increasing interest in extreme natural events. Many articles on extreme values in data modelling consider unimodal data. This work introduces an appropriate regression for extreme values to detect the presence of bimodality by means of systematic components of two parameters of the odd log-logistic log-normal distribution. The global influence is addressed to verify the model robustness and to find possible influential points. Quantile residuals are proposed to detect distribution deficiencies and outliers in the new regression. A real dataset from the electricity generation area is analysed, namely the Santo Antônio Hydroelectric Plant in the state of Rondônia (Brazil), to illustrate the potential of the new regression. The main results indicate that the proposed regression can identify changes in the means and variability of the power generation between extreme events, that is, between the months of June and December.

1 INTRODUCTION

The theory of extreme values has been widely used to model rare events, that is, events that have a low probability of occurrence. Generally the interest of researchers is to construct models that will be used to help support decisions on response or preventive measures of rare events. Some distributions exist to model rare events. The best known is the maximum extreme value (or Gumbel) distribution. It is widely used to model climatic events, such as heavy rainfall, extreme temperatures, water levels of rivers, ocean storm surges, wind speed and wave height, among others. Studies have also been conducted associated with pollution levels, materials engineering, traffic control, economics, insurance premiums and uncommon events in the financial market. Various generalisations of the extreme value distribution can be found in the literature. Kotz and Nadarajah (2000) [1] discussed various applications of the generalised extreme value (GEV) distribution; Cordeiro et al. (2012) [2] proposed the Kumaraswamy–Gumbel distribution; Cordeiro et al. (2013) [3] defined the exponentiated generalised Gumbel (GGu) distribution; and Pescim et al. (2012) [4] studied the Kummer beta GGu distribution. However, none of these distributions are able to model bimodal data, and not all of these authors studied the effect of covariables on the response variable, that is, constructed regression models.

Conscious consumption of electricity is essential for sustainable development. For example, Halvorsen and Larsen (2001) [5] analysed data in order to find factors that influence residential electricity consumption in Norway, Filippini and Pachauri (2004) [6] estimated the price and income elasticities of residential demand for electricity of all urban areas in India and Arisoy and Ozturk (2014) [7] estimated the price and income elasticities of demand for industrial and residential electricity in Turkey for the period 1960 and 2008.

This paper analyses the average daily potential to generate electricity (measured in megawatt-hours) by the Santo Antônio Hydroelectric Plant in Rondônia (Brazil). For this purpose, it proposes a new regression for bimodal data. Figure 1a shows that the histogram of the energy data has a bimodal shape. The tests prove that this regression is appropriate for these data. Figure 1b shows the presence of extreme events with similar behaviour in June and December, and also similarity between...
July and November. It rains very little in June and July in relation to the previous months and there is evidence of a large accumulation of water in the reservoir between January and May. Thus, in the months of drought (June and July), the plant continues to generate energy similar to November and December, since the rains occur regularly in November with a considerable increase in December. The proposed regression is also applied to analyse extreme events.

In recent years, regression analysis has been investigated carefully in extreme value theory. The location-scale regression (Lawless, 2003) [8] is frequently used for weather events. This paper proposes a regression model under the odd log-logistic log-normal (OLLLN) distribution (Ozel et al., 2018) [9]. In fact, this distribution is very flexible to model extreme events and bimodal data. The parameters are estimated via maximum likelihood. Unless otherwise stated, the results in the paper are new and original. They can encourage further research on the new regression. Monte Carlo simulations are conducted to study the accuracy of the OLLLN regression in terms of variances and mean squared error (MSE) measures.

Some assumptions are verified to detect influential observations in the regression. Quantile residuals (qrs) are defined to check the underlying distribution and to detect outliers in the regression.

This rest of the article is structured as follows. Section 2 reviews some characteristics of the OLLLN distribution. Section 3 defines the OLLLN regression for extreme events with two systematic components and discuss the maximum likelihood estimators (MLEs) of the parameters and qrs. Section 4 generates data based on selected parameter values for three simulation studies to evaluate the accuracy of the estimators in the OLLLN distribution and its associated regression and the adequacy of the normal approximation for the qrs. Section 5 contains an empirical demonstration of the utility of the new regression. Finally, Section 6 contains the concluding remarks.

