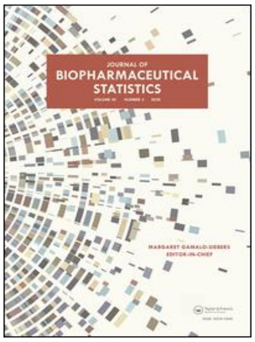

Taylor & Francis
Taylor & Francis Group

Journal of Biopharmaceutical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lbps20

A collection of parametric modal regression models for bounded data

André F. B. Menezes, Josmar Mazucheli & Subrata Chakraborty

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
https://doi.org/10.1080/10543406.2021.1918141

Taylor & Francis
Taylor & Francis Group

REVIEW

A collection of parametric modal regression models for bounded data

André F. B. Menezes[a], Josmar Mazucheli[b], and Subrata Chakraborty[c]

[a]Departamento De Estatística, Universidade Estadual De Campinas, Campinas, Brasil; [b]Departamento De Estatística, Universidade Estadual De Maringá, Maringá, Brasil; [c]Department of Statistics, Dibrugarh University, Assam, India

ABSTRACT
Modal regression is an alternative approach for investigating the relationship between the most likely response and covariates and can hence reveal important structure missed by usual regression methods. This paper provides a collection of parametric mode regression models for bounded response variable by considering some recently introduced probability distributions with bounded support along with the well-established Beta and Kumaraswamy distribution. The main properties of the distributions are highlighted and compared. An empirical comparison between the considered modal regression is demonstrated through the analysis of three data sets from health and social science. For reproducible research, the proposed models are freely available to users as an R package unitModalReg.

ARTICLE HISTORY
Received 14 October 2020
Accepted 13 April 2021

KEYWORDS
Parametric modal regression; Beta distribution; Kumaraswamy distribution

1. Introduction

Regression analysis is carried out to describe the influence of covariates on the response variable under investigation. In usual and most commonly applied regression analysis, the conditional mean of the response is modeled to describe the existing average relationship between the response and the covariates. However, despite the widespread applications, the mean regression might at times be found wanting specifically, in the presence of outliers and/or heavy-tailed noise in the response.

The earliest references to modal regressions are found in Sager and Thisted (1982) and Lee (1989). The former discussed non-parametric maximum likelihood estimation of isotonic modal regression function as a minimization of zero-one loss function and emphasized the role of mode as an objective statistical measure of interest. The authors also gave situational justification of preferring mode regression over mean or median regression. The authors gave three reasons (i) mode is the only parameter which makes sense while dealing with ordinal categorical response citing an example from study of treatment efficacy, (ii) when primary distribution of a location is contaminated by another distribution, mode regression is found to be more robust than the mean or median regression and (iii) when one's interest lies on mode itself Lee (1989) investigated mode regression stating how mode minimizes an expected loss function which arises in risk function approach to Bayesian estimation. They showed how mode regression may have major application in dealing with truncated asymmetric response under homogeneity.

Conditional mode regression is known to provide better and more informative description of the situation in hand than conditional mean or median regression models when the conditional distribution of response variable given covariate is heavy-tailed or skewed. Moreover, mode-based prediction interval tends to have a higher coverage probability than a mean-based prediction interval (Yao and Li 2014). The following narration from Chen et al. (2016) puts mode regression in proper perspective is

CONTACT Josmar Mazucheli jmazucheli@gmail.com Departamento De Estatística, Universidade Estadual De Maringá, Brasil

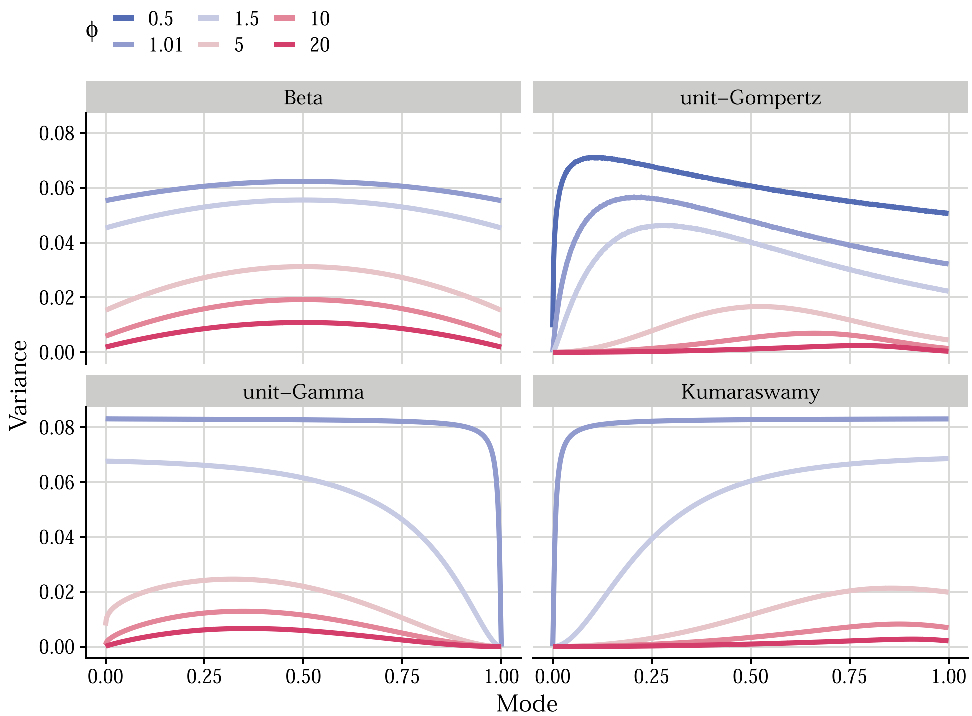

Figure 1. Mode and variance relationship for different values of ϕ.

worth recalling. "*Why would we ever use modal regression in favor a conventional regression method? The answer, at a high-level, is that conditional modes can reveal structure that is missed by the conditional mean. Figure 1 gives a definitive illustration of this point: we can see that, for the data examples in question, the conditional mean both fails to capture the major trends present in the response, and produces unnecessarily wide prediction bands. Modal regression is an improvement in both of these regards (better trend estimation and narrower prediction bands)*".

