ISSN 0104-0499

PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

DEs-UFSCar e SME-ICMC-USP

A QUANTILE PARAMETRIC REGRESSION MODEL FOR BOUNDED RESPONSE VARIABLES WITH AN APPLICATION IN THE ANALYSIS OF ATTITUDES SCALE

Cristian L. Bayes
Mário de Castro
Jorge L. Bazan

RELATÓRIO TÉCNICO

TEORIA E MÉTODO – SÉRIE A

Fevereiro/2014
nº 258

A quantile parametric regression model for bounded response variables with an application in the analysis of attitudes scale

Cristian L. Bayes* Mário de Castro** and Jorge L. Bazán**

[1]

* Departamento de Ciencias, Pontificia Universidad Católica del Perú, Lima, Perú

** Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos-SP, Brasil

[1] *Address for correspondence:* Cristian L. Bayes. Av Universitaria 1801, San Miguel, Lima 32, PUCP, Peru. E-mail: cbayes@pucp.edu.pe.

A quantile parametric regression model for bounded response variables with an application in the analysis of attitudes scale

Abstract

Bounded response variables are common in several applications where the responses are obtained through surveys or tests and in the particular case of score of attitudes, obtained using Likert scales. The usual normal regression model to explain this type of response ignores this fact and consequently new regression models were proposed recently to model the relationship among one or more covariates and the conditional mean of a response variable given the covariates considering the Beta distribution or mixture of them. However, when we are interested in know how covariates are influent for different levels of the score of attitudes, that is, how regression coefficients of a given covariate change for different quantiles of an attitude quantil, regression models should be considered. A new quantile parametric regression model for bounded response variables (and proportions as a particular case) is presented by considering the distribution introduced by Kumaraswamy (1980). A Bayesian approach is adopted for inference using Markov Chain Monte Carlo (MCMC) methods. Model comparison criteria and model assessment are also discussed. The method can be easily programmed and then easily used for data modeling. Furthermore, an application with a data set on attitudes of teachers toward Statistics is given. Specifically, we show that the quantile parametric regression model proposed here is an alternative modeling for bounded response variables. In our proposal, the effects of the covariates on the quantiles of the response variable can be directly assessed.

Keywords: proportions ,rates, Kumaraswamy distribution, Bayesian inference, link function, MCMC methods

1 Introduction

Bounded response variables are common in several applications where the responses are obtained through surveys or tests. This is frequent in Economics and Psychology. A particular case of this situation is the scale using the Likert construction, which is frequently used in the measurement of attitudes. For example, Estrada *et al.* (2010) studied the attitudes toward Statistics of teachers of primary education considering a Lickert scale of attitudes that consists of 25 items comprising five points ranging from “strongly disagree” (level 1) to “strongly agree” (level 5). The responses of a subject are added together to form a score with values in the set $\{5, 6, \ldots, 125\}$.

Traditional analysis to explore the relationship between attitude and covariates can be done using the normal regression model with the score as the response variable. However, this analysis ignores the fact that the score is a limited variable. That is, a response variable Y takes values in the $[c, d]$ interval, $c < d$, where c is the minimum value of Y and d is the maximum value of Y, $c > -\infty$ and $d < \infty$. Then, the usual practice when normal regression models are considered in order to explain this kind of response variable is to ignore this fact and assume that the response variable takes values in $\mathbb{R}$ or $\mathbb{R}^+$. However, it is easy to see that the transformed variable $(Y - d)/(d - c)$ takes values in the in the $[0, 1]$ interval. Hence, regression models for bounded response variables in the $[0, 1]$ interval turn out to be more convenient.

Regression models for response variables in the unity interval, including regression models for percentages, proportions, and fractions or rates, have been introduced recently in the literature. We mention the beta regression model introduced by Kieschnick & McCullough (2003) and Ferrari & Cribari-Neto (2004) and the beta-rectangular regression model proposed by Bayes *et al.* (2012).

Some examples of these situations can be given, such as the percentage of time devoted to an activity during a certain period of time, the fraction of income spent on food, the unemployment rate, the poverty rate, the score achieved in a test, the Gini index, the fraction of "good" cholesterol (HDL/total cholesterol), the proportion of sand in the soil, and the fraction of surface covered by vegetation.

In the beta regression model, the regression parameters are interpretable in terms of the mean response, and in many aspects are similar to generalized linear models. Estimation can be performed by maximum likelihood (Ferrari & Cribari-Neto, 2004) or by Bayesian methods (Branscum *et al.*, 2007). The beta regression model is sufficiently documented by several publications such as Espinheira *et al.* (2008a), Espinheira *et al.* (2008b), Ferrari *et al.* (2011), and Cribari-Neto & Zeileis (2010) and several applications as, for example, Kelly *et al.* (2007) and Wallis *et al.* (2009).

In addition, the beta rectangular model proposed by Bayes *et al.* (2012) is more robust than the beta regression model (Ferrari & Cribari-Neto, 2004). This new model includes the beta regression model and the variable dispersion beta regression model (Ferrari *et al.*, 2011) as particular cases. Another extension considering mixture of beta distributions is presented in Smithson *et al.* (2011).

In the above cited literature the authors consider only the relationship between one or more covariates and the conditional mean of a response variable given the covariates. However, in some applications the quantiles of the response variable are of central interest. For example, we can be interested in know how covariates are influent for different levels of score of attitudes, that is, how regression coefficients of a given covariate change for different quantiles of attitude. Thus, we can be interested in quantile regression models.

Quantile regression, introduced by Koenker & Bassett (1978), is particularly useful when the rate of change in the conditional quantile, expressed by the regression coefficients, depends on the quantile and the main advantage is its flexibility for modeling data with heterogeneous conditional distributions. Data of this type occur in many fields, including Econometrics, Survival Analysis, and Ecology (see, for example, Koenker & Hallock, 2001) but is inusual in Education and Psychology. In general, quantile regression models are commom for response variables taking values in $\mathbb{R}$ or $\mathbb{R}^+$ but to our knowledge, there is no quantile regression model for variables with bounded response.

Thus, a convenient quantile parametric regression model for response variables in the unity interval is proposed in this paper, which can be easily extended to bounded response variables.

We consider that the model proposed here for quantile regression models can be useful to model the relationship between the covariates and the conditional quantiles of the response variable given the covariates. Quantile regression also provides a more complete picture of the conditional distribution of the response variable given the covariates. Consider, for example, a model for the quantiles of a socioeconomic level or the achievement in an educational test. The interest might rest on the upper quantiles.

