A QUANTILE REGRESSION MODEL FOR BOUNDED RESPONSES BASED ON THE EXPONENTIAL-GEOMETRIC DISTRIBUTION*

Authors: PEDRO JODRÁ
– Departamento de Métodos Estadísticos,
Universidad de Zaragoza,
Zaragoza, Spain (pjodra@unizar.es)

MARÍA DOLORES JIMÉNEZ-GAMERO
– Departamento de Estadística e Investigación Operativa,
Universidad de Sevilla,
Sevilla, Spain (dolores@us.es)

Abstract:

- The paper first introduces a new two-parameter continuous probability distribution with bounded support from the extended exponential-geometric distribution. Closed-form expressions are given for the moments, moments of the order statistics and quantile function of the new law; it is also shown that the members of this family of distributions can be ordered in terms of the likelihood ratio order. The parameter estimation is carried out by the method of maximum likelihood and a closed-form expression is given for the Fisher information matrix, which is helpful for asymptotic inferences. Then, a new regression model is introduced by considering the proposed distribution, which is adequate for situations where the response variable is restricted to a bounded interval, as an alternative to the well-known beta regression model, among others. It relates the median response to a linear predictor through a link function. Extensions for other quantiles can be similarly performed. The suitability of this regression model is exemplified by means of a real data application.

Key-Words:

- *Exponential-geometric distribution; bounded support; regression model.*

AMS Subject Classification:

- 60E05, 62J02.

*The opinions expressed in this text are those of the authors and do not necessarily reflect the views of any organization.

1. INTRODUCTION

The development of new parametric probability distributions attracts a great deal of attention with the aim of providing useful models in many different areas. Some recent contributions can be found in Bakoban and Abu-Zinadah [7], Gómez-Déniz et al. [18] and Jodrá et al. [24], among others. With respect to models with bounded support, considerable effort has been focussed on providing alternatives to the beta distribution. A prominent alternative is the two-parameter Kumaraswamy distribution introduced by Kumaraswamy [28] and thoroughly studied by Jones [25]. Other less known two-parameter models are the transformed Leipnik distribution (see Jorgensen [26, pp. 196–197]) and the recently introduced Log–Lindley law (see Gómez-Déniz et al. [17] and Jodrá and Jiménez-Gamero [23]). There are more proposals such as the four-parameter Kumaraswamy Weibull distribution (Cordeiro et al. [10]) and the five-parameter Kumaraswamy generalized gamma distribution (Pascoa et al. [35]), that present the drawback of having a high number of parameters and in these cases the parameter estimation often presents some difficulties.

This paper introduces a new two-parameter probability distribution with bounded support derived from the extended exponential-geometric (EEG) distribution. The EEG law is a continuous probability distribution studied by Adamidis et al. [2] to model lifetime data. More precisely, a random variable Y is said to have an EEG distribution if the probability density function (pdf) is given by

$$f_Y(y;\alpha,\beta) = \frac{\alpha(1+\beta)e^{-\alpha\, y}}{(1+\beta\, e^{-\alpha\, y})^2}, \qquad y>0, \quad \alpha>0, \ \beta>-1,$$

where α and β are the model parameters. In particular, the case $\alpha > 0$ and $\beta \in (-1, 0)$ corresponds to the exponential-geometric distribution proposed by Adamidis and Loukas [3]. A generalization of the EEG law is the three-parameter Weibull-geometric distribution introduced by Barreto-Souza et al. [8].

From the EEG distribution, we define a new random variable X with support in the standard unit interval $(0, 1)$ by means of the transformation $X = \exp(-Y)$. It is easy to check that X has the following pdf and cumulative distribution function (cdf),

$$f(x;\alpha,\beta) = \frac{\alpha(1+\beta)x^{\alpha-1}}{(1+\beta\, x^{\alpha})^2}, \qquad 0<x<1, \quad \alpha>0, \ \beta>-1, \tag{1.1}$$

and

$$F(x;\alpha,\beta) = \frac{(1+\beta)x^{\alpha}}{1+\beta\, x^{\alpha}}, \qquad 0<x<1, \quad \alpha>0, \ \beta>-1,$$

respectively. In the sequel, the random variable defined by (1.1) will be referred to as the Log-extended exponential-geometric (LEEG) distribution. The LEEG distribution presents an advantage with respect to the beta distribution since it does not include special functions in its formulation. Figure 1 represents the pdf of X for several values of the parameters. It is interesting to note that the special case $\beta = 0$ corresponds to the power function distribution, which includes the uniform distribution for $\alpha = 1$.

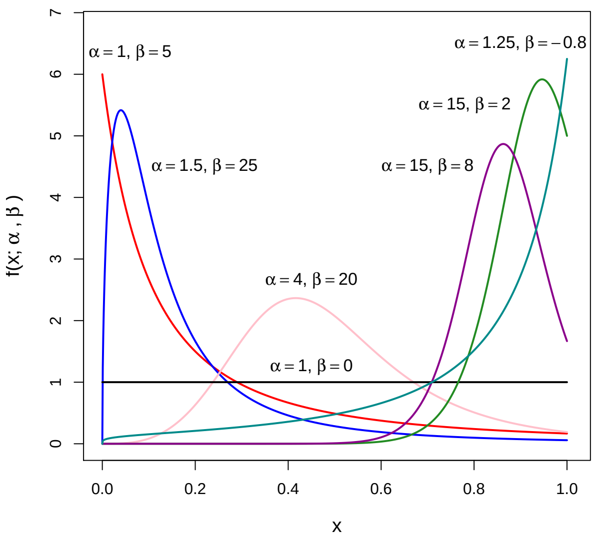

Figure 1: $f(x; \alpha, \beta)$ for different values of α and β.

Clearly, the LEEG distribution can be used to model real data taking values in the unit interval. Furthermore, as a linear transformation $(b - a)X + a$ moves a random variable X defined on $(0, 1)$ to any other bounded support (a, b), with $a < b$, the LEEG law can be extended to any bounded domain in a straightforward manner, so there is no need to explain such an extension.

On the basis of the proposed distribution, we introduce a new regression model which assumes that the response variable takes values in the standard unit interval, as an alternative to the well-known beta regression model (see Ferrari and Cribari-Neto [15]). Other regression models for bounded responses can be found in [33, 34, 36]. Regression models usually express a location measure of a distribution as a function of covariates. The location measure is commonly taken the mean (which is the case of classical regression models) or some quantile (which is the case of quantile regression, see, for example, the book by Koenker [27]). With this aim, it is noted that the LEEG distribution can be easily reparametrized in terms of any of its quantiles. As the median is a robust central tendency measure, we choose to reparametrize the LEEG law with its median and construct the associated regression model, which relates the median response to a linear predictor through a link function. Nevertheless, it will become evident that any other quantile could be used.

The literature on parametric quantile regression is rather scarce. An example is the parametric regression quantile in Noufaily and Jones [32], designed for a positive response, while our proposal is for a bounded response. In addition to this evident distinctive feature, the main difference between our approach and that in [32] lies in the following: Noufaily and Jones [32] assume a distribution for the response (specifically, the generalized gamma with three parameters) and consider parametric forms for the dependence of the parameters (or some subset of them) on the covariate (they only assume a unique covariate, although their proposal can be extended to more covariates); then they replace the parameters in the expression of the quantile function of the assumed model by the fitted regression equations for the parameters. By contrast, we reparametrize the distribution in terms of the median (although we could consider any other quantile) and assume a parametric form for the dependence of the median on the covariates (we do not limit the number of covariates). In our proposal, only one of the parameters is allowed to depend on the covariates, but it would be an obvious extension to express both of them as functions of the covariates. Note that our strategy is closer, in spirit, to Koenker [27], which assumes a regression model for a quantile; if the quantile is changed then the regression model also changes. In our scheme, if the distribution is parametrized in terms of another quantile (different from the median), the model parameters will change. On the contrary, in Noufaily and Jones [32] the model parameters are the same for each quantile since they do not fit a genuine quantile regression model, they just allow the distribution parameters to vary with the covariates and then replace them in the expression of the quantile function.