Chen (2018) made a nice review of mode regression using kernel density estimation. Wang et al. (2017) while investigating the association between the cognitive assessment of an individual with their neuroimaging features have shown how the mean regression analysis failed to reveal true picture due to the presence of heavy-tailed and skewed noise presented in data. They further stated the applicability of modal regression in the other field such as econometric, astronomy, traffic engineering, etc.

Most of the early works related to mode regression revolved around non-parametric and semi-parametric set up. The attention to the parametric mode regression was comparatively negligible this despite semi-parametric and non-parametric mode regression methods having a slow rate of convergence and dependency to bandwidth selection, hence minimal practical utility (Aristodemou, 2014). Pan et al. (2020) in their introductory section described the reason why parametric mode regression can be advantageous option for the machine learning community by proposing a parametric mode regression model using implicit function capable of analyzing multi-modal data, especially with large bandwidth limit.

Of late there are renewed interest in derivation of new probability distribution with bounded support especially in $(0, 1)$ as an alternative the famous Beta and Kumaraswamy distributions and also propose corresponding regression model to compete with the Beta regression. Some of these distributions are unit-Weibull (Mazucheli et al. 2020b, 2018c), unit-Birnbaum-Saunders (Mazucheli et al. 2018b), unit-Lindley (Mazucheli et al. 2019b), unit-Inverse Gaussian (Ghitany et al. 2019), unit-Gompertz (Mazucheli et al. 2019a) and new unit-Lindley (Mazucheli et al. 2020a). Its worth mentioning here that not all the distributions have compact analytical expression for mode, which is a prerequisite for modal regression.

Aristodemou (2014) presented a modal regression based on gamma distributed using a new parameterization to bring in the mode and precision parameters. Very recently, Zhou and Huang (2020) proposed a parametric modal regression model by considering the Beta distribution and presented an application with data from the Alzheimer's Disease Neuroimaging Initiative. With above background, we found enough motivation to present a collection of mode regression models based on some recently proposed probability distribution with support in (0,1) and having the unimodality property. For the sake of ranking the models, we considered three carefully chosen data sets for modeling using the proposed mode regression along with the Beta and Kumaraswamy based models.

The first data set which is about the relationship between stress (response variable) and anxiety is unimodal, and the response having increasing trend with respect to the covariate anxiety. The next data are about the proportion of votes received by a candidate in the second turn of presidential elections of 2010 in the municipals of Rio de Janeiro (RJ) and Ceara (CE) states collected with an aim to ascertain the impact of MHDI and region on the proportion of votes. In these data also the response variable is unimodal and decreasing with MHDI and state-wise disparity of proportion of votes is high. The last data set is a small one with fewer observations and is purposefully chosen to evaluate the flexibility of the models for small sample. These data are related to the migration of birds and intended to study the association between the proportion of birds that successfully reach the winter grounds and average mass (in grams), average wingspan (in cm) and the distance traveled (in km). In these data, maximum proportion that is the mode is the measure of interest. For all the data sets, we applied all four different mode regression models and our findings clearly suggest that there is no clear consensus on a particular model as the best in all cases. Instead, our results point toward the importance of having different models for different problems.

This paper is unfold as follows. Section 2 describes the main properties of the distributions and explores their peculiarities. A general definition of parametric modal regression, parameter estimation, and model adequacy are discussed in Section 3. The real data applications are presented in Section 4. Finally, a discussion of the important results is presented in Section 5.

2. Background

This section introduces the bounded distributions that are the basis of the proposed regression models. The focus is on the main properties and a new parametrization in terms of the mode. Furthermore, a theoretical comparison between the shapes of the distributions is also presented.

Without loss of generality, let Y be a random variable bounded on the unit interval with distribution $\mathcal{F}(a, b)$, where $a, b > 0$ are shape parameters. If the transformation $(a, b) \rightarrow (\mu, \phi)$ is one-to-one, where $\mu \in (0, 1)$ is the mode of $\mathcal{F}$ and $\phi > 0$ is a nuance parameter; hence, it is possible to obtain a parametrization based on the mode and subsequently the parametric modal regression for bounded data.

2.1. The Beta distribution

The Beta distribution is certainly the first choice to describe data on $(0, 1)$. It has been extensively discussed and several applied works using the Beta distribution are presented in the literature. A random variable Y follows a Beta distribution with shape parameters $a, b > 0$, if its probability density function (p.d.f.) can be written as

$$f(y|a, b) = \frac{1}{B(a, b)} y^{a-1} (1 - y)^{b-1}, \qquad 0 < y < 1$$

where $B_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1}dt$ denotes the incomplete Beta function and $B(a,b) = \int_0^1 t^a(1-t)^{b-1}dt$ is the complete Beta function. When $a, b > 1$ there is a unique mode for the Beta distribution, which is given by

$$\text{Mode}(Y) = \frac{a-1}{a+b-2}. \tag{1}$$

Recently, Zhou and Huang (2020) proposed a new parametrization of Beta distribution based on the mode. In particular, they set $a = 1 + \phi\mu$ and $b = 1 + \phi(1-\mu)$, where $\phi > 0$ is the shape/precision parameter and $\mu \in (0,1)$ is the mode.

Under this parametrization, the p.d.f. and variance of Beta are given, respectively, by

$$f(y|\mu,\phi) = \frac{y^{\mu\phi}(1-y)^{(1-\mu)\phi}}{B(\mu\phi+1,(1-\mu)\phi+1)} \qquad 0 < y < 1 \tag{2}$$

and

$$\text{Var}(Y) = \frac{(1+\phi\mu)(1+\phi(1-\mu))}{(2+\phi)^2(3+\phi)}. \tag{3}$$

2.2. The Kumaraswamy distribution

Kumaraswamy (1980) introduced a new bounded distribution which has known by his named and recently received great attention. In particular, the Kumaraswamy distribution can be obtained by using the transformation $\exp(-Y)$, when Y follows the generalized exponential distribution (Gupta and Kundu 1999).