2 | THE OLLLN DISTRIBUTION

The log-normal (LN) distribution has many applications in different areas. It is characterised by a strong positive asymmetry and the occurrence of a large number of low values and a small number of high to very high values. Its cumulative distribution function (cdf) and probability density function (pdf) are:

\[ G_{\mu, \sigma}(y) = \Phi \left( \frac{\log(y) - \mu}{\sigma} \right), \quad y > 0, \tag{1} \]

and

\[ g_{\mu, \sigma}(y) = \frac{1}{\sqrt{2\pi\sigma^2y}} \exp \left\{ -\frac{1}{2} \left[ \frac{\log(y) - \mu}{\sigma} \right]^2 \right\}, \tag{2} \]

respectively. Here, the location parameter is \( \mu \in \mathbb{R} \), the scale parameter is \( \sigma > 0 \) and \( \Phi(\cdot) \) is the standard normal cdf.

Let \( W \sim LN(\mu, \sigma) \) be a random variable having density function (2). The mean and variance of \( W \) are

\[ E(W) = e^{\mu+\sigma^2/2} \quad \text{and} \quad V(W) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1), \]

respectively.

In recent years, various methods for generalising or modifying this distribution have been investigated. By integrating the log-logistic density function, Ozel et al. (2018) [9] defined the OLLLN cdf (for \( y > 0 \)):

\[ F(y) = F(y; \mu, \sigma, \nu) = \int_0^{G_{\mu, \sigma}(y)} \frac{\nu x^{\nu-1}}{(1 + x^\nu)^2} dx \]

\[ = \frac{G_{\mu, \sigma}(y)^\nu}{G_{\mu, \sigma}(y)^\nu + G_{\mu, \sigma}(y)^\nu}, \tag{3} \]

where \( 0 < \nu < 1 \).
where \( \hat{c}_{\mu,\sigma}(y) = 1 - \hat{G}_{\mu,\sigma}(y) \), and \( \nu > 0 \) is a shape parameter. By setting \( \nu = 1 \) in Equation (3), one obtains the cdf of the LN distribution in Equation (1).

Henceforth, \( \eta(y) = \hat{G}_{\mu,\sigma}(y) \) and the OLLLN density is written as (for \( y > 0 \)):

\[
f(y) = f(y; \mu, \sigma, \nu) = \frac{\nu \exp \left\{ -\frac{1}{2} \left( \frac{\log(y) - \mu}{\sigma} \right)^2 \right\} \left\{ \Phi \left( \frac{\log(y) - \mu}{\sigma} \right) \left[ 1 - \Phi \left( \frac{\log(y) - \mu}{\sigma} \right) \right] \right\}^{\nu - 1}}{\sqrt{2\pi\sigma^2} \left\{ \Phi \left( \frac{\log(y) - \mu}{\sigma} \right) \right\}^\nu + \left[ 1 - \Phi \left( \frac{\log(y) - \mu}{\sigma} \right) \right]^{\nu/2} \left\{ \Phi \left( \frac{\log(y) - \mu}{\sigma} \right) \right\}^{\nu/2} \}. \tag{4}
\]

Henceforth, the random variable \( Y \) has the OLLLN density (4). Clearly, the LN distribution follows when \( \nu = 1 \). The asymmetry and kurtosis of \( Y \) are more flexible than those of LN quantities. Another motivation for the OLLLN distribution is to have bimodality. Equation \( \nu = 0 \) produces parameter independent observables. Figure 2a,b show the behaviour of the pdf of \( Y \) when varying \( \sigma \) and \( \nu \), respectively. Figure 2a reveals that the parameter \( \sigma \) is a kind of dispersion parameter. In Figure 2c, the behaviour of the scale parameter \( \mu \) is clear. This distribution is more flexible than the LN distribution especially in terms of bimodality when \( 0 < \nu < 0.5 \). The bimodality property was not investigated by Ozcel et al. (2018) [9]. However, the OLLLN distribution can model bimodal and asymmetric data.

