The new family of distributions and applications in heteroscedastic regression analysis

Gauss M. Cordeiro

Department of Statistics, Federal University of Pernambuco
Recife, PE, 50740-540, Brazil
gausscordeiro@gmail.com

Thiago G. Ramires

Department of Exact Sciences, University of São Paulo
Piracicaba, SP, Brazil
Interuniversity Institute for Biostatistics and statistical Bioinformatics, University of Hasselt
Hasselt, Belgium
thiagogentil@gmail.com

Edwin M.M. Ortega[*]

Department of Exact Sciences, University of São Paulo
Piracicaba, SP, Brazil
edwin@usp.br

Morad Alizadeh

Department of statistics, Persian Gulf University
751691-3798, Bushehr, Iran
moradalizadeh78@gmail.com

First we introduce and study some general mathematical properties of a new generator of continuous distributions with two extra shape parameters called the odd generalized half-Cauchy family. A second goal, we introduce the new log-generalized odd half-Cauchy heteroscedastic regression model with censored data, which represents a parametric family of models that includes as sub-models several widely known regression models that can be applied to data with no homogeneity of variance. Maximum likelihood estimation of the model parameters with censored data as well as a simulation study are investigated. We also prove empirically the flexibility of the new family by means real data sets.

Keywords: Censored data; generalized half-Cauchy; generated family; heteroscedastic regression models; maximum likelihood.

2000 Mathematics Subject Classification: 22E46, 53C35, 57S20

---

[*]Corresponding author.

1. Introduction

In recent years there has been an increased interest in defining new generators for univariate continuous distributions by introducing one or more additional shape parameter(s) to the baseline model. This induction of parameter(s) has been proved useful in exploring tail properties and also for improving the goodness-of-fit of the proposed generator family. Such distributions have been constructed by adding new parameters to a baseline cumulative distribution function (cdf) to obtain a new family of asymmetric distributions that are analytically more flexible. Some well-known generators are the beta-G by Eugene *et al.* (2002), gamma-G by Zografos and Balakrishanan (2009), Kumaraswamy-G by Cordeiro and de Castro (2011), McDonald-G by Alexander *et al.* (2012), the log-beta generalized half-normal by Pescim *et al.* (2013), the Zografos-Balakrishnan-G family by Nadarajah *et al.* (2014), flexible models generated by gamma random variables by Ortega *et al.*(2015) and generalized Weibull family of distributions by Cordeiro *et al.* (2015).

First, we introduce a new family of distributions based on the half-Cauchy (HC) distribution. The *Extended Generalized Odd Half-Cauchy-G* ("EGOHC-G for short") generator with two extra positive parameters α and β is defined by the cdf

$$F(x;\alpha,\beta,\boldsymbol{\xi}) = \frac{2}{\pi}\arctan\left[\frac{G(x;\boldsymbol{\xi})^{\alpha}}{1-G(x;\boldsymbol{\xi})^{\beta}}\right], \tag{1.1}$$

where α and β are positive shape parameters and $\boldsymbol{\xi}$ is a vector of parameters for the baseline G.

Let $g(x;\boldsymbol{\xi}) = dG(x;\boldsymbol{\xi})/dx$ be the baseline probability density function (pdf). The density function corresponding to (1.1) is given by

$$f(x;\alpha,\beta,\boldsymbol{\xi}) = \frac{2\,g(x;\boldsymbol{\xi})\,G(x;\boldsymbol{\xi})^{\alpha-1}\left[\alpha+(\beta-\alpha)G(x;\boldsymbol{\xi})^{\beta}\right]}{\pi\left\{G(x;\boldsymbol{\xi})^{2\alpha}+\left[1-G(x;\boldsymbol{\xi})^{\beta}\right]^{2}\right\}}. \tag{1.2}$$

If $\alpha = \beta$ we obtain as a special case the *Generalized Odd Half-Cauchy-G* (GOHC-G) family (Cordeiro *et al.*, 2017). Furthemore, the basic motivations for using the EGOHC-G family in practice are the following:

- to make the kurtosis more flexible compared to the baseline model;
- to produce a skewness for symmetrical distributions;
- to construct heavy-tailed distributions that are not longer-tailed for modeling real data;
- to generate distributions with symmetric, left-skewed, right-skewed and reversed-J shaped;
- to define special models with all types of the hazard rate function (hrf);
- to provide consistently better fits than other generated models under the same baseline distribution.

For a given probability $u \in (0,1)$, we obtain the corresponding quantile of X, say $Q(u)$, by inverting $F(x) = u$ in equation (1.1). Then, for each u, we have to solve numerically the nonlinear equation given by

$$\frac{G(x;\boldsymbol{\xi})^{\alpha}}{1-G(x;\boldsymbol{\xi})^{\beta}} = \tan\left(\frac{\pi u}{2}\right). \tag{1.3}$$

In practical applications, the response variable is affected by explanatory variables, regression models can be proposed in different forms in survival analysis. Among them, the location-scale regression model (Lawless, 2003) is distinguished and it is frequently used in clinical trials. In the

context, some distributions have been used to analyze censored data. For example, Ortega *et al.* (2014) defined a log-linear regression model for the odd Weibull distribution with censored data, Braga *et al.* (2016) proposed the odd log-logistic normal distribution with applications in analysis of experiments and da Cruz *et al.* (2016) proposed the log-odd log-logistic Weibull regression model. Thus, using the same approach adopted in this work, a distribution obtained from a generated EGOHC-G family will be expressed in the form of models belonging to the location-scale models.

Violation of the homogeneity of the error variances can have adverse consequences for the efficiency of estimators. In this way, we also propose a heteroscedastic regression model on the basis of the log-EGOHC-G distribution (parameterized in a convenient way), where both location and dispersion parameters vary across observations through regression structures. The log-EGOHC-G regression model can be used for modeling problems with censoring and uncensoring data. It should be mentioned that censored data is very common in lifetime because of time limits and other restrictions on data collection.

The paper is organized as follows. Section 2 provides some special distributions in the EGOHC family. In Section 3, we derive explicit expressions for the ordinary and incomplete moments, moment generating function (mgf) and mean deviations. Some inferential tools are discussed in Section 4. The performance of the maximum likelihood estimators (MLEs) for a special model is also investigated by a simulation study. In Section 5, we present a generalization of heteroscedastic regression models based on the EGOHC family. In Section 6, we fit some EGOHC-G distributions to three real data sets to prove empirically the potentiality of the new family. Finally, Section 7 ends with some conclusions.

2. Special EGOHC-G distributions

In this section, we present four special EGOHC-G models. The cdf (1.1) and pdf (1.2) of X will be most tractable when $G(x)$ and $g(x)$ have simple analytic expressions in all examples below, α and β are positive shape parameters.

2.1. *The EGOHC-W model*

Taking $G(x)$ to be the Weibull cdf with scale parameter $\lambda > 0$ and shape parameter $c > 0$, say $G(x) = 1 - \exp[-(x/\lambda)^c]$, the EGOHC-W pdf follows as

$$f(x; \alpha, \beta, \lambda, c) = \frac{2cx^{c-1}e^{-(\frac{x}{\lambda})^c}\left[1 - e^{-(\frac{x}{\lambda})^c}\right]^{\alpha-1}\left\{\alpha + (\beta - \alpha)\left[1 - e^{-(\frac{x}{\lambda})^c}\right]^{\beta}\right\}}{\lambda^c \pi \left\{\left[1 - e^{-(\frac{x}{\lambda})^c}\right]^{2\alpha} + \left\{1 - \left[1 - e^{-(\frac{x}{\lambda})^c}\right]^{\beta}\right\}^2\right\}}. \quad (2.1)$$