If Y follows the Kumaraswamy distribution, its p.d.f. is given by

$$f(y|a,b) = aby^{a-1}(1-y^a)^{b-1}, \qquad 0 < y < 1$$

where $a, b > 0$ are shape parameters.

As discussed by Jones (2009) the Kumaraswamy distribution is unimodal when $a, b > 1$ and its mode is given by

$$\text{Mode}(Y) = \left(\frac{a-1}{ab-1}\right)^{1/a}. \tag{4}$$

In order to propose a mode-type parametrization for Kumaraswamy distribution, we write its p.d.f. in terms of $b = \phi^{-1}\left[1 + \mu^{-\phi}(\phi-1)\right]$ and $a = \phi$, where $\phi > 0$ is the shape parameter and $\mu \in (0,1)$ is the mode. Hence, the p.d.f. and variance of Kumaraswamy can be written, respectively, as

$$f(y|\mu,\phi) = \left[1 + (1-\phi)\mu^{-\phi}\right]y^{\phi-1}(1-y^\phi)^{\phi^{-1}\left[1+\mu^{-\phi}(\phi-1)\right]-1}, \qquad 0 < y < 1 \tag{5}$$

and

$$\text{Var}(Y) = bB(1+2/\phi, b) - \{bB(1+1/\phi, b)\}^2. \tag{6}$$

2.3. The unit-Gamma distribution

Grassia (1977) proposed the unit-Gamma (UGa) distribution, which has been neglected in the statistical literature until recently, Mazucheli et al. (2018a) deriving bias corrections for its parameters. Since then several works related to unit-Gamma have been presented.

If Y follows the UGa distribution, its p.d.f. is given by

$$f(y|a,b) = \frac{b^a}{\Gamma(a)} y^{b-1} (-\log y)^{a-1}, \qquad 0 < y < 1$$

where $a, b > 0$ are shape parameters.

According to Grassia (1977) the UGa distribution is unimodal when $a, b > 1$ and its mode is given by

$$\text{Mode}(Y) = \exp\left(\frac{1-a}{b-1}\right). \tag{7}$$

In order to introduce a mode-type parametrization for UGa distribution let assume that $b = [1 + \log(\mu) - \phi] \log(\mu)^{-1}$, where $\phi = a > 0$ is the shape parameter and $\mu \in (0,1)$ is the mode. Thus, the p.d.f. and variance of UGa are given, respectively, by

$$f(y|\mu,\phi) = \frac{1}{\Gamma(\phi)} [1 + \log(\mu) - \phi]^{\phi} \log(\mu)^{-\phi} y^{(1-\phi)\log(\mu)} (-\log y)^{\phi-1}, \qquad 0 < y < 1 \tag{8}$$

and

$$\text{Var}(Y) = \left(\frac{\phi}{\phi+2}\right)^{[1+\log(\mu)-\phi]\log(\mu)^{-1}} - \left(\frac{\phi}{\phi+1}\right)^{2[1+\log(\mu)-\phi]\log(\mu)^{-1}}. \tag{9}$$

2.4. The unit-Gompertz distribution

Recently, Mazucheli et al. (2019a) by considering an appropriate transformation on Gompertz distribution introduced the unit-Gompertz (UGz) distribution, for which the p.d.f. is given by

$$f(y|a,b) = a\,b\,y^{-(b+1)} \exp\left[-a\left(y^{-b} - 1\right)\right], \qquad 0 < y < 1$$

where $a, b > 0$ are shape parameters.

The authors showed that UGz distribution is log-concave and unimodal for all $a, b > 0$. The mode is reached at

$$\text{Mode}(Y) = \left(\frac{ab}{b+1}\right)^{1/b}. \tag{10}$$

Similar to other distributions, the mode-type parametrization for UGz is obtained by considering $a = \phi^{-1}\mu^{\phi}(\phi+1)$ and $b = \phi$, where $\phi > 0$ is the shape parameter and $\mu \in (0,1)$ is the mode. Thus, the p.d. f. and variance of UGz are given, respectively, by

$$f(y|\mu,\phi) = \mu^{\phi}(\phi+1) y^{-(\phi+1)} \exp\left[-\phi^{-1}\mu^{\phi}(\phi+1)\left(y^{-\phi} - 1\right)\right], \qquad 0 < y < 1 \tag{11}$$

and

$$\text{Var}(Y) = \left[\phi^{-1}\mu^{\phi}(\phi+1)\right]^{2/\phi} e^{2\phi^{-1}\mu^{\phi}(\phi+1)} \left\{\Gamma\left(1 - \frac{2}{\phi}, \phi^{-1}\mu^{\phi}(\phi+1)\right) - \left[\Gamma\left(1 - \frac{1}{\phi}, \phi^{-1}\mu^{\phi}(\phi+1)\right)\right]^2\right\}. \tag{12}$$

where $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function.