The quantile function (qf) takes the simple form:

\[
y = Q_{\text{LN}} \left( \frac{n^{1/\nu}}{\mu^{1/\nu} + (1 - n)^{1/\nu}} \right), \tag{5}
\]

where \( Q_{\text{LN}}(u) = \hat{G}_{\mu,\sigma}^{-1}(u) = \exp[\mu + \sigma \Phi^{-1}(u)] \) is the LN qf and \( Q_{\nu}(u) = \Phi^{-1}(u) \) is the normal qf.

Generated values for \( Y \) can be simulated as follows:

- Generate \( \nu \sim U(0, 1) \).
- Set \( z = n^{1/\nu}/[\mu^{1/\nu} + (1 - n)^{1/\nu}] \) and obtain \( y = Q_{\text{LN}}(z) \) from Equation (5).

Some shapes of the density function (4) for fixed parameters are given in Figure 3. The main characteristic of the OLLLN distribution is that it can model bimodal data. These plots confirm that the generated values agree with the OLLLN distribution.

### 3 | THE OLLLN REGRESSION

In several real problems from different fields, it is important to verify the relationship of two or more variables. This type of modelling is called regression analysis, and determines how the explanatory variables explain the variability of the response variable. The data collection allows understanding the dependence between variables and carrying out studies for formulation of the regression, estimation and inference of the parameters, diagnostic and residual analysis.

Some studies of regression models in extreme events can be found in the literature. For example, Barreto-Souza and Vasconcellos (2011) [10] introduced an extended extreme value regression, Pinheiro and Ferrari (2016) [11] compared some extensions of the Gumbel extreme value model related to wind speed data. Prataviera et al. (2018) [12] defined the generalised odd log-logistic flexible Weibull regression model with applications in repairable systems and Hashimoto et al. (2019) [13] introduced the zero-spiked regression models with application in the resin oil production. More recently, Hashimoto et al. (2019) [14] proposed the Burr XII gamma-Weibull regression model with random effects and censored data, Prataviera et al. (2019) [15] formulated the exponentiated power exponential regression and Souza Vasconcellos et al. (2019) [16] studied the odd-log-logistic regression based on a generalised inverse Gaussian distribution.

Following these ideas, here a regression from the OLLLN distribution for bimodal data is defined using likelihood techniques. A new regression is introduced for extreme events by modelling the scale and dispersion parameters. The OLLLN regression is defined by the distribution (4) of the response variable \( Y \), where the parameters \( \mu \) and \( \sigma \) are related to a known vector \( x = (x_1, \ldots, x_p)^T \) of covariates through the logarithm link functions (for \( i = 1, \ldots, n \))

\[
\mu_i = \exp(x_i^T \beta_1), \quad \sigma_i = \exp(x_i^T \beta_2), \tag{6}
\]

respectively, where \( x_i^T = (x_{i1}, \ldots, x_{ip}) \) and \( \beta_1 = (\beta_{11}, \ldots, \beta_{1p})^T \) and \( \beta_2 = (\beta_{21}, \ldots, \beta_{2p})^T \).

Let \((y_1, x_{11}), \ldots, (y_n, x_{np})\) be a sample of \( n \) independent observations. The total log-likelihood function for the unknown vector \( \theta = (\beta_1^T, \beta_2^T, \nu)^T \) is determined from Equations (4) and (6)
FIGURE 2  The OLLLN density: (a) Varying the parameter $\sigma$. (b) Varying the parameter $\nu$. (c) Varying the parameter $\mu$