The remainder of this paper is organized as follows. In Section 2, some statistical properties of the LEEG distribution are studied. Precisely, it is shown that the LEEG law can be derived as the distribution of the minimum or maximum of a geometric random number of independent random variables with power function distribution, the moments, as well as the moments of the order statistics, can be expressed analytically in terms of the Lerch transcendent function, the quantile function can be given in closed form and the members of the new family of distributions can be ordered in terms of the likelihood ratio order. For the sake of clarity, the proofs of this section are deferred to Appendix B. Section 3 deals with the parameter estimation problem. Specifically, the method of maximum likelihood is theoretically and numerically studied. In addition, an explicit expression for the Fisher information matrix is obtained, which is useful for asymptotic inferences on the parameters. The proof of these results is deferred to Appendix C. Some numerical results studying the finite sample performance of the maximum likelihood estimators as well a real data set application are also displayed in this section. Section 4 shows how to construct a regression model for bounded responses on the basis of the LEEG distribution. A real data application demonstrates that such model may be more appropriate than others previously proposed. For the sake of completeness, Appendix A presents a known result concerning a logarithmic integral, which is used to provide unified proofs in Appendices B and C.

2. Statistical properties

This section studies some statistical properties of the LEEG distribution. Specifically, an stochastic representation is provided together with the shape of the pdf, the computation of moments, the computer-generation of pseudo-random data and the computation of moments of the order statistics. In all cases, closed-form expressions are given. Additionally, it is shown that the new family of distributions can be ordered in terms of the likelihood ratio order.

2.1. Stochastic representation

The LEEG distribution has been defined in (1.1) via an exponential transformation of the EEG distribution. It should be noted that the LEEG law can also be derived as follows. Let N be a random variable having a geometric distribution with probability mass function (pmf) given by

$$P(N=n)=\left(1-\frac{1}{1+\beta}\right)^{n-1}\frac{1}{1+\beta},\qquad n=1,2,\ldots,$$

with $\beta>0$. Let M be a random variable having a geometric distribution with pmf given by

$$P(M=m)=(-\beta)^{m-1}(1+\beta),\qquad m=1,2,\ldots,$$

with $\beta\in(-1,0)$. Let $T_1,T_2,\ldots$ be independent identically distributed random variables having T_i a power function distribution with parameter $\alpha>0$, that is, its cdf is given by $F_{T_i}(t;\alpha)=t^{\alpha}$, $0<t<1$. Assume that N and M are independent of T_i, $i=1,2,\ldots$.

Proposition 2.1. (i) *The random variable* $V=\min\{T_1,T_2,\ldots,T_N\}$ *has a LEEG distribution with parameters* $\alpha>0$ *and* $\beta>0$. (ii) *The random variable* $W=\max\{T_1,T_2,\ldots,T_M\}$ *has a LEEG distribution with parameters* $\alpha>0$ *and* $\beta\in(-1,0)$.

2.2. Shape and mode

As it can be seen from Figure 1, the pdf of the LEEG distribution has a wide variety of shapes. The next result characterizes the shape of the pdf in terms of the parameter values.

Proposition 2.2. *Let* X *be a LEEG distribution with parameters* $\alpha>0$ *and* $\beta>-1$.

(i) *For any* $\alpha > 1$, *if* $\beta > (\alpha - 1)/(1 + \alpha)$ *then* X *has a mode at* $x = \left(\frac{\alpha - 1}{(1+\alpha)\beta}\right)^{1/\alpha}$ *and if* $\beta \in (-1, (\alpha-1)/(1+\alpha)]$ *then* (1.1) *is an increasing function.*

(ii) *For any* $0 < \alpha < 1$, *if* $\beta \in (-1, (\alpha-1)/(1+\alpha))$ *then* (1.1) *has a minimum at* $x = \left(\frac{\alpha - 1}{(1+\alpha)\beta}\right)^{1/\alpha}$ *and if* $\beta \geq (\alpha-1)/(1+\alpha)$ *then* (1.1) *is a decreasing function.*

(iii) *If* $\alpha = 1$ *and* $\beta = 0$, *then* (1.1) *is the pdf of the uniform distribution on* $(0, 1)$.

2.3. Moments

The moments of X can be expressed in closed form in terms of the Lerch transcendent function, Φ. Remind that Φ is defined as the analytic continuation of the series

$$\Phi(z, \lambda, v) = \sum_{i=0}^{\infty} \frac{z^i}{(i+v)^\lambda},$$

which converges for any real number $v > 0$ if z and λ are any complex numbers with either $|z| < 1$ or $|z| = 1$ and $\mathrm{Re}(\lambda) > 1$ (see Apostol [5] for further details).

Proposition 2.3. *Let* X *be a LEEG distribution with parameters* $\alpha > 0$ *and* $\beta > -1$. *The moments of* X *are given by*

$$E[X^k] = 1 - \frac{(1+\beta)k}{\alpha}\,\Phi\left(-\beta, 1, 1 + \frac{k}{\alpha}\right), \qquad k = 1, 2, \ldots. \tag{2.1}$$

It is interesting to note that the Lerch transcendent function is available in computer algebra systems such as Maple (function `LerchPhi(z,λ,v)`) and Mathematica (function `LerchPhi[z,λ,v]`). Accordingly, usual statistical measures involving $E[X^k]$ can be efficiently computed from Eq. (2.1).

2.4. Quantile function

An interesting advantage of the LEEG distribution with respect to the beta distribution is that the cdf of X is readily invertible.

Proposition 2.4. *The quantile function of the LEEG distribution with parameters* $\alpha > 0$ *and* $\beta > -1$ *is given by*

$$F^{-1}(u; \alpha, \beta) = \left(\frac{u}{1 + \beta - \beta\, u}\right)^{1/\alpha}, \qquad 0 < u < 1.$$

From Proposition 2.4, the quartiles of the LEEG law are given by

$$Q_1 = \left(\frac{1}{4+3\beta}\right)^{1/\alpha}, \quad Q_2 = \left(\frac{1}{2+\beta}\right)^{1/\alpha}, \quad Q_3 = \left(\frac{3}{4+\beta}\right)^{1/\alpha}.$$

The explicit expression in Proposition 2.4 is helpful in simulation studies because pseudo-random data from the LEEG distribution can be generated by computer using the inverse transform method.

2.5. Order statistics

Next, analytical expressions to compute the moments of the order statistics are provided. To this end, it is shown that the moments of the largest order statistic of the LEEG law can be given in terms of a finite sum involving the Lerch transcendent function Φ and the generalized Stirling numbers of the first kind R_n^j (see Appendix A for the definition and calculation of these numbers).

Let $X_1, \dots, X_n$ be a random sample of size n from the LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. Let $X_{1:n} \le X_{2:n} \le \dots \le X_{n:n}$ be the order statistics obtained by arranging X_i, $i = 1, \dots, n$, in non-decreasing order of magnitude. For any $n = 1, 2, \dots$ and $k = 1, 2, \dots$, denote by $E[X_{r:n}^k]$ the kth moment of $X_{r:n}$, $r = 1, \dots, n$.

Proposition 2.5. *Let $X_1, \dots, X_n$ be a random sample of size n from a LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. Let $X_{n:n}$ be the largest order statistic. Then*

$$E[X_{n:n}^k] = \frac{(1+\beta)^n}{\Gamma(n)} \sum_{j=0}^{n} R_n^j(k/\alpha, 1)\Phi\left(-\beta, 1-j, n+\frac{k}{\alpha}\right), \qquad k = 1, 2, \dots .$$

The result in Proposition 2.5 is useful to evaluate the moments of $X_{r:n}$, for $r = 1, \dots, n-1$, thanks to the following well-known formula (see, for example, David and Nagaraja [13, p. 45])

$$E[X_{r:n}^k] = \sum_{j=r}^{n} (-1)^{(j-r)} \binom{j-1}{r-1} \binom{n}{j} E[X_{j:j}^k], \qquad r = 1, \dots, n-1.$$

2.6. Stochastic orderings

To conclude Section 2, it is shown that the members of the new distribution can be ordered in terms of the likelihood ratio order, which is defined as follows (see, for example, Shaked and Shanthikumar [40, Chapter 1]).

Definition 2.1. Let X_1 and X_2 be two continuous random variables with pdfs f_1 and f_2, respectively, such that $f_2(x)/f_1(x)$ is non-decreasing over the union of the supports of X_1 and X_2. Then X_1 is said to be smaller than X_2 in the likelihood ratio order, denoted by $X_1 \leq_{LR} X_2$.

The likelihood ratio order is stronger than the hazard rate order and the usual stochastic order, which are defined as follows.

Definition 2.2. Let X_1 and X_2 be two random variables with cdfs F_1 and F_2 and hazard rates h_1 and h_2, respectively. Then

(i) X_1 is said to be stochastically smaller than X_2, denoted by $X_1 \leq_{ST} X_2$, if $F_1(x) \geq F_2(x)$ for all x.

(ii) X_1 is said to be smaller than X_2 in the hazard rate, denoted by $X_1 \leq_{HR} X_2$, if $h_1(x) \leq h_2(x)$ for all x.