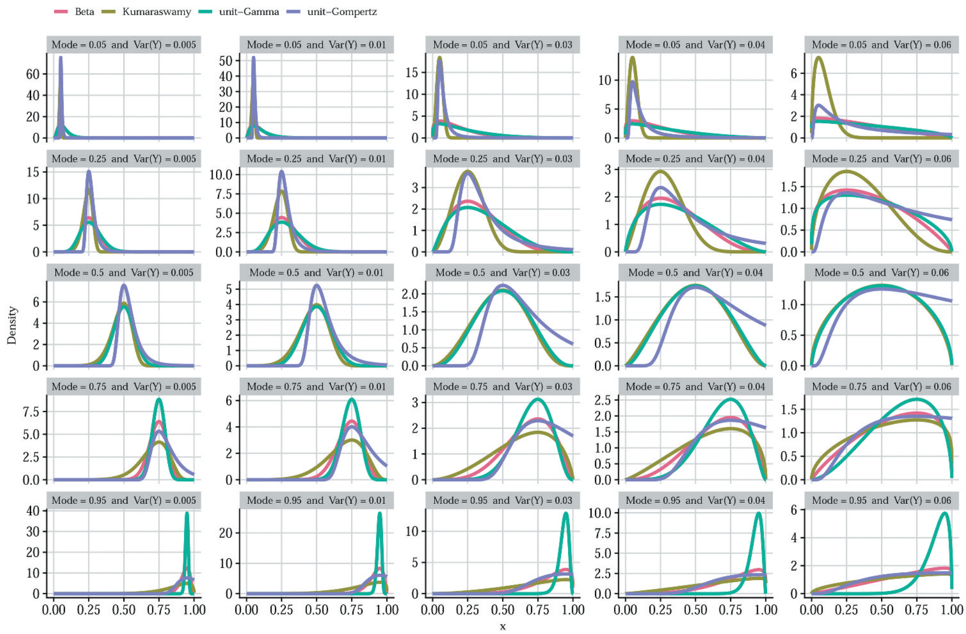

Figure 2. Probability density functions for distinct values of variance (fixed for $\mu = 0.5$) and mode.

2.5. Comparing bounded distribution

To explore and compare the flexibility of the aforementioned models to deal with unimodal bounded data, Figures 1 and 2 present the behavior of variance in terms of mode and the shape of p.d.f. according to variance and mode.

Figure 1 shows the mode and variance relationship of the aforementioned distribution for different values of ϕ. For unit-Gompertz, we also consider $\phi = 0.5$, since the distribution is unimodal for all $\phi > 0$, as discussed above. Interestingly, to point out that:

- All four distributions have important differences in their shapes.
- unit-Gamma and Kumaraswamy have complementary behavior. Their maximum variance is approximately 0.08 for $\mu \to 0$ and $\mu \to 1$, respectively.
- Beta presents a symmetric relationship between mode and variance. Its maximum variance is around 0.06 attained at $\mu = 0.5$.
- unit-Gompertz has a singular behavior and the maximum variance is around 0.07.

To compare the shape of distributions used in regression analysis, we fixed the values of ϕ to have different variance when $\mu = 0.5$. The p.d.f. presented in Figure 2 shows that the distributions have different shapes for the same value of variance and mode. This theoretical results emphasize the distinct characteristics of the distributions, which in practice means that one distribution can fit better than the other depending on the observed data set. Therefore, we humbly recommend that practitioners should check all distributions available in order to provide more reliability of inference conclusions.

3. Regression modeling

Suppose that a random experiment (designed or observational) is conducted and the primary outcome can be describe by a continuous and bounded random variable Y_i, for $i = 1, \ldots, n$, conditionally to fixed-effects or non-random variable x_i. We shall assume that the observed variable y_i is a set of

Table 1. Summary of the considered modal regression models for bounded data.

Distribution	Notation	Unimodal	Variance
Beta	Beta(μ_i, ϕ)	$\phi > 1$	$\frac{(1+\phi\mu_i)(1+\phi(1-\mu_i))}{(2+\phi)^2(3+\phi)}$
Kumaraswamy	Kum(μ_i, ϕ)	$\phi > 1$	See (6)
unit-Gamma	UGa(μ_i, ϕ)	$\phi > 1$	See (9)
unit-Gompertz	UGz(μ_i, ϕ)	$\phi > 0$	See (12)

independent realizations of Y_i according to the distributions $\mathcal{F}(\cdot,\cdot)$. Thus, the mode regression models based on the aforementioned distributions are defined by

$$Y_i|x_i \sim \mathcal{F}(\mu_i, \phi), \quad \text{with} \quad \mu_i = g^{-1}\left(x_i^{\mathrm{T}}\beta\right), \tag{13}$$

where ϕ is a precision/shape parameter of the distribution $\mathcal{F}$, μ_i denotes the mode of $\mathcal{F}$ and $g(\cdot)$ is an appropriate link function.

Since the mode of all distribution lies on $(0,1)$ the most useful well-known link functions for $g(\cdot)$ are logit: $g(\mu_i) = \log\left(\mu_i/(1-\mu_i)\right)$; probit: $g(\mu_i) = \Phi^{-1}(\mu_i)$, where $\Phi^{-1}(\cdot)$ is the standard normal quantile function; and complementary log-log: $g(\mu_i) = \log\left[-\log(1-\mu_i)\right]$.

Due to the direct interpretation of the parameters, in this paper we consider the *logit* link function. Thus, we have that

$$\frac{\mu_i}{1-\mu_i} = \exp\left(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}\right), \tag{14}$$

which leads to the following interpretations:

- If x_{1i} is continuous, for a unit increase, the percentage change in the mode response is $100\% \cdot (e^{\widehat{\beta}_1} - 1)$, keeping the other predictors fixed.
- If x_{1i} is an indicator variable then $100\% \cdot e^{\widehat{\beta}_1}$ represents the percentage change in the mode response for $x_{1i} = 1$ to $x_{1i} = 0$, keeping the other predictors fixed.

As discussed in Section 2, the aforementioned distributions allow heteroscedasticity, that is, the conditional variance depends on the covariates. Table 1 summarizes the main properties of the bounded distributions considered and establishes the notation.

4. Estimation

Parameter estimation and inference are conducted under the classical approach. Let $y = (y_1, \ldots, y_n)$ be a random sample from (13), the maximum likelihood of the parameter vector $\theta = (\beta, \phi)$ is given by

$$\widehat{\theta} = \arg\left[\sup_{\theta\in\Theta} \ell(\theta|y)\right] \tag{15}$$

where $\ell(\cdot|\cdot)$ denotes the log-likelihood of distribution $\mathcal{F}(\cdot,\cdot)$ and Θ is the parameter space of θ.

