

Generalized Odd Power Cauchy Family and Its Associated Heteroscedastic Regression Model

Emrah Altun^{1,*}, Morad Alizadeh², Thiago G. Ramires³, Edwin M.M. Ortega⁴

¹*Department of Mathematics, Bartin University, Bartin 74100, Turkey*

²*Department of Statistics, Persian Gulf University, Bushehr, 75169, Iran*

³*Department of Mathematic, Federal University of Technology - Paraná, Brazil*

⁴*Department of Exact Science, ESALQ-USP, University of São Paulo, Brazil*

Abstract This study introduces a generalization of the odd power Cauchy family by adding one more shape parameter to gain more flexibility modeling the complex data structures. The linear representations for the density, moments, quantile, and generating functions are derived. The model parameters are estimated employing the maximum likelihood estimation method. The Monte Carlo simulations are performed under different parameter settings and sample sizes for the proposed models. In addition, we introduce a new heteroscedastic regression model based on the special member of the proposed family. Three data sets are analyzed with competitive and proposed models.

Keywords Generated family; Heteroscedastic regression model; Maximum likelihood; Moment; Power Cauchy.

AMS 2010 subject classifications 60E05, 62N02

DOI:10.19139/soic-2310-5070-765

1. Introduction

Rooks *et al.* (2010) proposed the power Cauchy (PC) distribution. The probability density function (pdf) and cumulative distribution function (cdf) of the PC distribution are

$$r(t) = \frac{2}{\pi} \frac{\alpha t^{\alpha-1}}{\sigma^{2\alpha} + t^{2\alpha}}, \quad (1)$$

and

$$R(t) = \frac{2}{\pi} \arctan \left[\left(\frac{t}{\sigma} \right)^\alpha \right], \quad (2)$$

where $\alpha > 0$ is a shape parameter and $\sigma > 0$ is a scale parameter. Using the PC distribution as a generator distribution, we propose a new family of distributions named as *generalized odd power Cauchy-G* (GOPC-G). The cdf of the GOPC-G family is given by

$$F_{GOPC-G}(x; \alpha, \beta, \kappa) = \frac{2}{\pi} \arctan \left\{ \left[\frac{G(x; \kappa)^\alpha}{1 - G(x; \kappa)^\alpha} \right]^\beta \right\}, \quad (3)$$

*Correspondence to: Emrah Altun (Email: emrahaltun@bartin.edu.tr). Department of Mathematics, Bartin University, Bartin 74100, Turkey.

where $G(x; \kappa)$ is the cdf of the baseline distribution and κ represents the parameter vector of the baseline distribution. The parameters $\alpha > 0$ and $\beta > 0$ are the shape parameters. The pdf to (3) is

$$f_{\text{GOPC-G}}(x; \alpha, \beta, \kappa) = \frac{2\alpha\beta g(x; \kappa) G(x; \kappa)^{\alpha\beta-1} [1 - G(x; \kappa)^\alpha]^{\beta-1}}{\pi \left\{ G(x; \kappa)^{2\alpha\beta} + [1 - G(x; \kappa)^\alpha]^{2\beta} \right\}}, \quad (4)$$

where $g(x; \kappa)$ is the pdf of the baseline distribution. The GOPC-G family contains some of the G-class distributions as its submodels. For instance, when $\alpha = 1$, the GOPC-G family reduces to the odd power cauchy (OPC-G) family (Alizadeh *et al.*, 2018). When $\beta = 1$, the GOPC-G family reduces to generalized odd half-cauchy (GOHC-G) family (Cordeiro *et al.*, 2017a). Additionally, when $\alpha = \beta = 1$, we have odd half-Cauchy (OHC-G) family. Henceforward, the density in (4) is denoted as $X \sim \text{GOPC-G}(\alpha, \beta, \kappa)$. The hazard rate function (hrf) of X is given by

$$h_{\text{GOPC-G}}(x; \alpha, \beta, \kappa) = \frac{2\alpha\beta g(x; \kappa) G(x; \kappa)^{\alpha\beta-1} [1 - G(x; \kappa)^\alpha]^{\beta-1}}{\left\{ \pi - 2 \arctan \left\{ \left[\frac{G(x; \kappa)^\alpha}{1 - G(x; \kappa)^\alpha} \right]^\beta \right\} \right\} \left\{ G(x; \kappa)^{2\alpha\beta} + [1 - G(x; \kappa)^\alpha]^{2\beta} \right\}}. \quad (5)$$

Now, some possible relations of the GOPC-G family with other families are given.

1. If $Y = G(X; \kappa)^\alpha \implies F_Y(y) = \frac{2}{\pi} \arctan \left[\left(\frac{y}{1-y} \right)^\beta \right]$ for $0 < y < 1$;
2. If $Y = \frac{G(X; \kappa)^\alpha}{1 - G(X; \kappa)^\alpha} \implies Y \sim PC(\beta)$.

The other main purpose is to provide a new use of the *generalized odd power Cauchy Weibull* (GOPC-W) distribution in a framework of regression model, where both location and dispersion parameters of a regression model based on the logarithm of the GOPC-W random variable vary across observations through regression structures. The log-transform of the random variable having the GOPC-W density is used to construct a new regression model. The proposed regression model is appropriate for both modeling the censored and uncensored response variable. This approach is very common in constructing regression models in survival analysis, e.g., Liu (2014) studied survival models on unobserved heterogeneity and their applications, Ortega *et al.* (2017) proposed the heteroscedastic log-exponentiated Weibull regression model and Cordeiro *et al.* (2017b) defined the extended generalized odd half-Cauchy family and applications in heteroscedastic regression analysis.

The other parts of the presented study are organized as follows. Some of the special cases of the GOPC-G family are given in Section 2. In Section 3, the statistical properties of the GOPC-G family are discussed in detail. In Section 4, the parameter estimation issue of the GOPC-G family is addressed based on the maximum likelihood method. The simulation study is given to evaluate the performance of the estimation method for finite sample sizes. The heteroscedastic regression model is defined in Section 5. In Section 6, three data sets are analyzed to prove the importance of the proposed models in real life problems. Section 7 contains conclusions of the presented study.