From a Bayesian perspective, Yu & Moyeed (2001) proposed to assume an asymmetric Laplace distribution (ALD) in a parametric quantile regression model for an unbounded response variable. Kozumi & Kobayashi (2011) provided an useful stochastic representation for the ALD that facilitates the implementation of a Gibbs sampling scheme for this model. This approach has been extended to longitudinal data by Geraci & Bottai (2007). Our proposal is similar to Yu & Moyeed (2001), but for a response variable in the unit interval and using as parametric model the Kuramaswamy distribution.

Since the cumulative distribution function (cdf) of the beta distribution does not have a closed form, quantile regression models built upon this distribution poses some difficulties. In contrast, the Kumaraswamy distribution (Kumaraswamy, 1980; Jones, 2009) is a family of continuous probability distributions defined on the $(0, 1)$ interval that is similar to the beta distribution, but much simpler due to the simple closed form of both its probability density function (pdf) and cdf. This distribution was originally proposed by P. Kumaraswamy for variables that are bounded below and above.

The purpose of this paper is to show the advantages of using a parametric quantile regression model for proportions considering the Kumarawasmy distribution proposed here and how this model can be used for the study of attitudes. Without loss of generality, this model is equally valid for all bounded response variables.

The paper is organized as follows. In Section 2 we present a short account of the Kumaraswamy distribution and a new parametrization is introduced. Our regression model is formulated in Section 3. A Bayesian approach to this model is developed in Section 4 including model assesment and model comparison criteria. Section 5 presents an application, which is developed to show the usefulness of our regression model. Final comments are presented in Section 6.

2 The Kumaraswamy distribution

A random variable Y follows a Kumaraswamy distribution if its pdf is given by

$$f(y|\alpha,\beta) = \alpha\beta y^{\alpha-1}(1-y^{\alpha})^{\beta-1}, \quad 0<y<1,\ \alpha,\beta>0. \tag{1}$$

Hence, the mean and variance are expressed by

$$E(Y|\alpha,\beta) = \beta B\left(1+\frac{1}{\alpha},\beta\right) \quad \text{and} \quad Var(Y|\alpha,\beta) = \beta B\left(1+\frac{2}{\alpha}\right) - \beta^2 B^2\left(1+\frac{1}{\alpha}\right), \tag{2}$$

where $B(\cdot,\cdot)$ denotes the beta function.

As pointed out by Mitnik & Baek (2013), the expressions for $E(Y)$ and $Var(Y)$ make a mean-variance based reparametrization unfeasible. However, we can find a simple expression for the quantile function, given by

$$\kappa(q) = F^{-1}(q) = \{1 - (1-q)^{\frac{1}{\beta}}\}^{\frac{1}{\alpha}}, \quad 0 < q < 1. \tag{3}$$

As a particular case the median is $\kappa(0.5) = (1 - 0.5^{\frac{1}{\beta}})^{\frac{1}{\alpha}}$.

For a quantile regression analysis, we consider a reparametrization of the Kumaraswamy distribution in terms of the q-quantile and the shape parameter following the ideas presented in Mitnik & Baek (2013). In order to obtain a more appropriate regression structure for the Kumaraswamy distribution, we take

$$\kappa = \{1 - (1-q)^{\frac{1}{\beta}}\}^{\frac{1}{\alpha}} \quad \text{and} \quad \phi = -\log\left(1 - (1-q)^{\frac{1}{\beta}}\right) \tag{4}$$

as a new parametrization. In this case, q is assumed known and the parameter space of $(\kappa, \phi)^T$ is given by $(0,1) \times (0, \infty)$.

Under this parametrization, the pdf and the cdf of the Kumaraswamy distribution turn out to be

$$f(y|\kappa,\phi) = -\frac{\log(1-q)\phi}{\log\left(1-e^{-\phi}\right)\log(\kappa)} y^{-\frac{\phi}{\log(\kappa)}-1}\left\{1 - y^{-\frac{\phi}{\log(\kappa)}}\right\}^{\frac{\log(1-q)}{\log\left(1-e^{-\phi}\right)}-1} \tag{5}$$

and

$$F(y|\kappa,\phi) = 1 - \left\{1 - y^{-\frac{\phi}{\log(\kappa)}}\right\}^{\frac{\log(1-q)}{\log\left(1-e^{-\phi}\right)}}. \tag{6}$$

We consider the notation $Y \sim \mathrm{K}(\kappa, \phi, q)$ with quantile parameter $\kappa \in (0,1)$, shape parameter $\phi > 0$, and $0 < q < 1$ is assumed known.

Figure 1 depicts the pdf in (5) for different values of κ and ϕ. We pick the first decile, the median, and the last decile. When κ is fixed, we note that ϕ is a parameter that controls the precision and the kurtosis of the distribution. For the largest values of ϕ we observe less dispersion and high values of kurtosis. In general, we interpret ϕ as a shape parameter. On the other hand, when ϕ is fixed we note that κ acts as a parameter that controls the location and the skewness of the distribution, so that for larger values of ϕ we observe that the mode moves to the right and the distribution has negative skewness. In general, since κ is the q-quantile of Y, we interpret that this parameter is a location parameter in the range of values of the variable being modeled.

3 The Kuramaswamy quantile regression model

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ be a vector of observed responses that take values in the $(0,1)^n$ hypercube. The Kuramaswamy quantile regression model is given by

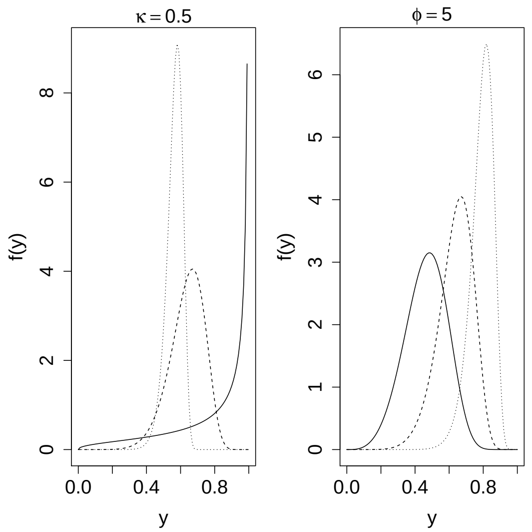
$$\mathrm{Y}_i \overset{\text{indep.}}{\sim} \mathrm{K}(\kappa_i, \phi, q) \quad \text{and} \quad g_q^{-1}(\kappa_i) = \eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}, \tag{7}$$

for $i = 1, \ldots, n$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ is a vector of regression coefficients associated with thee location and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})^T$ is a vector of k covariates. Here, ϕ is considered a parameter to be estimated and q is assumed known. Moreover, the quantile function $\kappa = \kappa(q)$ is identifiable when q is specified and $g_q^{-1}(\cdot)$ is a strictly monotone and twice differentiable function that maps $(0,1)$ into $\mathbb{R}$.