The LEEG family can be ordered in the following way.

Proposition 2.6. *Let X_1 and X_2 be two random variables having a LEEG distribution with parameters (α, β_1) and (α, β_2), respectively, for some $\alpha > 0$, $\beta_1, \beta_2 > -1$. If $\beta_1 \geq \beta_2$ then $X_1 \leq_{LR} X_2$.*

As an immediate consequence of Proposition 2.6 and the well-known fact that

$$X_1 \leq_{LR} X_2 \Rightarrow X_1 \leq_{HR} X_2 \Rightarrow X_1 \leq_{ST} X_2,$$

the following corollary is stated.

Corollary 2.1. *Let X_1 and X_2 be two random variables having a LEEG distribution with parameters (α, β_1) and (α, β_2), respectively, for some $\alpha > 0$, $\beta_1, \beta_2 > -1$. If $\beta_1 \geq \beta_2$ then*

(i) $E(X_1^k) \leq E(X_2^k)$, $\forall k > 0$.

(ii) $h_1(x) \leq h_2(x)$, $\forall x \in (0, 1)$.

As a special case of Corollary 2.1 (i) it follows that, for fixed $\alpha > 0$, the mean of the LEEG distribution decreases as β increases.

3. Parameter estimation

This section considers the estimation of the parameters of the LEEG distribution. Specifically, Subsection 3.1 describes the maximum likelihood (ML) method. A closed-form expression for the Fisher information matrix is provided in Subsection 3.2. The performance of the ML method is evaluated via a Monte Carlo simulation study in Subsection 3.3. Finally, a real data application is presented in Subsection 3.4.

3.1. Maximum likelihood method

Let $X_1, \ldots, X_n$ be a random sample of size n from a LEEG distribution with unknown parameters $\alpha > 0$ and $\beta > -1$ and denote by $x_1, \ldots, x_n$ the observed values. From the likelihood function, $L(\alpha, \beta) = \prod_{i=1}^{n} f(x_i; \alpha, \beta)$, the log-likelihood function is given by

$$\log L(\alpha,\beta) = n \log \alpha + n \log(1+\beta) + (\alpha - 1) \sum_{i=1}^{n} \log x_i - 2 \sum_{i=1}^{n} \log(1 + \beta\, x_i^{\alpha}). \tag{3.1}$$

The ML estimates of α and β are the values $\hat{\alpha}$ and $\hat{\beta}$ that maximize $\log L(\alpha, \beta)$. The partial derivatives of $\log L(\alpha, \beta)$ with respect to each parameter are the following:

$$\frac{\partial}{\partial \alpha} \log L(\alpha,\beta) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log x_i - 2\beta \sum_{i=1}^{n} \frac{x_i^{\alpha} \log x_i}{1 + \beta\, x_i^{\alpha}}, \tag{3.2}$$

$$\frac{\partial}{\partial \beta} \log L(\alpha,\beta) = \frac{n}{1+\beta} - 2 \sum_{i=1}^{n} \frac{x_i^{\alpha}}{1 + \beta\, x_i^{\alpha}}. \tag{3.3}$$

The ML estimates of the parameters satisfy the system that results from equating to 0 the equations (3.2) and (3.3). Nevertheless, since such system does not have an explicit solution, in order to obtain the ML estimates it is preferable to maximize the function (3.1). Subsection 3.3 will deal with this practical issue.

Another practical point is the possible presence of extreme values in the data. Although we are assuming that the data are continuous, which implies that the probability of observing the values zero and one is null, in applications, due to rounding errors, these extreme cases may appear in the observations. By looking at the expression of the log-likelihood (3.1), the presence of ones involves no problem; on the other hand, the presence of zeroes implies that the log-likelihood cannot be calculated. In such a case, we recommend replacing all zeroes by a positive small quantity.

3.2. Fisher information matrix

Below, an analytical expression for the Fisher information matrix is given, which let us explicitly calculate the asymptotic covariance matrix of the ML estimators. To this end, the polylogarithm function, which is a particular case of the Lerch transcendent function (see Appendix A), plays an important role.

Proposition 3.1. *Let $X_1, \ldots, X_n$ be a random sample of size n from a LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. For $\alpha > 0$ and $\beta \in (-1, 0) \cup (0, \infty)$ the Fisher information matrix is given by*

$$I(\alpha, \beta) = \begin{bmatrix} \frac{n}{\alpha^2} - \frac{2n}{3\alpha^2\beta}\left\{(1+\beta)\mathrm{Li}_2(-\beta) + \beta\right\} & \frac{n(1+\beta)}{3\alpha\beta}\left(\frac{1}{(1+\beta)^2} - \frac{\log(1+\beta)}{\beta}\right) \\ \frac{n(1+\beta)}{3\alpha\beta}\left(\frac{1}{(1+\beta)^2} - \frac{\log(1+\beta)}{\beta}\right) & \frac{n}{3(1+\beta)^2} \end{bmatrix},$$

where Li_2 *denotes the polylogarithm function of order two. For $\alpha > 0$ and $\beta = 0$,*

$$I(\alpha, 0) = \begin{bmatrix} \frac{n}{\alpha^2} & -\frac{n}{2\alpha} \\ -\frac{n}{2\alpha} & \frac{n}{3} \end{bmatrix}.$$

As it is well-known, it is useful to have an explicit expression for $I(\alpha, \beta)$ since by inverting this matrix we get the asymptotic covariance matrix of the ML estimators and it can be used to approximate their standard errors. Denote by N_2 a bivariate normal distribution and by $\xrightarrow{d}$ the convergence in distribution.

Proposition 3.2. *Let $X_1, \ldots, X_n$ be a random sample of size n from a LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. Let $\widehat{\boldsymbol{\theta}}$ denote the ML estimator of $\boldsymbol{\theta} = (\alpha, \beta)$. Then,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}_2(\mathbf{0}, \Sigma),$$

where $\Sigma = \Sigma(\alpha, \beta)$ is such that for $\beta \neq 0$

$$\Sigma(\alpha, \beta) = \begin{bmatrix} -\frac{3\alpha^2\beta^4}{(1+\beta)c(\beta)} & -\frac{3\alpha\beta^2[(1+\beta)^2\log(1+\beta) - \beta]}{c(\beta)} \\ -\frac{3\alpha\beta^2[(1+\beta)^2\log(1+\beta) - \beta]}{c(\beta)} & \frac{3\beta^3(1+\beta)[2(1+\beta)\mathrm{Li}_2(-\beta) - \beta]}{c(\beta)} \end{bmatrix},$$

with

$$c(\beta) = (1+\beta)^3\log^2(1+\beta) - 2\beta(1+\beta)\log(1+\beta) + \beta^3[2\mathrm{Li}_2(-\beta) - 1] + \beta^2$$

and Li_2 *stands for the polylogarithm function of order two, and for $\beta = 0$*

$$\Sigma(\alpha, 0) = \begin{bmatrix} 4\alpha^2 & 6\alpha \\ 6\alpha & 12 \end{bmatrix}.$$

3.3. Simulation study

As discussed in Subsection 3.1, in order to obtain the ML estimates of the parameters the following optimization problem is solved

$$\begin{aligned} &\max\ \log L(\alpha,\beta) \\ &\text{s.t.}\quad \alpha > 0 \\ &\qquad\quad \beta > -1, \end{aligned} \tag{3.4}$$

where $\log L(\alpha,\beta)$ is given in Eq. (3.1). In our simulations, problem (3.4) was solved by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, available in the function `constrOptim` of the R programming language [37]. We chose the BFGS algorithm because (3.4) is an optimization problem with linear inequality constraints. The BFGS algorithm requires a starting point, which must be in the interior of the feasible region, together with the gradient function of $\log L(\alpha,\beta)$. As starting point we tried several options with little or no effect on the final solution. All numerical results in this paper were obtained by using as starting point the pair $(1,1)$.

The performance of the ML estimators was assessed via a Monte Carlo simulation study. The following notation was used. The number of random samples generated is denoted by N and the size of each random sample is denoted by n. The following quantities were computed for the simulated estimates $\hat{\alpha}_j$, $j = 1,\dots,N$:

(i) The mean: $\bar{\alpha} = (1/N)\sum_{j=1}^{N}\hat{\alpha}_j$.

(ii) The bias: $\text{Bias}(\hat{\alpha}) = \bar{\alpha} - \alpha$.

(iii) The mean-square error: $\text{MSE}(\hat{\alpha}) = (1/N)\sum_{j=1}^{N}(\hat{\alpha}_j - \alpha)^2$.