A random variable having pdf (2.1) is denoted by $X \sim$ GOHC-W$(\alpha, \beta, \lambda, c)$. For $c = 1$, we obtain the generalized odd half-Cauchy-exponential distribution. Plots of the pdf and hrf of the EGOHC-W model for selected parameter values are displayed in Figure 1. We provide in Figures 2a-b a numerical investigation to identify how the parameter values change the shapes of the pdf of X for some parameter ranges. To identify the shapes of the pdf, the signals of the numeric derivatives were obtained determining whether the pdf is increasing or decreasing in a range, being possible to determine the modal, bimodal, decreasing and other shapes. Based on these plots, we can obtain
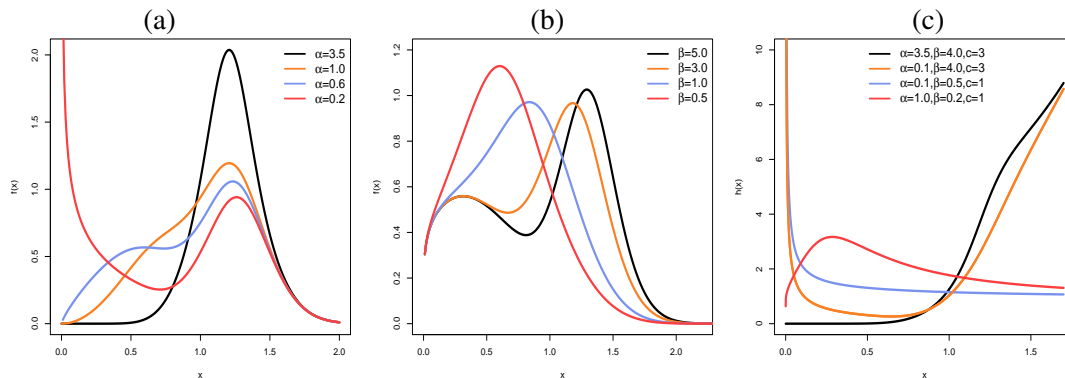
Fig. 1. The EGOHC-W model: (a) Density functions for $\beta = 4$, $\lambda = 1$, $c = 3$ and some values of α. (b) Density functions for $\alpha = 0.4$, $\lambda = 1$, $c = 3$ and some values of β. (c) Hazard function for $\lambda = 1$ and some values of α, β and c.

bimodal shapes for the pdf of the EGOHC-W distribution for large values of β parameter. However, large values of β parameter is necessary to obtain this characteristic when the α and c parameters increases. Similar analyzes can be made with the other new models.
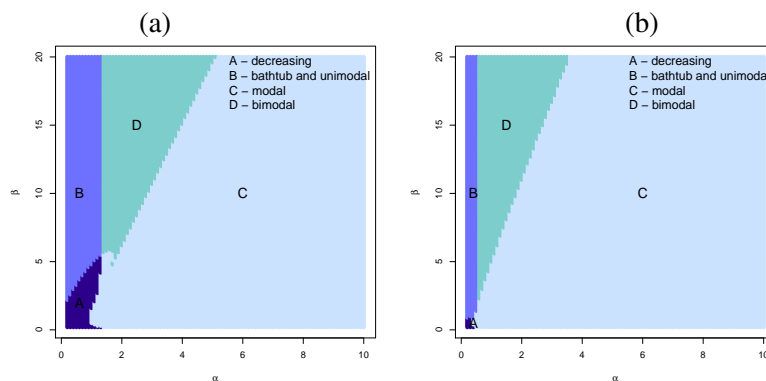

Fig. 2. The EGOHC-W pdf shapes as functions of α and β for $\lambda = 2$ (a) $c = 1$ and (b) $c = 2$.

2.2. *The EGOHC-Gu model*

The cdf of the Gumbel model is given by (for $x \in \mathbb{R}$) $G(x) = \exp\{-\exp[-(x-\mu)/\sigma]\}$. The GOHC-Gu density can be expressed as

$$f(x; \alpha, \beta, \mu, \sigma) = \frac{2\exp(-z - \alpha e^{-z})[\alpha + (\beta - \alpha)\exp(-\beta e^{-z})]}{\sigma\pi\{\exp(-2\alpha e^{-z}) + [1 - \exp(-\beta e^{-z})]\}},$$

where $z = (x - \mu)/\sigma$, $\mu \in \mathbb{R}$ is the location parameter and $\sigma > 0$ is the scale parameter. Plots of the EGOHC-Gu density function for some parameter values are displayed in Figure 3.

2.3. *The EGOHC-Ga model*

The gamma cdf (for $x > 0$) with shape parameter $a > 0$ and scale parameter $b > 0$ is $G(x) = \gamma(a, bx)/\Gamma(a)$, where $\Gamma(a) = \int_0^\infty w^{a-1} e^{-w} dw$ is the gamma function and $\gamma(a, x) = \int_0^x w^{a-1} e^{-w} dw$
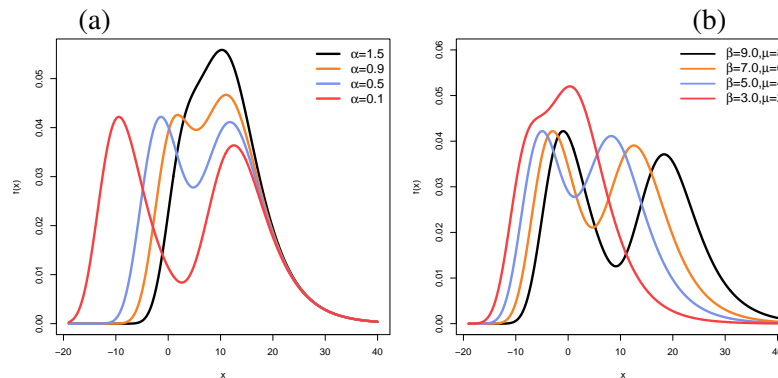
Fig. 3. The EGOHC-Gu density function: (a) For $\beta = 5$, $\mu = 3$, $\sigma = 5$ and some values of α. (b) For $\alpha = 0.2$, $\sigma = 5$ and some values of β and μ.

is the incomplete gamma function. The EGOHC-Ga pdf becomes (for $x > 0$) is given by

$$f(x;\alpha,\beta,a,b) = \frac{2b^a x^{a-1} e^{-bx} \gamma(a,bx)^{\alpha-1} \left[\alpha \Gamma(a)^\beta + (\beta - \alpha)\gamma(a,bx)^\beta\right]}{\pi \Gamma(a)^{\alpha+\beta} \left\{ \frac{\gamma(a,bx)^{2\alpha}}{\Gamma(a)^{2\alpha}} + \frac{[\Gamma(a)^\beta - \gamma(a,bx)^\beta]^2}{\Gamma(a)^{2\beta}} \right\}}. \tag{2.2}$$

2.4. *The EGOHC-LL model*

Consider the log-logistic (LL) distribution with shape parameter $a > 0$ and scale parameter $\gamma > 0$, where the pdf and cdf (for $x > 0$) are

$$g(x) = \frac{\gamma}{a^\gamma} x^{\gamma-1} \left[1 + \left(\frac{x}{a}\right)^\gamma\right]^{-2} \qquad \text{and} \qquad G(x) = 1 - \frac{1}{1 + \left(\frac{x}{a}\right)^\gamma},$$

respectively. By inserting these expressions into (1.2) gives the EGOHC-LL density function

$$f(x;\alpha,\beta,a,\gamma) = \frac{2\gamma x^{\gamma\alpha-1}}{\pi a^{\gamma\alpha}} \frac{\alpha + (\beta - \alpha)\left(\frac{x}{a}\right)^{\gamma\beta}\left[1 + \left(\frac{x}{a}\right)^\gamma\right]^{-\beta}}{\frac{\left(\frac{x}{a}\right)^{2\gamma\alpha}}{\left[1 + \left(\frac{x}{a}\right)^\gamma\right]^{\alpha-1}} + \frac{\left\{\left[1 + \left(\frac{x}{a}\right)^\gamma\right]^\beta - \left(\frac{x}{a}\right)^{2\gamma\beta}\right\}^2}{\left[1 + \left(\frac{x}{a}\right)^\gamma\right]^{2\beta-\alpha-1}}}. \tag{2.3}$$

A random variable having pdf (2.3) is denoted by $X \sim \text{GOHC-LL}(\alpha,\beta,a,\gamma)$. Plots of the density function (2.3) and hrf for some parameter values are displayed in Figure 4.

