SORT 45 (2) July-December 2021, 181-200 DOI: 10.2436/20.8080.02.115

Exponentiated power Maxwell distribution with quantile regression and applications

Francisco A. Segovia[1], Yolanda M. Gómez[2] and Diego I. Gallardo[3]

Abstract

In this paper we introduce an extension of the power Maxwell distribution. We also discuss a reparametrized version of this model applied to quantile regression. Some properties of the model and estimation based on the maximum likelihood estimation method are studied. We also present a simulation study to assess the performance of estimators in such finite samples, and two applications to real data sets to illustrate the model.

MSC: 62E10, 62J02.

Keywords: Maxwell distribution, exponentiated distributions, maximum likelihood, quantile regression.

1. Introduction

Lehmann (1953) and Durrans (1992) introduced a family of distributions named exponentiated distributions. Their cumulative distribution function (CDF) is defined as

$$\varphi_F(w;\gamma) = F(w)^\gamma, \quad w \in \mathbb{R}, \gamma > 0 \tag{1}$$

where $F(w)$ is the CDF for a certain random variable. It follows directly that the probability density function (PDF) is

$$\varphi_f(w;\gamma) = \gamma f(w) F(w)^{\gamma-1}, \tag{2}$$

where $f(w)$ is the PDF related to $F(w)$. Durrans (1992) considered this methodology by using the normal distribution, i.e., $F = \Phi$ and $f = \phi$, the normal CDF and PDF of

[1] Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile.

[2] Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile, e-mai: yolanda.gomez@uda.cl

[3] Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile. e-mail: diego.gallardo@uda.cl

Received: March 2021
Accepted: November 2021

the standard normal distribution, respectively. This model was also discussed in more detail in Gupta and Gupta (2007), Pewsey, Gómez and Bolfarine (2012) and Rêgo, Cintra and Cordeiro (2012). Gupta and Kundu (1999) used this methodology to introduce the generalized exponential distribution, setting $F(w)$ as the CDF of the exponential model. Gómez and Bolfarine (2015) consider the case where $F(w)$ is the CDF of a half-normal distribution, resulting in a distribution which belongs to the family of beta generalized half-normal distributions. Other extensions using this methodology include the exponentiated Weibull (Mudholkar and Srivastava, 1993; Mudholkar, Srivastava and Freimer, 1995), the exponentiated Pareto (Gupta, Gupta and Gupta, 1998), exponentiated Gumbel (Nadarajah, 2005), exponentiated log-normal (Kakde and Shirle, 2006), exponentiated gamma (Nadarajah and Gupta, 2007) and power piecewise exponential (Gómez, Gallardo and Arnold, 2017). The Maxwell (M) distribution was proposed by Maxwell (1860) in order to model velocities among gas molecules. Maxwell's research was generalized by Boltzmann (1871a,c,b), to develop the distribution of energies among molecules. This distribution has diverse applications in the areas of physics, chemistry, and physical chemistry, (see Dunbar (1982)). Singh et al. (2018) introduced the power Maxwell (PM) distribution, based on taking the power of a random variable that has Maxwell distribution. Segovia et al. (2020) introduced the slashed power Maxwell (SPM) distribution and use it for outlier data modelling. However they do not use those extensions of the PM distribution considering a regression structure. We consider the specific parametrization considered in Huang and Chen (2015), where the CDF and PDF of the variable are given by

$$F_W(w;\psi,\beta) = G\left(\frac{w^{2\beta}}{2\psi^2}, \frac{3}{2}\right), \quad w \geq 0 \tag{3}$$

$$f_W(w;\psi,\beta) = \frac{4\beta}{(2\psi^2)^{3/2}\sqrt{\pi}} w^{3\beta-1} \exp\left\{-\frac{1}{2\psi^2} w^{2\beta}\right\},$$

respectively, where $\psi, \beta > 0$, and $G(\cdot, a)$ denotes the CDF for the gamma distribution with shape and scale parameters equal to a and 1, respectively. On the other hand, Galarza et al. (2017) used the skewed distributions family (SKD) in order to introduce quantile regression, where one parameter represents the quantile of the distribution. Gómez et al. (2019) introduced the Gamma-sinh Cauchy (GSC) distribution aiming at applying the model to quantile regression. The resulting model can be either unimodal or bimodal depending on the combinations of two parameters, where one of them is fixed and depends on the modelled quantile. Gallardo et al. (2020a) introduced a novel parametric quantile regression model for asymmetric response variables, where the response variable follows a power skew-normal distribution. Gallardo, Gómez-Déniz and Gómez (2020b) presented a discrete distribution by discretizing a generalized half-normal distribution, which can be reparametrized for use in a regression model based on the median. Sánchez et al. (2020) use a model based on the Birnbaum-Saunders distribution in order to perform quantile regression.

The aim of this paper is to introduce an extension of the PM distribution using the methodology presented in equation (1), aiming to perform quantile regression. The resulting PDF can be either strictly increasing or unimodal. The manuscript is organized as follows. In Section 2 we introduce the exponentiated power Maxwell (EPM) distribution, and we propose the reparametrized EPM (REPM) distribution with some properties such as its CDF, hazard function (HF) and moments. In Section 3, we discuss the inference for the REPM regression model based on the maximum likelihood (ML) estimation. In Section 4 we present a simulation study in finite samples, focusing our attention on parameter recovery. In Section 5 we present two applications to real data, fitting the REPM distribution to two real data sets. Finally, in Section 6 we present the main conclusions of the work.

2. Exponentiated power Maxwell distribution

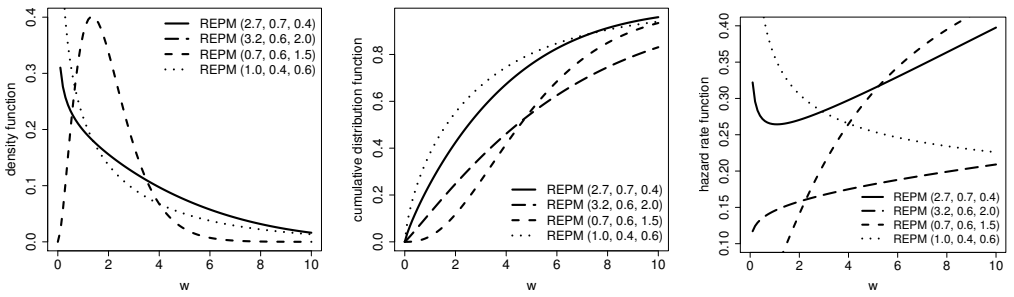
Following the methodology related to equation (1), we introduce the following extension of the PM model.