FIGURE 3  (a–c) Histograms and the OLLLN densities
\[ l(\theta) = \frac{n}{2} \log(\nu) - \frac{1}{2} \sum_{i=1}^{n} \left( \frac{\log(y_i) - \mu_i}{\sigma_i} \right)^2 \]
\[ - \frac{1}{2} \sum_{i=1}^{n} \log(2\pi\sigma_i^2) + (\nu - 1) \]
\[ \times \sum_{i=1}^{n} \log \left[ \Phi \left( \frac{\log(y_i) - \mu_i}{\sigma_i} \right) \right] \left[ 1 - \Phi \left( \frac{\log(y_i) - \mu_i}{\sigma_i} \right) \right] \]
\[ - \sum_{i=1}^{n} \log(y_i) - 2 \sum_{i=1}^{n} \log \left[ \Phi \left( \frac{\log(y_i) - \mu_i}{\sigma_i} \right) \right] \]
\[ + \left[ 1 - \Phi \left( \frac{\log(y_i) - \mu_i}{\sigma_i} \right) \right]^\nu \]. (7)

The MLE \( \hat{\theta} \) of \( \theta \) can be calculated numerically by maximising (7) using the \texttt{gamlass} package (Stasinopoulos and Rigby, 2007) [17] of the \texttt{R} software. This process can be executed for several initial values since it often leads to more than one maximum. The final estimate corresponds to the largest of the maxima.

Confidence intervals for the parameters (in large samples) can be constructed via the normal approximation for the estimates. The global influence and residual analysis are explored after fitting the regression. The generalised Cook distance \( GD_i(\hat{\theta}) = \left( \hat{\theta}_i - \hat{\theta}_i^0 \right)^2 | -L_i(\hat{\theta}) | \hat{\theta}_i - \hat{\theta} \) determines the influential observations. Another measure is given by \( LD_i(\hat{\theta}) = 2 | L_i(\hat{\theta}) - L_i(\hat{\theta}_i) | \), where the notation “(i)" refers to the deletion of the \( i \)th observation (Cook, 1977) [18].

Residual analysis investigates the suitability of the OLLLN regression based on residuals. Some graphic techniques for waste analysis are: residuals versus adjusted value plots; residual versus order of data plots and simulated residual envelope plots. The well-known qrs can be expressed as
\[ q_{r_i} = \Phi^{-1} \left( \frac{\eta(y_i)^\nu}{\eta(y_i)^\nu + [1 - \eta(y_i)]^\nu} \right). \] (8)

4 SIMULATION STUDIES

Three Monte Carlo simulation studies are demonstrated. First the precision of the estimates in the current distribution is checked. Second, the behaviour of the estimates for the OLLLN regression defined in the previous section is studied. Third, the normal approximation for the empirical distribution of the residuals is analysed.

4.1 The OLLLN distribution

Some properties of the MLEs of the OLLLN distribution are evaluated under a classical analysis using Monte Carlo simula-

| TABLE 1 Results for the OLLLN distribution |
|-------------------------------------------|

| Parameter | AE | Bias | MSE | AE | Bias | MSE |
|-----------|----|------|-----|----|------|-----|
| \( \mu \) | 0.967 | -0.033 | 0.197 | 0.971 | -0.029 | 0.101 |
| \( \sigma \) | 0.602 | 0.522 | 0.309 | 0.617 | 0.537 | 0.306 |
| \( \nu \) | 0.404 | 0.234 | 0.069 | 0.401 | 0.231 | 0.059 |

| n = 50 | n = 100 |
|--------|--------|
| Parameter | AE | Bias | MSE | AE | Bias | MSE |
| \( \mu \) | 0.982 | -0.018 | 0.030 | 0.980 | -0.019 | 0.009 |
| \( \sigma \) | 0.619 | 0.539 | 0.297 | 0.621 | 0.541 | 0.295 |
| \( \nu \) | 0.396 | 0.226 | 0.053 | 0.394 | 0.224 | 0.051 |

| n = 300 | n = 1000 |
|--------|---------|
| Parameter | AE | Bias | MSE | AE | Bias | MSE |
| \( \mu \) | 0.982 | -0.018 | 0.030 | 0.980 | -0.019 | 0.009 |
| \( \sigma \) | 0.619 | 0.539 | 0.297 | 0.621 | 0.541 | 0.295 |
| \( \nu \) | 0.396 | 0.226 | 0.053 | 0.394 | 0.224 | 0.051 |

- The sample sizes \( n = 50, 100, 300 \) and 1000 correspond to four scenarios;
- Calculate the estimates \( \hat{\mu}, \hat{\sigma} \) and \( \hat{\nu} \);
- Repeat the previous step 1000 times to obtain the average estimates (AEs), biases and MSEs.