3. Moments, generating function and other measures

We hardly need to emphasize the necessity and importance of moments in any statistical analysis especially in applied work. Some of the most important features and characteristics of a distribution can be studied through moments (e.g., tendency, dispersion, skewness and kurtosis). The useful relation with the exponentiated class, used in the section, can be obtained directly from the authors.
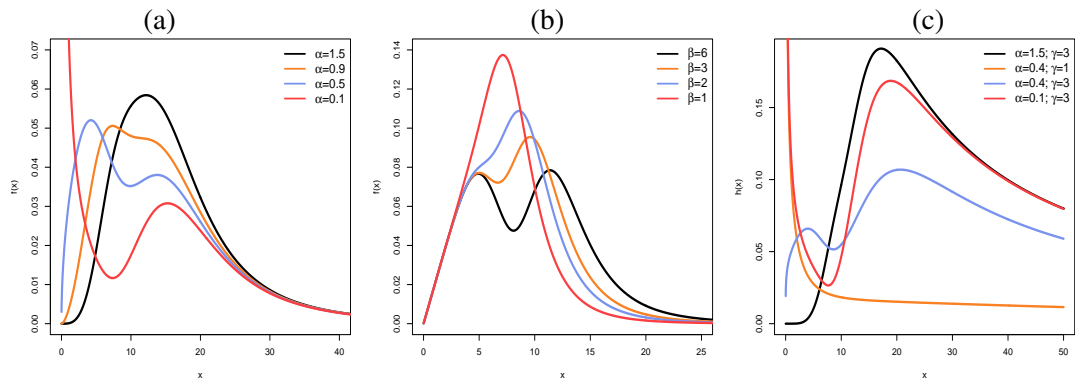
Fig. 4. The EGOHC-LL model: (a) Density function for $a = 10, \gamma = 3, \beta = 4$ and some values of α. (b) Density function for $a = 10, \gamma = 3, \alpha = 0.4$ and some values of β. (c) Hazard function for $\beta = 4, a = 10$ and some values of α and γ.

3.1. *Moments*

Henceforth, let $Y_{k+1} \sim$ exp-G$(k+1)$. The nth moment of X can be expressed as

$$\mu'_n = E(X^n) = \sum_{k=0}^{\infty} v_k E(Y_{k+1}^n). \tag{3.1}$$

Explicit expressions for the moments of several exponentiated distributions are given by Nadarajah and Kotz (2006). They can be used to produce μ'_n. Here, we give two examples. First, we consider the Gumbel distribution with cdf $G(x) = 1 - \exp\left\{ -\exp\left(\frac{x-\mu}{\sigma}\right) \right\}$. The moments of the exponentiated Gumbel distribution with power parameter $(k+1)$ can be obtained from Nadarajah and Kotz (2006). The nth moment of the EGOHC-Gu distribution reduces to

$$E(X^n) = \sum_{k=0}^{\infty} v_k (k+1) \sum_{i=0}^{n} \binom{n}{i} \mu^{n-i} (-\sigma)^i \left(\frac{\partial}{\partial p} \right)^i \left[(k+1)^{-p} \Gamma(p) \right] \Bigg|_{p=1}.$$

Second, we consider the EGOHC-standard logistic (EGOHC-SL) distribution, where $G(x) = (1 + e^{-x})^{-1}$. A result from Prudnikov *et al.* (1986, Section 2.6.13, equation 4) we obtain

$$E(X^n) = \sum_{k=0}^{\infty} v_k \left(\frac{\partial}{\partial t} \right)^n B(t+k+1, 1-t) \Bigg|_{t=0},$$

where $B(a,b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ is the beta function.

When the moments do not exist, for example, for the Cauchy, Lévy and Pareto distributions, alternative measures for the skewness and kurtosis, based on qfs, are sometimes more appropriate. Recently, Ramires *et al.* (2016) studied the skewness and kurtosis of the log-sinh Cauchy distribution using the Galton's skewness and Morrs's kurtosis, respectively. The measures of skewness $\mathscr{B}$ and kurtosis $\mathscr{M}$ are given by

$$\mathscr{B} = \frac{Q(6/8) + Q(2/8) - 2Q(4/8)}{Q(6/8) - Q(2/8)} \quad \text{and} \quad \mathscr{M} = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)},$$

respectively.

We illustrate these measures by selecting the EGOHC-W distributions. Figure 5 displays some plots of the $\mathscr{B}$ and $\mathscr{M}$ measures as functions of α and β. The additional parameters α and β have substantial effect on the skewness and kurtosis of X.
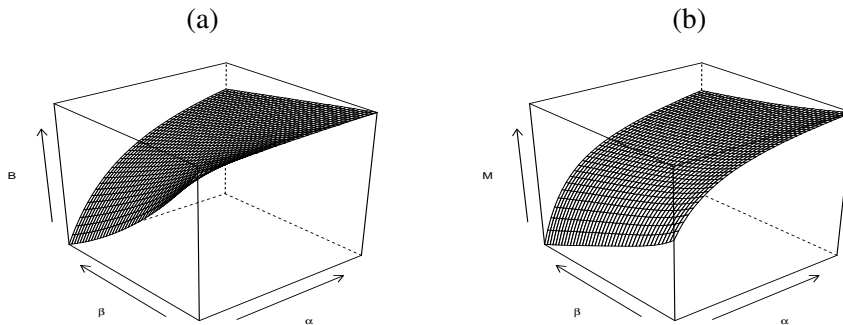

Fig. 5. Effects of the parameters α and β on the measures: (a) $\mathscr{B}$ and (b) $\mathscr{M}$ for the EGOHC-W$(\alpha, \beta, 1, 3)$ model.

3.2. *Generating function and mean deviations*

We provide two formulae for the mgf $M(t) = E(e^{tX})$ of X. A first formula for is obtained using the relation with the exponentiated class as

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

where $M_{k+1}(t)$ is the mgf of Y_{k+1}. A second formula for can also be derived using the relation with the exponentiated class, but in terms of the present qf as

$$M(t) = \sum_{k=0}^{\infty} (k+1) v_k \rho(t,k), \tag{3.2}$$

where $\rho(t,a)$ can be evaluated (at least numerically) by

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

We can obtain the mgfs of several EGOHC-G distributions directly from equation (3.2). For example, the mgf of the EGOHC-SL distribution (for $t < 1$) is given by

$$M(t) = \sum_{k=0}^{\infty} (k+1) B(t+k+1, 1-t) v_k.$$

The mean deviations about the mean ($\delta_1 = E(|X - \mu_1'|)$) and about the median ($\delta_2 = E(|X - M|)$) of X can be expressed as

$$\delta_1 = 2\mu_1' F(\mu_1') - 2m_1(\mu_1') \qquad \text{and} \qquad \delta_2 = \mu_1' - 2m_1(M), \tag{3.3}$$

respectively, where $\mu_1' = E(X)$, $M = Median(X)$ is the median given in Section 4, $F(\mu_1')$ is easily evaluated from (1.1) and $m_1(z) = \int_{-\infty}^{z} x f(x) dx$ is the first incomplete moment. To compute δ_1 and

δ_2, the expression for $m_1(z)$ can be derived from the relation with the exponentiated class as

$$m_1(z) = \sum_{k=0}^{\infty} v_k J_{k+1}(z), \qquad \text{where} \qquad J_{k+1}(z) = \int_{-\infty}^{z} x h_{k+1}(x) dx. \tag{3.4}$$

Equation (3.4) is the basic quantity to determine the mean deviations of the exp-G distributions. Hence, the mean deviations in (3.3) depend only on the exp-G mean deviations. As an alternative, $m_1(z)$ can be derived by setting $u = G(x)$ in (3.4)

$$m_1(z) = \sum_{k=0}^{\infty} (k+1) v_k T_{k+1}(z), \qquad \text{where} \qquad T_{k+1}(z) = \int_{0}^{G(z)} Q_G(u) u^k du. \tag{3.5}$$

In a similar way, the mean deviations of any EGOHC-G distribution can be determined from equation (3.5). For example, the mean deviations of the EGOHC-SL distribution are obtained by using the generalized binomial expansion from the function

$$T_{k+1}(z) = \frac{1}{\Gamma(z)} \sum_{m=0}^{\infty} \frac{(-1)^m \Gamma(k+m+1) [1 - \exp(-mz)]}{(m+1)!}.$$

4. Maximum likelihood estimation

In this section, we determine the maximum likelihood estimates (MLEs) of the EGOHC-G parameters from complete samples only. Let $x_1, \ldots, x_n$ be a random sample of size n from the EGOHC-G$(\alpha, \beta, \boldsymbol{\xi}^{\top})^{\top}$ distribution. The log-likelihood function for the vector of parameters $\theta = (\alpha, \beta, \boldsymbol{\xi}^{\top})^{\top}$ can be expressed as

$$l(\theta) = n \log(2) + \sum_{i=1}^{n} \log g(x_i; \boldsymbol{\xi}) + (\alpha - 1) \sum_{i=1}^{n} \log G(x_i; \boldsymbol{\xi}) + \sum_{i=1}^{n} \log \left[\alpha + (\beta - \alpha) G(x_i; \boldsymbol{\xi})^{\beta} \right]$$
$$- n \log(\pi) - \sum_{i=1}^{n} \log \left\{ G(x_i; \boldsymbol{\xi})^{2\alpha} + \left[1 - G(x_i; \boldsymbol{\xi})^{\beta} \right]^2 \right\},$$

where $g(x_i; \boldsymbol{\xi})$ and $G(x_i; \boldsymbol{\xi})$ are defined in Section 2.