2. Special Models

Three special members of the proposed family are provided.

2.1. Generalized odd power Cauchy-normal (GOPC-N)

The GOPC-N distribution opens new opportunities to generate uni or bimodal and skew-symmetric normal distributions. Its density is given by

$$f_{\text{GOPC-N}}(x; \alpha, \beta, \mu, \sigma) = \frac{2\alpha\beta\phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right)^{\alpha\beta-1} [1 - \Phi\left(\frac{x-\mu}{\sigma}\right)^\alpha]^{\beta-1}}{\sigma\pi \left\{ \Phi\left(\frac{x-\mu}{\sigma}\right)^{2\alpha\beta} + [1 - \Phi\left(\frac{x-\mu}{\sigma}\right)^\alpha]^{2\beta} \right\}},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, $\alpha > 0$ and $\beta > 0$ are shape parameters, $\mu \in \mathbb{R}$ is a location parameter and σ is scale parameter.

2.2. Generalized odd power Cauchy-log-logistic (GOPC-LL)

Let X follows the log-logistic distribution with pdf and cdf ($x > 0$),

$$g(x) = \frac{\gamma x^{\gamma-1}}{\lambda^\gamma [1 + (\frac{x}{\lambda})^\gamma]^2} \quad \text{and} \quad G(x) = \frac{1}{1 + (\frac{x}{\lambda})^{-\gamma}},$$

respectively, where $\lambda > 0$ and $\gamma > 0$. Inserting these expressions in (4), we obtain the GOPC-LL density function given by

$$f_{GOPC-LL}(x; \alpha, \beta, \lambda, \gamma) = \frac{2\alpha\beta \left\{ \frac{(\gamma/\lambda)(x/\lambda)^{\gamma-1}}{[1+(x/\lambda)^\gamma]^2} \right\} \left[\frac{1}{1+(x/\lambda)^{-\gamma}} \right]^{\alpha\beta-1} \left\{ 1 - \left[\frac{1}{1+(x/\lambda)^{-\gamma}} \right]^\alpha \right\}^{\beta-1}}{\pi \left\{ \left[\frac{1}{1+(x/\lambda)^{-\gamma}} \right]^{2\alpha\beta} + \left\{ 1 - \left[\frac{1}{1+(x/\lambda)^{-\gamma}} \right]^\alpha \right\}^{2\beta} \right\}}.$$

2.3. Generalized odd power Cauchy-Weibull (GOPC-W)

The cdf of the Weibull distribution is $G(x) = 1 - e^w$, where $w = -(\frac{x}{b})^a$ and $a, b > 0$. The GOPC-W pdf can be expressed from (4) as

$$f_{GOPC-W}(x; \alpha, \beta, a, b) = \frac{2\alpha\beta a x^{a-1} e^w (1 - e^w)^{\alpha\beta-1} [1 - (1 - e^w)^\alpha]^{\beta-1}}{b^a \pi \{(1 - e^w)^{2\alpha\beta} + [1 - (1 - e^w)^\alpha]^{2\beta}\}}.$$

3. Main Properties

In this section, we study some of the statistical properties of the GOPC-G family.

3.1. Quantile Function

The quantile function (qf) is a solution of $F(x) = u$ where $u \sim U(0, 1)$ and it is denoted as $Q_G(u)$. The qf of the GOPC-G family is given by

$$X = Q(u) = Q_G \left(\left(\frac{[\tan(\frac{\pi u}{2})]^{1/\beta}}{1 + [\tan(\frac{\pi u}{2})]^{1/\beta}} \right)^{1/\alpha} \right). \quad (6)$$

The qf in (6) is very useful to generate random variables from the GOPC-G family for a given baseline distribution. Additionally, Bowley's skewness (Bowley, 1901) and Moors's kurtosis (Moors, 1986) are calculated based on the quantiles. Therefore, the qf is also useful to investigate the shape of the GOPC-G family. The required formulas of these measures are given by

$$\text{Skewness} = \frac{Q(1/4) + Q(3/4) - 2Q(1/2)}{Q(3/4) - Q(1/4)},$$

and

$$\text{Kurtosis} = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)}.$$

These measures are less sensitivity to outliers. Moreover, they can be used as an alternative measures when the moments of the distributions do not have a closed form. The results of the Bowley's skewness and Moors's kurtosis of the GOPC-W distribution are summarized in Table 1. These results show the effects of the parameters α and β on the skewness and kurtosis measures.

Table 1. The skewness and kurtosis value of the GOPC-W distribution for $a = 2$ and $b = 2$.

Parameters		Skewness	Kurtosis
α	β		
0.1	0.1	0.976	0.929
	0.5	0.870	1.897
	1	0.669	2.014
	1.5	0.512	1.819
	2	0.407	1.673
0.5	0.1	0.624	0.742
	0.5	0.218	1.089
	1	0.111	1.265
	1.5	0.074	1.321
	2	0.056	1.343
2	0.1	0.329	0.769
	0.5	0.091	1.196
	1	0.047	1.314
	1.5	0.032	1.346
	2	0.024	1.358

3.2. Linear representation

By using Taylor and generalized binomial expansions, the pdf (4) of X can be expressed as

$$f(x) = \frac{2\alpha\beta g(x)}{\pi} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} \frac{G(x)^{\alpha(\beta+i)-1}}{G(x)^{2\alpha\beta} + [1 - G(x)^\alpha]^{2\beta}}.$$