The figures of the simulation reported in Table 1 clearly indicate that the MLEs are accurate. In addition, the MSEs decrease to zero when \( n \) increases.

4.2 The OLLLN regression

The second simulation study examines the precision of the estimates in the OLLLN regression for \( n = 100, 300 \) and 500. Ten thousand samples are simulated in two scenarios \( (\nu = 0.2 \) and \( \nu = 1.3 \) by considering \( \mu_i = 0.5 + 1.9 x_{i1} \) and \( \sigma_i = 0.4 + 0.1 x_{i3} \) (for \( i = 1, \ldots, n \)). The observations are generated from \( Y_i \sim \text{OLLLN}(\mu_i, \sigma_i, \nu) \) and \( x_{i3} \sim \text{Bernoulli}(6, 0.5) \).

The results of the simulations in Table A.1 (Section A) indicate that:

- The MSEs tend to zero when \( n \) increases.
- The AEs are in agreement with the true parameters for \( n \) large.

So, the finite sample distribution of the estimators can be well approximated by the normal distribution.

4.3 Empirical distribution of the residuals

Here samples of 1000 observations are generated based on the same scheme of the previous simulation study. Figures 4 and 5 provide normal probability plots of the qrs. They indicate that the empirical distribution of these residuals is close to the normal distribution. The normal probability plot can be adopted for these residuals with simulated envelopes.
ENERGY GENERATION DATA

Here, the OLLLN regression is applied to evaluate the mean daily energy generation potential of the Santo Antônio Hydroelectric Plant in Porto Velho (Brazil). It has 50 bulb-type turbines for power generation with a capacity of about 71.6 megawatts (MW) each. It is the fourth largest hydroelectric plant in operation in Brazil and one of the largest in the world. This plant is extremely important for supply of power to the national grid in Brazil, besides contributing significantly to serve regional demand for electricity in the states of Rondônia and Acre. The data were collected in 2016 and refer to $n = 213$ daily averages of electricity generation, covering the period from 1 June to 31 December. An OLLLN regression is constructed with nonlinear functions for $\mu$ and $\sigma$ in the presence of bimodality. The purpose is to detail the relationship between the daily mean and dispersion of generation by the plant with other explanatory factors, thus shedding some light on the behaviour of the response variable during this time frame. The following variables are considered in this study:

- $y_i$: mean daily electricity generation in megawatt-hours (MWh);
Table 2 lists some descriptive statistics of these data, which have positive asymmetry and negative kurtosis.

The GGu distribution, very popular to model extreme events, is an alternative approach for modelling these data. The cdf and pdf are

\[ G(y; \mu, \sigma, \nu) = \left(1 + \frac{1}{\nu} \exp \left[-\frac{(y - \mu)}{\sigma}\right]\right)^{-\nu}, y \in \mathbb{R}, \]

(9)

and

\[ g(y; \mu, \sigma, \nu) = \frac{1}{\sigma} \left(1 + \frac{1}{\nu} \exp \left[-\frac{(y - \mu)}{\sigma}\right]\right)^{-\nu-1} \]

\[ \times \exp \left[-\frac{(y - \mu)}{\sigma}\right], \]

(10)

respectively, where \( \mu \in \mathbb{R}, \sigma > 0 \) and \( \nu > 0 \).

For four distributions, the estimates and their standard errors (SEs) (between parentheses) are calculated, along with the Akaike information criterion (AIC), Bayesian information criterion (BIC) and global deviance (GD), which are reported in Tables 3 and 4. The figures in Table 4 show that the OLLLN distribution can be selected as the appropriate model.

The OLLLN distribution includes as sub-model the LN distribution, which allows their relative comparison. The LR statistic in Table 5 supports the OLLLN distribution for these data.