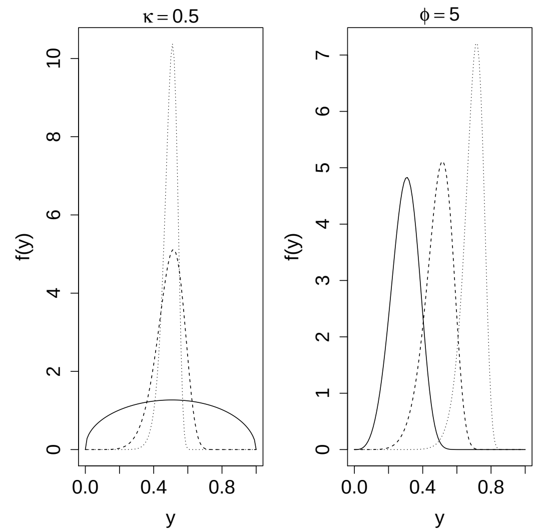
In general, $g_q(\cdot)$ can be any cdf corresponding to a continuous distribution where the inverse function is called the link function relating the quantile κ_i to the covariates $\boldsymbol{x}_i$. Some examples of link function are the logit, probit, and complementary log-log functions. In this paper we adopt the logit link, but of course other link functions might be explored.

Under the parameterization in (5), the likelihood function can be written as

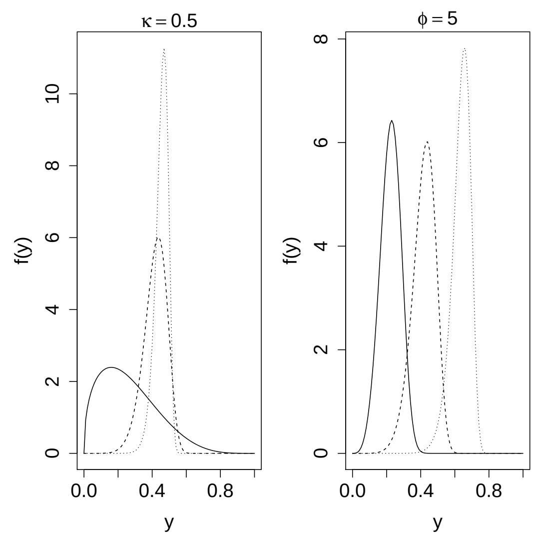
$$L(\boldsymbol{\theta}|\boldsymbol{Y}) = \prod_{i=1}^{n} f(y_i|\kappa_i, \phi) = \prod_{i=1}^{n} \frac{-\log(1-q)\phi}{\log\left(1-e^{-\phi}\right)\log(\kappa_i)} y_i^{-\frac{\phi}{\log(\kappa_i)}-1} \left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}^{\frac{\log(1-q)}{\log\left(1-e^{-\phi}\right)}-1}, \tag{8}$$



(a) 0.1-quantile



(b) 0.5-quantile



(c) 0.9-quantile

Figure 1: Kumaraswamy pdf for different values of κ and ϕ. Left panel: $\kappa = 0.5$ and different values of ϕ: 1 (solid line), 5 (dashed line), and 10 (dotted line). Right panel: $\phi = 5$ and different values of κ: 0.3 (solid line), 0.5 (dashed line), and 0.7 (dotted line).

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ and κ_i is defined in (7). Therefore, the log-likelihood function, denoted by $\ell(\boldsymbol{\theta}|\boldsymbol{Y})$, is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}|\boldsymbol{Y}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\boldsymbol{Y}) = \sum_{i=1}^{n} \Bigg[&\log(-\log(1-q)) - \log\left(-\log\left(1-e^{-\phi}\right)\right) - \log(-\log(\kappa_i)) \\ &+ \log(\phi) - \left\{\frac{\phi}{\log(\kappa_i)} + 1\right\} \log(y_i) + \left\{\frac{\log(1-q)}{\log\left(1-e^{-\phi}\right\}} - 1\right\} \log\left(1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right) \Bigg].\end{aligned} \tag{9}$$

The first and second derivatives of $\ell(\boldsymbol{\theta}|\boldsymbol{Y})$ with respect to $\boldsymbol{\theta}$ are presented in Appendix A.

4 Bayesian inference

With independent data, the likelihood function for the Kumaraswamy quantile regression model is given by (8) and (7). In this way, the posterior distribution of $\boldsymbol{\theta}$, denoted by $p(\boldsymbol{\theta}|\boldsymbol{Y})$, is obtained as

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto L(\boldsymbol{\theta}|\boldsymbol{Y}) p(\boldsymbol{\theta}), \tag{10}$$

where $p(\boldsymbol{\theta})$ stands for the prior distribution of $\boldsymbol{\theta}$. To complete the Bayesian specification of the model, we assume that the elements of the parameter vector are *a priori* independent, that is,

$$p(\boldsymbol{\theta}) = p(\phi)p(\boldsymbol{\beta}) = p(\phi) \prod_{j=1}^{k} p(\beta_j). \tag{11}$$

We adopt $\boldsymbol{\beta}_j \sim \mathrm{N}(0, \sigma_{0j}^2)$ and $\log(\phi) \sim \mathrm{N}(0, \sigma_1^2)$, where σ_{0j}^2, $j = 1, \ldots, k$, and σ_1^2 are set to ensure vague prior knowledge. After plugging the likelihood function (8) and the prior distribution into (10), we get

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto \prod_{i=1}^{n} -\frac{\log(1-q)\phi}{\log\left(1-e^{-\phi}\right)\log(\kappa_i)} y_i^{-\frac{\phi}{\log(\kappa_i)}-1} \left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}^{\frac{\log(1-q)}{\log(1-e^{-\phi})}-1} p(\phi) \prod_{j=1}^{k} p(\beta_j). \tag{12}$$

The posterior mean is taken as the point estimator.