For the proposed models, it is not possible to derive analytical solution for the MLE $\widehat{\theta}$, thus numerical solution using optimization algorithm such as Newton-Raphson and quasi-Newton is performed. In particular, we considered the BFGS algorithm available in the optim function from **R**. Furthermore, inference regarding the model parameters is conducted using the large sample theory of the maximum likelihood (Cox and Hinkley 1974), where the observed Fisher information matrix given by

$$\mathbf{J}(\widehat{\theta}) = -\frac{\partial}{\partial\theta^{\mathrm{T}}\partial\theta^{T}} \ell(\theta|y)\bigg|_{\theta=\widehat{\theta}}$$

is used to estimated the standard errors of parameter.

The proposed models are freely available to users through the open-source **R** package unitModalReg at https://github.com/AndrMenezes/unitModalReg, with a similar interface to the standard glm (R Core Team 2020) function for fitting generalized linear models.

For sake of simplicity, we are assuming that the nuance parameter does not vary across the covariates. This is justified by the fact that the four regression models accommodate heteroscedasticity without the need to model directly the nuance parameter, see the expressions of variance in Table 1. Nonetheless, it is noteworthy that a strictly positive link function relating the nuance parameter ϕ with covariates w_i, not necessarily equal to x_i, can be considered.

4.1. Model adequacy

In order to evaluate the departures from model assumptions, the randomized quantile residuals introduced by Dunn and Smyth (1996) are considered. Generally, they are defined as follows

$$r_i = \Phi^{-1}\left[F(y_i, |\widehat{\mu}_i, \widehat{\phi})\right], \quad i = 1, \ldots, n,$$

where $\Phi(\cdot)$ is the standard normal distribution function, $F(\cdot|\cdot,\cdot)$ is the cumulative distribution function of $\mathcal{F}$ distribution and $\widehat{\mu}_i = g^{-1}(x_i^{\mathrm{T}}\widehat{\beta})$.

Apart from the variability due to the estimates of parameters, these residuals have standard normal distribution if the proposed model is correctly specified (Dunn and Smyth 1996). Thus, to check if the model assumption is adequate, we can examine the residuals plots with simulated envelope proposed by Atkinson (1981). If a large proportion of the observations lies outside the envelope, then one has evidence against the adequacy of the fitted model (Oliveira et al. 2019).

5. Applications

In this section, three real applications are presented in order to explore the potentiality of each regression model to solve empirical problems. To perform a discrimination between the models, two likelihood-based criteria are considered. The AIC criterion proposed by Akaike (1974) and defined as $\text{AIC} = 2p - 2\log\widehat{L}$, where L is the likelihood evaluated at the MLE and p is the number of parameters in the model. The decision rule is favorable to the model with the lowest value. In addition, we performed a formal test based on the Vuong likelihood ratio test for non-nested models (Vuong 1989). The Vuong statistic to compare two regression models is defined by

$$T = \frac{1}{\widehat{\omega}^2\sqrt{n}} \sum_{i=1}^{n} \log\frac{f(y_i|x_i,\widehat{\theta})}{g(y_i|x_i,\widehat{\gamma})}$$

where

$$\widehat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f(y_i|x_i,\widehat{\theta})}{g(y_i|x_i,\widehat{\gamma})}\right)^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f(y_i|x_i,\theta)}{g(y_i|x_i,\widehat{\gamma})}\right)\right]^2$$

is an estimator for the variance of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\log\frac{f(y_i|x_i,\widehat{\theta})}{g(y_i|x_i,\widehat{\gamma})}$, $f(y_i|x_i,\widehat{\theta})$, and $g(y_i|x_i,\widehat{\gamma})$ are the corresponding rival densities evaluated at the maximum likelihood estimates. When $n \to \infty$ we have that $T \xrightarrow{D} N(0,1)$. Therefore, at $\alpha\%$ level of significance the null hypothesis of the equivalence of the competing distributions is rejected if $|T| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $\alpha/2$ quantile of standard normal distribution. In practical terms, $f(y_i|x_i,\widehat{\theta})$ is better (worse) than $g(y_i|x_i,\widehat{\gamma})$ if $T > z_{\alpha/2}$ (or $T < -z_{\alpha/2}$).

We also provide in Appendix A empirical graphics and formal statistical test to check misspecification of link function for all models in each applications.

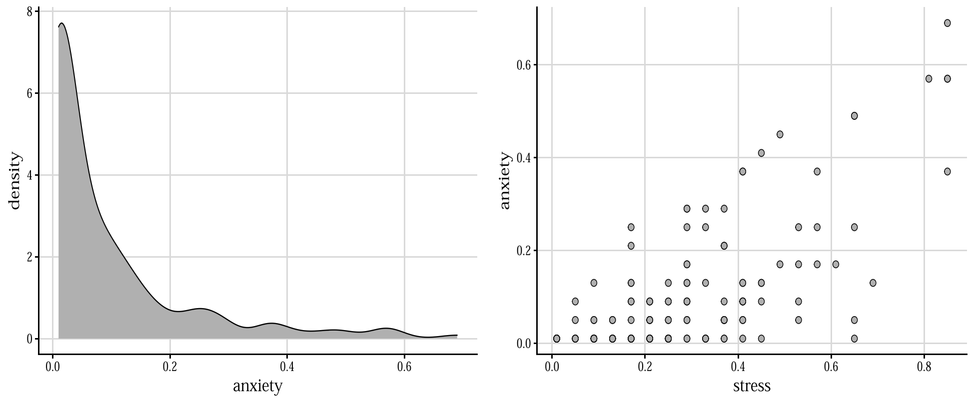

Figure 3. Kernel density plot of anxiety and scatter plot of anxiety versus stress.

Table 2. Summary of the fitted models – Dependency of anxiety on stress.