The quantities $\bar{\beta}$, $\text{Bias}(\hat{\beta})$ and $\text{MSE}(\hat{\beta})$ are analogously defined and were also computed. In particular, we generated $N = 10,000$ random samples of different sizes n for several values of α and β. Some simulation results are shown in Table 1, where it is included the mean, bias and MSE of the simulated estimates together with the asymptotic variance of the estimators calculated directly from the diagonal elements of $(1/n)\Sigma(\alpha,\beta)$, with $\Sigma(\alpha,\beta)$ given by Proposition 3.2, and denoted by $\text{Var}[\hat{\alpha}]$ and $\text{Var}[\hat{\beta}]$. From the obtained results, it can be concluded that the ML method provides acceptable estimates of the parameters, although it should be noted that the ML method tended to slightly overestimate the value of both parameters in the cases considered in the present study.

	$\alpha = 0.25$				$\beta = -0.25$				$\alpha = 1.25$				$\beta = -0.80$			
	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]
$n = 50$	0.2759	0.0259	0.0078	0.0060	−0.0663	0.1836	0.2755	0.1429	1.5817	0.3317	0.7677	0.4335	−0.7176	0.0823	0.0436	0.0170
$n = 75$	0.2664	0.0164	0.0047	0.0040	−0.1350	0.1150	0.1507	0.0952	1.4641	0.2141	0.4293	0.2890	−0.7479	0.0520	0.0214	0.0113
$n = 100$	0.2614	0.0114	0.0034	0.0030	−0.1702	0.0798	0.0994	0.0714	1.4136	0.1636	0.2892	0.2167	−0.7605	0.0394	0.0138	0.0085
$n = 200$	0.2562	0.0062	0.0016	0.0015	−0.2087	0.0411	0.0429	0.0357	1.3281	0.0781	0.1271	0.1083	−0.7812	0.0187	0.0055	0.0042
$n = 500$	0.2525	0.0025	0.0006	0.0006	−0.2341	0.0158	0.0153	0.0142	1.2798	0.0298	0.0468	0.0433	−0.7931	0.0068	0.0019	0.0017
	$\alpha = 1.0$				$\beta = 5.0$				$\alpha = 1.5$				$\beta = 10.0$			
	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]
$n = 50$	1.0333	0.0333	0.0335	0.0310	6.2783	1.2783	18.0034	8.4895	1.5515	0.0515	0.0643	0.0565	12.8102	2.8102	77.0472	31.1192
$n = 75$	1.0251	0.0251	0.0224	0.0206	5.8739	0.8739	9.9312	5.6596	1.5341	0.0341	0.0412	0.0376	11.8078	1.8078	37.2463	20.7461
$n = 100$	1.0173	0.0173	0.0162	0.0155	5.6031	0.6031	6.0647	4.2447	1.5230	0.0230	0.0295	0.0282	11.1966	1.1966	23.6258	15.5596
$n = 200$	1.0081	0.0081	0.0079	0.0077	5.2955	0.2955	2.6045	2.1223	1.5126	0.0126	0.0143	0.0141	10.6480	0.6480	9.7978	7.7798
$n = 500$	1.0044	0.0044	0.0031	0.0031	5.1294	0.1294	0.9255	0.8489	1.5044	0.0044	0.0057	0.0056	10.2248	0.2248	3.4507	3.1119
	$\alpha = 15.0$				$\beta = 2.0$				$\alpha = 15.0$				$\beta = 10.0$			
	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]
$n = 50$	15.6584	0.6584	10.9093	9.4940	2.6075	0.6075	4.0518	2.0166	15.4493	0.4493	6.2381	5.6530	12.6339	2.6339	71.0531	31.1192
$n = 75$	15.4822	0.4822	7.0518	6.3293	2.4081	0.4081	2.1360	1.3444	15.2776	0.2776	4.0164	3.7687	11.6095	1.6095	35.2727	20.7461
$n = 100$	15.3295	0.3295	5.1931	4.7470	2.2870	0.2870	1.4440	1.0083	15.2294	0.2294	3.0211	2.8265	11.2339	1.2339	24.5033	15.5596
$n = 200$	15.1849	0.1849	2.4365	2.3735	2.1432	0.1432	0.5975	0.5041	15.1187	0.1187	1.4387	1.4132	10.5759	0.5759	9.6097	7.7798
$n = 500$	15.0682	0.0682	0.9674	0.9494	2.0565	0.0565	0.2204	0.2016	15.0389	0.0389	0.5740	0.5653	10.2267	0.2267	3.3672	3.1119

Table 1: ML estimates of α and β.

3.4. A real data application

In this subsection, a real data set illustrates the practical usefulness of the LEEG distribution by showing that it may be a more appropriate model than other distributions with support in the standard unit interval.

The data set is available from the personal website of Professor E.W. Frees[1] and consists of 73 observations on 7 variables. The data were collected from a questionnaire carried out with the purpose of relating cost effectiveness to management philosophy of controlling the company's exposure to various property and casualty losses, after adjusting for company effects such as size and industry type. These data have been previously analyzed by Schmit and Roth [38], Frees [16, Chapter 6], Gómez-Déniz et al. [17] and Jodrá and Jiménez-Gamero [23].

In this section, interest is centered on the variable FIRMCOST (divided by 100), which is a measure of the cost effectiveness of the risk management practices of the firm. Based on Subsection 3.1, the LEEG law was fitted to the variable FIRMCOST/100. The ML estimates obtained were $\hat{\alpha} = 1.4322$ and $\hat{\beta} = 52.1069$. It can also be checked that the correlation coefficient between the theoretical and the empirical cumulative probabilities is 0.9956.

Additionally, we applied the following goodness-of-fit tests based on the empirical cdf: the Cramér von Mises statistic W^2, the Watson statistic U^2, the Anderson–Darling statistic A^2 and the Kolmogorov–Smirnov statistic D. A detailed definition together with simple formulae for computing these statistics can be found in D'Agostino and Stephens[12, Chapter 4]. To get the p-values we applied a parametric bootstrap generating 10,000 bootstrap samples (see Stute et al. [41] and Babu and Rao [6] for full details). We also applied two test based on the empirical characteristic function [19, 20] by using the integral transformation, as proposed in Meintanis et al. [30], taking as weight functions: the standard normal law, FC_1, and the pdf $w(t) = \{1 - \cos(t)\}/\pi t^2$, which is the choice recommended in Epps and Pulley [14] (see also Section 4 in [20]), FC_2. The results are shown in Table 3.4 and suggest that the LEEG law provides a satisfactory fit.

	W^2	U^2	A^2	D	FC_1	FC_2
Statistic value:	0.0571	0.0571	0.5133	0.0626	0.0011	0.1142
p-value:	0.2610	0.2610	0.1363	0.5320	0.1164	0.2663

Table 2: Goodness-of-fit tests.

The LEEG fitting was compared to the ones provided by other two-parameter

[1] `http://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/data.html`, filename: RiskSurvey.

distributions used to model data in the unit interval. Specifically, we considered the beta, Kumaraswamy, Log–Lindley and transformed Leipnik distributions. In order to compare these models, we calculated the Akaike information criterion AIC (see Akaike [4]), the consistent Akaike information criterion CAIC (see Bozdogan [9]) and the Bayesian information criterion BIC (see Schwarz [39]), which are defined as follows, $\text{AIC} = 2m - 2\log L$, $\text{CAIC} = m(1 + \log n) - 2\log L$ and $\text{BIC} = m[\log n - \log(2\pi)] - 2\log L$, respectively, where m is the number of parameters, n is the sample size and L denotes the maximized value of the likelihood function. As it is well-known, the model with lowest values of AIC, CAIC and BIC is preferred. For each fitted distribution, Table 3.4 shows the ML estimated parameters together with the log-likelihood, AIC, CAIC and BIC values. Looking at Table 3.4, the LEEG distribution provides the best fit. Moreover, the Vuong test [42] was applied to compare the LEEG model to the beta, Kumaraswamy, Log–Lindley and transformed Leipnik distributions. In the four cases the Vuong statistic was very close to 0, so suggesting that all these distributions can be considered equally close to the data. In this regard, we consider the LEEG distribution an attractive alternative to the aforesaid models.