The MLE $\widehat{\theta}$ of θ can be determined by maximizing the log-likelihood presented in last equation. Iterative techniques such as the Newton-Raphson type-algorithms are commonly adopted to estimate θ. We employ the numerical NLMixed in SAS procedure.

We can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain likelihood ratio (LR) statistics for testing some special models of the proposed family. Tests of the hypotheses of the type $H_0 : \psi = \psi_0$ *versus* $H : \psi \neq \psi_0$, where ψ is a subset of parameters of θ, can be performed through LR statistics in the usual way.

4.1. *Simulation study: EGOHC-W model*

An easy way to validate the approximate normal distribution for $\widehat{\theta}$ is by simulating a specific distribution of the new family of distribution. Here, the EGOHC-W model is selected as an example. We use equation (1.3) to simulate the EGOHC-W$(\alpha = 0.4, 5, \beta = 0.8, 5, c = 3, \lambda = 5)$ model by taking u as a uniform random variable in $(0, 1)$ for $n = 50, 150$ and 300. For each sample size, we evaluate the MLEs of the parameters. Then, we repeat this process 1,000 times and compute the averages of the estimates (AEs), biases and mean squared errors (MSEs). The simulation results are reported in Table 1.

Table 1. The AEs, biases and MSEs based on 1,000 simulations of the EGOHC-W distribution when $\alpha = 0.4, 2, \beta = 0.8, 5$ c=3 and λ=5, and n=50, 150 and 300.

| | | $\alpha = 0.4$ and $\beta = 5$ | | | | $\alpha = 2.0$ and $\beta = 5$ | | |
n	Parameter	AE	Bias	MSE	Parameter	AE	Bias	MSE
50	α	0.4113	0.0113	0.0173	α	2.1502	0.1502	0.6403
	β	5.2399	0.2399	2.0056	β	5.2706	0.2706	2.6097
	c	3.1544	0.1544	0.4415	c	3.1459	0.1459	0.4631
	λ	5.0115	0.0115	0.1982	λ	4.9999	-1e-04	0.1655
150	α	0.403	0.003	0.0048	α	2.0507	0.0507	0.2267
	β	5.1892	0.1892	1.4248	β	5.3081	0.3081	2.3582
	c	3.0509	0.0509	0.1214	c	3.066	0.0660	0.1486
	λ	5.0038	0.0038	0.0784	λ	4.9936	-0.0064	0.0709
300	α	0.4027	0.0027	0.0022	α	2.0439	0.0439	0.1167
	β	5.1102	0.1102	0.8070	β	5.2017	0.2017	0.9785
	c	3.0287	0.0287	0.0555	c	3.0209	0.0209	0.0697
	λ	5.0006	6e-04	0.0402	λ	4.9859	-0.0141	0.0418
		$\alpha = 0.4$ and $\beta = 0.8$				$\alpha = 2.0$ and $\beta = 0.8$		
n	Parameter	AE	Bias	MSE	Parameter	AE	Bias	MSE
50	α	0.4076	0.0076	0.0129	α	2.0994	0.0994	0.5502
	β	0.9496	0.1496	0.5377	β	1.1029	0.3029	0.4494
	c	3.1881	0.1881	0.2841	c	3.2073	0.2073	0.5705
	λ	5.0217	0.0217	0.2672	λ	4.8846	-0.1154	0.101
150	α	0.3943	-0.0057	0.0043	α	2.0915	0.0915	0.2273
	β	0.8203	0.0203	0.1104	β	0.9401	0.1401	0.1516
	c	3.121	0.121	0.1199	c	3.0355	0.0355	0.1854
	λ	5.0538	0.0538	0.1301	λ	4.9228	-0.0772	0.0496
300	α	0.392	-0.008	0.0021	α	2.0593	0.0593	0.122
	β	0.7888	-0.0112	0.044	β	0.9243	0.1243	0.0846
	c	3.0925	0.0925	0.0587	c	3.0124	0.0124	0.0911
	λ	5.0571	0.0571	0.0719	λ	4.9185	-0.0815	0.0387

The figures in Table 1 indicate that the MSEs and the biases of $\hat{\alpha}$, $\hat{\beta}$, $\hat{c}$ and $\hat{\lambda}$ decay toward zero when the sample size increases for all settings of α and β, as expected under first-under asymptotic theory. When n increases, the AEs of the parameters tend to be closer to the true parameter values. This fact supports that the asymptotic normal distribution provides an adequate approximation to the finite sample distribution of the MLEs. Figure 6 displays the true densities at selected parameter values and the density functions evaluated at the average values of the MLEs given in Table 1 for $n = 50$. These plots are in agreement with the first-order asymptotic theory for the MLEs and they show a the fast convergence even for small sample sizes.

5. The heteroscedastic regression model

The last decade is full of works on generalized classes of regression models, which are always precious for applied statisticians. In many practical applications, the lifetimes are affected by explanatory variables such as the cholesterol level, blood pressure, weight and many others. Parametric models to estimate univariate survival functions and for censored data regression problems are widely used. A regression model that provides a good fit to lifetime data tends to yield more precise estimates of the quantities of interest. In the other hand, a standard assumption in regression
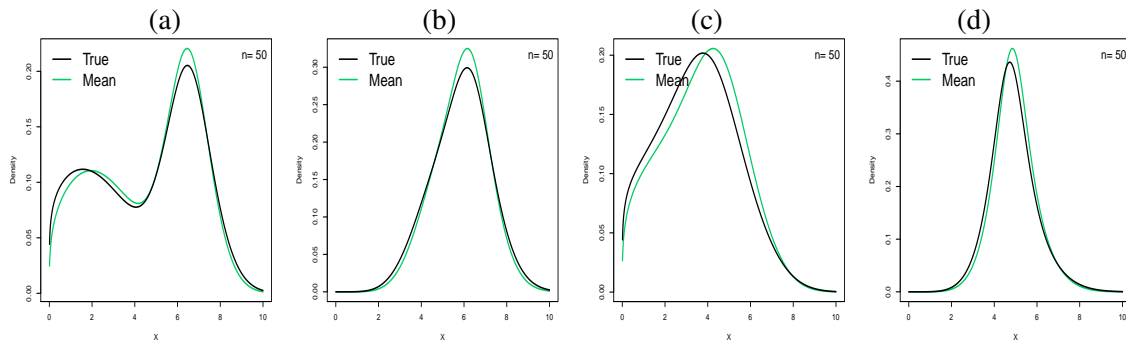
Fig. 6. Density functions of the EGOHC-W distribution at the true parameter values and at the AEs for $n = 50$, $c = 3$, $\lambda = 5$ and: (a) α=0.4 and $\beta = 5$; (b) α=2 and $\beta = 5$; (c) α=0.4 and $\beta = 0.8$; (d) α=2 and $\beta = 0.8$.

analysis is the homogeneity of the error variances. Violation of this assumption can have adverse consequences for the efficiency of estimators, and then it is important to check for heteroscedasticity whenever it is considered a possibility.