Since the posterior distribution in (12) has an analytically intractable expression, we resort to the Gibbs sampler (see, e.g., Chen *et al.*, 2000) in order to sample from this posterior distribution. The full conditional distributions of β_j and ϕ, denoted by $p(\beta_j|\cdot)$ and $p(\phi|\cdot)$, are easily obtained from (12). Furthermore, since these distributions neither have a known form nor are log-concave, in each cycle of the Gibbs sampler we perform Metropolis steps. Let $\ell(\beta_j|\cdot) = \log\left(p(\beta_j|\cdot)\right)$. Following Chen *et al.* (2000, Section 2.2), a proposal value $\beta_j^* \sim \mathrm{N}(\widehat{\beta}_j, \widehat{\sigma}_j^2)$ for β_j is generated, where $\widehat{\beta}_j$ maximizes $\ell(\beta_j|\cdot)$ and $\widehat{\sigma}_j^2$ is minus the inverse of $d^2\ell(\beta_j|\cdot)/d\beta_j^2$ evaluated at $\beta_j = \widehat{\beta}_j$. The maximization of $\ell(\beta_j|\cdot)$ can be carried out with the Nelder-Mead algorithm implemented by O'Neill (1971). A move from β_j to β_j^* takes place with probability

$$\min\left(\frac{\pi(\beta_j^*|\cdot)\phi\big((\beta_j - \widehat{\beta}_j)/\widehat{\sigma}_j\big)}{\pi(\beta_j|\cdot)\phi\big((\beta_j^* - \widehat{\beta}_j)/\widehat{\sigma}_j\big)}, 1\right),$$

where $\phi(\cdot)$ denotes the standard normal pdf. Samples from the full conditional distribution of ϕ are drawn in a similar fashion with $d^2\ell(\phi|\cdot)/d\phi^2$. The second derivatives $d^2\ell(\beta_j|\cdot)/d\beta_j^2$, $j = 1, \ldots, k$, and $d^2\ell(\phi|\cdot)/d\phi^2$ are showed in (A1) and (A2), respectively.

The MCMC computations were implemented using the FORTRAN language. The computational codes are available on request from the authors. We have also written BUGS codes in WinBUGS (Lunn *et al.*, 2000) (see the printout in Appendix B), which give a relatively straightforward implementation of the proposed model.

The convergence of the chains was monitored by the Geweke's statistic (Geweke, 1992) and graphical inspection of the chains. The highest posterior density (HPD) intervals were estimated following the steps described in Chen *et al.* (2000, Section 7.3.1). Once $\boldsymbol{\beta}$ is estimated, consequently κ is estimated and considering that this is a location parameter in the range of values of the variable being modeled, κ_i for a particular value q is an estimate of the variable of interest.

4.1 Model assessment

In this section a device for the assessment of model fitting is presented. Taking into account the cdf in (6), it follows that $F(Y_i|\kappa_i,\phi) \overset{\text{iid}}{\sim} \text{Uniform}(0,1)$, where $\kappa_i = 1/\{1+\exp(-\boldsymbol{x}_i^T\boldsymbol{\beta})\}$, $i = 1,\ldots,n$. Therefore, under the postulated model in (7), $Q(\boldsymbol{Y},\boldsymbol{\theta}) = -\sum_{i=1}^n \log\big(F(Y_i|\boldsymbol{\theta})\big)$ is distributed as a Gamma$(n,1)$ random variable. The pivotal quantity $Q(\boldsymbol{Y},\boldsymbol{\theta})$ is a key element for model checking. If $\boldsymbol{\theta}_0$ denotes the data generating value of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{\text{post}}$ is drawn from the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Y}$, Johnson (2007) proved that $Q(\boldsymbol{Y},\boldsymbol{\theta}_0)$ and $Q(\boldsymbol{Y},\boldsymbol{\theta}_{\text{post}})$ have the same distribution. As advocated by Johnson (2007), a useful tool for model assessment can be based on graphical comparisons of the posterior distribution of this pivotal quantity to its reference distribution. This graphical diagnostic may reveal model inadequacy.

4.2 Model comparison criteria

There are several criteria for comparing different models fitted to a given data set and for selecting the one that best fits the data. Two of the often used criteria are based on the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) and the conditional predictive ordinates (CPO) (Gelfand *et al.*, 1992). The DIC is built upon the deviance $\mathcal{D}(\boldsymbol{\theta}) = -2\ell(\boldsymbol{\theta}|\boldsymbol{Y})$, with $\ell(\boldsymbol{\theta}|\boldsymbol{Y})$ as in (9). From G samples $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_G$ generated by the Gibbs sampler, the DIC is computed as $DIC = \mathcal{D}(\overline{\boldsymbol{\theta}}) + 2p_D$, where $p_D = \overline{\mathcal{D}}(\boldsymbol{\theta}) - \mathcal{D}(\overline{\boldsymbol{\theta}})$ is termed the effective number of parameters, with $\overline{\mathcal{D}}(\boldsymbol{\theta}) = \sum_{g=1}^G \mathcal{D}(\boldsymbol{\theta}_g)/G$ and $\overline{\boldsymbol{\theta}} = \sum_{g=1}^G \boldsymbol{\theta}_g/G$. Given a set of candidate models, the model yielding the smallest value of the DIC is the one that best fits the data. For each observation, CPO_i can be approximated by $\widehat{CPO}_i = [\{\sum_{g=1}^G 1/L(\boldsymbol{\theta}_g|y_i)\}/G]^{-1}$, where $L(\boldsymbol{\theta}|y_i)$ comes from (8). An omnibus measure of fit based on the $\widehat{CPO}_i$'s is provided by the log-pseudomarginal likelihood ($LPML$) (Geisser & Eddy, 1979) with expression $LPML = \sum_{i=1}^{n} \log(\widehat{CPO}_i)$. The larger the value of $LPML$, the better the fit of the model.

5 Application

Estrada *et al.* (2010) studied the attitudes toward Statistics of 146 teachers of primary education taking into account some characteristics of them as the country where the teachers live (Spain, $n = 66$, baseline and Peru, P, $n = 80$), specialty (Sciences, $n = 43$, baseline, Social Sciences, SS, $n = 75$, and Elementary School, ES, $n = 28$), and sex (female, $n = 85$, baseline and male, M, $n = 61$). The scale of attitudes consists of 25 items comprising five points ranging from "strongly disagree" (level 1) to "strongly agree" (level 5). The responses of a subject are added together to form a score with values in the set $S \in \{5, 6, \ldots, 125\}$. For this data set, the $S_i, i = 1, 2, \ldots 146$ score ranges from 48 to 102, with mean and standard deviation equal to 77.9 and 11.0, respectively. Traditional analyses to explore the relationship between attitude and the above covariates can be done using the normal regression model with the score as the response variable. Furthermore, a beta regression model can be fitted by taking the transformed score y, given by $Y_i = (S_i - 25)/(125 - 25)$, as the response variable, ranging from 0.23 to 0.77, with mean and standard deviation equal to 0.53 and 0.11, respectively. We begin fitting the Bayesian normal and beta regression models. For the normal model $S_i \overset{\text{indep.}}{\sim} \text{N}(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2)$, $i = 1, \ldots, n$, we put independent priors $\beta_k \sim \text{N}(0, 10^4)$, $j = 1, \ldots, k$, and $1/\sigma^2 \sim \text{Gamma}(0.01, 0.01)$. For the beta model, $y_i \overset{\text{indep.}}{\sim} \text{Beta}(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2)$, $i = 1, \ldots, n$, we put independent priors $\beta_k \sim \text{N}(0, 10^4)$, $j = 1, \ldots, k$, and $1/\sigma^2 \sim \text{Gamma}(0.01, 0.01)$. We ran the BUGS code furnished by Branscum *et al.* (2007). In the Gibbs sampler, the first 5000 iterations were discarded. Then, we performed 25000 additional iterations with thinning equal to 5, leading to 5000 samples for each parameter. Results from both analyses are showed in