Definition 1. *A random variable W follows an exponentiated power Maxwell distribution with scale parameter ψ and shape parameters β and γ, if its CDF, PDF and HF are given, respectively, by:*

$$\begin{aligned}
F_Y(w;\psi,\beta,\gamma) &= \left[G\left(\frac{w^{2\beta}}{2\psi^2},\frac{3}{2}\right)\right]^{\gamma}, \quad w>0 \\
f_Y(w;\psi,\beta,\gamma) &= \gamma\left[G\left(\frac{w^{2\beta}}{2\psi^2},\frac{3}{2}\right)\right]^{\gamma-1}\frac{\beta w^{2\beta-1}}{\psi^2}\, g\left(\frac{w^{2\beta}}{2\psi^2},\frac{3}{2}\right), \quad w>0, \\
h_W(w;\psi,\beta,\gamma) &= \frac{\gamma\left[G\left(\frac{w^{2\beta}}{2\psi^2},\frac{3}{2}\right)\right]^{\gamma-1}\beta w^{2\beta-1}\, G\left(\frac{w^{2\beta}}{2\psi^2},\frac{3}{2}\right)}{\psi^2\left\{1-\left[G\left(\frac{w^{2\beta}}{2\psi^2},\frac{3}{2}\right)\right]^{\gamma}\right\}}, \quad w>0
\end{aligned} \tag{4}$$

where $\psi,\beta,\gamma>0$ and $g(\cdot,a)$ is the PDF related to $G(\cdot,a)$.

In Figure 1, we illustrate the PDF, CDF, and HF of the REPM distribution. It is interesting to point out that the HF can be strictly increasing, strictly decreasing, or have a bathtub shape. The equation for finding the mode is immediately obtained from calculating the first derivative of the density. However, we consider a parametrization for this model based on (μ,β,γ), where $\mu=\psi^{\frac{1}{\beta}}$. We denote this as REPM$(\mu,\beta,\gamma)$. The main object of this parametrization will be justified later.



(a) **(b)** **(c)**

Figure 1. *Plots of the PDF (a), CDF (b) and HF (c) for different combinations of parameters of the REPM*(ψ,β,γ) *distribution.*

Proposition 1. *If* $W \sim REPM(\mu,\beta,\gamma)$*, the rth non-central moment of* W *can be calculated as*

$$E(W^r) = \int_0^1 \frac{1}{\beta g\left(\dfrac{w^{2\beta}}{2\mu^{2\beta}}, \dfrac{3}{2}\right)} r w^{r-2\beta} \mu^{2\beta} (1-u)^{\gamma} du$$

for $r \geq 1$*, where* $w = [2\mu^{2\beta} G^{-1}(u, 3/2)]^{1/(2\beta)}$*,* G^{-1} *is the inverse function of* $G(\cdot, a)$*.*

Proof. By using the definition of expectation and making the substitution $u = G\left(\dfrac{w^{2\beta}}{2\psi^2}, \dfrac{3}{2}\right)$, the result is immediate ■.

The gamma distribution is very useful to express both the CDF and the PDF of the REPM distribution. However, usual quantities of interest such as he mean and mode of the model do not have closed form. Therefore, in order to perform regression analysis in the model, other alternatives should be studied, as we illustrate in the following proposition.

Proposition 2. *If* $W \sim REPM(\mu,\beta,\gamma)$*, then* $100 \times \rho$*-th, the* ρ*-th quantile* $0 < \rho < 1$*, is given by*

$$p_\rho = \left[2\mu^{2\beta} G^{-1}\left(\rho^{1/\gamma}, \frac{3}{2}\right)\right]^{1/2\beta}, \tag{5}$$

Proof. It is immediate using the definition of quantile ■.

Corollary 1. *From proposition 2, it follows directly that the median of the REPM distribution is*

$$Me(w) = \left[2\mu^{2\beta} G^{-1}\left(0.5^{1/\gamma}, \frac{3}{2}\right)\right]^{1/2\beta}.$$

Table 1 shows the mean, variance, median and mode for different values of μ, β and γ. Note that the mean, variance and median increase as γ increases; all four quantities increase as μ increases. It is also interesting to point out that the variance grows extremely as β decreases ($\beta < 1$). On the other hand

Table 1. *Mean, variance, median and mode for the REPM model with different combination of parameters.*

(μ, β, γ)	Mean	Variance	Median	Mode
$(1.3, 1.5, 0.5)$	1.403	1.386	1.365	0.347
$(1.3, 1.5, 1.0)$	1.738	1.732	1.724	0.254
$(1.3, 1.5, 1.5)$	1.912	1.904	1.891	0.207
$(1.3, 1.5, 2.0)$	2.024	2.015	1.997	0.179
$(2.3, 0.5, 1.5)$	8.566	7.230	4.545	34.909
$(2.3, 1.0, 1.5)$	4.189	4.078	3.848	2.154
$(2.3, 1.5, 1.5)$	3.382	3.369	3.346	0.648
$(2.3, 2.0, 1.5)$	3.055	3.063	3.081	0.305
$(0.6, 1.5, 1.5)$	0.882	0.879	0.872	0.044
$(1.0, 1.5, 1.5)$	1.471	1.465	1.455	0.122
$(1.3, 1.5, 1.5)$	1.912	1.904	1.891	0.207
$(1.6, 1.5, 1.5)$	2.353	2.344	2.328	0.313

$$F_W(\mu; \mu, \beta, \gamma) = \left[G\left(\frac{1}{2}, \frac{3}{2}\right)\right]^{\gamma} = C^{\gamma}, \tag{6}$$

with $C = G(1/2, 3/2) = 2\Phi(1) - 2\phi(1) - 1 \approx 0.199$. In equation (6), we note that the CDF evaluated in μ depends only on the value of γ. As C^{γ} is a strictly decreasing function for γ and $0 < C < 1$, the equation $F_W(\mu; \mu, \beta, \gamma) = \rho$, (for $0 < \rho < 1$) has a unique solution for γ. Specifically,

$$F_W(\mu; \mu, \beta, \gamma) = \rho \Leftrightarrow \gamma = \frac{\log(\rho)}{\log(C)}.$$

For a fixed ρ, if we set $\gamma = \gamma(\rho) = \log(\rho)/\log(C)$ as fixed, then μ represents directly the ρth quantile of the distribution. Table 2 shows some values for $\gamma(\rho)$ with different values for ρ. Henceforth, we will use the notation REPM(μ, β, γ) to refer to this alternative parametrization. This is a very useful result, because in practice many characteris-

Table 2. *Value of* $\gamma(\rho)$ *for some values of* ρ.