Using again the generalized binomial expansion and power series for the ratio of two power series, we have

$$\frac{G(x)^{\alpha(\beta+i)-1}}{G(x)^{2\alpha\beta} + [1 - G(x)^\alpha]^{2\beta}} = \frac{\sum_{k=0}^{\infty} \rho_k^{(i)} G(x)^k}{\sum_{k=0}^{\infty} \omega_k G(x)^k} = \sum_{k=0}^{\infty} \nu_k^{(i)} G(x)^k,$$

where

$$\rho_k^{(i)} = s_k(\alpha(\beta+i) - 1), \quad \omega_k = s_k(2\alpha\beta) + \sum_{j=0}^{\infty} (-1)^j \binom{2\beta}{j} s_k(j\alpha),$$

$$s_k = s_k(\alpha) = \sum_{i=k}^{\infty} (-1)^{i+k} \binom{\alpha}{i} \binom{i}{k}, \quad \nu_k^{(i)} = \nu_k(\alpha, \beta, i) = \frac{1}{\omega_0} \left(\rho_k^{(i)} - \frac{1}{\omega_0} \sum_{r=1}^k \omega_r \nu_{k-r}^{(i)} \right), \text{ for } r \geq 1,$$

and $\nu_0^{(i)} = \rho_0^{(i)}/\omega_0$. Then, we can write

$$f(x) = \sum_{k=0}^{\infty} b_k h_{k+1}(x), \tag{7}$$

where

$$b_k = \frac{2\alpha\beta}{\pi(k+1)} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} \nu_k(\alpha, \beta, i),$$

and $h_a(x) = a g(x) G(x)^{a-1}$ denotes the exponentiated-G (“exp-G”) density function. Several studies, such as Mudholkar *et al.* (1995) and Nadarajah and Kotz (2006), have been studied properties of the exp-G densities. Equation (7) confirms that the GOPC-G density function can be expressed as a linear combination of the exp-G densities.

3.3. Moments

Let Y_k be a random variable having exp-G density $h_{k+1}(x)$. The n th moment of X can be determined from (7) as

$$E(X^n) = \sum_{k=0}^{\infty} b_k E(Y_k^n) = \sum_{k=0}^{\infty} (k+1) b_k \psi_{n,k}, \quad (8)$$

where $\psi_{n,k} = \int_{-\infty}^{\infty} x^n G(x)^k g(x) dx = \int_0^1 Q_G(u)^n u^k du$. Following the work of Nadarajah and Kotz (2006), $E(X^n)$ can be obtained. Besides, $\psi(n, k)$ was obtained for several distributions by Cordeiro and Nadarajah (2011).

3.4. Generating function

Let $M_X(t) = E(e^{tX})$ be the moment generating function (mgf) of X . We obtain from (7)

$$M_X(t) = \sum_{k=0}^{\infty} b_k M_k(t) = \sum_{i=0}^{\infty} (k+1) b_k \rho_k(t), \quad (9)$$

where $\rho_k(t) = \int_{-\infty}^{\infty} e^{tx} G(x)^k g(x) dx = \int_0^1 \exp[t Q_G(u)] u^k du$ and $M_k(t)$ is the mgf of Y_{k+1} . So, $M_X(t)$ can be determined from the exp-G generating function. It is possible to derive the mgfs of the special members of the GOPC-G family from (9).

Theorem 1

If $G(x)$ has a mgf, then, $F(x)$ has a mgf.

Proof

Let $m = \inf\{x | G(x) \geq 0.5\}$, then

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{\infty} e^{tx} \times \frac{2\alpha\beta g(x) G(x)^{\alpha\beta-1} [1 - G(x)^\alpha]^{\beta-1}}{\pi [G(x)^{2\alpha\beta} + (1 - G(x)^\alpha)^{2\beta}]} dx \\ &\leq \int_{-\infty}^{\infty} e^{tx} \times \frac{2\alpha\beta g(x)}{\pi [G(x)^{2\alpha\beta} + (1 - G(x)^\alpha)^{2\beta}]} dx \\ &= \int_{-\infty}^m e^{tx} \times \frac{2\alpha\beta g(x)}{\pi [G(x)^{2\alpha\beta} + (1 - G(x)^\alpha)^{2\beta}]} dx + \int_m^{\infty} e^{tx} \times \frac{2\alpha\beta g(x)}{\pi [G(x)^{2\alpha\beta} + (1 - G(x)^\alpha)^{2\beta}]} dx. \end{aligned}$$

The first integral in the last line is finite and the second integral is no greater than

$$\int_m^{\infty} e^{tx} \frac{2\alpha\beta g(x)}{\pi G(x)^{2\alpha\beta}} dx.$$

For $x > m$, we have $G(x) \geq 0.5$, so that

$$\int_m^{\infty} e^{tx} \frac{2\alpha\beta g(x)}{\pi G(x)^{2\alpha\beta}} dx < \frac{\alpha\beta 2^{2\alpha\beta+1}}{\pi} \int_m^{\infty} e^{tx} g(x) dx < \infty.$$

Then, $M_X(t) < \infty$. □

Corollary 1

Every distribution in the GOPC-G family has exactly the same number of moments of $G(x)$.

3.5. Mean deviations

The mean deviations about the mean, δ_1 , and median, δ_2 , are, respectively

$$\delta_1(X) = 2\mu'_1 F(\mu'_1) - 2m_1(\mu'_1) \quad \text{and} \quad \delta_2(X) = \mu'_1 - 2m_1(M), \quad (10)$$

where $\mu'_1 = E(X)$ and $M = \text{Median}(X)$ defined by $M = Q(0.5)$. The quantity $F(\mu'_1)$ can be calculated from (3) and $m_1(z)$ is the first incomplete moment obtained from (11) for $n = 1$.

$$m_n(y) = \int_0^y x^n f(x) dx = \sum_{k=0}^{\infty} (k+1) b_k \int_0^{G(y)} Q_G(u)^n u^k du. \quad (11)$$