Let X be a random variable having the pdf (1.2) with $x > 0$. A class of heteroscedastic regression models is characterized by the fact that the random variable $Y = \log(X)$ has a distribution wherein the location $\mu(\mathbf{v})$ and scale $\sigma(\mathbf{w})$ parameters depend only on the explanatory variable vectors $\mathbf{v}$ and $\mathbf{w}$, which may be identical or not. The main idea of heteroscedastic models is to capture the effect of explanatory variables in the mean and variability of the response variable. If these explanatory variables affect only the response variable, the class of location regression models appear as a special case where only the location $\mu(\mathbf{v})$ parameter is modeled by explanatory variables.

The heteroscedastic regression model is defined by

$$Y = \mu(\mathbf{v}) + \sigma(\mathbf{w})Z,$$

where Z has the distribution that does not depend on $\mathbf{v}$ and $\mathbf{w}$. The random variable Y (for $y \in \Re$) has density function given by

$$f(y; \alpha, \beta, \mu, \sigma) = \frac{2 g\left(\frac{y-\mu(\mathbf{v})}{\sigma(\mathbf{w})}\right) G^{\alpha-1}\left(\frac{y-\mu(\mathbf{v})}{\sigma(\mathbf{w})}\right) \left[\alpha + (\beta - \alpha)G^\beta\left(\frac{y-\mu(\mathbf{v})}{\sigma(\mathbf{w})}\right)\right]}{\sigma \pi \left\{G^{2\alpha}\left(\frac{y-\mu(\mathbf{v})}{\sigma(\mathbf{w})}\right) + \left[1 - G^\beta\left(\frac{y-\mu(\mathbf{v})}{\sigma(\mathbf{w})}\right)\right]^2\right\}}, \quad (5.1)$$

where the functions $G(\cdot)$ and $g(\cdot)$ are defined in Section 2.

We refer to equation (5.1) as the *Log-extended Generalized Odd Half-Cauchy-G* (LEGOHC-G) family, say $Y \sim \text{LGOHC-G}(\alpha, \beta, \mu, \sigma)$. The new family in the form of location and scale is able to define new regression models, which can be used in survival analysis and reliability.

For illustrative purposes, let X be a random variable having the EGOHC-W density function defined in Section 2.1. The random variable $Y = \log(X)$, re-parameterized in terms of $\mu = \log(\lambda)$ and $\sigma = 1/c$, has density function given by

$$f(y; \alpha, \beta, \mu, \sigma) = \frac{2 e^{\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}} \left(1 - e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}}\right)^{\alpha-1} \left[\alpha + (\beta - \alpha)\left(1 - e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}}\right)^\beta\right]}{\sigma \pi \left\{\left(1 - e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}}\right)^{2\alpha} + \left[1 - \left(1 - e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}}\right)^\beta\right]^2\right\}}, \quad (5.2)$$

where $\alpha > 0$ and $\beta > 0$ are shape parameters, $\mu \in \Re$ is the location parameter and $\sigma > 0$ is the scale parameter. We refer to equation (5.2) as the *log-extended generalized odd half-Cauchy-Weibull* (LEGOHC-W) distribution, say $Y \sim$ LEGOHC-W$(\alpha, \beta, \mu, \sigma)$. If $X \sim$ EGOHC-W$(\alpha, \beta, \lambda, c)$, then $Y = \log(X) \sim$ LEGOHC-W$(\alpha, \beta, \mu, \sigma)$.

The survival function of Y is given by

$$S(y) = 1 - \frac{2}{\pi} \arctan \left[\frac{\left(1 - e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\alpha}}{1 - \left(1 - e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\beta}} \right]. \tag{5.3}$$

Now, we define the standardized random variable $Z = (Y - \mu)/\sigma$ having the density function

$$f(z; \alpha, \beta) = \frac{2 e^{z - e^z} \left(1 - e^{-e^z}\right)^{\alpha - 1} \left[\alpha + (\beta - \alpha)\left(1 - e^{-e^z}\right)^{\beta}\right]}{\pi \left\{ \left(1 - e^{-e^z}\right)^{2\alpha} + \left[1 - \left(1 - e^{-e^z}\right)^{\beta}\right]^2 \right\}}, \tag{5.4}$$

Next, we propose a linear heteroscedastic regression model linking the response variable y_i and the explanatory variable vectors $\mathbf{v}_i^T = (v_{i1}, \ldots, v_{ip_1})$ and $\mathbf{w}_i^T = (w_{i1}, \ldots, w_{ip_2})$ as follows

$$y_i = \mu_i + \sigma_i z_i, \ i = 1, \ldots, n, \tag{5.5}$$

where the random error z_i has density function (5.4), $\mu_i = \mathbf{v}_i^T \tau_1$, $\sigma_i = \exp(\mathbf{w}_i^T \tau_2)$ and $\tau_1 = (\tau_{11}, \ldots, \tau_{1p_1})^T$, $\tau_2 = (\tau_{21}, \ldots, \tau_{2p_2})^T$, $\alpha > 0$ and $\beta > 0$ are unknown parameters. The parameters μ_i and σ_i are the location and scale parameters of y_i, respectively. The location $\mu = (\mu_1, \ldots, \mu_n)^T$ and scale $\sigma = (\sigma_1, \ldots, \sigma_n)^T$ parameter vectors are represented by linear models $\mu = \mathbf{V}\tau_1$ and $\sigma = \mathbf{W}\tau_2$, respectively, where $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)^T$ and $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)^T$ are known model matrices. The LEGOHC-W model (5.5) opens new possibilities for fitting many different types of data.

Consider a sample $(y_1, \mathbf{v}_1, \mathbf{w}_1), \ldots, (y_n, \mathbf{v}_n, \mathbf{w}_n)$ 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. Standard likelihood estimation techniques can be applied here. The log-likelihood function for the vector of parameters $\theta = (\alpha, \beta, \tau_1^T, \tau_2^T)^T$ from model (5.5) has the form

$$l(\theta) = r[\log(2) - \log(\pi)] + \sum_{i \in F}(z_i - u_i) + (\alpha - 1)\sum_{i \in F}\log[1 - \exp(-u_i)] - \sum_{i \in F}(\mathbf{w}_i^T \tau_2) +$$

$$\sum_{i \in F}\log\{\alpha + (\beta - \alpha)[1 - \exp(-u_i)]^{\beta}\} + \sum_{i \in C}\log\left\{1 - \frac{2}{\pi}\arctan\left[\frac{[1 - \exp(-u_i)]^{\alpha}}{1 - [1 - \exp(-u_i)]^{\beta}}\right]\right\}$$

$$- \sum_{i \in F}\log\left\{[1 - \exp(-u_i)]^{2\alpha} + \left[1 - \{1 - \exp(-u_i)\}^{\beta}\right]^2\right\},$$

where $u_i = \exp(z_i)$, $z_i = (y_i - \mu_i)/\sigma_i$, $\mu_i = \mathbf{v}_i^T \tau_1$, $\sigma_i = \exp(\mathbf{w}_i^T \tau_2)$ and r is the number of uncensored observations (failures). The maximum likelihood estimate (MLE) $\widehat{\theta}$ of the vector of unknown parameters can be determined by maximizing the log-likelihood (5.6). We use the NLMixed procedure in SAS to calculate the estimate $\widehat{\theta}$. Initial values for τ_1 are taken from the fit of the log-Weibull regression model.