Table 1: Posterior results for the normal model and the beta regression model with logit link for the mean and identity link for the precision (ϕ).

		Full model			Reduced model		
Model	Parameter	Mean	SD	95% HPD interval	Mean	SD	95% HPD interval
Normal	Intercept	85.21	1.78	(81.78, 88.79)	83.94	1.17	(81.62, 86.20)
	Specialty: SS	-2.96	1.95	(-6.77, 0.85)			
	Specialty: ES	-6.94	2.79	(-12.42, -1.44)	-4.49	2.25	(-8.91, -0.10)
	Country: P	-8.46	1.92	(-12.27, -4.76)	-9.39	1.78	(-12.9, -5.93)
	Sex: M	0.39	1.69	(-2.95, 3.66)			
	σ^2	90.60	11.00	(70.01, 112.42)	90.80	10.93	(70.39, 112.57)
	DIC	1077.2			1076.4		
Beta	Intercept	0.41	0.07	(0.27, 0.56)	0.36	0.05	(0.26, 0.45)
	Specialty: SS	-0.12	0.08	(-0.28, 0.03)			
	Specialty: ES	-0.28	0.11	(-0.50, -0.06)	-0.18	0.09	(-0.34, -0.01)
	Country: P	-0.34	0.08	(-0.50, -0.19)	-0.38	0.07	(-0.51, -0.24)
	Sex: M	0.01	0.07	(-0.13, 0.14)			
	ϕ	26.37	3.08	(20.70, 32.73)	26.31	3.02	(20.37, 32.19)
	DIC	-267.2			-269.1		

Table 1. The reduced models include only the significant coefficients at a 5% level. A coefficient is significant when its 95% HPD interval does not contain 0. For these models, the baseline category of specialty merges Sciences and Social Sciences.

In Table 1 we see that specialty and country are the significant covariates. Since the full and reduced models yield similar values of DIC, for the sake of simplicity we select the reduced model as our working model. From this model, we conclude that Elementary School teachers and Peruvian teachers present significant lower attitudes toward Statistics in comparison with the baseline categories (Sciences/Social Sciences teachers and Spanish teachers, respectively).

Next the Kumaraswamy quantile regression model was fitted to the 0.25, 0.5, and 0.75 quantiles of attitude. We ran the Gibbs sampler under the same conditions of the normal and beta regression models. The hyperparameters in (11) were set at the same values of the simulation study previously conducted (not shown) to evaluate the code used to implement the proposed model. That is, the hyperparameters in (11) were set at $\sigma_{01}^2 = \sigma_{02}^2 = \sigma_1^2 = 10^4$.

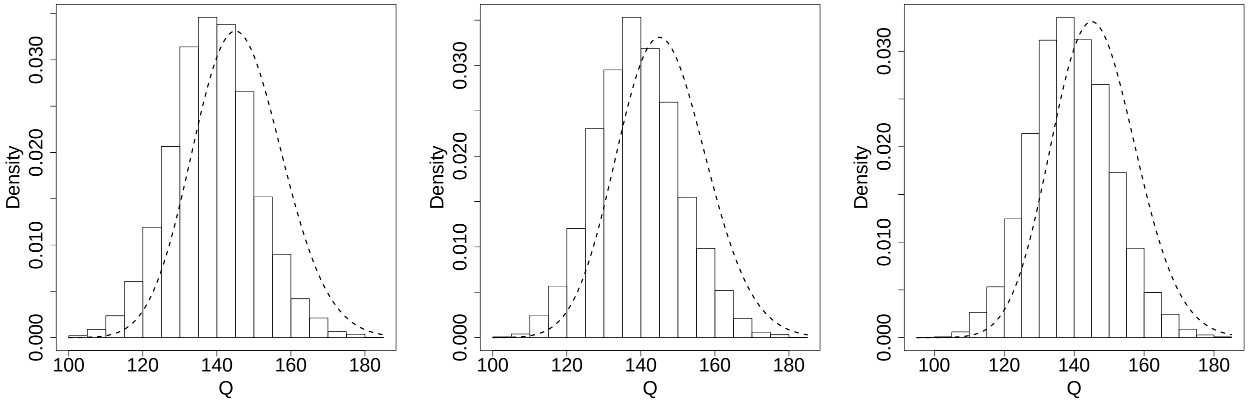
Table 2: Posterior results for the Kumaraswamy quantile regression model.

Quantile	Parameter	Full model Mean	SD	95% HPD interval	Reduced model Mean	SD	95% HPD interval
0.25	Intercept	0.15	0.09	(-0.02, 0.32)	0.10	0.06	(-0.01, 0.20)
	Specialty: SS	-0.15	0.09	(-0.32, 0.03)			
	Specialty: ES	-0.35	0.14	(-0.60, -0.08)	-0.23	0.11	(-0.43, 0.01)
	Country: P	-0.33	0.09	(-0.50, -0.15)	-0.38	0.08	(-0.54, -0.22)
	Sex: M	0.04	0.08	(-0.11, 0.20)			
	ϕ	4.58	0.22	(4.16, 5.01)	4.56	0.22	(4.14, 4.98)
	p_D	6.12			4.02		
	DIC	-263.4			-264.6		
	$LPML$	1110.2			1111.6		
0.5	Intercept	0.43	0.08	(0.28, 0.58)	0.38	0.05	(0.29, 0.48)
	Specialty: SS	-0.14	0.08	(-0.30, 0.02)			
	Specialty: ES	-0.32	0.13	(-0.57, -0.08)	-0.21	0.11	(-0.41, 0.00)
	Country: P	-0.31	0.08	(-0.48, -0.16)	-0.35	0.08	(-0.50, -0.20)
	Sex: M	0.04	0.08	(-0.11, 0.19)			
	ϕ	3.72	0.22	(3.29, 4.12)	3.69	0.21	(3.28, 4.12)
	p_D	6.05			4.03		
	DIC	-263.6			-268.6		
	$LPML$	1109.3			1110.8		
0.75	Intercept	0.68	0.08	(0.54, 0.83)	0.64	0.05	(0.55, 0.74)
	Specialty: SS	-0.13	0.08	(-0.29, 0.02)			
	Specialty: ES	-0.31	0.12	(-0.54, -0.08)	-0.20	0.10	(-0.39, -0.01)
	Country: P	-0.29	0.08	(-0.45, -0.15)	-0.34	0.07	(-0.48, -0.19)
	Sex: M	0.03	0.07	(-0.10, 0.18)			
	ϕ	3.04	0.21	(2.59, 3.43)	3.01	0.21	(2.59, 3.41)
	p_D	6.01			4.04		
	DIC	-263.7			-268.6		
	$LPML$	1108.8			1110.6		