4. Estimation and Simulation study

The maximum likelihood estimation method is used to get the maximum likelihood estimates (MLEs) of the parameters of the GOPC-G density. Let x_1, \dots, x_n be a sample from the GOPC-G density and $\Theta = (\alpha, \beta, \boldsymbol{\kappa}^\top)^\top$ is the parameter vector. The log-likelihood function of the GOPC-G density is

$$\begin{aligned} \ell_n(\Theta) = & n \log\left(\frac{2\alpha\beta}{\pi}\right) + \sum_{i=1}^n \log g(x_i; \boldsymbol{\kappa}) + (\alpha\beta - 1) \sum_{i=1}^n \log G(x_i; \boldsymbol{\kappa}) \\ & + (\beta - 1) \sum_{i=1}^n \log [1 - G(x_i; \boldsymbol{\kappa})^\alpha] - \sum_{i=1}^n \log \left\{ G(x_i; \boldsymbol{\kappa})^{2\alpha\beta} + [1 - G(x_i; \boldsymbol{\kappa})^\alpha]^{2\beta} \right\}. \end{aligned} \quad (12)$$

The direct maximization of (12) gives the MLE of Θ . This can be done by using statistical or mathematical software such as R, MATLAB or SAS. Here, we use our choice from R software and its function called as `optim` which is commonly used for optimization purpose for a given initial vector.

Simulation Study

It is important to investigate the asymptotic behaviours of the MLEs for the GOPC-G density under a given baseline distribution. For this purpose, we implement a simulation study to see the performance of the MLEs for a finite sample size. The simulation results are evaluated based on the following metrics: bias, mean square error (MSE), estimated average length (AL) and coverage probability (CP). The required formulas for the computation of these measures are given below.

$$\widehat{Bias}_\epsilon(n) = \frac{1}{N} \sum_{i=1}^N (\hat{\epsilon}_i - \epsilon) \quad \text{and} \quad \widehat{MSE}_\epsilon(n) = \frac{1}{N} \sum_{i=1}^N (\hat{\epsilon}_i - \epsilon)^2,$$

$$CP_\epsilon(n) = \frac{1}{N} \sum_{i=1}^N I(\hat{\epsilon}_i - 1.95996s_{\hat{\epsilon}_i}, \hat{\epsilon}_i + 1.95996s_{\hat{\epsilon}_i}) \quad \text{and} \quad AL_\epsilon(n) = \frac{3.919928}{N} \sum_{i=1}^N s_{\hat{\epsilon}_i}.$$

where $\epsilon = \alpha, \beta, a, b$ and $(s_{\hat{\alpha}_i}, s_{\hat{\beta}_i}, s_{\hat{a}_i}, s_{\hat{b}_i})$ represent the standard errors of the parameters. We choose the Weibull distribution as a baseline distribution of the GOPC-G distribution. So, the GOPC-W distribution is used. The simulation is replicated $N = 10,000$ times for each sample sizes. The sample sizes are increased by 5 and started from $n = 50$ and ended by $n = 1,000$. The parameters of the GOPC-W distribution is selected as $\alpha = 0.5, \beta = 0.5, a = 2, b = 2$. Figure 1 displays the simulation results. The following results are obtained: (i) the biases and MSEs decrease when the sample size increases, (ii) CPS are near the desired value 0.95 and ALs are decreasing function of the sample size, as expected.

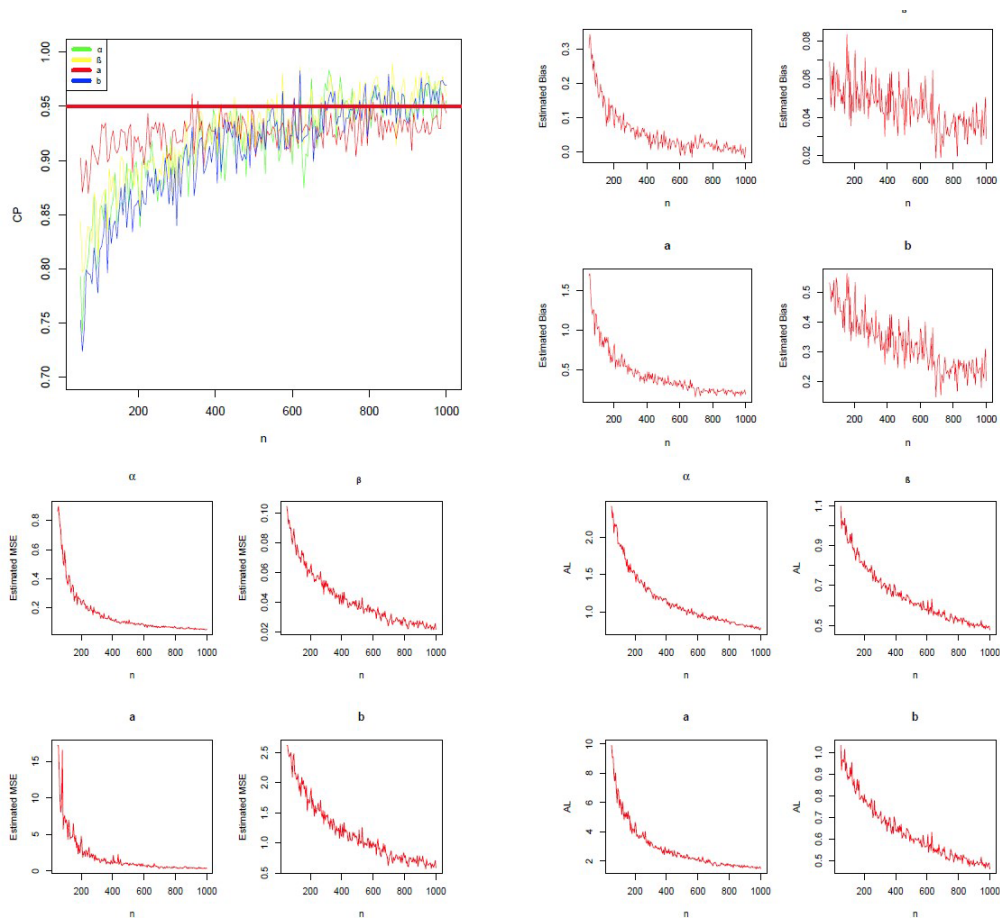


Figure 1. Simulation results of the GOPC-W distribution.