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\gamma(\rho)$	1.425	0.996	0.745	0.567	0.429	0.316	0.221	0.138	0.065

tics inherent to each observation are available. For this reason, we introduce a regression framework for applying the model to any quantile of the distribution. This also allows a more detailed relation among the covariates and the response variable than is possible using the regression in a single measure such as mean or median. To be more specific, for a non-homogeneous population, we consider that $w_i(\rho)$, the ρ-th quantile of the response variable, are independent and are such that $w_i(\rho) \sim \text{REPM}(\mu_i(\rho), \beta(\rho), \gamma(\rho))$, $i = 1, \ldots, n$, where the quantile of such variable is related to a set of covariates, say $\mathbf{x}_i^{\top} = (x_{i1}, ..., x_{ip})$, through the logarithmic link as

$$\log \mu_i(\rho) = \mathbf{x}_i^{\top} \boldsymbol{\tau}(\rho), \quad i = 1, \ldots, n, \tag{7}$$

where $\boldsymbol{\tau}(\rho) = (\tau_0(\rho), \ldots, \tau_p(\rho))^{\top}$ are the regression coefficients. These can be interpreted as follows: $\exp(\tau_0(\rho))$ represents the value of the ρ-th quantile of the response variable when all covariates are fixed at 0; and $\exp(\tau_j(\rho))$, $j = 1, \ldots, p$, represents the percentage increment (or decrement) in the ρ-th quantile for the response variable when the j-th covariate is increased by one unit and the rest of the covariates are fixed.

To avoid overloading the notation, hereinafter we use simply μ_i, β and γ instead of $\mu_i(\rho), \beta(\rho), \gamma(\rho)$ to specify the parameters, but understanding that in a regression model context, we are interested in modelling the ρ-th quantile.

3. Inference

In this section, we discuss the ML estimation for the REPM regression model under a classical approach. Let $W_i(\rho) \sim \text{REPM}(\mu_i, \beta, \gamma)$ independent variables, where the ith observation is related to a set of covariates $\mathbf{x}_i$ as in equation (7) and $\gamma = \gamma(\rho) =$

$\log(\rho)/\log(C)$ is fixed. The log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\tau}^\top, \beta, \gamma)^\top$ is

$$\ell(\boldsymbol{\theta}) = n\log(\gamma) + (\gamma-1)\sum_{i=1}^{n}\log\left(G\left(\frac{w_i^{2\beta}}{2\mu_i^{2\beta}}, \frac{3}{2}\right)\right) + n\log(\beta) - 2\sum_{i=1}^{n}\log(\mu_i^{\beta}) + \\ + (2\beta-1)\sum_{i=1}^{n}\log(w_i) - n\log\left(\frac{\Gamma(3/2)}{\sqrt{2}}\right) - \sum_{i=1}^{n}\log(\mu_i^{\beta}) + \beta\sum_{i=1}^{n}\log(w_i) - \frac{1}{2^{2\beta}}\sum_{i=1}^{n}\left(\frac{w_i}{\mu_i}\right)^{2\beta}. \tag{8}$$

The ML estimators can be obtained by maximizing equation (8), using numerical procedures such as the Newton-Raphson algorithm. As an alternative, we use the `optim` routine in the R software (R Core Team, 2021) for the `L-BFGS-B` method, which is a limited memory modification for the traditional Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS), a constrained Quasi-Newton type algorithm which avoids the computation of the hessian matrix for the objective function and its respective inverse. The asymptotic variance of the ML estimators (say $\widehat{\boldsymbol{\theta}}$) can be estimated as follows $\widehat{\mathrm{Var}(\boldsymbol{\theta})} = \mathrm{diag}(-\mathbf{I}(\widehat{\boldsymbol{\theta}})^{-1})$, where $\mathbf{I}(\widehat{\boldsymbol{\theta}})$ is observed Fisher information evaluated in $\widehat{\boldsymbol{\theta}}$, that is

$$\mathbf{I}(\widehat{\boldsymbol{\theta}}) = -\left.\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Details about the components of this matrix can be found in appendix A. The asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ is $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \mathbf{N}(0, \mathbf{I}(\widehat{\boldsymbol{\theta}})^{-1})$, as $n \to +\infty$.

In order to perform a residual analysis, we can use the quantile residuals (see Dunn and Smith (1996)) defined as

$$r_i = \Phi^{-1}[F_W(w_i; \widehat{\boldsymbol{\theta}})], \quad i = 1, 2, ..., n,$$

where $F_W(w_i; \widehat{\boldsymbol{\theta}})$ is the CDF of the REPM model evaluated in the ML estimate of $\boldsymbol{\theta}$. As the ML estimator is a consistent estimator (when $n \to +\infty$), and if the model is appropriate for the data, $r_1, r_2, ..., r_n$ should be a random sample from the standard normal distribution. Also note that the independent observation assumption implies that the quantile residuals are also independent. The normality assumption can be tested, for instance, by a normality test such as the Kolmogorov-Smirnov (KS) (see Kolmogorov (1993) and Smirnov (1939)), Shapiro-Wilks (SW) (see Shapiro and Wilks (1965)) and Anderson-Darling (AD) (see Anderson and Darling (1952)) tests.