Table 2 collects some posterior summaries. Based on the DIC and $LPML$ values, hereafter we report only results for the reduced models.

For the reduced models in Table 2, in Figure 2 we show the histograms of the posterior samples of $Q(\boldsymbol{Y}, \boldsymbol{\theta})$ in Section 4.1 and the Gamma$(146, 1)$ density function. Since the histograms and the density function overlap each other in a great extent, these plots do not suggest serious departures from the postulated model. The same pattern was observed for the full models in

Table 2.



(a) 0.25-quantile (b) 0.5-quantile (c) 0.75-quantile

Figure 2: Histograms of the posterior samples of $Q(\boldsymbol{Y}, \boldsymbol{\theta})$ for the reduced models and the Gamma$(146, 1)$ density function.

Figure 3 displays some plots for the 0.75-quantile model. According to the trace plots, we see that the chains show a good mixing. The posterior densities are approximately symmetric, so that the effective number of parameters p_D closely matches the actual number of parameters in Table 2.

Not surprisingly, the posterior means in Tables 1 and 2 have the same sign. We observe that larger the quantile, smaller the estimate of the shape parameter ϕ, indicating less precision and higher kurtosis for the quantile of attitude. We also note that the estimates for the 0.5-quantile model and the beta regression model in Table 1 are similar. The effect of specialty slightly changes from non-significant to significant. This point can be seen as a novelty and a strength of our analysis, for the role of a covariate is not necessarily important at different levels (quantiles) of the response variable.

From (7) we obtain a closed form expression for the quantil of attitude toward Statistics for each teacher, that is $\kappa_i = g_q(\boldsymbol{x}_i^T \boldsymbol{\beta})$, where $\boldsymbol{x}_i$ is the vector of covariates for the i-th teacher, $g_q(\cdot)$ is cdf of the logistic distribution, and $\boldsymbol{\beta}$ is the coefficient vector associated with the covariates.

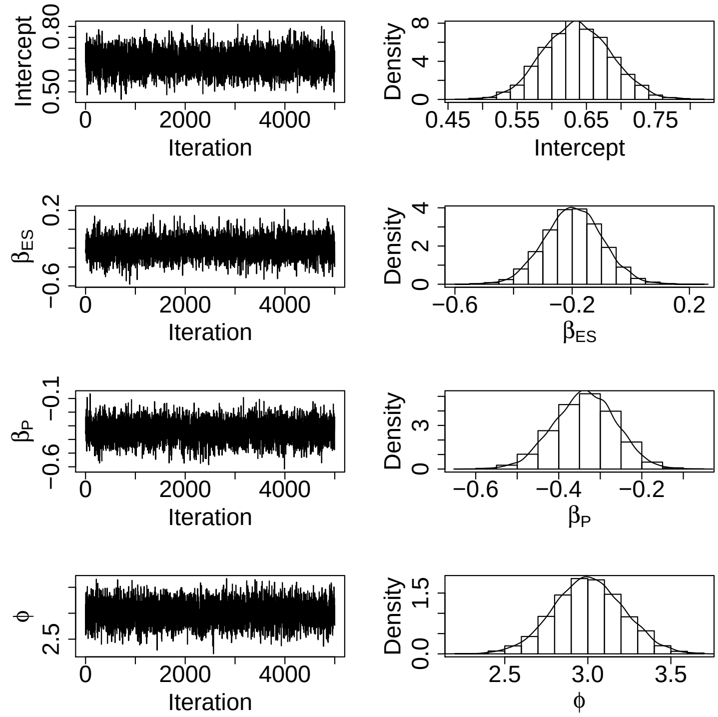


Figure 3: Trace plots of the chains, histograms, and approximate marginal posterior density functions of the parameters for the 0.75-quantile reduced model.

For a given q, using the output of the Gibbs sampler we get samples of the q-quantile $\kappa_i(q)$ for each individual. Since in this application x_1 (country) and x_2 (specialty) in the reduced models are dummy binary variables, four groups can be formed by the combination of the values of x_1 and x_2 and, consequently, posterior quantiles of attitude estimates for different groups can be obtained and compared.

Figure 4 shows posterior summaries for the 0.25, 0.5, and 0.75 quantiles of attitude estimated for different groups considered by the combination of covariates values in this study. This figure synthesizes the effects of the covariates on the quantiles. There are four different combinations of the levels of specialty and country. Note that for Spanish teachers, comparing the teachers of Sciences/Social Sciences and Elementary School teachers, the three 95% HPD intervals of attitude do not intersect. Note also the higher variability in the attitude for Peruvian teachers. As pointed out by Estrada *et al.* (2010), this can be explained, at least partially, since in Spain there is a greater effort in Statistics teachers' formation, curriculum organization, and didactics. We show that covariates are influent for different levels of score of attitudes, that is, we show how regression coefficients of a given covariate change for different quantiles of the attitude toward Statistics.