5. The Heteroscedastic LGOPC-W Regression Model

Here, we introduce the heteroscedastic regression model based on the GOPC-W density, given in Section 2.3. Applying $Y = \log(X)$ transformation and $a = 1/\sigma, b = e^\mu$ re-parametrizations on the GOPC-W density, we have

$$f(y) = \frac{\frac{2\alpha\beta}{\sigma} \exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right] \{1 - \exp[-\exp\left(\frac{y-\mu}{\sigma}\right)]\}^{\alpha\beta-1} \left[1 - \{1 - \exp[-\exp\left(\frac{y-\mu}{\sigma}\right)]\}^\alpha\right]^{\beta-1}}{\pi \left\{ \{1 - \exp[-\exp\left(\frac{y-\mu}{\sigma}\right)]\}^{2\alpha\beta} + \left[1 - \{1 - \exp[-\exp\left(\frac{y-\mu}{\sigma}\right)]\}^\alpha\right]^{2\beta} \right\}}, \quad (13)$$

where $\mu \in \mathfrak{R}$ is the location of Y and σ is the scale parameter. The parameters $\alpha > 0$ and $\beta > 0$ controls the shape of the density. The density in (13) is called as log-GOPC-W and denoted as $Y \sim \text{LGOPC-W}(\alpha, \beta, \mu, \sigma)$. The survival function of (13) is

$$S(y) = 1 - \frac{2}{\pi} \arctan \left\{ \left[\frac{\{1 - \exp[-\exp\left(\frac{y-\mu}{\sigma}\right)]\}^\alpha}{1 - \{1 - \exp[-\exp\left(\frac{y-\mu}{\sigma}\right)]\}^\alpha} \right]^\beta \right\} \quad (14)$$

Using the transformation $Z = (Y - \mu)/\sigma$, the standardized log-GOPC-W density is

$$f(z) = \frac{2\alpha\beta \exp[(z) - \exp(z)] \{1 - \exp[-\exp(z)]\}^{\alpha\beta-1} [1 - \{1 - \exp[-\exp(z)]\}^\alpha]^{\beta-1}}{\pi \left\{ \{1 - \exp[-\exp(z)]\}^{2\alpha\beta} + [1 - \{1 - \exp[-\exp(z)]\}^\alpha]^{2\beta} \right\}}. \quad (15)$$

In many practical applications, the lifetimes are affected by explanatory variables. Parametric models to estimate univariate survival functions and for censored data regression problems are widely used. Standard regression models require the assumption of homogeneity of error variances. Therefore, it is necessary to check the possible heteroscedasticity.

Let y_i be a response variable following the LGOPC-W density and $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip_1})$, $\mathbf{w}_i^\top = (w_{i1}, \dots, w_{ip_2})$ are independent variable vectors for the location and scale parameters of the LGOPC-W regression model. The regression model is given by

$$y_i = \mu_i + \sigma_i z_i, \quad i = 1, \dots, n, \quad (16)$$

where z_i is in (15). The independent variables are linked to the location parameter μ_i by identity link function, given as $\mu_i = \mathbf{x}_i^\top \boldsymbol{\eta}$. The heteroscedasticity is modeled by means of the scale parameter of the response variable and log-link function is used to link the independent variables to the scale parameter, given as $\sigma_i = \exp(\mathbf{w}_i^\top \boldsymbol{\tau})$ where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{p_1})^\top$ and $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{p_2})^\top$ are the regression parameter vectors. Note that when $\sigma_i = \sigma$, the regression model (16) reduce to the homoscedastic regression model, where only μ is modeled using explanatory variables.

The general formula for the log-likelihood function of the regression model for the parameter vector $\Theta = (\alpha, \beta, \boldsymbol{\eta}^\top, \boldsymbol{\tau}^\top)^\top$ from model (16) is $l(\Theta) = \sum_{i \in F} l_i(\Theta) + \sum_{i \in C} l_i^{(c)}(\Theta)$, where $l_i(\Theta) = \log[f(y_i)]$, $l_i^{(c)}(\Theta) = \log[S(y_i)]$, $f(y_i)$ is the density (13) and $S(y_i)$ is the survival function (14) of Y_i . The response variable is defined as $y_i = \min\{\log(x_i), \log(c_i)\}$ where x_i is the observed lifetime and c_i is the censoring time. Under these specifications, the log-likelihood function of the LGOPC-W regression model is

$$\begin{aligned} \ell(\Theta) &= r \log[2\alpha\beta] + \sum_{i \in F} (z_i - u_i) - \sum_{i \in F} (\mathbf{w}_i^\top \boldsymbol{\tau}) + (\alpha\beta - 1) \sum_{i \in F} \log\{1 - \exp[-u_i]\} + (\beta - 1) \\ &\times \sum_{i \in F} \log[1 - \{1 - \exp[-u_i]\}^\alpha] - \sum_{i \in F} \log \left[\pi \left\{ \{1 - \exp[-u_i]\}^{2\alpha\beta} + [1 - \{1 - \exp[-u_i]\}^\alpha]^{2\beta} \right\} \right] \\ &+ \sum_{i \in C} \log \left[1 - \frac{2}{\pi} \arctan \left\{ \left[\frac{\{1 - \exp[-u_i]\}^\alpha}{1 - \{1 - \exp[-u_i]\}^\alpha} \right]^\beta \right\} \right], \end{aligned} \quad (17)$$

where $u_i = \exp(z_i)$, $z_i = (y_i - \mu_i)/\sigma_i$, and r is the number of uncensored observations. The unknown parameter vector Θ is obtained by maximizing the equation in (17) by using the optim function of R software. Further, we can use the LR statistic for testing if the dispersion is constant for different ranges/levels of the explanatory variables (similar to the assumption of homogeneity of variance).