Distribution	ML estimates	$\log L$	AIC	CAIC	BIC
LEEG(α, β)	$\hat{\alpha} = 1.4322$				
$f(x;\alpha,\beta) = \dfrac{\alpha(1+\beta)x^{\alpha-1}}{(1+\beta x^{\alpha})^2}$	$\hat{\beta} = 52.1069$	93.63	−183.26	−176.68	−182.35
Beta(a, b)	$\hat{a} = 0.6125$				
$f(x;a,b) = \dfrac{1}{\mathrm{B}(a,b)}\, x^{a-1}(1-x)^{b-1}$	$\hat{b} = 3.7978$	76.11	−148.23	−141.65	−147.32
Kumaraswamy(a, b)	$\hat{a} = 0.6648$				
$f(x;a,b) = abx^{a-1}(1-x^a)^{b-1}$	$\hat{b} = 3.4407$	78.65	−153.30	−146.72	−152.40
Log–Lindley(a, b)	$\hat{a} = 0.6906$				
$f(x;a,b) = a[b + a(b-1)\log x]x^{a-1}$	$\hat{b} = 0.0231$	76.60	−149.20	−142.62	−148.30
Transformed Leipnik(μ, λ)	$\hat{\mu} = 0.0261$				
$f(x;\mu,\lambda) = \dfrac{[x(1-x)]^{-\frac{1}{2}}}{\mathrm{B}\left(\frac{\lambda+1}{2}, \frac{1}{2}\right)}\left(1 + \dfrac{(x-\mu)^2}{x(1-x)}\right)^{-\frac{\lambda}{2}}$	$\hat{\lambda} = 6.4061$	80.51	−157.02	−150.43	−156.11

Table 3: Fitted distribution, ML estimates, log-likelihood, AIC, CAIC and BIC.

4. A regression model for bounded responses

Regression models are commonly used to model the mean of a response variable as a function of a set of covariates (also called independent variables or regressors). As shown in Proposition 2.3, the moments of the LEEG distribution can be expressed in terms of the Lerch transcendent function, which implies that the mean does not possess a simple expression. This fact makes difficult to build a regression model which relates the mean response with covariates. By contrast, the expression of the quantiles of the LEEG distribution is quite tractable, so our proposal is to use them to construct a regression model. In principle, we could choose any quantile, but since the median is a robust measure of location and, in this regard, it is considered as a competitor of the mean, we will choose the median.

As a first step towards the construction of the regression model, the LEEG distribution is reparametrized in terms of the median Q_2 by equating Q_2 to a new parameter θ and solving the resultant equation for β. The resulting pdf is

$$f(x;\alpha,\theta)=\frac{\alpha\theta^{\alpha}(1-\theta^{\alpha})x^{\alpha-1}}{[\theta^{\alpha}+(1-2\theta^{\alpha})x^{\alpha}]^2},\qquad 0<x<1,\ \alpha>0,\ 0<\theta<1. \tag{4.1}$$

It should be noted that all properties studied for the parametrization (1.1) carry over for the above one with $\beta=(1-2\theta^{\alpha})/\theta^{\alpha}$.

Let $X_1,\dots,X_n$ be n independent random variables and denote by x_1, $\dots$, x_n the observed values. Assume that each X_i has pdf $f(x;\alpha,\theta_i)$ given by (4.1). Suppose that the median of X_i satisfies $\theta_i=g(z_i^t\gamma)$, $i=1,\dots,n$, where $z_i=(z_{i1},\dots,z_{ik})^t$ is the vector of covariates associated to the response x_i, $\gamma=(\gamma_1,\dots,\gamma_k)$ is an unknown vector of regression coefficients and g is the link function. It is assumed that the link function g is a strictly monotonic and twice differentiable function. There are several possible choices for g satisfying the required conditions, such as the logit, probit, log-log, Cauchy, etc.

From Eq. (4.1), the log-likelihood function of the model with covariates is given by

$$\begin{aligned}\ell(\alpha,\gamma) &= n\log\alpha+(\alpha-1)\sum_{i=1}^{n}\log x_i+\alpha\sum_{i=1}^{n}\log\theta_i+\sum_{i=1}^{n}\log(1-\theta_i^{\alpha})\\ &\quad -2\sum_{i=1}^{n}\log(\theta_i^{\alpha}+x_i^{\alpha}-2\theta_i^{\alpha}x_i^{\alpha}).\end{aligned}$$

The derivatives of $\ell(\alpha,\gamma)$ with respect to each parameter, which are required to compute the ML estimates of the parameters, are given by

$$\begin{aligned}\frac{\partial}{\partial\alpha}\ell(\alpha,\gamma) &= \frac{n}{\alpha}+\sum_{i=1}^{n}\log x_i+\sum_{i=1}^{n}\log\theta_i-\sum_{i=1}^{n}\frac{\theta_i^{\alpha}\log\theta_i}{1-\theta_i^{\alpha}}\\ &\quad -2\sum_{i=1}^{n}\frac{\theta_i^{\alpha}\log\theta_i+x_i^{\alpha}\log x_i-2x_i^{\alpha}\theta_i^{\alpha}(\log\theta_i+\log x_i)}{\theta_i^{\alpha}+(1-2\theta_i^{\alpha})x_i^{\alpha}},\\ \frac{\partial}{\partial\gamma_r}\ell(\alpha,\gamma) &= \alpha\sum_{i=1}^{n}\frac{1}{\theta_i}\frac{\partial}{\partial\gamma_r}\theta_i-\alpha\sum_{i=1}^{n}\frac{\theta_i^{\alpha-1}}{1-\theta_i^{\alpha}}\frac{\partial}{\partial\gamma_r}\theta_i-2\alpha\sum_{i=1}^{n}\frac{(1-2x_i^{\alpha})\theta_i^{\alpha-1}}{\theta_i^{\alpha}+(1-2\theta_i^{\alpha})x_i^{\alpha}}\frac{\partial}{\partial\gamma_r}\theta_i,\end{aligned}$$

for $r=1,\dots,k$. The derivative $\frac{\partial}{\partial\gamma_r}\theta_i$ will depend on the chosen link function. For example, if it is considered the logit link, which is given by

$$\theta_i=\frac{\exp(z_i^t\gamma)}{1+\exp(z_i^t\gamma)},$$

then

$$\frac{\partial}{\partial \gamma_r} \theta_i = \theta_i(1-\theta_i) z_{ir}, \qquad i = 1, \ldots, n, \; r = 1, \ldots, k.$$

As in most regression models, for the proposed model it is possible to evaluate the marginal effects that each covariate has on the conditional median, given the covariates, by calculating (see, for example, [36, § 2.2.3])

$$\delta_{ij} = \frac{\partial \theta_i}{\partial z_{ij}} = \theta_i(1-\theta_i)\gamma_j, \qquad i = 1, \ldots, n, \; j = 1, \ldots, k. \tag{4.2}$$

This marginal effect indicates that a small change in the jth covariate, say ν, increases or decreases the conditional median θ_i by a quantity $\delta_{ij}\nu + o(\nu)$. As a summary measure of all these $k \times n$ effects, one can calculate the average marginal effects that each covariate has on the conditional median by evaluating the above derivative at $\bar{\theta} = \theta(\bar{z})$, obtaining

$$\bar{\delta_j} = \frac{\partial \bar{\theta}}{\partial z_{ij}} = \bar{\theta}(1-\bar{\theta})\gamma_j, \qquad j = 1, \ldots, k.$$

For the practical use of these quantities, all parameters must be replaced by estimators.

As an application, we analyze the data set considered in Subsection 3.4. The full data set consists of 73 observations on 7 variables: FIRMCOST, previously studied; ASSUME, the per occurrence retention amount as a percentage of total assets; CAP, which indicates that the firm owns a captive insurance company; SIZELOG, the logarithm of total assets; INDCOST, a measure of the firm industry risk; CENTRAL, a measure of the importance of the local managers in choosing the amount of risk to be retained; and SOPH, a measure of the degree of importance in using analytical tools.

As response variable we took x =FIRMCOST/100 and the other variables were considered as covariates. An intercept was also included in the regression model. The data were analyzed using the beta regression model and the LEEG regression model presented in this paper. Following [17], the logit link was considered in all cases. This data set was also analyzed in [17] by using the Log–Lindley regression model. Nevertheless, due to the problems observed in [23], we will not consider such model in our study. The response variables x and $1-x$ were both studied. For the analysis of the beta regression model we used the package `betareg` (see [11]) of the R programming language [37]; to obtain the ML estimates of the parameters in the LEEG regression model we used the function `optim` of the R language. Table 4 reports the value of the log-likelihood function for the models under consideration.

As expected, the values of the log-likelihood function for x and $1-x$ for the beta fitting are identical, since if a random variable X has a beta law with parameters a and b, then $1-X$ has a beta law with parameters b and a. On the other hand, the values of the log-likelihood for x and $1-x$ for the LEEG

	x	$1-x$
Beta	87.72	87.72
LEEG	122.48	103.33

Table 4: Values the of the log-likelihood with covariates for the responses x and $1-x$.

fittings differ, since these laws do not possess the aforementioned property of the beta distribution. Hence, if the value of the log-likelihood function is used as a criterion for comparison, we see that the best fit is obtained for the LEEG regression model for the response variable x.