Parameter	Beta	UGz	UGa	Kum
β_0	−6.940 (0.590)	−5.549 (0.171)	−7.726 (0.637)	−8.441 (0.893)
β_1	8.472 (0.831)	4.063 (0.490)	9.432 (0.868)	6.291 (0.621)
ϕ	2.649 (0.100)	0.983 (0.071)	7.717 (0.756)	1.014 (0.012)
AIC	−521.070	−619.737	−588.402	−605.412

Table 3. Vuong tests for all model combinations – Dependency of anxiety on stress.

Comparison	Vuong	*p*-value
Beta versus UGz	−9.042	< 0.000
Beta versus UGa	−7.068	< 0.000
Beta versus Kum	−3.833	< 0.000
UGz versus UGa	2.371	0.009
UGz versus Kum	0.905	0.183
UGa versus Kum	−1.120	0.131

5.1. Dependency of anxiety on stress

This data set is from Smithson and Verkuilen (2006) and consists of a sample of 166 nonclinical women in Townsville, Queensland, Australia. The variables are linearly transformed scales from the Depression Anxiety Stress Scales, which normally range from 0 to 42. In Figure 3, left-hand plot depicts the density function of the response variable (anxiety) while in the right, it shows the scatter plotted the anxiety against stress. From this figure, its easy to see that the density is unimodal, also as the stress score increases the anxiety also increase. In these data, our focus is on the modal anxiety and how it is being caused by stress. Here, modeling the conditional mode is obviously more appropriate than the conditional mean or median due to the corresponding risk to public health.

Here, we are regressing the mode of anxiety in terms of the stress predictor, i.e.,

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \text{stress}_i, \quad i = 1, \ldots, 166.$$

Table 2 reports the parameter estimates, their standard errors (in parentheses) and the AIC criterion by model. Also, Table 3 shows the Voung test results for all possible model comparisons. Based on the AIC criterion and Voung test the UGz distribution provides the best fit for this data set. However, it is observed that the Voung test do not reject the null hypothesis of model equivalence between UGz and Kum regression models. The estimated coefficient associated to stress effect (β_1) of UGz model is quite different comparing to other distribution specification, especially for Beta and

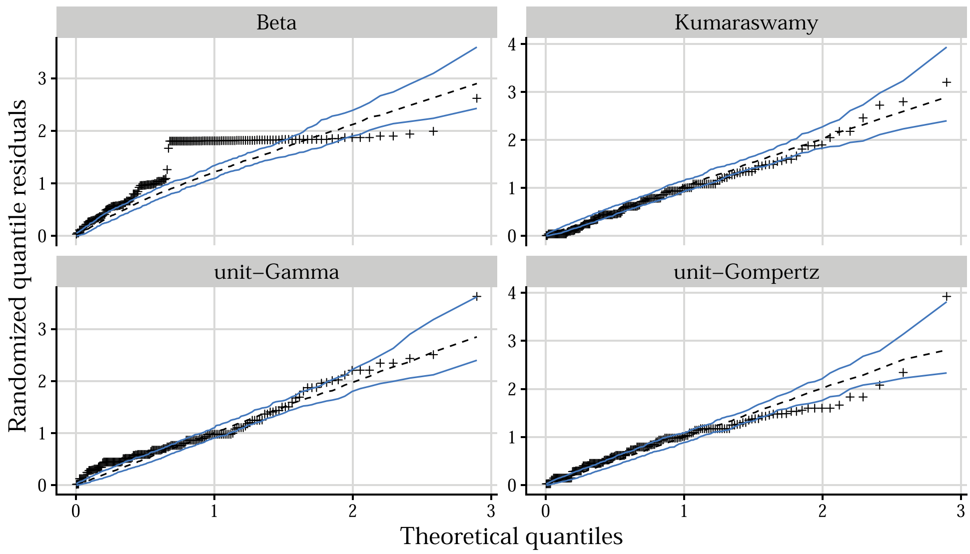

Figure 4. Half-normal plot with simulated envelope – Dependency of anxiety on stress.

UGa regression models. For instance, the β_1 estimate of Beta is approximately 52% large than the UGz regression.

The half-normal plot with simulated envelope presented in Figure 4 reinforce that Beta distribution do not provide a good fit to these data set, motivating the use of different distributions assumption to analyze an observed data set.

This empirical application shows the importance of providing diagnostic plots (the Half-normal plots) beside the goodness-of-fit measures. In fact, the Voung test conducted shows that the UGz and Kum model are statistically equivalent, since the p-value is not significantly small (see Table 3).

5.2. Proportion of votes

In this analysis, we considered the proportion of votes that the Dilma Roussef of the Partido dos Trabalhadores (PT) received in the second turn of presidential elections of 2010 in the municipals of Rio de Janeiro (RJ) and Ceará (CE) states. Our goal is to measure the impact of MHDI and region (southeast and northeast) on the proportion of votes achieved by Dilma in 2010. The data set has 270 observations, 86 cities of RJ and 184 from CE. The source of data is *Tribunal Superior Eleitoral* (http://www.tse.jus.br/) and *Atlas do Desenvolvimento Humano no Brasil* (http://www.atlasbrasil.org.br/2013/pt/).

In any election process, the parties more often are interested to check the determinants (covariates) of the modal proportion in their favor for obvious reason to learn voters behavior to formulate their future poll strategies. Averages of proportion of votes is not of much interests compared to the modal proportion.

For an empirical investigation the covariate MHDI is categorized based on the quantiles using 3 bins the resulting classes are $(0.54, 0.61)$, $(0.611, 0.659)$ and $(0.66, 0.837)$. Hence, the density of the response variable (proportion of votes) is plotted in Figure 5 from left to right for these three classes according to the state CE (in blue) and RJ (in red). In fact, a visual inspection indicates that the response variable conditional to covariate MHDI is unimodal except for the state RJ in second category where few smaller more peaks around the the main mode can be seen. We further observe from the plots that (i) the greater the MHDI lesser the proportion of votes and (ii) there is a bigger difference of proportion of votes between the states.