5.1. Simulation study

Here, the performance of the MLEs of the parameters in the heteroscedastic LGOPC-W regression model is discussed by means of a simulation study. We generate $N = 1,000$ samples of sizes $n = 50, 250, 500$ and 1000 from model (16) considering the structures $\mu = \eta_0 + \eta_1 x_1$ and $\sigma = \exp(\tau_0 + \tau_1 x_1)$, where x_1 is generated from a binomial $(n, 0.5)$ distribution. The values of the parameters are taken as: $\alpha = 2, \beta = 2, \eta_0 = 2, \eta_1 = 2, \tau_0 = 2, \tau_1 = 0.5$, and the response variable Y is generated using the inverse transform method. The simulation results are given in Table 2. The results are interpreted based on the biases, average of estimates (AEs) and MSEs. As expected, the biases and MSEs decrease as function of sample size. Also, the AEs are near the true values of the parameters.

Table 2. The AEs, biases and MSEs based on 1,000 simulations for the heteroscedastic LGOPC-W regression model with parameters: $\alpha = 2$, $\beta = 2$, $\eta_0 = 2$, $\eta_1 = 2$, $\tau_0 = 2$ and $\tau_1 = 0.5$ for $n = 50$, $n = 250$, $n = 500$, $n = 1000$

Sample size	Parameters	α	β	η_0	η_1	τ_0	τ_1
n=50	AE	2.37122	2.37474	1.71614	1.85354	1.76171	0.49527
	Bias	0.37122	0.37474	-0.28386	-0.14646	-0.23829	-0.00473
	MSE	1.29224	6.79284	6.77942	3.69430	0.72357	0.07655
n=250	AE	2.09860	2.19632	1.90296	1.98140	1.96501	0.49707
	Bias	0.09860	0.19632	-0.09704	-0.01860	-0.03499	-0.00293
	MSE	0.44065	1.65357	3.39365	1.28981	0.20506	0.01310
n=500	AE	2.08187	2.17488	1.84429	1.89706	2.01897	0.50047
	Bias	0.08187	0.17488	-0.15571	-0.10295	0.01897	0.00047
	MSE	0.24524	0.79209	1.93738	0.80711	0.11012	0.00633
n=1000	AE	1.99703	2.09559	2.09914	2.06646	1.99788	0.49819
	Bias	-0.00297	0.09559	0.09914	0.06646	-0.00212	-0.00181
	MSE	0.11694	0.42081	0.87750	0.43457	0.06011	0.00281

6. Applications

Three data sets are analyzed to prove the importance of the GOPC-G family in real data modeling. The GOPC-G family is compared with the below families.

1. Odd log-logistic-G (OLL-G) by Gleaton and Lynch (2006)
2. Exponentiated half-logistic-G (EHL-G) by Cordeiro *et al.* (2014)
3. Exponentiated generalized-G (EG-G) by Cordeiro *et al.* (2013)
4. Beta-G (B-G) by Eugene *et al.* (2002)
5. Generalized odd half-Cauchy-G (GOHC-G) by Cordeiro *et al.* (2017a)
6. Extended generalized half-Cauchy-G (EGOHC-G) by Cordeiro *et al.* (2017b)

To decide the best model, we use the model selection criteria such as minimized negative log-likelihood ($-\hat{\ell}$) values, Akaike Information Criterion (AIC), Cramer von Mises (W^*) and Anderson Darling (A^*) statistics (see, Chen and Balakrishnan, 1995). The model having the smallest values of these statistics is the best for the data used.

6.1. Data: Strengths of glass fibers

The first data set is about the breaking stress of carbon fibres (in Gba). It can be downloaded from the R package, `gamlls.data`. Recently, Cordeiro *et al.* (2017a) used these data and compared the performance of GOHC-LL distribution with more than ten competitive models. The results of the first application are listed in Table 3. As seen from these results, the GOPC-LL distribution has the lowest values of the model selection statistics. So, GOPC-LL is the best choice for the data used.

Table 3. MLEs with their SEs of the fitted models and goodness-of-fit statistics for strengths of glass fibers data set

Models	α	β	λ	γ	$-\hat{\ell}$	AIC	A^*	W^*
LL			0.1260 0.0139	0.4220 0.0267	22.7890	49.5790	2.7490	0.4970
OLL-LL	1.3950 13.1530		0.1760 1.6590	0.4220 0.0260	22.7890	51.5790	2.7480	0.4960
EHL-LL	0.1510 0.0238	0.6108 0.0227	0.0332 0.0025	0.5469 0.0222	11.3213	30.6425	0.5356	0.0968
EG-LL	0.4234 0.0142	0.1785 0.0401	0.0352 0.0065	0.5339 0.0138	11.7965	31.5929	0.6177	0.1088
B-LL	0.0578 0.0447	0.1693 0.1487	0.0129 0.0099	0.5291 0.0291	10.9129	29.8250	0.4744	0.0839
GOHC-LL	0.2004 0.0689		0.0503 0.0136	0.6145 0.0278	11.3606	28.7211	0.5236	0.0943
EGOCH-LL	0.2125 0.0657	0.8327 0.7485	0.0545 0.0105	0.5137 0.0644	10.0270	28.0539	0.2990	0.0511
OPC-LL	1.0460 9.0006		0.1510 1.2995	0.4371 0.0242	20.6521	47.3041	2.2523	0.4058
GOPC-LL	0.2867 0.0822	0.2470 0.1665	0.0173 0.0106	0.4840 0.0301	9.8323	27.6647	0.2595	0.0424

The LR test is used to compare the GOPC-LL distribution with its sub-models. The null hypothesis $H_0 : \beta = 1$ is tested against the hypothesis $H_1 : \beta \neq 1$ for the comparison purpose of the GOPC-LL and GOHC-LL models. Similarly, testing the hypothesis $H_0 : \alpha = 1$ against the $H_0 : \alpha \neq 1$ is equivalent to comparison of the GOPC-LL and OPC-LL models. The results are listed in Table 4 which reveals that the GOPC-LL distribution provides better fit than its sub-models at 10% significance level for the data used.