The plots of the estimated densities and cdfs of the OLLLN, LN, N and GGu distributions and the histogram and the empirical cdf (black curve) for the energy data are shown in Figure 6. Once more, it is proven that the OLLLN distribution is suitable for these data.

Figure 6a reveals the bimodality shape of the energy generation data, which the LN, N and GGu distributions do not have.
Estimation in the OLLLN regression

The parameters $\mu$ and $\sigma$ take non-linear equations (for $i = 1, \ldots, 213$)

$$
\mu_i = \exp \left( \beta_{10} + \sum_{j=1}^{6} \beta_{1j} d_{ij} \right),
$$

and

$$
\sigma_i = \exp \left( \beta_{20} + \sum_{j=1}^{6} \beta_{2j} d_{ij} \right).
$$

The estimates, SEs and $p$-values are reported in Table 6. The covariable “month” is significant for both components in $\mu$ and $\sigma$ at a 5% significance level. Some final interpretations of the fitted regression are addressed at the end of this section.

Table 7 lists the goodness-of-fit statistics for three fitted regressions, showing that the OLLLN regression gives the smallest measures for them.

The LR statistic for comparing two fitted regressions in Table 8 supports the OLLLN regression. The rejection of the null model is significant at 5% level, thus giving clear evidence that the shape parameter $\nu$ is necessary when modelling data of this type.

The quantities $LD_i(\Theta)$ and $GD_i(\Theta)$ (Section 3) are obtained and their plots are reported in Figure 7. They show that only the observation $\#187$ is a possible influential case. Further, Figure 8a displays plots of the residuals for the fitted OLLLN regression, which reveal that they have a random behaviour and that just the observation $\#187$ is out of the range $[-3, 3]$. Thus, no indications contrary the current assumptions for the fitted regression. The normal plot for the residuals with simulated envelope in Figure 8b indicates only one external point outside the envelope. Hence, the OLLLN regression is adequate for the energy data. Figure 8c reports the estimated cdfs based on the fitted OLLLN regression and the empirical cumulative function in specified months.
5.1 Interpretations of the systematic component for $\mu$

The covariable month has June as reference for $\mu$.

- The covariable $d_{6}$ is not significant at the 5% level, so there is no difference in the means for June and December.
- A relevant difference exists between June on the one hand and July, August, September, October and November on the other. Since the estimates are negative, there is evidence that the average daily energy generation declined from July to November in relation to June.
- Figure 8c shows the absence of a significant difference among the months of August, September and October. The same applies to the months of July and November as well as June and December. The OLLLN regression manages to detect similar behaviour between June and December and also between July and November (extreme events). Thus, there are three clusters, namely “June and December”, “July and November”, and “August, September and October”.
- Regarding these three clusters, the highest energy generation occurs in June and December, followed by July and November. In turn, the lowest generation is in the cluster “August, September and October”.

5.2 Interpretations of the systematic part of $\sigma$

For the systematic component of the dispersion parameter $\sigma$, the estimates related to August ($d_{2}$) and September ($d_{3}$) are significantly negative, thus indicating a lower variability in energy generation in these two months in relation to June as noted in Figure 8c. So, the OLLLN regression is accepted for these data.

6 CONCLUDING REMARKS

This paper presents the odd OLLLN regression for extreme events and the presence of bimodality with two systematic components. The proposed regression can be a valuable tool, especially for bimodal data as well as extreme energy values. The parameters are estimated via maximum likelihood and influence diagnostics and model performance are investigated based on qrs. The new regression is suitable to explain daily average power generation by the Santo Antônio Hydroelectric Plant located in Rondônia, Brazil.

ACKNOWLEDGEMENTS

This work was supported by CNPq, CAPES and the management of the Santo Antônio Hydroelectric Plant.