The fit of the LEGOHC-W model gives the estimated survival function for y_i

$$S(y_i; \hat{\alpha}, \hat{\beta}, \hat{\tau}_1^T, \hat{\tau}_2^T) = \log \left\{ 1 - \frac{2}{\pi} \arctan \left[\frac{[1 - \exp(-\hat{u}_i)]^{\hat{\alpha}}}{1 - [1 - \exp(-\hat{u}_i)]^{\hat{\beta}}} \right] \right\},$$

where

$$\hat{u}_i = \exp(\hat{z}_i), \qquad \hat{z}_i = \frac{y_i - \mathbf{v}_i^T \hat{\tau}_1}{\exp(\mathbf{w}_i^T \hat{\tau}_2)}.$$

Note that in the healing models already there is heterogeneity in the data because we always have two subpopulations, one formed by the failure data and another for censored data. However, we can test the assumption of homogeneity of variance, for the LEGOHC-W regression model based on the LR statistic. Following (5.5) and (5.6), we generalized the scale parameter σ by σ_i, where $\sigma_i = \exp(\mathbf{w}_i^T \tau_2)$ and $\mathbf{w}_i$ is a vector of explanatory variable values. We assume that there exists a unique value σ_0, then $\sigma_i = \sigma_0$ and Y_i's have constant variance. Hence, the LR test for the homogeneity of scalar parameter becomes the test of hypothesis expressed by $\mathcal{H}_0 : \sigma_i = \sigma_0$ against $\mathcal{H}_a : \sigma_i \neq \sigma_0$, which is provided by $\Lambda = 2[\ell(\hat{\alpha}, \hat{\beta}, \hat{\tau}_1, \hat{\tau}_2) - \ell(\tilde{\alpha}, \tilde{\beta}, \tilde{\tau}_1, \sigma_0)]$. Here $\tilde{\alpha}, \tilde{\beta}$ and $\tilde{\tau}_1$ denotes the restricted MLEs of α, β and τ_1, respectively, obtained from the maximization of (5.6) under $\mathcal{H}_0 : \sigma_i = \sigma_0$. We can also use the LR statistic for comparing some special models with the LEGOHC-W model.

5.1. *Simulation study: heteroscedastic LEGOHC-W regression model*

Here, we evaluate the performance of the MLEs of the parameters in the heteroscedastic LEGOHC-W regression model given by (5.5) by means of a simulation study, by taking different sample sizes $n = 50, 150$ and 300. We consider model (5.5) where the model parameters are: $\alpha = 4$, $\beta = 5$, $\mu = -14 - 5x_1$ and $\sigma = 3 + 0.5x_1$, where the variable x_1 is generated from a binomial $(n,0.5)$ distribution. This configuration considers that the response variable Y is affected by the effects of two groups $x_{1i} = 0, 1$, and each group has different values of the location "μ" and scale "σ" parameters. The values of the response variable Y, denoted by $y_1, \ldots, y_n$, are generated from the LEGOHC-W distribution using the quantile function (1.3) and, for each configuration of n, all results are obtained from 1,000 Monte Carlo replications. We present the results, where the location and scale parameters are defined by $\mu_i = \tau_{10} + \tau_{11} x_{1i}$, and $\sigma_i = \exp(\tau_{20} + \tau_{21} x_{1i})$, respectively. The results of the Monte Carlo study in Table 2 indicate that the MSEs of the MLEs of the parameters decay toward zero when the sample size increases, as expected under first-order asymptotic theory.

6. Applications

In this section, we provide three applications to real data to prove empirically the flexibility of the EGOHC-G family. For the first two examples, the computations are performed using the goodness.fit subroutine in the AdequacyModel script of the R package. In the third application for censored data, the computations are done using the NLMixed subroutine of the SAS software.

6.1. *Data: Actuarial sciences*

It is important for the Mexican Institute of Social Security (MISS) to study the distributional behaviour of the mortality of retired people on disability because it enables the calculation of long

Table 2. The AEs, biases and MSEs based on 1,000 simulations for the heteroscedastic LEGOHC-W regression model $\alpha = 4$, $\beta = 5$, τ_{10}=-14, τ_{11}=-5, τ_{20}=3 and τ_{21}=0.5, n=50, 150 and 300.

Parameter	$n = 50$			$n = 150$			$n = 300$		
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
α	4.308	0.308	1.381	4.292	0.292	0.566	4.014	0.014	0.490
β	8.540	3.540	30.327	7.467	2.467	15.689	7.068	2.068	12.197
τ_{10}	-13.820	0.180	14.671	-14.200	-0.200	6.187	-14.067	-0.067	5.246
τ_{11}	-4.666	0.334	20.449	-5.130	-0.130	10.089	-5.084	-0.084	6.431
τ_{20}	2.847	-0.153	0.070	2.913	-0.087	0.020	2.925	-0.075	0.012
τ_{21}	0.516	0.016	0.043	0.499	-0.001	0.014	0.495	-0.005	0.007

and short term financial estimation, such as the assessment of the reserve required to pay the "minimum pensions". The data set refers to 280 lifetimes (in years) of retired women with temporary disabilities, which are incorporated in the Mexican insurance public system and who died during 2004. Recently, Balakrishnan *et al.* (2009) reported and analysed these data using the mixture inverse Gaussian (MIG) distribution. Cordeiro *et al.* (2014c) fitted the Kummer beta generalized gamma (KBG-GG) distribution to the current data and conclude that it provides a better fit than the beta generalized gamma (BGG) (Cordeiro *et al.*, 2013), exponentiated generalized gamma (EGG) (Cordeiro *et al.*, 2011), generalized gamma (GG) and Kumaraswamy generalized gamma (Kw-GG)(Pascoa *et al.*, 2011) distributions.

Next, we compare the KBG-GG and some of its sub-models with the EGOHC-Ga model to fit these data. We also present the fit of the special GOCH-Ga model when $\alpha = \beta$. Table 3 provides the MLEs (and the corresponding standard errors in parentheses) of the model parameters and the values of the statistics AIC and BIC for the fitted models.

Table 3. MLEs of the model parameters for the actuarial data, the corresponding SEs (given in parentheses) and the AIC and BIC statistics.

Model	Estimates						AIC	BIC
EGOHC-Ga(α,β,a,b)	1.324	10.655	21.496	0.540			*2095.7*	*2110.2*
	(0.777)	(7.817)	(7.183)	(0.138)				
GOHC-Ga(α,a,b)	0.268	42.613	0.703				2118.9	2129.8
	(0.160)	(21.380)	(0.302)					
KBG-GG(α,β,k,a,b,c)	7.449	1.856	35.679	0.162	0.431	-0.948	2104.0	2125.8
	(0.002)	(0.002)	(0.002)	(0.016)	(0.028)	(0.263)		
BGG(α,β,k,a,b)	32.724	1.970	3.998	0.987	3.836		2115.1	2133.2
	(16.593)	(0.874)	(10.644)	(0.350)	(3.133)			
EGG(α,β,k,a)	33.838	2.818	3.100	0.905			2113.3	2127.9
	(30.242)	(1.503)	(10.773)	(0.969)				
Kw-GG($\alpha,\tau,k,\lambda,\phi$)	33.534	1.429	2.652	2.237	9.210		2114.8	2133.0
	(10.897)	(0.938)	(0.685)	(0.768)	(7.141)			

The values of the AIC and BIC statistics reveal that the EGOHC-Ga model provides the best fits compared to other nine fitted models. Additional information can be obtained more visually from the fitted pdf, cdf and the histogram of the data in Figure 7. In fact, for the current data, the EGOHC-Ga model provides a better fit then the others fitted models.
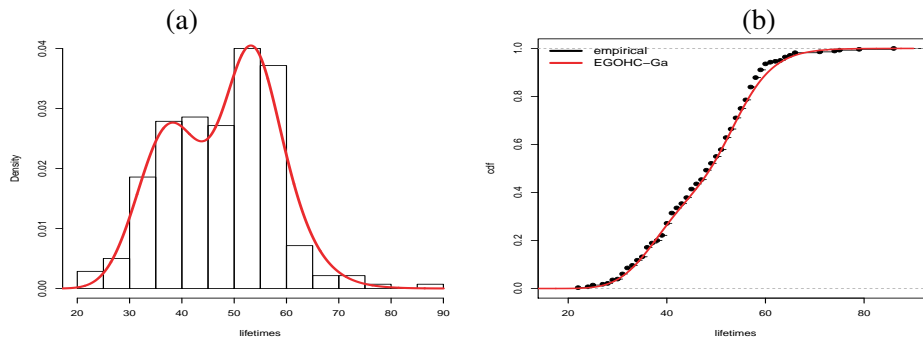
Fig. 7. For actuarial data: (a) Estimated EGOHC-Ga density. (b) Estimated EGOHC-Ga cdf and the empirical cdf.