Table 3. *Simulation study for the REPM model.*

true values					$n = 50$				$n = 100$				$n = 200$			
ρ	β	τ_0	τ_1	estimator	mean	SE	CP	RMSE	mean	SE	CP	RMSE	mean	SE	CP	RMSE
0.50	2.0	2.0	0.5	β	2.049	0.176	1.000	0.086	2.017	0.081	0.686	0.168	2.007	0.057	0.642	0.120
				τ_0	2.049	0.176	1.000	0.086	2.023	0.120	1.000	0.065	2.010	0.087	1.000	0.045
				τ_1	0.505	0.150	0.940	0.136	0.503	0.101	0.969	0.110	0.502	0.073	0.945	0.072
			1.5	β	2.038	0.115	0.655	0.263	2.017	0.081	0.686	0.168	2.007	0.057	0.674	0.115
				τ_0	2.049	0.174	1.000	0.093	2.020	0.123	1.000	0.065	2.011	0.085	1.000	0.045
				τ_1	1.529	0.170	0.975	0.154	1.520	0.125	0.969	0.110	1.504	0.084	0.976	0.075
	2.0	0.5	2.0	β	2.096	0.116	0.658	0.264	2.041	0.082	0.675	0.167	2.019	0.057	0.696	0.114
				τ_0	0.498	0.102	0.98	0.089	0.499	0.079	0.979	0.066	0.499	0.052	0.981	0.044
				τ_1	2.005	0.184	0.985	0.140	2.003	0.133	0.989	0.103	2.002	0.092	0.985	0.070
		1.5		β	2.082	0.116	0.633	0.270	2.043	0.081	0.661	0.175	2.018	0.057	0.658	0.121
				τ_0	1.494	0.157	0.988	0.110	1.503	0.104	0.997	0.066	1.500	0.073	0.997	0.043
				τ_1	2.005	0.199	0.977	0.166	1.998	0.128	0.992	0.102	2.000	0.092	0.986	0.072
	0.5	2.0	2.0	β	0.525	0.116	0.995	0.069	0.512	0.082	0.999	0.043	0.506	0.057	1.000	0.03
				τ_0	1.983	0.350	0.915	0.386	2.011	0.237	0.932	0.248	2.004	0.169	0.940	0.175
				τ_1	2.034	0.638	0.933	0.651	1.987	0.441	0.956	0.415	1.999	0.309	0.974	0.284
	1.5			β	1.575	0.116	0.748	0.205	1.531	0.081	0.807	0.127	1.518	0.057	0.797	0.089
				τ_0	1.997	0.163	0.991	0.116	2.003	0.123	0.991	0.094	1.999	0.084	0.992	0.060
				τ_1	2.004	0.188	0.956	0.175	1.999	0.15	0.957	0.143	2.003	0.108	0.965	0.102
0.75	2.0	2.0	0.5	β	2.096	0.126	0.645	0.287	2.052	0.088	0.671	0.190	2.019	0.062	0.667	0.129
				τ_0	1.987	0.202	0.988	0.135	1.989	0.149	0.995	0.104	1.997	0.101	0.994	0.064
				τ_1	0.498	0.218	0.928	0.241	0.509	0.165	0.942	0.169	0.500	0.107	0.962	0.102
			1.5	β	2.106	0.126	0.604	0.306	2.05	0.088	0.663	0.187	2.023	0.062	0.702	0.122
				τ_0	1.986	0.209	0.992	0.140	1.996	0.144	0.993	0.093	1.998	0.103	0.994	0.069
				τ_1	1.502	0.258	0.958	0.245	1.497	0.167	0.957	0.161	1.497	0.123	0.952	0.117
	2.0	0.5	2.0	β	2.117	0.126	0.605	0.310	2.050	0.088	0.637	0.192	2.029	0.062	0.660	0.131
				τ_0	0.484	0.150	0.951	0.149	0.488	0.103	0.947	0.102	0.497	0.074	0.968	0.069
				τ_1	2.008	0.272	0.972	0.25	2.007	0.181	0.963	0.164	1.999	0.130	0.972	0.115
		1.5		β	2.103	0.126	0.636	0.289	2.039	0.088	0.677	0.187	2.022	0.062	0.685	0.125
				τ_0	1.483	0.18	0.985	0.134	1.491	0.126	0.986	0.094	1.500	0.091	0.991	0.067
				τ_1	2.006	0.257	0.973	0.222	2.000	0.181	0.969	0.160	1.992	0.133	0.973	0.118
	0.5	2.0	2.0	β	0.528	0.126	0.997	0.074	0.511	0.088	1.000	0.047	0.507	0.062	0.999	0.032
				τ_0	1.961	0.532	0.939	0.575	1.971	0.355	0.936	0.367	1.990	0.250	0.939	0.257
				τ_1	1.947	0.987	0.931	1.001	1.983	0.632	0.952	0.627	2.009	0.448	0.951	0.436
	1.5			β	1.568	0.126	0.778	0.216	1.540	0.088	0.776	0.143	1.519	0.062	0.783	0.101
				τ_0	1.994	0.198	0.978	0.161	1.99	0.151	0.976	0.126	1.994	0.104	0.976	0.087
				τ_1	1.974	0.295	0.939	0.300	2.002	0.225	0.951	0.224	2.000	0.157	0.961	0.153

4. Simulation study

In this section, we present a simulation study in order to assess the performance of the ML estimators for the REPM regression model. We considered one covariate, i.e., $\mu_i = \tau_0 + \tau_1 x_i$, $\gamma(\rho)$ as fixed, and the covariates $x_1, \ldots, x_n$ were simulated from the standard uniform distribution. We considered six vectors for (β, τ_0, τ_1): (2, 2, 0.5), (2, 2, 1.5), (2, 0.5, 2), (2, 1.5, 2), (0.5, 2, 2), (1.5, 2, 2); three values for the sample size: 50, 100 and 200; and two values for the modelled quantile: 0.50 and 0.75, totalling 36 combinations of parameters, sample size and quantile. Each scenario was replicated 1,000 times. To simulate values from the REPM model, we can use the following algorithm based on the inverse transform method:

- Generate $U_i \sim U(0,1), \quad i = 1, 2, ..., n.$
- Compute $W_i = \left[2\mu^{2\beta} G^{-1}\left(U_i^{1/\gamma}, \frac{3}{2} \right) \right]^{1/2\beta}$.

For each sample, we compute the ML estimates and the estimated standard errors based on the estimated hessian matrix. Table 3 summarizes the results, considering the mean of the ML estimations, their standard errors (SE), the 95% coverage probability (CP) based on the asymptotic normality for the ML estimators and the estimated root mean squared error (RMSE). Note that as the sample size increases, the mean of the ML estimators is closer to the true value of the parameters, while the RMSE decreases, suggesting than the estimators are consistent for the REPM model even in a finite sample size. Results also suggest that the CP terms converge to the nominal values with which they were built, suggesting that the asymptotic normality of the estimators is also reasonable in finite samples for the REPM model.

5. Application

In this section we illustrate our proposal with two real data sets, comparing it with other proposals in the literature. In the first application we fit the REPM model without covariates. We compare the results with the M, PM and gamma (G) distributions. In the second application we fit our proposal considering covariates, comparing results with the GSC, skewed Laplace (SKL) and skewed Student-t (SKT) models. Codes in R software (R Core Team, 2021) are avaliable as supplementary material.