ORCID

Julio Cezar Souza Vasconcelos https://orcid.org/0000-0001-6794-3175
REFERENCES
1. Kotz, S., Nadarajah, S.: Extreme Value Distributions: Theory and Applications. Imperial College Press, London (2000)
2. Cordeiro, G.M., et al.: The Kumaraswamy Gumbel distribution. Statistical Methods and Applications 21(2), 139–168 (2012)
3. Cordeiro, G.M., et al.: The exponentiated generalized class of distributions. Journal of Data Science 11(1), 1–27 (2013)
4. Pescim, R.R., et al.: The new class of Kummer beta generalized distributions. Statistics and Operations Research Transactions 36(2), 153–180 (2012)
5. Halvorsen, B., Larsen, B.: The flexibility of household electricity demand over time. Resource and Energy Economics 23, 1–18 (2001)
6. Filippini, M., Pachauri, S.: Elasticities of electricity demand in urban Indian households. Energy Policy 32, 429–436 (2004)
7. Ansoby, I., Ozturk, I.: Estimating industrial and residential electricity demand in Turkey: A time varying parameter approach. Energy 66, 959–964 (2014)
8. Lawless, J.F.: Statistical Models and Methods of Lifetime Data, 2nd edn. John Wiley and Sons, New Jersey (2003)
9. Ozel, G., et al.: The odd log-logistic log-normal distribution with theory and applications. Advances in Data Science and Adaptive Analysis 10(4), 1–20 (2018)
10. Barreto-Souza, W., Vasconcellos, K.L.P.: Bias and skewness in a general extreme-value regression model. Comput. Stat. Data Anal. 55(3), 1379–1393 (2011)
11. Pinheiro, E.C., Ferrari, S.L.P.: A comparative review of generalizations of the Gumbel extreme value distribution with an application to wind speed data. J. Stat. Comput. Simul. 86(11), 2241–2261 (2016)

APPENDIX A
Table A.1 refers to the simulation study of the OLLLN regression.

### TABLE A.1 Results for the OLLLN regression

| Parameter | $n = 100$ | $n = 300$ | $n = 500$ |
|-----------|----------|----------|----------|
| $\beta_{10}$ | 0.596 | 0.538 | 0.552 | 0.096 | 0.038 | 0.052 | 0.168 | 0.002 | 0.028 |
| $\beta_{11}$ | 1.914 | 1.915 | 1.905 | 0.015 | 0.015 | 0.003 | 2.039 | 0.015 | 0.034 |
| $\beta_{20}$ | 0.425 | 0.432 | 0.435 | 0.025 | 0.032 | 0.035 | 1.662 | 1.391 | 1.362 |
| $\beta_{21}$ | 0.103 | 0.100 | 0.099 | 0.003 | 0.000 | 0.001 | 1.021 | 1.012 | 1.010 |
| $\gamma$ | 0.231 | 0.216 | 0.214 | 0.031 | 0.016 | 0.014 | 0.012 | 0.002 | 0.002 |

**Scenario 2**

| Parameter | $n = 100$ | $n = 300$ | $n = 500$ |
|-----------|----------|----------|----------|
| $\beta_{10}$ | 0.497 | 0.502 | 0.504 | -0.003 | 0.002 | 0.004 | 0.168 | 0.002 | 0.028 |
| $\beta_{11}$ | 1.902 | 1.899 | 1.899 | 0.002 | -0.001 | -0.001 | 0.019 | 0.000 | 0.003 |
| $\beta_{20}$ | 0.294 | 0.385 | 0.398 | -0.106 | -0.015 | -0.002 | 2.039 | 1.602 | 1.498 |
| $\beta_{21}$ | 0.101 | 0.099 | 0.099 | 0.001 | -0.001 | -0.001 | 1.023 | 1.012 | 1.009 |
| $\gamma$ | 1.340 | 1.348 | 1.342 | 0.040 | 0.048 | 0.042 | 0.622 | 0.239 | 0.169 |

How to cite this article: Vasconcelos JCS, Cordeiro GM, Ortega EMM, Gabriel JG. A regression model for extreme events and the presence of bimodality with application to energy generation data. IET Renew Power Gener. 2021;15:452–461. https://doi.org/10.1049/rpg2.12043