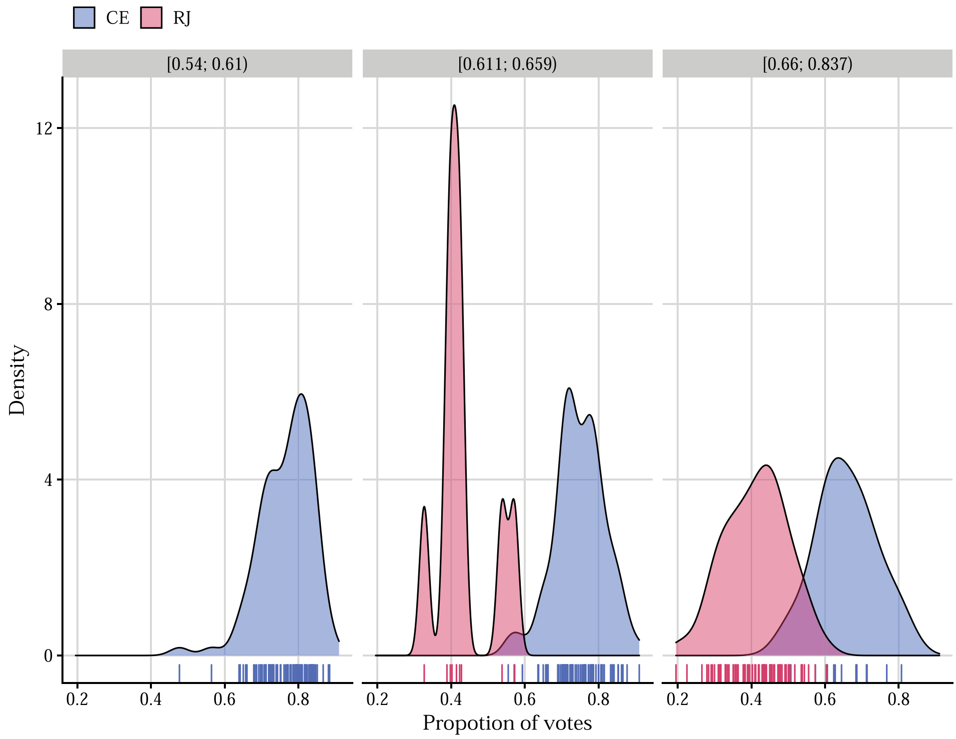

Figure 5. Kernel density plots of proportion of votes of Dilma according to the categorized MHDI (from left to right $(0.54, 0.61)$, $(0.611, 0.659)$ and $(0.66, 0.837)$) and colored by the states.

Table 4. Dip statistic and *p*-value to test the unimodality of proportion of votes conditional to the categorized MHDI and states.

State	MHDI	Statistic	*p*-value
CE	[0.54; 0.61)	0.0341	0.6808
CE	[0.611; 0.659)	0.0499	0.1134
RJ	[0.611; 0.659)	0.0856	0.8446
CE	[0.66; 0.837)	0.0870	0.5238
RJ	[0.66; 0.837)	0.0263	0.9884

Table 5. Summary of the fitted models – Proportion of votes.

Parameter	Beta	UGz	UGa	Kum
β_0	3.211 (0.438)	1.439 (0.527)	3.288 (0.425)	3.298 (0.365)
β_1	−3.295 (0.706)	−0.686 (0.834)	−3.444 (0.689)	−3.389 (0.587)
β_2	−1.237 (0.081)	−1.613 (0.099)	−1.214 (0.085)	−1.069 (0.073)
ϕ	3.550 (0.090)	4.587 (0.263)	11.184 (0.949)	8.686 (0.464)
AIC	−642.321	−477.106	−629.813	−597.885

We considered the dip statistics introduced by Hartigan (1985) to test the null hypothesis $\mathcal{H}_0$: distribution is unimodal versus the alternative hypothesis $\mathcal{H}_1$: the distribution is non-unimodal, that is, at least bimodal. As shown in Table 4 for all the categories, the dip test do not reject the null hypothesis of unimodality, that is, that the distribution of proportion of votes of Dilma is unimodal.

To analyze the impact of MHDI and state on the proportion of votes of Dilma, we consider the following linear predictor for the mode of the aforementioned distributions

Table 6. Vuong tests for all model combinations – Proportion of votes.

Comparison	Vuong	*p*-value
Beta versus UGz	15.057	< 0.000
Beta versus UGa	1.337	0.091
Beta versus Kum	2.953	0.002
UGz versus UGa	−5.594	< 0.000
UGz versus Kum	−5.011	< 0.000
UGa versus Kum	1.432	0.076

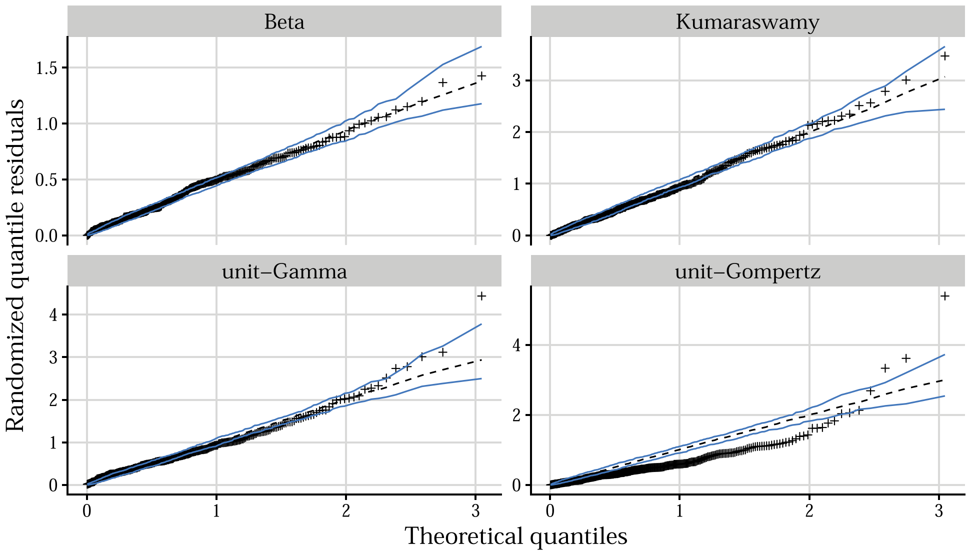

Figure 6. Half-normal plot with simulated envelope.