5.1. Reinfection time data

In certain populations the occurrence of sexually transmitted diseases blue is a major problem. Even those that are not lethal represent a threat that must be taken into account. Specifically, gonorrhea and chlamydia are a focus of investigation because they are often asymptomatic in females. As a result they are often left untreated, which can lead to complications such as sterility. The following data set corresponds to the time to reinfection of 887 individuals by either gonorrhea or chlamydia, where the subject had already been infected with one of these diseases previously (see Klein and Moeschberger (2003)). This data set can be found in the std data included in the KMSurv R package (Klein, Moeschberger and Yan, 2012).

Table 4. *Descriptive analysis for the reinfection time data.*

mean	s.d.	median	interquartile range	min.	max.	skewness	kurtosis
369.5	370.1	247.0	501.0	1.0	1529.0	1.2	3.5

Table 4 shows a descriptive analysis for this data set. Note that 50% of the individuals were reinfected within the first 8 months. The times also present a positive skewness and a kurtosis slightly greater than normal distribution. Figure 5 shows the ML estimates for the parameters of the M, PM, G and REPM distributions. For each model we also present the AIC criteria, which suggest that the REPM model gives a better fit than the rest of the models. Figure 2 depicts the histogram with the estimated PDF and comparing the empirical CDF with the estimated CDF for the models discussed, showing that the REPM model presents a better fit for this data. Finally, Figure 3 shows the quantile-quantile (QQ) plots for the REPM, PM and G distributions. Note that the QQ plots suggest that, of the three models tested, the REPM is the most appropriate for this data set.

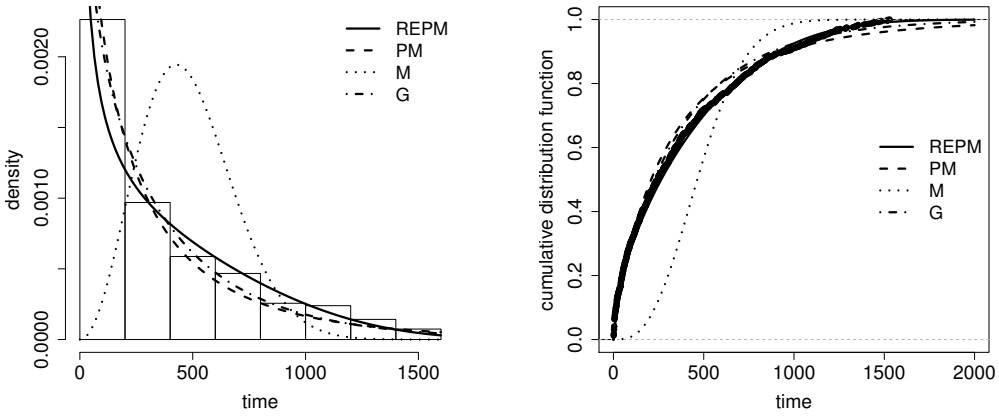
5.2. Clotting data

This data set presents measurements of the clotting time of blood (`time`, in seconds) for normal plasma diluted to nine different percentage concentrations with prothrombin-free plasma (`lconc`, in logarithm scale) for 18 patients. It must also be considered that the clotting time was induced by two lots of thromboplastin (`lot2`, categorized as 0 and 1). The data (see MLGdata R package) are available in McCullagh and Nelder (1989) (p. 302) (see R code below).

```
clotting<-data.frame(time=c(118, 58, 42, 35, 27, 25, 21, 19,
    18, 69, 35, 26, 21, 18, 16, 13, 12, 12),
    lconc=c(1.609, 2.303, 2.708, 2.996, 3.401, 3.689, 4.094,
    4.382, 4.605, 1.609, 2.303, 2.708, 2.996, 3.401, 3.689,
    4.094, 4.382, 4.605),
    lot=factor(c(rep(0, 9), rep(1, 9))))
```

Table 5. *Maximum likelihood estimates for the data with it's respective standard deviation in parenthesis for the infection time data*

Parameter	M	PM	G	REPM
α	$< 0.001(0.028)$	0.038 (0.004)	0.796 (0.027)	—
β	—	0.321 (0.009)	$0.002(< 0.001)$	**1.079 (0.158)**
μ	—	—	—	**578.576(0.150)**
γ	—	—	—	**0.177(0.195)**
log-likelihood	-7593.0	-6053.0	-6033.8	$-$**6013.3**
AIC	15188.0	12109.9	12071.6	**12032.7**

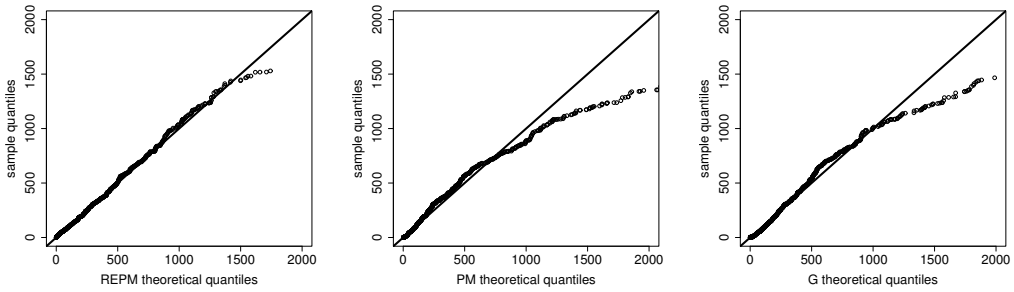


(a) *Histogram* **(b)** *Empirical plot*

Figure 2. *Histogram and empirical plot for the reinfection time data.*

We aim to model the clotting `time` for the i-th individual using `lconc`, `lot2` and the interaction between those covariates. We considered $\texttt{time}(\rho) \sim \text{REPM}(\mu_i, \beta, \gamma)$, where $\gamma = \gamma(\rho) = \log(\rho)/\log(C)$ is fixed and

$$\mu_i = \mu_i(\rho) = \exp\left(\tau_0 + \tau_1 \texttt{lconc}_i + \tau_2 \texttt{lot2}_i + \tau_3 \texttt{lconc}_i \times \texttt{lot2}_i\right), \quad i = 1, ..., 18,$$



(a) *QQ-plot, REPM.* **(b)** *QQ-plot, PM.* **(c)** *QQ-plot, G.*

Figure 3. *Q-Q plot for the REPM, PM and G models for the reinfection time data.*

Table 6 presents a descriptive analysis for the `global time`, `time` for `lot` $= 0$ ($\texttt{time}_0$), `time` for `lot` $= 1$ ($\texttt{time}_1$) and `lconc` for the clotting data set. We can verify that the global `time` has a significant standard deviation and is positively skewed, with a considerable kurtosis coefficient. Moreover, Figure 4 shows the plots for `time` versus `lconc` separated by `lot`.