In addition, we applied the Vuong test [42] for testing the null hypothesis that both models are equally close to the actual model, against the alternative that one model is closer than the other. The test rejected the null hypothesis in favor of the hypothesis that the LEEG regression model is closer than the beta regression model (the p-value is 0.0012).

We also compared the Pearson residuals of both models. Figure 2 displays them.

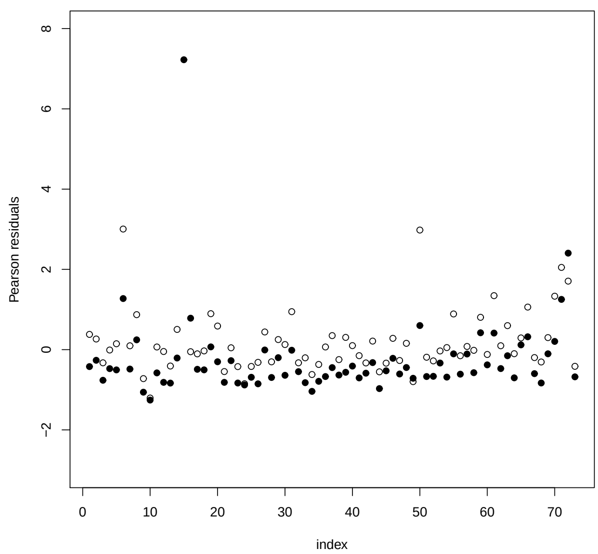

Figure 2: Pearson residuals for the beta regression model (black) and the LEEG regression model (white).

Table 5 displays the estimation results for the LEEG regression model with response variable x. The standard errors of the parameter estimates were approximated by means of the square root of the diagonal elements of the negative of the observed information matrix, that is, the matrix whose entries are the second order derivatives of the log-likelihood (its expression is omitted for the sake of brevity). The p-values of the Wald test for testing the nullity of each parameter were calculated by using the normal approximation. From these results, it can be inferred that the covariates SIZELOG and INDCOST have a significant non-null effect on the response variable. These two covariates have the largest average marginal effects, negative for SIZELOG, indicating that an increase in SIZELOG diminishes the median of the response variable, and positive for INDCOST, indicating that an increase in INDCOST increases the median of the response variable.

Before ending this section we would like to remark that the lack of a simple expression for the quantiles of the classic beta distribution hampers the development of a quantile regression based on it.

Parameter	Estimate	S.E.	t-Wald	p-value	a.m.e.
α	2.20257	0.22661	9.71975	0.0000	
Intercept	3.98741	1.21128	3.29191	0.0010	
ASSUME	−0.01234	0.01216	−1.01482	0.3102	−0.00080
CAP	−0.05257	0.22327	−0.23545	0.8139	−0.00340
SIZELOG	−0.90907	0.12466	−7.29242	0.0000	−0.05884
INDCOST	2.34318	0.62296	3.76138	0.0002	0.15166
CENTRAL	−0.13648	0.08385	−1.62766	0.1036	−0.00883
SOPH	0.00932	0.01965	0.47398	0.6355	0.00060

Table 5: Parameter estimates for the LEEG regression model with response x and average marginal effects (a.m.e.).

Appendix A

This appendix is devoted to present a known result concerning a logarithmic integral. Such result will be used to solve in a unified manner the integrals arising in Appendices B and C.

For any real numbers $a \geq 0$, $s \geq 1$ and $z > -1$, denote by

$$\Gamma_n(z,s,a) = \int_0^1 \frac{u^a \log^{s-1}(1/u)}{(1+zu)^{n+1}}\, du, \qquad n = 1, 2, \ldots. \tag{4.3}$$

Jodrá and Jiménez-Gamero [22] showed that $\Gamma_n(z,s,a)$ can be expressed as a finite sum involving the Lerch transcendent function together with the generalized Stirling numbers of the first kind. To be more precise, Mitrinović [31] defined the generalized Stirling numbers of the first kind, $R_n^j(\rho,\tau)$, by means of the following generating function

$$\prod_{j=0}^{n-1}(w-\rho-\tau j)=\sum_{j=0}^{n} R_n^j(\rho,\tau)w^j,$$

where n is a non-negative integer and ρ, τ are complex numbers with $\tau \neq 0$. Mitrinović [31] expressed these numbers in terms of the best-known signed Stirling numbers of the first kind $R_n^j(0,1)$ (see Abramowitz and Stegun [1, p. 824])

$$R_n^j(\rho,\tau)=\sum_{k=0}^{n-j}\binom{j+k}{k}(-1)^k\rho^k\tau^{n-j-k}R_n^{j+k}(0,1), \qquad \rho\neq 0, \tag{4.4}$$

which is important from a computational point of view since the numbers $R_n^j(0,1)$ are available in most computer algebra systems. Jodrá and Jiménez-Gamero [22, Theorem 2.1] established that for any $a \geq 0$, $s \geq 1$ and $z > -1$,

$$\Gamma_n(z,s,a)=\frac{\Gamma(s)}{\Gamma(n+1)}\sum_{j=0}^{n} R_n^j(a-n+1,1)\Phi(-z,s-j,a+1), \qquad n=1,2,\ldots, \tag{4.5}$$

which in the special case $z=0$ becomes $\Gamma_n(0,s,a)=\Gamma(s)/(a+1)^s$. Additionally, (4.5) can be expressed in terms of the polylogarithm function if $a=0,1,\ldots,n-1$ (see [22, Corollary 2.6] and also [21]), specifically,

$$\Gamma_n(z,s,a)=\frac{\Gamma(s)}{(-z)^{a+1}\Gamma(n+1)}\sum_{j=1}^{n} R_n^j(a-n+1,1)\mathrm{Li}_{s-j}(-z). \tag{4.6}$$

It is interesting to note that the Lerch transcendent function includes as a particular case the polylogarithm function, more precisely, $\mathrm{Li}_\lambda(z)=z\Phi(z,\lambda,1)$ (see Apostol [5]). In particular, the case $\lambda=1$ corresponds to the natural logarithm, $\mathrm{Li}_1(z)=-\log(1-z)$, and the case $\lambda=2$ is known as dilogarithm or polylogarithm function of order two.

Appendix B

Here, we give the proofs of the results stated in Section 2.

Proof of Proposition 2.1: The conditional cdf of the random variable $V|N=n$ is $F_{V|N=n}(v;\alpha)=1-(1-v^\alpha)^n$, with $0<v<1$, $\alpha>0$ and $n=1,2,\ldots$. Then, it is clear the following

$$P(V\leq v, N=n)=[1-(1-v^\alpha)^n]\left(1-\frac{1}{1+\beta}\right)^{n-1}\frac{1}{1+\beta},$$

where $\beta>0$. Hence, part (i) follows from the fact that the marginal cdf of V is

$$F_V(v;\alpha,\beta)=\sum_{n=1}^{\infty}P(V\leq v,N=n)=\frac{(1+\beta)v^\alpha}{1+\beta\, v^\alpha}, \quad 0<v<1,\ \alpha>0,\ \beta>0.$$

The proof of part (ii) follows a similar pattern. The conditional cdf of $W|M = m$ is $F_{W|M=m}(w;\alpha) = w^{\alpha\, m}$, with $0 < w < 1$, $\alpha > 0$ and $m = 1, 2, \ldots$. Therefore, $P(W \leq w, M = m) = w^{\alpha\, m}(-\beta)^{m-1}(1+\beta)$, where $\beta \in (-1, 0)$. Finally, considering that $F_W(w;\alpha,\beta) = \sum_{m=1}^{\infty} P(W \leq w, M = m)$ the result is obtained. □

Proof of Proposition 2.2: The first derivative of (1.1) is given by

$$\frac{\partial}{\partial x} f(x;\alpha,\beta) = -\frac{\alpha(1+\beta)}{(1+\beta x^\alpha)^3}\,[\beta(1+\alpha)x^\alpha - (\alpha-1)]. \tag{4.7}$$

The solution of the equation $(\partial/\partial x) f(x;\alpha,\beta) = 0$ is $x_0 = \left(\dfrac{\alpha-1}{(1+\alpha)\beta}\right)^{1/\alpha}$. Moreover, after some calculations, it can be checked that

$$\frac{\partial^2}{\partial x^2} f(x;\alpha,\beta)\bigg|_{x=x_0} = -\frac{(1+\beta)(1+\alpha)^2(\alpha-1)^2}{8\alpha\beta}.$$