6 Final comments

In this paper a new quantile parametric regression model for bounded response variables is proposed. Our model is built upon the distribution introduced by Kumaraswamy (1980). A reparametrization of this distribution in terms of a given quantile and the shape parameter enables us to link any quantile of the distribution to covariates. Inference is based on a Bayesian approach with proper (and vague) prior distributions. Since the posterior distribution is not amenable to analytical treatment, we rely on Markov Chain Monte Carlo methods. Results

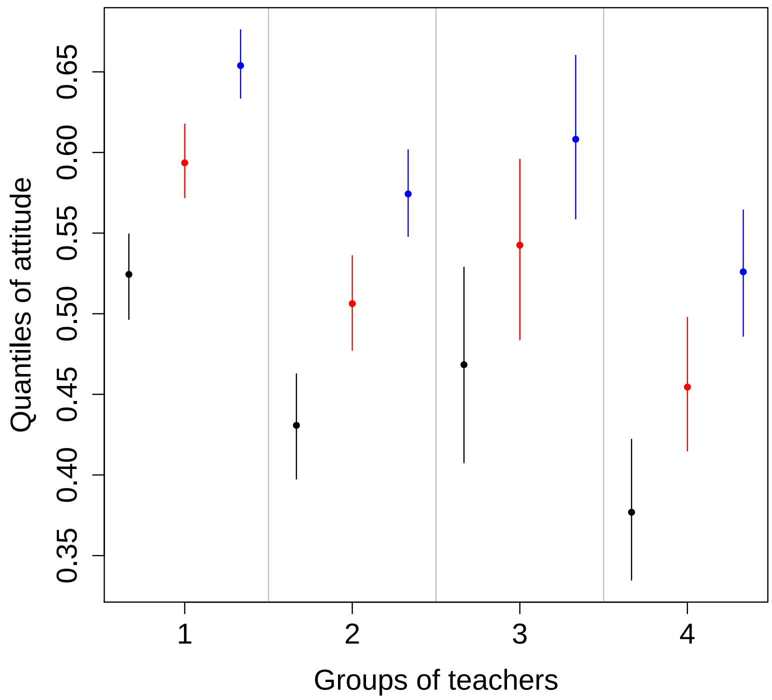


Figure 4: Posterior means and 95% HPD intervals for the 0.25- (black), 0.50- (red), and 0.75-quantiles (blue) of attitude estimated for different groups considered, from left to right. 1 ($x_1 = 0, x_2 = 0$): Spanish teachers of Sciences/Social Sciences, 2 ($x_1 = 0, x_2 = 1$): Elementary School Spanish teachers, 3 ($x_1 = 1, x_2 = 0$): Peruvian teachers of Sciences/Social Sciences, and 4 ($x_1 = 1, x_2 = 1$): Elementary School Peruvian teachers.

from a simulation study shows that even in case of extreme quantiles our Bayesian proposal yields estimators with a good performance. Furthermore, an application with a data set on attitudes of Peruvian and Spanish teachers toward Statistics is given.

We envision future works exploring different link functions in (7), possibly asymmetric ones. Bayesian diagnostic tools (Peng & Dey, 1995) are also of interest. Models for zero-inflated and one-inflated data sets would extend the present paper, as well as extensions to longitudinal data (Geraci & Bottai, 2007), to clustered data (Reich *et al.*, 2010), and to censored data (Wang *et al.*, 2013), noticing that in these works the response variable is unbounded.

Finally, the method can be easily programmed, as can be seen in Appendix B, and then easily used for data modeling, so that the quantile parametric regression model proposed here is an altenartive modeling for bounded response variables.

References

Bayes, C. L., Bazán, J. L. & García, C. (2012). A new robust regression model for proportions. *Bayesian Analysis*, 7, 771–796.

Bernardo, J. M., Berger, J. O., Dawid, A. P. & Smith, A. F. M., editors (1992). *Bayesian Statistics, 4*, New York. Oxford.

Branscum, A. J., Johnson, W. O. & Thurmond, M. C. (2007). Bayesian beta regression: Application to household data and genetic distance between foot-and-mouth disease viruses. *Australian & New Zealand Journal of Statistics*, 49, 287–301.

Chen, M.-H., Shao, Q.-M. & Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York.

Cribari-Neto, F. & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34, 1–24.

Espinheira, P., Ferrari, S. L. P. & Cribari-Neto, F. (2008a). Influence diagnostics in beta regression. *Computational Statistics & Data Analysis*, 52, 4417–4431.

Espinheira, P., Ferrari, S. L. P. & Cribari-Neto, F. (2008b). On beta regression residuals. *Journal of Applied Statistics*, 35, 407–419.

Estrada, A., Bazán, J. L. & Aparicio, A. (2010). Un estudio comparativo de las actitudes hacia la estadística en profesores españoles y peruanos (in Spanish). *Unión – Revista Iberoamericana de Educación Matemática*, 24, 45–56.

Ferrari, S. L. P. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31, 799–815.

Ferrari, S. L. P., Espinheira, P. L. & Cribari, F. (2011). Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica*, 65, 337–351.

Geisser, S. & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153–160.

Gelfand, A. E., Dey, D. K. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In Bernardo *et al.* (1992), pages 147–167.

Geraci, M. & Bottai, M. (2007). Quantile regression for longitudinal data using asymmetric laplace distribution. *Biostatistics*, 8, 140–154.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo *et al.* (1992), pages 169–193.

Johnson, V. E. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Analysis*, 2, 719–734.

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

Kelly, G. O., Garabed, R., Branscum, A., Perez, A. & Thurmond, M. (2007). Prediction model for sequence variation in the glycoprotein gene of infectious hematopoietic necrosis virus in California, USA. *Diseases of Aquatic Organisms*, 78, 97–104.

Kieschnick, R. & McCullough, B. D. (2003). Regression analysis of variates observed on (0,1): percentages, proportions and fractions. *Statistical Modelling*, 3, 193–213.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.

Koenker, R. & Hallock, K. F. (2001). Quantile regression: An introduction. *Journal of Economic Perspectives*, 15, 143–156.

Kozumi, H. & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81, 1565–1578.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random process. *Journal of Hidrology*, 46, 79–88.

Lunn, D. J., Thomas, A., Best, N. & Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325–337.

Mitnik, P. A. & Baek, S. (2013). The Kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54, 177–192.

O'Neill, R. (1971). Algorithm AS 47: Function minimization using a simplex procedure. *Applied Statistics*, 20, 338–345.

Peng, F. & Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics*, 23, 199–213.

Reich, B. J., Bondell, H. D. & Wang, H. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, 11, 337–352.

Smithson, M., Merkle, E. C. & Verkuilen, J. (2011). Beta regression finite mixture models of polarization and priming. *Journal of Educational and Behavioral Statistics*, 36, 804–831.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B*, 64, 583–639.

Wallis, E., Mac Nally, R. & Lake, S. (2009). Do tributaries affect loads and fluxes of particulate organic matter, inorganic sediment and wood? Patterns in an upland river basin in south-eastern Australia. *Hydrobiologia*, 636, 307–317.

Wang, H., Zhou, J. & Li, Y. (2013). Variable selection for censored quantile regression. *Statistica Sinica*, 23, 145–167.

Yu, K. & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, 54, 437–447.