Table 6. *Descriptive analysis for the clotting data.*

variable	mean	s.d.	median	interquartile range	min.	max.	skewness	kurtosis
`global time`	32.500	26.440	23.000	17.000	12.000	118.000	2.127	7.185
$\texttt{time}_0$	40.333	31.851	27.000	21.000	18.000	118.000	1.805	5.078
$\texttt{time}_1$	24.667	18.248	18.000	13.000	12.000	69.000	1.780	5.012

Table 7 shows the AIC values and p-values obtained in the K-S test for the quantile residuals, for the SKL, SKT, GSC and REPM quantile regression models different quantile values. Note that the AIC for the REPM is the lowest value of all the models (except for $\rho = 0.1$); the K-S test does not reject the null hypothesis that quantile residuals for this model are a random sample from the standard normal distribution (except for $\rho = 0.9$) with any significance level, suggesting that the model is appropriate for all the modelled quantiles (except for $\rho = 0.90$).

Figure 5 shows the ML estimator for the regression coefficients with their their respective asymptotic 95% confidence intervals. Note that `lconc` and `lot2` are significant in explaining all the quantiles modelled. Figure 6 shows the profile density for the ρ-th quantile of `time` for $\rho = 0.5$ and $\rho = 0.75$. Note how the distribution of the `time` according to our model seems to differ from the other distributions, showing a better representation of the population. Regarding the interpretation of the coefficients,

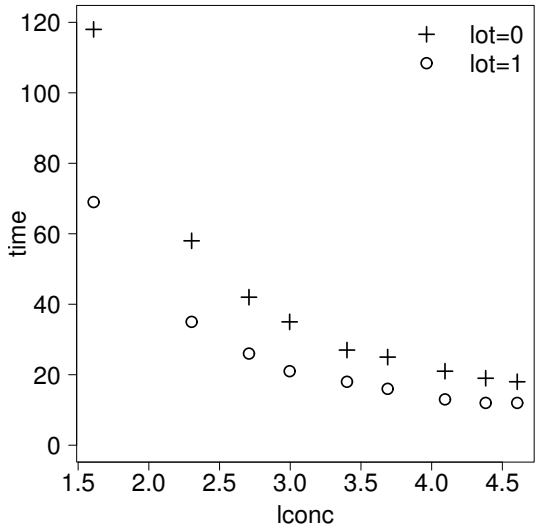


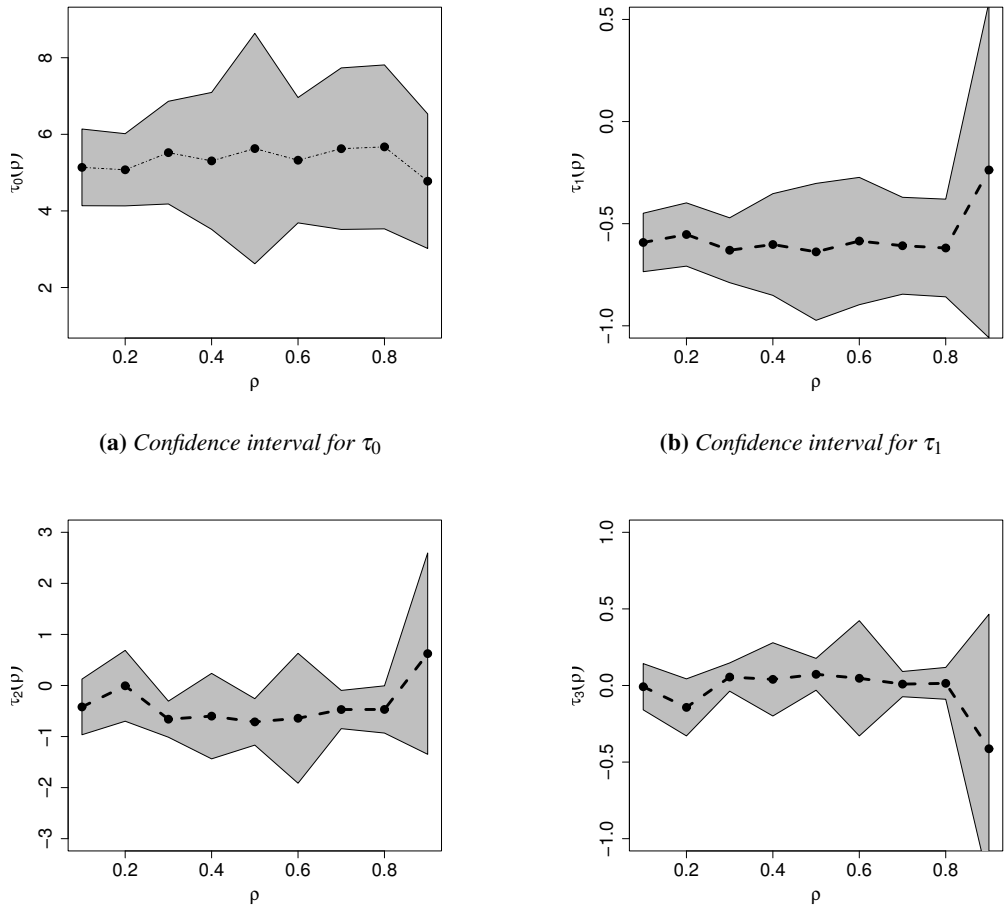
Figure 4. *Plot for clotting data.*

for example, we can conclude that

- For $\rho = 0.5$ (the median case) we obtain $\exp(\widehat{\tau}_1) = 0.528$. This means that for a fixed type of thromboplastin, the median of the clotting `time` decreases by 47.2% for each unit increase in the `lnonc`.
- For $\rho = 0.5$ (the median case) $\exp(\widehat{\tau}_2) = 0.490$. This implies that for a fixed `lnonc`, the median of the clotting `time` decreases by 51.0% when the type of thromboplastin is changed from `lot2` $= 1$ to `lot2` $= 0$.