On the one hand, if $\alpha > 1$ and $\beta > (\alpha-1)/(1+\alpha)$ then $x_0 \in (0,1)$ and $\frac{\partial^2}{\partial x^2} f(x;\alpha,\beta)\big|_{x=x_0} < 0$ which implies that x_0 is the mode of X. In addition, from (4.7), it can be seen that (1.1) is an increasing function if $\alpha > 1$ and $\beta \in (-1, (\alpha-1)/(1+\alpha)]$ since $(\partial/\partial x) f(x;\alpha,\beta) > 0$. This proves part (i). On the other hand, if $0 < \alpha < 1$ and $\beta < (\alpha-1)/(1+\alpha)$ then $x_0 \in (0,1)$ and $\frac{\partial^2}{\partial x^2} f(x;\alpha,\beta)\big|_{x=x_0} > 0$ which implies that (1.1) achieves a minimum at x_0. It can also be checked that (1.1) is a decreasing function if $0 < \alpha < 1$ and $\beta \geq (\alpha-1)/(1+\alpha)$. This proves part (ii). Part (iii) is directly obtained from (1.1). □

Proof of Proposition 2.3: For any $k = 1, 2, \ldots$, the k-th moment of X can be computed as follows

$$E[X^k] = \int_0^1 x^k f(x;\alpha,\beta)dx = \int_0^1 x^k \frac{\alpha(1+\beta)x^{\alpha-1}}{(1+\beta\, x^\alpha)^2}\, dx = (1+\beta)\int_0^1 \frac{u^{k/\alpha}}{(1+\beta\, u)^2}\, du,$$

where in the last equality we have made the change of variable $x^\alpha = u$. Hence, the k-th moment of X can be rewritten as below

$$E[X^k] = (1+\beta)\int_0^1 \frac{u^{k/\alpha}}{(1+\beta\, u)^2}\, du = (1+\beta)\Gamma_1(\beta, 1, k/\alpha),$$

where Γ_1 is given by Eq. (4.3). Using Eq. (4.5), we have

$$\Gamma_1(\beta, 1, k/\alpha) = R_1^1(k/\alpha, 1)\Phi\left(-\beta, 0, 1+\frac{k}{\alpha}\right) + R_1^0(k/\alpha, 1)\Phi\left(-\beta, 1, 1+\frac{k}{\alpha}\right).$$

By virtue of (4.4), $R_1^1(k/\alpha, 1) = 1$ and $R_1^0(k/\alpha, 1) = -k/\alpha$ since $R_1^0(0,1) = 0$ and $R_1^1(0,1) = 1$. Moreover, $\Phi(-\beta, 0, 1+k/\alpha) = 1/(1+\beta)$. Hence, the result is obtained. □

Proof of Proposition 2.4: The result is obtained directly by solving the equation $F(x;\alpha,\beta)=u$, $0<u<1$, with respect to the variable x. □

Proof of Proposition 2.5: For any $n=1,2,\ldots$, the k-th moment of the largest order statistic $X_{n:n}$ is given by

$$E[X_{n:n}^k]=n\int_0^1 x^k\,[F(x;\alpha,\beta)]^{n-1}\,f(x;\alpha,\beta)dx=n(1+\beta)^n\int_0^1\frac{u^{k/\alpha+n-1}}{(1+\beta u)^{n+1}}\,du,$$

where in the second equality we have made the change of variable $u=x^\alpha$. Now, taking into account Eq. (4.3), $E[X_{n:n}^k]$ can be written as follows

$$E[X_{n:n}^k]=n(1+\beta)^n\Gamma_n\left(\beta,1,\frac{k}{\alpha}+n-1\right).$$

Finally, the claimed result follows by applying Eq. (4.5) in the above equation. □

Proof of Proposition 2.6: Let us denote $v(x)=\frac{\partial}{\partial x}\log\left(\frac{f(x;\alpha,\beta_2)}{f(x;\alpha,\beta_1)}\right)=\frac{num}{den}$, where $den=x(1+\beta_1x^\alpha)(1+\beta_2x^\alpha)$ and $num=2\alpha x^\alpha(\beta_1-\beta_2)$. It can be checked that $den>0$ for any $x\in(0,1)$, $\alpha>0$ and $\beta_1,\beta_2>-1$ and also that $num\geq 0$ for any $x\in(0,1)$ and $\alpha>0$ if and only if $\beta_1\geq\beta_2$. Since $v(x)\geq 0$ implies that $\frac{f(x;\alpha,\beta_2)}{f(x;\alpha,\beta_1)}$ is non-decreasing in x, the result follows. □

Appendix C

Here, we give the proofs of the results presented in Subsection 3.2.

Proof of Proposition 3.1: The Hessian matrix of $\log L(\alpha,\beta)$ is defined by

$$H(\alpha,\beta)=\begin{bmatrix}\dfrac{\partial^2\log L(\alpha,\beta)}{\partial\alpha^2} & \dfrac{\partial^2\log L(\alpha,\beta)}{\partial\alpha\partial\beta}\\ \dfrac{\partial^2\log L(\alpha,\beta)}{\partial\beta\partial\alpha} & \dfrac{\partial^2\log L(\alpha,\beta)}{\partial\beta^2}\end{bmatrix},$$

with

$$\frac{\partial^2}{\partial\alpha^2}\log L(\alpha,\beta)=-\frac{n}{\alpha^2}-2\beta\sum_{i=1}^n\frac{x_i^\alpha(\log x_i)^2}{(1+\beta\,x_i^\alpha)^2},\tag{4.8}$$

$$\frac{\partial^2}{\partial\alpha\partial\beta}\log L(\alpha,\beta)=-2\sum_{i=1}^n\frac{x_i^\alpha\log x_i}{(1+\beta\,x_i^\alpha)^2},\tag{4.9}$$

$$\frac{\partial^2}{\partial\beta^2}\log L(\alpha,\beta)=-\frac{n}{(1+\beta)^2}+2\sum_{i=1}^n\frac{x_i^{2\alpha}}{(1+\beta\,x_i^\alpha)^2}.\tag{4.10}$$

From (4.8)–(4.10), the Fisher information matrix, $I(\alpha,\beta) = -E[H(\alpha,\beta)]$, is given by

$$I(\alpha,\beta) = \begin{bmatrix} \dfrac{n}{\alpha^2} + 2\beta n \displaystyle\int_0^1 \frac{x^\alpha (\log x)^2}{(1+\beta\, x^\alpha)^2}\, f(x)dx & 2n \displaystyle\int_0^1 \frac{x^\alpha \log x}{(1+\beta\, x^\alpha)^2}\, f(x)dx \\ 2n \displaystyle\int_0^1 \frac{x^\alpha \log x}{(1+\beta\, x^\alpha)^2}\, f(x)dx & \dfrac{n}{(1+\beta)^2} - 2n \displaystyle\int_0^1 \frac{x^{2\alpha}}{(1+\beta\, x^\alpha)^2}\, f(x)dx \end{bmatrix},$$

where we have used the notation $f(x)$ instead of $f(x;\alpha,\beta)$ for brevity. Below, we consider each integral expression in the elements of $I(\alpha,\beta)$. Let us first assume that $\beta \neq 0$. Making the change of variable $u = x^\alpha$ and taking into account (4.3), those integrals can be expressed as follows

$$\begin{aligned} \int_0^1 \frac{x^\alpha (\log x)^2}{(1+\beta\, x^\alpha)^2} f(x)dx &= \frac{1+\beta}{\alpha^2} \int_0^1 \frac{u(\log(1/u))^2}{(1+\beta\, u)^4}\, du = \frac{1+\beta}{\alpha^2}\, \Gamma_3(\beta,3,1), \\ \int_0^1 \frac{x^\alpha \log x}{(1+\beta\, x^\alpha)^2} f(x)dx &= -\frac{1+\beta}{\alpha} \int_0^1 \frac{u\log(1/u)}{(1+\beta\, u)^4}\, du = -\frac{1+\beta}{\alpha}\, \Gamma_3(\beta,2,1), \\ \int_0^1 \frac{x^{2\alpha}}{(1+\beta\, x^\alpha)^2} f(x)dx &= (1+\beta) \int_0^1 \frac{u^2}{(1+\beta\, u)^4}\, du = (1+\beta)\Gamma_3(\beta,1,2). \end{aligned}$$

Now, by virtue of (4.6) and after some calculations we get

$$\begin{aligned} \Gamma_3(\beta,3,1) &= -\frac{1}{3\beta}\left(\frac{\mathrm{Li}_2(-\beta)}{\beta} + \frac{1}{1+\beta}\right), \\ \Gamma_3(\beta,2,1) &= \frac{1}{6\beta}\left(\frac{\log(1+\beta)}{\beta} - \frac{1}{(1+\beta)^2}\right), \\ \Gamma_3(\beta,1,2) &= \frac{1}{3(1+\beta)^3}, \end{aligned}$$

where Li_2 denotes the polylogarithm function of order two. Now, the stated result is obtained by substituting in the elements of $I(\alpha,\beta)$ the value of the corresponding integrals.