The empirical scaled TTT transform can be used to identify the shape of the hazard function. The TTT plot for the actuarial data given in Figure 8(a) shows an hrf. On the other hand, Figure 8(b) shows that the estimated hrf of the EGOHC-Ga model is also increasing, indicating that this distribution is appropriate to fit these data.
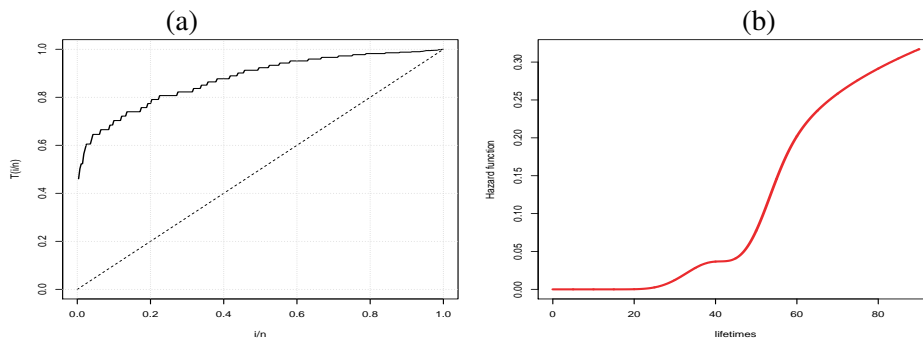

Fig. 8. For the actuarial data: (a) TTT plot; (b) Estimated hrf of the EGOHC-Ga distribution.

6.2. *Data: Ozone level*

The data correspond to daily ozone level measurements (in ppb = ppm$\times$1000) that were collected in New York during May-September, 1973. These data were taken from Nadarajah (2008) and have been provided by the New York State Department of Conservation.

Similar to the previous application, we compare the KBG-GG model and some of its sub-models with the EGOHC-LL model fitted to the current data. The fit of the special case GOCH-LL model, when $\alpha = \beta$, is also reported. Table 5 provides the MLEs (and the corresponding standard errors in parentheses) of the model parameters and the values of the statistics AIC and BIC for the fitted models.

The fitted pdf, cdf given in Figure 9 as well as the TTT plot and estimated hrf presented in Firure 10 indicates that the EGOHC-LL model provides a good t for these data.

Table 4. MLEs of the model parameters for the ozone data, the corresponding SEs (given in parentheses) and the AIC and BIC statistics.

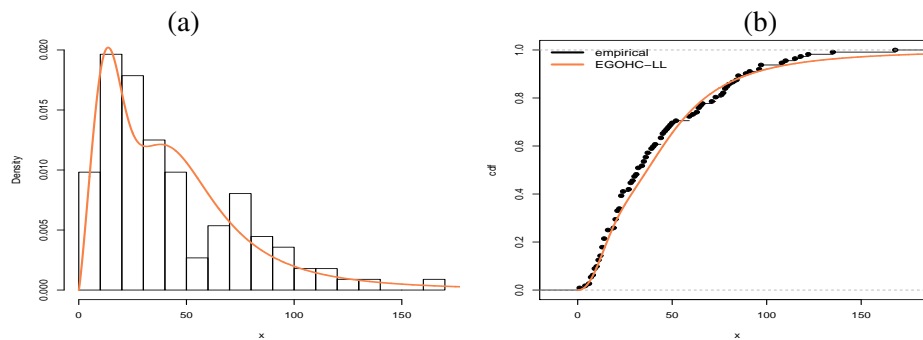| Model	Estimates						AIC	BIC
EGOHC-LL(α,β,a,b)	0.830	6.572	22.995	2.647			*1062.5*	*1073.4*
	(0.278)	(3.598)	(5.259)	(0.472)				
GOHC-LL(α,λ,c)	0.554	51.496	2.126				1066.6	1074.7
	(0.342)	(22.286)	(0.657)					
KBG-GG(α,β,k,a,b,c)	3.040	1.076	20.242	0.080	0.159	-0.215	1067.1	1083.6
	(0.027)	(0.051)	(0.010)	(0.012)	(0.059)	(0.413)		
BGG(α,β,k,a,b)	4.077	1.137	17.593	0.092	0.174		1087.7	1101.5
	(0.183)	(0.082)	(0.266)	(0.010)	(0.089)			
EGG(α,β,k,a)	3.703	0.637	4.959	0.728			1090.2	1101.2
	(2.156)	(0.096)	(1.528)	(0.371)				
Kw-GG($\alpha,\tau,k,\lambda,\phi$)	0.600	0.550	11.200	0.405	0.749		1091.9	1105.7
	(0.161)	(0.022)	(1.665)	(0.110)	(0.463)			
GG(α,β,k)	3.129	0.592	4.344				1088.3	1096.6
	(0.567)	(0.036)	(0.292)					

Fig. 9. For ozone data: (a) Estimated EGOHC-LL density. (b) Estimated EGOHC-LL cdf and the empirical cdf.
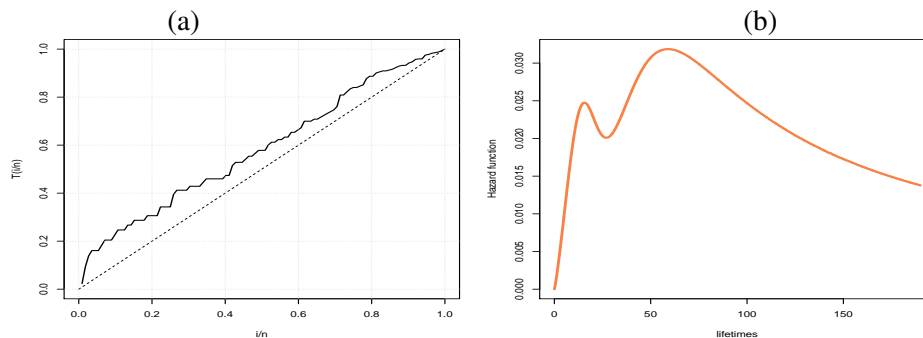

Fig. 10. For the ozone data: (a) TTT plot; (b) Estimated hazard rate function of the EGOHC-LL model.

6.3. *Data: Voltage data*

Lawless (2003) reports an experiment in which specimens of solid epoxy electrical-insulation were studied in an accelerated voltage life test. The sample size is $n = 60$, the percentage of censored

observations was 10% and are considered three levels of voltage 52.5, 55.0 and 57.5. The variables involved in the study are: x_i - failure times for epoxy insulation specimens (in min); c_i - censoring indicator (0 =censoring, 1 =lifetime observed); v_{i1} - voltage (kV). Before to propose the regression model, we present in Figure 11 the box plot for each levels of voltage considering only the uncensored observations ($c = 1$), to check the assumption of the homogeneity of variance. We can conclude that the variability for each voltage level v_{i1} is different, and then a heteroscedastic model is required.
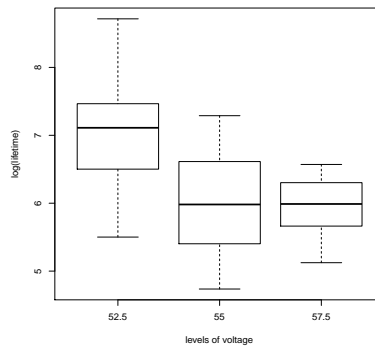

Fig. 11. For the voltage data the box plot for each levels of voltage.

Next, we present results by fitting the heteroscedastic and the location models given by

$$y_i = \tau_{10} + \tau_{11}v_{i1} + \exp(\tau_{20} + \tau_{21}v_{i1})z_i \qquad \text{and} \qquad y_i = \tau_{10} + \tau_{11}v_{i1} + \sigma z_i,$$

respectively, where the random variable Y_i follows the LEGOHC-W distribution given in (5.2). For comparison, we also consider that Y_i follows the log-Weibull (LW) distribution. The MLEs of the model parameters, the asymptotic standard errors of these estimates and the values of the AIC and BIC measures to compare the LEGOHC-W and LW heteroscedastic and location regression models are listed in Table 5.

Table 5. MLEs of the parameters to the voltage data for the LEGOHC-W and L-W heteroscedastic and location regression models, the corresponding SEs (given in parentheses), p-values in [·] and the statistics AIC and BIC.