Table 7. *AIC and p-values for the K-S test of SKT, SKL, GSC, and REPM model for the clotting data.*

	AIC				K-S			
ρ	SKT	SKL	GSC	**REPM**	SKT	SKL	GSC	**REPM**
0.10	121.130	127.916	110.820	**111.367**	0.003	0.003	0.186	**0.431**
0.25	125.554	132.958	118.335	**109.253**	0.004	0.001	0.119	**0.428**
0.50	133.049	143.110	129.903	**111.556**	0.002	< 0.001	0.250	**0.247**
0.75	151.402	155.568	144.733	**113.330**	0.092	0.500	0.018	**0.190**
0.90	149.034	150.596	167.269	**130.857**	0.125	0.200	< 0.001	< 0.001



(a) *Confidence interval for* τ_0

(b) *Confidence interval for* τ_1

(c) *Confidence interval for* τ_2

(d) *Confidence interval for* τ_3

Figure 5. *ML estimation of the regression coefficientes (with their respectively asymptotic* 95% *confidence interval), for the different values of the* ρ*-th quantile for the clotting data*

6. Conclusions

Exponentiated distributions have been used to extend a variety of well-known distribution models, resulting in flexible distributions that can be applied in a greater diversity of scenarios. This paper proposes the REPM distribution as an alternative model by which to introduce covariates, obtaining interpretations related to the quantile of the distribution. Nowadays there is a reasonable set of classic distributions with positive support, such as the exponential, gamma, Weibull, log-normal (LN), etc. So the question naturally arises "Why consider the REPM model instead of the common distribution that works well?". While it is true that models like LN and G have proved to be flexible

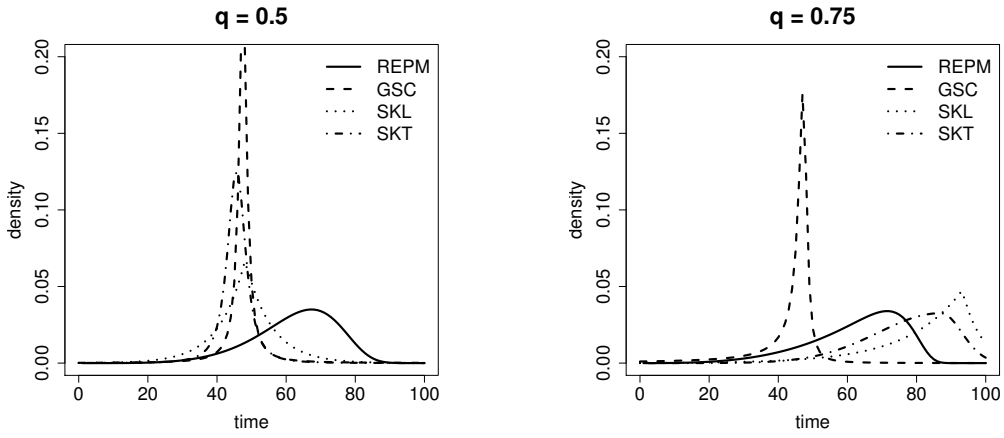


Figure 6. *Distribution for 0.5 (a) and 0.75 (b) quantiles of time considering* `lconc` *and* `lot2` *equal to 2.3 and 0, respectively. Curves in solid, dashed, dotted and dot-dash line represent the density functions estimated by the REPM, GSC, SKL and SKT models, respectively, for the clotting data*

enough to cover many situations, there are a few factors that must be borne in mind. For example, the LN distribution has a hazard rate function that may be unrealistic in some contexts, such as lifetimes data sets, since it is decreasing for long values. On the other hand, the G distribution, although it has a less strict hazard rate function, is not as flexible as the corresponding REPM model; moreover it does not have a closed function for the ρ-th quantile, i.e. quantile regression cannot be applied simply in this model. The real data applications above show that the REPM is a competent alternative to such traditional models.

Funding. *The research of Francisco A. Segovia was supported by Vicerrectoría de Investigación y Postgrado de la Universidad de Atacama. The research of Yolanda M. Gómez was supported by proyecto DIUDA programa de inserción NÂº 22367 of the Universidad de Atacama.*

A. Appendix: Score function and observed Fisher information

We devote this section to express the components of $\mathbf{I}(\widehat{\boldsymbol{\theta}})$ discussed in Section 3.

If $W \sim REPM(\boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\mu, \beta, \gamma)^{\intercal}$, then we can $\partial^2 \log f_W(w;\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\intercal}$, as follows

$$
\begin{aligned}
\frac{\partial^2 \log f_W(w;\boldsymbol{\theta})}{\partial\beta^2} &= (\gamma-1)\left\{-\left[\frac{1}{G(\cdot)}\frac{\partial G(\cdot)}{\partial\beta}\right]^2+\frac{1}{g(\cdot)}\left[\frac{\partial \log G(\cdot)}{\partial\beta}\right]\left[\log\left(\frac{w}{\mu}\right)g(\cdot)+\frac{\partial g(\cdot)}{\partial\beta}\right]\right\},\\
\frac{\partial^2 \log f_W(w;\boldsymbol{\theta})}{\partial\mu^2} &= (1-\gamma)\left\{\left[\frac{1}{G(\cdot)}\frac{\partial G(\cdot)}{\partial\mu}\right]^2+\frac{1}{g(\cdot)}\left[\frac{\partial \log G(\cdot)}{\partial\mu}\right]\left[\frac{\partial g(\cdot)}{\partial\mu}-(2\beta-1)u^{-1}g(\cdot)\right]\right\},\\
\frac{\partial^2 \log f_W(w;\boldsymbol{\theta})}{\partial\gamma^2} &= -\frac{1}{\gamma^2},\\
\frac{\partial^2 \log f_W(w;\boldsymbol{\theta})}{\partial\beta\partial\mu} &= -(\gamma-1)w^{2\beta}\mu^{-2\beta-1}\left[1+2\beta\log\left(\frac{w}{\mu}\right)\right]\left\{\left(\frac{w}{\mu}\right)^{2\beta}\log\left(\frac{w}{\mu}\right)+\frac{g(\cdot)}{G(\cdot)}\right\},\\
\frac{\partial^2 \log f_W(w;\boldsymbol{\theta})}{\partial\beta\partial\gamma} &= \frac{1}{G(\cdot)}\frac{\partial G(\cdot)}{\partial\beta},\\
\frac{\partial^2 \log f_W(w;\boldsymbol{\theta})}{\partial\mu\partial\gamma} &= \frac{1}{G(\cdot)}\frac{\partial G(\cdot)}{\partial\mu},
\end{aligned}
$$