Table 4. The LR test results for strengths of glass fibers data set.

	Hypotheses	LR	p -value
GOPC-LL versus OPC-LL	$H_0 : \alpha = 1$	21.6396	< 0.001
GOPC-LL versus GOHC-LL	$H_0 : \beta = 1$	3.0566	0.0804

6.2. Data: Ozone level

The second data, available in Nadarajah (2008), is about the daily ozone level measurements (in ppb = ppm \times 1000). In this application, we investigate the performance of the GOPC-W distribution and compared it performance with the other generalizations of the Weibull distribution. The results are given in Table 5. From obtained results, we conclude that the GOPC-W distribution is the best among others since it has the smallest values of the model selection criteria.

Table 5. MLEs and their SEs of the fitted models and goodness-of-fit statistics for ozone level data set

Models	α	β	a	b	$-\ell$	AIC	A^*	W^*
W			1.340 0.095	46.059 3.373	542.610	1089.221	0.966	0.170
OLL-W	1.310 0.505		1.067 0.369	47.660 4.990	542.285	1090.570	0.822	0.139
EHL-W	3.127 1.653	2.619 15.055	0.634 0.152	42.692 386.916	541.601	1091.201	0.551	0.085
EG-W	2.062 41.865	2.582 1.608	0.834 0.242	50.365 1225.053	541.203	1090.405	0.541	0.086
B-W	3.124 2.895	2.853 9.677	0.722 0.386	49.576 168.443	541.179	1090.358	0.536	0.085
GOHC-W	2.182 1.781		0.768 0.301	22.877 19.632	543.324	1092.648	0.792	0.125
EGOCH-W	2.739 2.337	3.180 3.684	0.702 0.262	17.042 17.414	543.228	1094.450	0.759	0.118
OPC-W	0.933 0.309		1.215 0.348	45.237 3.997	543.839	1093.680	1.047	0.178
GOPC-W	7.931 2.970	0.220 0.086	1.473 0.192	18.409 0.503	538.980	1085.960	0.250	0.042

The results of the LR test are given in Table 6 which shows that the GOPC-W distribution exhibits better performance than its sub-models at 5% significance level.

Table 6. The LR test results for strengths of glass fibers data set.

	Hypotheses	LR	p -value
GOPC-W versus OPC-W	$H_0 : \alpha = 1$	9.718	0.001
GOPC-W versus GOHC-W	$H_0 : \beta = 1$	8.688	0.003

6.3. Data: Voltage

The third data is about an experiment on the specimens of solid epoxy electrical-insulation and its application on the accelerated voltage life test (see Lawless, 2003). The sample size is $n = 60$ and the voltage levels are: 52.5, 55.0, 57.5. The dependent variable y_i is the log-failure times of the epoxy insulation specimens and the independent variable x_{i1} is the voltage level. Bartlett test is used to explore possible heteroscedasticity. The p -value of the Bartlett test is obtained as 0.0241 which ensures that homogeneity of the variance assumption is violated. The heteroscedastic and homoscedastic regression models fitted to the voltage data set are given, respectively, by

$$y_i = \eta_0 + \eta_1 x_{i1} + \exp(\tau_1 + \tau_2 x_{i1}) z_i \quad \text{and} \quad y_i = \eta_0 + \eta_1 x_{i1} + \sigma z_i, \quad (18)$$

where y_i follows the density in (13). The results of the regression models are given in Table 7. The results of the log-Weibull (LW), log-OPC-W (LOPC-W) and log-EGOHC-W (LEGOHC-W) regression models are also given as competitive models. The best model is selected based on the AIC and Bayesian Information Criteria (BIC) statistics. From Table 7, the heteroscedastic LGOPC-W regression model is chosen the best model among others since it has the lowest values of the AIC and BIC statistics.

Table 7. The results of the fitted regression models with corresponding SEs (in parentheses), p-values in [] and the AIC and BIC statistics.

Model	α	β	η_0	η_1	σ	τ_1	τ_2	AIC	BIC
<i>Heteroscedastic</i>									
LGOPC-W	1850.383	0.568	-51.768	0.891	-	8.161	-0.122	160.5	173.1
	3.211	0.106	0.001	0.085	-	0.315	0.023		
			[<0.001]	[<0.001]	-	[<0.001]	[<0.01]		
LOPC-W		64.488	67.295	-0.728	-	5.423	-0.025	168.1	178.6
		28.448	23.349	0.266	-	0.479	0.0006		
			0.004	0.006	-	[<0.001]	[<0.001]		
LEGOHC-W	104557	90428	-126.880	1.934	-	6.782	-0.08	161.3	173.9
	0.001	0.001	0.015	0.049	-	0.469	0.01		
			[<0.001]	[<0.001]	-	[<0.001]	[<0.001]		
LW			20.699	-0.250	-	6.147	-0.115	171.4	179.8
			3.114	0.055	-	3.171	0.057		
			[<0.001]	[<0.001]	-	[0.057]	[0.049]		
<i>Homoscedastic</i>									
LGOPC-W	1815.526	0.830	3.372	-0.181	6.321	-	-	164.0	174.5
	5.009	0.397	0.633	0.061	2.835	-	-		
			[<0.001]	[0.003]	-	-	-		
LOPC-W		7.811	20.110	-0.202	6.885	-	-	167.3	175.6
		9.684	4.157	0.054	8.502	-	-		
			[<0.001]	[<0.001]	-	-	-		
LEGOHC-W	936670	1534457	-14.702	-0.184	11.742	-	-	165.6	176.1
	0.001	0.001	4.253	0.057	1.185	-	-		
			[0.001]	[0.002]	-	-	-		
LW			22.000	-0.274	0.845	-	-	173.4	179.7
			3.046	0.055	0.09	-	-		
			[<0.001]	[<0.001]	-	-	-		

The homoscedastic variance assumption is tested with LR test. It is also used to compare the homoscedastic and heteroscedastic LGOPC-W regression models. The null hypothesis $H_0 : \tau_2 = 0$ is tested. The test statistic value is $w = 5.499$ and its p-value is 0.019 which is rejected at 5% level. It means that the heteroscedastic LGOPC-W regression model is better than the homoscedastic LGOPC-W regression model.