Model	α	β	τ_{10}	τ_{11}	τ_{20}	τ_{21}	AIC	BIC
Heteroscedastic	104557	90428	-126.880	1.934	6.782	-0.080	*161.3*	*173.9*
LEGOHC-W	(0.001)	(0.001)	(0.015)	(0.049)	(0.469)	(0.010)		
			[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]		
Heteroscedastic	-	-	20.699	-0.250	6.147	-0.115	171.4	179.8
LW	-	-	(3.114)	(0.055)	(3.171)	(0.057)		
			[<0.0001]	[<0.0001]	[0.0573]	[0.0496]		

	α	β	τ_{10}	τ_{11}	σ	AIC	BIC
Location	936670	1534457	-14.702	-0.184	11.742	165.6	176.1
LEGOHC-W	(0.001)	(0.001)	(4.253)	(0.057)	(1.185)		
			[0.001]	[0.002]			
Location	-	-	22.000	-0.274	0.845	173.4	179.7
L-W	-	-	(3.046)	(0.055)	(0.090)		
			[<0.0001]	[<0.0001]			

Based on the figures in Table 5, we conclude that the fitted LEGOHC-W heteroscedastic regression model has the lower AIC and BIC values than the location regression model. We also conduct the formal tests to verify the homogeneity assumption, based on the LR statistic described in Section 5. Based in the LR statistic, we reject the null hypothesis $H_0 : \tau_{21} = 0$, at 5% of significance level, in favor of the LEGOHC-W heteroscedastic regression model. The value of the LR statistic is $w = 6.3$ with the p-value 0.012.

Finally, Figure 12 provides the plots of the empirical and estimated survival function for the heteroscedastic and location regression models. These plots indicate that the heteroscedastic regression model provides a good fit to these data.
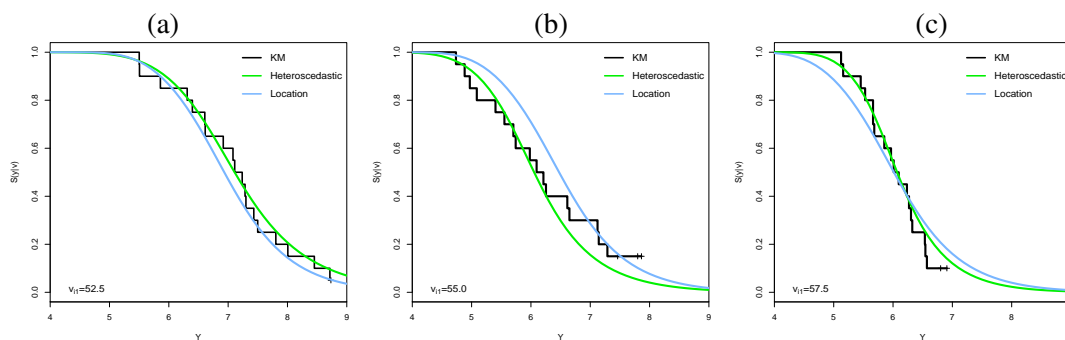

Fig. 12. Estimated LEGOHC-W survival function for the heteroscedastic and location regression model and empirical survival for the voltage data considering the voltage levels: (a) $v_{i1} = 52.5$; (b) $v_{i1} = 55.0$; (c) $v_{i1} = 57.5$.

7. Conclusions

We propose a new family of extended generalized odd Half-Cauchy-G (EGOHC-G) distributions which can include as special cases all classical continuous distributions. For any parent continuous distribution G, we can define the corresponding EGOHC-G distribution with two extra positive parameters. So, the new family extends several common distributions such as the gamma, Weibull, Gumbel and log-logistic distributions. The mathematical properties of the new family such as moments and generating functions are obtained for any EGOHC-G distribution. After, we introduce the called log-extended generalized odd Half-Cauchy-G (LEGOHC-G) distribution. Based on this distribution, we propose a very suitable heteroscedastic LEGOHC-G regression model for modeling lifetime data, which allows us to jointly model the location and dispersion parameters. The new class of regression models can serve as a good alternative for lifetime data analysis and it is much more flexible than the usual heteroscedastic regression model in analyzing lifetime data in many practical situations. The potentiality of the proposed family is illustrated by a simulations study and also using three real data sets. We show that some models in the new family can produce better fits than those corresponding generated models from the log-logistic, Weibull, gamma and GOHC families using the same baseline distribution.

Acknowledgment

References

[1] Alexander, C., Cordeiro, G.M., Ortega, E.M.M. and Sarabia. J.M. (2012). Generalized beta–generated distributions, *Computational Statistics and Data Analysis*, **56**, 1880-1897.

[2] Balakrishnan, N., Leiva, V., Sanhueza, A. and Cabrera, E. (2009). Mixture inverse Gaussian distributions and its transformations, moments and applications. *Statistics*, **43**, 91-104.

[3] Braga, A.S., Cordeiro, G.M., Ortega, E.M.M. and da Cruz, J.N. (2016). The odd log-logistic normal distribution: Theory and applications in analysis of experiments. *Journal of Statistical Theory and Practice*, **10**, 311-335.

[4] Cordeiro, G.M., Alizadeh, M., Ramires, T.G. and Ortega, E.M.M. (2017). The Generalized Odd Half-Cauchy Family of Distributions: Properties and Applications. *Communications in Statistics-Theory and Methods*, **46**, 5685-5705.

[5] Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, **81**, 883-898.

[6] Cordeiro, G.M., Ortega, E.M.M. and Silva, G.O. (2011). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of statistical computation and simulation*, **81**, 827-842.

[7] Cordeiro, G.M., Castellares, F., Montenegro, L.C. and de Castro, M. (2013). The beta generalized gamma distribution. *Statistics*, **47**, 888-900.

[8] Cordeiro, G.M., Pescim, R.R., Demétrio, C.G.B and Ortega, E.M.M. (2014c). The Kummer beta generalized gamma distribution. *Journal of Data Science*, **12**, 661-698.

[9] Cordeiro, G.M., Ortega, E.M.M. and Ramires, T.G. (2015). A new generalized Weibull family of distribuitons: mathematical properties and applications. *Journal of Statistical Distributions and Applications*, **2**, 2-25.

[10] da Cruz, J.N., Ortega, E.M.M. and Cordeiro, G.M. (2016). The log-odd log-logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis. *Journal of Statistical Computation and Simulation*, **86**, 1516-1538.

[11] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications, *Communication in Statististics– Theory Methods*, **31**, 497–512.

[12] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley: New York.

[13] Nadarajah, S. and Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*, **92**, 97-111.

[14] Nadarajah, S. (2008). A truncated inverted beta distribution with application to air pollution data. *Stochastic Environmental Research and Risk Assessment*, **22**, 285-289.

[15] Nadarajah, S., Cordeiro, G.M. and Ortega, E.M.M. (2014). The Zografos-Balakrishnan-G family of distributions: Mathematical properties and applications. *Communications in Statistics - Theory and Methods*, **44**, 186-215.

[16] Ortega, E.M.M., Cordeiro, G.M., Hashimoto, E.M. and Cooray, K. (2014). A log-linear regression model for the odd Weibull distribution with censored data. *Journal of Applied Statistics*, **41**, 1859-1880.

[17] Ortega, E.M.M., Lemonte, A.J., Silva, G.O. and Cordeiro, GM. (2015). New flexible models generated by gamma random variables for lifetime modeling. *Journal of Applied Statistics*, **42**, 2159-2179.

[18] Pascoa, M.A., Ortega, E.M.M. and Cordeiro, G.M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *tatistical methodology*, **8**, 411-433.

[19] Pescim, R.R., Ortega, E.M.M., Cordeiro, G.M., Demetrio, C.G.B. and Hamedani, G.G. (2013). The log-beta generalized half-normal regression model. *Journal of Statistical Theory and Applications*, **12**, 330-347.

[20] Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. (1986). *Integrals and Series*, volumes 1, 2 and 3. Gordon and Breach Science Publishers, Amsterdam.

[21] Ramires, G.R., Ortega, E.M.M., Cordeiro, G.M. and Hens, N. (2016). A bimodal flexible distribution for lifetime data. *Journal of Statistical Computation and Simulation*, **86**, 2450-2470.

[22] Zografos, K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, **6**, 344-362.