The result for $\beta = 0$ is derived by means of routine calculations, so we omit the details. □

Proof of Proposition 3.2: The result follows by using standard large sample theory results for ML estimators (for example, by applying Lehmann and Casella [29, Theorem 5.1, p. 463]). In particular, the asymptotic covariance matrix of the ML estimators, Σ, is obtained by inverting the expected Fisher information matrix $(1/n)I(\alpha,\beta)$, with $I(\alpha,\beta)$ provided in Proposition 3.1. □

ACKNOWLEDGMENTS

The authors thank the anonymous referee for his/her constructive comments, which led to an improvement of the paper. Research of P. Jodrá has been partially funded by grant of Diputación General de Aragón –Grupo E24-17R– and ERDF funds. Research of M.D. Jiménez-Gamero has been partially funded by grant MTM2017-89422-P of the Spanish Ministry of Economy, Industry and Competitiveness, ERDF support included.

REFERENCES

[1] Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.

[2] Adamidis, K., Dimitrakopoulou, T. and Loukas, S. (2005). On an extension of the exponential-geometric distribution, *Statistics & Probability Letters*, **73**, 3, 259–269.

[3] Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate, *Statistics & Probability Letters*, **39**, 1, 35–42.

[4] Akaike, H. (1974). A new look at statistical model identification, *IEEE Transactions on Automatic Control*, **19**, 6, 716–723.

[5] Apostol, T.M. (2010). *Zeta and related functions*. In: "NIST Handbook of Mathematical Functions" (Olver, F.W.F., Lozier, D.W., Boisvert, R.F. and Clark, C.W., Eds.), National Institute of Standards and Technology, Washington, DC, and Cambridge University Press, Cambridge, 601–616.

[6] Babu, G.J. and Rao, C.R. (2004). Goodness-of-fit tests when parameters are estimated, *Sankhyā*, **66**, 1, 63–74.

[7] Bakoban, R.A. and Abu-Zinadah, H.H. (2017). The beta generalized inverted exponential distribution with real data applications, *REVSTAT Statistical Journal*, **15**, 1, 65–88.

[8] Barreto-Souza, W., Lemos de Morais, A. and Cordeiro, G.M. (2011). The Weibull-geometric distribution, *Journal of Statistical Computation and Simulation*, **81**, 5, 645–657.

[9] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, **52**, 3, 345–370.

[10] Cordeiro, G.M., Ortega, E.M.M. and Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data, *Journal of the Franklin Institute. Engineering and Applied Mathematics*, **347**, 8, 1399–1429.

[11] Cribari-Neto, F. and Zeileis, A. (2010). Beta Regression in R, *Journal of Statistical Software*, **34**, 2, 1–24.

[12] D'Agostino, R.B. and Stephens, M.A. (Eds.) (1986). *Goodness-of-Fit-Techniques*, Marcel Dekker, New York.

[13] DAVID, H.A. and NAGARAJA, H.N. (2003). *Order Statistics*, 3rd edition, John Wiley & Sons, Hoboken, New Jersey.

[14] EPPS, T.W. and PULLEY, L.B. (1983). A test for normality based on the empirical characteristic function, *Biometrika* **70**, 3, 723–726.

[15] FERRARI, S.L.P. and CRIBARI-NETO F. (2004). Beta regression for modelling rates and proportions, *Journal of Applied Statistics*, **31**, 7, 799–815.

[16] FREES, E.W. (2010). *Regression Modeling with Actuarial and Financial Applications*, International Series on Actuarial Science, Cambridge University Press, Cambridge.

[17] GÓMEZ-DÉNIZ, E., SORDO, M.A. and CALDERÍN-OJEDA, E. (2014). The Log-Lindley distribution as an alternative to the beta regression model with applications in insurance, *Insurance: Mathematics & Economics*, **54**, 49–57.

[18] GÓMEZ-DÉNIZ, E., VÁZQUEZ-POLO, F.J. and GARCÍA, V. (2017). The Modified Borel-Tanner (MBT) regression model, *REVSTAT Statistical Journal*, **15**, 3, 425–442.

[19] JIMÉNEZ-GAMERO, M.D., ALBA-FERNÁNDEZ, M.V., JODRÁ, P. and BARRANCO-CHAMORRO, I. (2015). An approximation to the null distribution of a class of Cramér-von Mises statistics, *Mathematics and Computers in Simulation*, **118**, 258–272.

[20] JIMÉNEZ-GAMERO, M.D., ALBA-FERNÁNDEZ, M.V., MUÑOZ-GARCÍA, J. and CHALCO-CANO, Y. (2009). Goodness-of-fit tests based on empirical characteristic functions, *Computational Statistics & Data Analysis*, **53**, 12, 3957–3971.

[21] JODRÁ, P. (2008). On a connection between the polylogarithm function and the Bass diffusion model, *Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences*, **464**, 2099, 3081–3088.

[22] JODRÁ, P. and JIMÉNEZ-GAMERO, M.D. (2014). On a logarithmic integral and the moments of order statistics from the Weibull-geometric and half-logistic families of distributions, *Journal of Mathematical Analysis and Applications*, **410**, 2, 882–890.

[23] JODRÁ, P. and JIMÉNEZ-GAMERO, M.D. (2016). A note on the Log-Lindley distribution, *Insurance: Mathematics & Economics*, **71**, 186–194.

[24] JODRÁ, P., JIMÉNEZ-GAMERO, M.D. and ALBA-FERNÁNDEZ, M.V. (2015). On the Muth Distribution, *Mathematical Modelling and Analysis*, **20**, 3 291–310.

[25] JONES, M.C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages, *Statistical Methodology*, **6**, 1, 70–81.

[26] JORGENSEN, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London.

[27] KOENKER, R. (2005). *Quantile Regression*, Cambridge University Press, New York.

[28] KUMARASWAMY, P. (1980). Generalized probability density-function for double-bounded random-processes, *Journal of Hydrology*, **46**, 1-2, 79–88.

[29] LEHMANN, E.L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd edition, Springer Texts in Statistics, Springer-Verlag, New York.

[30] MEINTANIS,S.G., JIMÉNEZ-GAMERO, M.D. and ALBA-FERNÁNDEZ, M.V. (2014). A class of goodness-of-fit tests based on transformation, *Communications in Statistics. Theory and Methods*, **43**, 8, 1708–1735.

[31] Mitrinović, D.S. (1961). Sur une classe de nombres reliés aux nombres de Stirling, *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences de Paris*, **252**, 2354–2356.

[32] Noufaily, A. and Jones, M.C. (2013). Parametric quantile regression based on the generalized gamma distribution, *Journal of the Royal Statistical Society. Series C. Applied Statistics*, **62**, 5, 723–740.

[33] Papke, L.E. and Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401 (K) plan participation rates, *Journal of Applied Economerics*, **11**, 619–632.

[34] Papke, L.E. and Wooldridge, J.M. (2008). Panel data methods for fractional response variables with an application to test pass rates, *Journal of Econometrics*, **145**, 1-2, 121–133.

[35] Pascoa, M.A.R., Ortega, E.M.M. and Cordeiro, G.M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis, *Statistical Methodology*, **8**, 5, 411–433.

[36] Pérez-Rodríguez, J.V. and Gómez-Déniz, E. (2015). Spread component costs and stock trading characteristics in the Spanish Stock Exchange. Two flexible fractional response models, *Quantitative Finance*, **15**, 12, 1943–1962.

[37] R Development Core Team (2017). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org/`.

[38] Schmit, J.T. and Roth, K. (1990). Cost effectiveness of risk management practices, *The Journal of Risk and Insurance*, **57**, 3, 455–470.

[39] Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 2, 461–464.

[40] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, Springer-Verlag, New York.

[41] Stute, W., González Manteiga, W. and Presedo Quindimil, M. (1993). Bootstrap based goodness-of-fit tests, *Metrika* **40**, 3-4, 243–256.

[42] Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, **57**, 2, 307–333.