The residual analysis is performed based on the randomized quantile residuals which was proposed by Dunn and Smyth (1996). It is calculated with $\hat{r}_i = \Phi^{-1}(\hat{u}_i)$, where $\Phi^{-1}(\cdot)$ is the qf of the standard normal and $\hat{u}_i = F(t_i|\hat{\theta}_i)$. Figure 2 displays the quantile-quantile plots of the randomized quantile residuals for the LGOPC-W regression models for both heteroscedastic and homoscedastic. These figures confirm that the heteroscedastic LGOPC-W regression model is better than the homoscedastic LGOPC-W regression model.

7. Conclusions

A new family of distributions called generalized odd power Cauchy-G, shortly GOPC-G, is introduced. The proposed family is studied in detail. The heteroscedastic regression model based on a new generalization of the Weibull distribution is proposed. Two simulation studies are carried out to assess the suitability of the maximum likelihood method estimating the model parameters. Three applications are presented to induce the researchers in favor of the proposed models. The proposed family is hoped to attract attention from the different applied sciences.

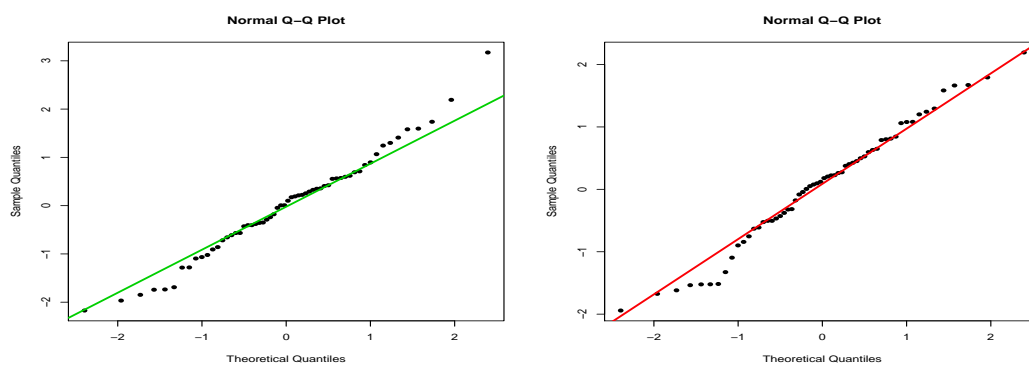


Figure 2. Q-Q plot for the fitted LGOPC-W heteroscedastic (right) and homoscedastic (left) regression models.

REFERENCES

1. Alizadeh, M., Altun, E., Cordeiro, G. M. and Rasekhi, M. (2018). The odd power cauchy family of distributions: properties, regression models and applications. *Journal of statistical computation and simulation*, **88**, 785-807.
2. Bowley, A. L. (1901). *Elements of Statistics*, P.S. King & Son, Laondon.
3. Chen, G. and Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, **27**, 154-161.
4. Cordeiro, G. M., Alizadeh, M. and Ortega, E. M. (2014). The exponentiated half-logistic family of distributions: Properties and applications. *Journal of Probability and Statistics*, **2014**, 1-21.
5. Cordeiro, G. M., Ortega, E. M. and da Cunha, D. C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, **11**, 1-27.
6. Cordeiro, G. M., Alizadeh, M., Ramires, T. G. and Ortega, E. M. (2017a). The Generalized Odd Half-Cauchy Family of Distributions: Properties and Applications. *Communications in Statistics-Theory and Methods*, **46**, 5685-5705.
7. Cordeiro, G. M., Ramires, T. G., Ortega, E. M. and Alizadeh, M. (2017b). The new family of distributions and applications in heteroscedastic regression analysis. *Journal of Statistical Theory and Applications*, **16**, 401-418.
8. Cordeiro, G. M. and Nadarajah, S. (2011). Closed-form expressions for moments of a class of beta generalized distributions. *Brazilian Journal of Probability and Statistics*, **25**, 14-33.
9. Dunn P.K and Smyth G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236-244.
10. Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, **31**, 497-512.
11. Gleaton, J.U. and Lynch, J.D. (2006) Properties of generalized log-logistic families of lifetime distributions, *Journal of Probability and Statistical Science*, **4**, 51-64.
12. Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (Vol. 362). John Wiley and Sons
13. Liu, X. (2014). Survival models on unobserved heterogeneity and their applications in analyzing large-scale survey data. *Journal of Biometrics and Biostatistics*, **5**, 1-23.
14. Moors, J. J. A. (1986). The meaning of kurtosis: Darlington reexamined. *The American Statistician*, **40**, 283-284.
15. Mudholkar, G.S., Srivastava, D.K. and Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*, **37**, 436-445.
16. Nadarajah, S. (2008). A truncated inverted beta distribution with application to air pollution data. *Stochastic Environmental Research and Risk Assessment*, **22**, 285-289.
17. Nadarajah, S. and Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*, **92**, 97-111.
18. Ortega, E.M.M., Lemonte, A.J., Cordeiro, G.M., Cancho, and Mialhe, F.L. (2017). Heteroscedastic log-exponentiated Weibull regression model. *Journal of Applied Statistics*, **45**, 384-408.
19. Rooks, B., Schumacher, A. and Cooray, K. (2010). The power Cauchy distribution: derivation, description, and composite models, NSF-REU Program Reports. Available from http://www.cst.cmich.edumathematics/research/REU_and_LURE.shtml