A Derivatives of the log-likelihood function

In this appendix we present the expressions of the first and second derivatives of the log-likelihood function of our model. These expressions are used in Section 4. From (9) and (7), we get, for $j = 1, \ldots, k$,

$$\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \kappa_i} g'(\eta_i) \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \kappa_i} g'(\eta_i) x_{ij},$$

where $g'(\eta_i) = \partial\kappa_i/\partial\eta_i$, with

$$\frac{\partial\ell_i(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial\kappa_i} = -\frac{1}{\kappa_i\log(\kappa_i)} - \frac{\phi\log(y_i)\left\{\frac{\log(1-q)}{\log\left(1-e^{-\phi}\right)} - 1\right\} y_i^{-\frac{\phi}{\log(\kappa_i)}}}{\kappa_i\log^2(\kappa_i)\left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}} + \frac{\phi\log(y_i)}{\kappa_i\log^2(\kappa_i)}$$

and

$$\frac{\partial\ell_i(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial\phi} = -\frac{e^{-\phi}\log(1-q)\log\left(1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right)}{(1-e^{-\phi})\log^2\left(1-e^{-\phi}\right)} + \frac{\log(y_i)\left\{\frac{\log(1-q)}{\log\left(1-e^{-\phi}\right)} - 1\right\} y_i^{-\frac{\phi}{\log(\kappa_i)}}}{\log(\kappa_i)\left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}}$$
$$- \frac{\log(y_i)}{\log(\kappa_i)} + \frac{1}{\phi} - \frac{e^{-\phi}}{(1-e^{-\phi})\log\left(1-e^{-\phi}\right)}.$$

The second derivative of $\ell(\boldsymbol{\theta}|\boldsymbol{Y})$ with respect to β_j is given by

$$\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial\beta_j^2} = \sum_{i=1}^{n}\left\{\frac{\partial^2\ell_i(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial\kappa_i}g'^2(\eta_i) + \frac{\partial\ell_i(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial\kappa_i}g''(\eta_i)\right\}x_{ij}^2, \tag{A1}$$

for $j,\ldots,k$, with

$$\frac{\partial^2\ell_i(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial\kappa_i^2} = \frac{1}{\kappa_i^2\log^2(\kappa_i)} + \frac{1}{\kappa_i^2\log(\kappa_i)} + \left\{\frac{\log(1-q)}{\log\left(1-e^{-\phi}\right)} - 1\right\}$$
$$\times\left[-\frac{\phi^2\log^2(y_i)y_i^{-\frac{2\phi}{\log(\kappa_i)}}}{\kappa_i^2\log^4(\kappa_i)\left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}^2} - \frac{\phi^2\log^2(y_i)y_i^{-\frac{\phi}{\log(\kappa_i)}}}{\kappa_i^2\log^4(\kappa_i)\left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}}\right.$$
$$\left. + \frac{2\phi\log(y_i)y_i^{-\frac{\phi}{\log(\kappa_i)}}}{\kappa_i^2\log^3(\kappa_i)\left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}} + \frac{\phi\log(y_i)y_i^{-\frac{\phi}{\log(\kappa_i)}}}{\kappa_i^2\log^2(\kappa_i)\left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}}\right]$$
$$- \phi\left\{\frac{2}{\kappa_i^2\log^3(\kappa_i)} + \frac{1}{\kappa_i^2\log^2(\kappa_i)}\right\}\log(y_i).$$

Finally, the second derivative of $\ell(\boldsymbol{\theta}|\boldsymbol{Y})$ with respect to ϕ is given by

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \phi^2} =& \sum_{i=1}^{n} - \frac{2e^{-\phi} \log(1-q) \log(y_i) y_i^{-\frac{\phi}{\log(\kappa_i)}}}{(1-e^{-\phi}) \log(\kappa_i) \log^2(1-e^{-\phi}) \left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}} \\
&+ \left\{ \frac{\log(1-q)}{\log(1-e^{-\phi})} - 1 \right\} \left[- \frac{\log^2(y_i) y_i^{-\frac{2\phi}{\log(\kappa_i)}}}{\log^2(\kappa_i) \left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}^2} - \frac{\log^2(y_i) y_i^{-\frac{\phi}{\log(\kappa_i)}}}{\log^2(\kappa_i) \left\{1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right\}} \right] \\
&+ \log(1-q) \left\{ \frac{2e^{-2\phi}}{(1-e^{-\phi})^2 \log^3(1-e^{-\phi})} + \frac{e^{-\phi}}{(1-e^{-\phi}) \log^2(1-e^{-\phi})} \right. \\
&+ \left. \frac{e^{-2\phi}}{(1-e^{-\phi})^2 \log^2(1-e^{-\phi})} \right\} \log\left(1 - y_i^{-\frac{\phi}{\log(\kappa_i)}}\right) + n \left\{ \frac{e^{-2\phi}}{(1-e^{-\phi})^2 \log^2(1-e^{-\phi})} - \frac{1}{\phi^2} \right. \\
&+ \left. \frac{e^{-\phi}}{(1-e^{-\phi}) \log(1-e^{-\phi})} + \frac{e^{-2\phi}}{(1-e^{-\phi})^2 \log(1-e^{-\phi})} \right\}.
\end{aligned} \tag{A2}$$

B BUGS code

The BUGS code developed in the WinBUGS framework (Lunn *et al.*, 2000) for the full models in Table 2 is given below. We adopted the logit link for the quantile parameter and the log link for the precision parameter. The hyperparameters in the prior distributions for all the parameters need to be specified by the user, as well as the probability q corresponding to the quantile of interest.

```
model
{
    for (i in 1:n) {
        L[i] <- exp(logLike[i])
        ones[i] <- 1
        p[i] <- L[i] / C
        ones[i] ~ dbern(p[i])
```

```
        logLike[i] <- log(a[i]) + log(b) + (a[i] - 1) * log(y[i]) +
            (b - 1) * log(1 - pow(y[i], a[i]))
        a[i] <- -phi / log(kappa[i])
        logit(kappa[i]) <- beta[1] + beta[2] * x1[i] + beta[3] *
            x2[i] + beta[4] * x3[i] + beta[5] * x4[i]
    }
    b <- log(1 - q) / log(1 - exp(-phi))
    for (j in 1:k) {
        beta[j] ~ dnorm(0.0, 1.0E-04)
    }
    logphi ~ dnorm(0.0, 1.0E-04)
    phi <- exp(logphi)
    C <- 1.0E+05
}
```

Acknowledgements

The second author acknowledges partial support from CNPq, Brazil.