where $G(\cdot) = G(w^{2\beta}/2\mu^{2\beta}, 3/2)$, $g(\cdot) = g(w^{2\beta}/2\mu^{2\beta}, 3/2)$, and

$$
\begin{aligned}
\frac{\partial G(\cdot)}{\partial\beta} &= g(\cdot)\left(\frac{w}{\mu}\right)^{2\beta}\log\left(\frac{w}{\mu}\right),\\
\frac{\partial G(\cdot)}{\partial\mu} &= -\beta\mu^{-2\beta-1}w^{2\beta}g(\cdot),\\
\frac{\partial g(\cdot)}{\partial\beta} &= g(\cdot)\log\left(\frac{w}{\mu}\right)\left[1-\left(\frac{w}{\mu}\right)^{2\beta}\right],\\
\frac{\partial g(\cdot)}{\partial\mu} &= \frac{\beta}{\mu}g(\cdot)\left[\left(\frac{w}{\mu}\right)^{2\beta}-1\right].
\end{aligned}
$$

References

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2), 193–212.

Boltzmann, L. (1871a). Analytischer beweis des zweiten haubtsatzes der mechanischen wärmetheorie aus den sätzen über das gleichgewicht der lebendigen kraft. *Wiener Berichte*, 63, 712–732.

Boltzmann, L. (1871b). Einige allgemeine sätze über wärmegleichgewicht. *Wiener Berichte*, 63, 679–711.

Boltzmann, L. (1871c). Über das wärmegleichgewicht zwischen mehratomigen gasmolekülen. *Wiener Berichte*, 63, 397–418.

Dunbar, R. (1982). Deriving the Maxwell distribution. *Journal of Chemical Education*, 59, 22–23.

Dunn, K. P. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236–244.

Durrans, R. (1992). Distributions of fractional order statistics in hydrology. *Water Resources Research*, 28, 1649–1655.

Galarza, C., Lachos, V., Cabral, C. and Castro, L. (2017). Robust quantile regression using a generalized class of skewed distributions. *Stat*, 6, 113–130.

Gallardo, D., Bourguignon, M., Galarza, C. and Gómez, H. (2020a). A parametric quantile regression model for asymmetric response variables on the real line. *Symmetry*, 12, 1938.

Gallardo, D., Gómez-Déniz, E. and Gómez, H. (2020b). Discrete generalized half-normal distribution and its applications in quantile regression. *SORT*, 44(2), 265–284.

Gómez, Y. and Bolfarine, H. (2015). Likelihood-based inference for the power half-normal distribution. *Journal of Statistical Theory and Applications*, 14(4), 383–398.

Gómez, Y., Gallardo, D. and Arnold, B. (2017). The power piecewise exponential model. *Journal of Statistical Computation and Simulation*, 88(5), 825–840.

Gómez, Y., Gómez-Déniz, E., Venegas, O., Gallardo, D. and Gómez, H. (2019). An asymmetric bimodal distribution with application to quantile regression. *Symmetry,* 11, 899.

Gupta, R. and Gupta, R. (2008). Analyzing skewed data by power normal model. *TEST*, 17, 197–210.

Gupta, R. and Kundu, D. (1999). Generalized exponential distributions. *Australian New Zealand Journal of Statistics*, 41(2), 173–188.

Gupta, R., Gupta, P. and Gupta, R. (1998). Modeling failure time data by Lehmann alternatives. *Communications in Statistics - Theory and Methods*, 27(4), 887–904.

Huang, J. and Chen, S. (2015). Tail behavior of the generalized Maxwell distribution. *Communications in Statistics - Theory and Methods*, 45(14), 4230–4236.

Kakde, C. and Shirle, D. (2006). On exponentiated lognormal distribution. *International Journal of Agricultural and Statistical Sciencies,* 2, 319–326.

Klein, J. and Moeschberger, M. (2003). *Survival Analysis Techniques for Censored and Trun Cated Data*, Springer-Verlag, New York, Vol. 2.

Klein, J., Moeschberger, M. and Yan, J. (2012). KMsurv: Data sets from Klein and Moeschberger (1997), Survival Analysis. https://CRAN.R-project.org/package= KMsurv, R package version 0.1-5.

Kolmogorov, A. (1993). Sulla determinazione empirica di uma legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, 83-91.

Lehmann, E. (1953). The power of rank tests. *Annals of Mathematical Statistics*, 24(1), 23–43.

Maxwell, J. (1860). Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres. *Philosophical Magazine*, Series 4, 19, 19–32.

McCullagh, P. and Nelder, J. (1989). Generalized Linear Models. *Monographs on Statistics and Applied Probability*, 37. Chapman & Hall.

Mudholkar, G. and Srivastava, D. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate date. *IEEE Transactions on Reliability*, 42, 299–302.

Mudholkar, G., Srivastava, D. and Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*, 37, 436–445.

Nadarajah, S. (2005) The exponentiated Gumbel distribution with climate application. *Environmetrics*, 17, 13–23.

Nadarajah, S. and Gupta, A. (2007). The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin*, 59, 29–54.

Pewsey, A., Gómez, H. and Bolfarine, H. (2012). Likelihood based inference for distributions of fractional order statistics. *TEST*, 21, 775–789.

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/

Rêgo, L., Cintra, R. and Cordeiro, G. (2012). On some properties of the beta normal distribution. *Communications in Statistics - Theory and Methods*, 41, 3722–3738.

Sánchez, L., Leiva, V., Galea, M. and Saulo, H. (2020). Birnbaum-Saunders quantile regression models with application to spatial data. *Mathematics*, 8, 1000.

Segovia, F., Gómez, Y., Venegas, O. and Gómez, H. (2020). A power Maxwell distribution with heavy tails and applications. *Mathematics*, 8, 1116.

Shapiro, S. and Wilks, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.

Singh, A., Bakouch, H., Kumar, S. and Singh, U. (2018). Power Maxwell distribution: Statistical properties, Estimation and Application. arXiv, 1807.01200v1.

Smirnov, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Mathématique de l'Université de Moscow*, 2, 3-11.