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \text{MHDI}_i + \beta_2 \text{State}_i, \quad i = 1, \ldots, 270$$

where $\text{State}_i = 1$ if the i city is from RJ and 0 if it is from CE.

Table 5 lists the parameter estimates, their standard errors (in parentheses) and the AIC criteria by model. For Beta, UGa and Kum regression model, the ϕ estimates are greater than one, indicating that the response variable conditional to the covariates are unimodal. The obtained results show a large difference, in terms of AIC values, in favor to Beta and UGa distributions, with a slight superiority of the former. However, the estimated coefficients are very similar, except for β_1 in UGz mode, which implies in related interpretations. This model in comparison corroborated with Voung tests performed for all model combinations and reported in Table 6.

To check the model assumption, the half-normal plot with simulated envelope of the proposed models is presented in Figure 6. These figures corroborate with likelihood criteria and show that the Beta and UGa following by Kumaraswamy models provide the best fit to these data.

Based on the inference results of the models, it can be concluded that (i) the mode of response variable decrease approximately by $3.6\%(e^{-3.3})$ as the MHDI increase, this means that cities with large MHDI provided less votes for Dilma and (ii) the RJ state has a mode $30\%(e^{-1.2})$ smaller than the CE, meaning that the cities of CE state provide more votes than RJ for Dilma.

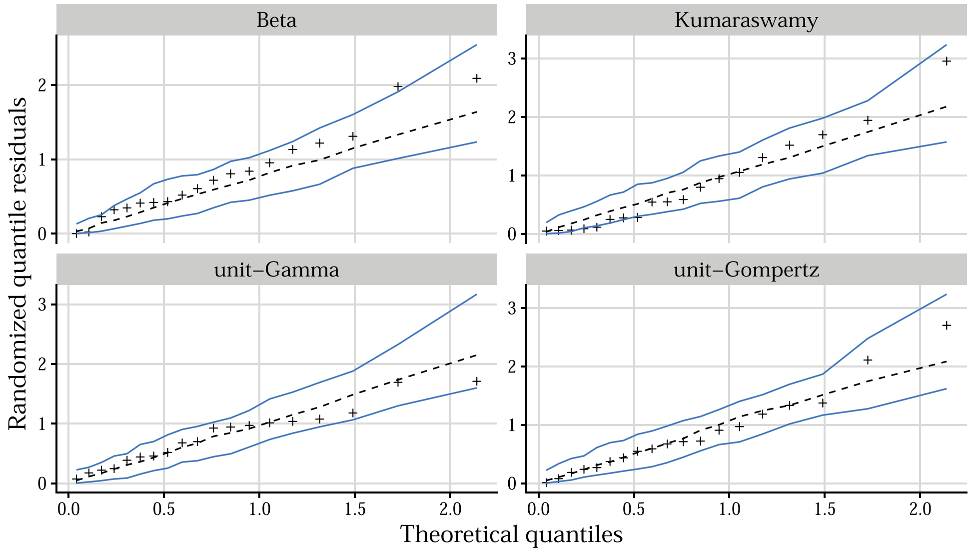

Figure 7. Half-normal plot with simulated envelope – Birds migration.

5.3. Birds migration

This data set has been taken from the recent book of Korosteleva (2019) and related to migration of birds. As mentioned by the author, the source of the data set came from consulting projects, which she had involved. In particular, a group of ornithologists ringed 19 flocks of migratory birds prior to migration and collected some measures. This analysis is important to evaluate the empirical model assumptions when the sample size is small.

Here, the aim of investigation is to seek the association between the proportion of birds that successfully reach the winter grounds and average mass (in grams), average wingspan (in cm) and the distance traveled (in km). Again, it is more adequate to consider modal proportion instead of mean as an investigator might ask how the covariates impact in determining the modal proportion of successful winter migration. Hence, we have regressed the modal proportion of birds that successfully reach the winter grounds on the covariates average mass (in grams), average wingspan (in cm) and the distance traveled (in km). We consider the following linear predictor for the mode parameter

$$\text{logit}(\mu_i) = \beta_0 + \beta_1\text{Mass} + \beta_2\text{Wingspan} + \beta_3\text{Distance}, \quad i = 1, \ldots, 19.$$

Table 7 shows the parameter estimates, their standard errors (in parentheses) and the AIC criteria for each model. For Beta, UGa and Kum regression models the ϕ estimates are greater than one, indicating that the response variable conditional to the covariates is unimodal. The results show a quite difference, in terms of AIC and Voung statistic (see Table 8), in favor to Kumaraswamy

Table 7. Summary of the fitted models – Birds migration.

Parameter	Beta	UGz	UGa	Kum
β_0	7.574 (3.423)	5.643 (5.069)	4.280 (1.384)	9.847 (3.359)
β_1	−0.263 (1.032)	−0.168 (1.279)	−0.587 (1.796)	0.084 (0.315)
β_2	−0.556 (1.826)	−1.473 (3.030)	5.002 (5.959)	−2.035 (1.026)
β_3	−3.260 (1.397)	−2.302 (1.927)	−3.838 (2.398)	−3.785 (1.331)
ϕ	1.009 (0.452)	0.615 (0.290)	1.273 (0.388)	3.083 (0.737)
AIC	−2.909	3.245	−4.383	−6.730

Table 8. Vuong tests for all model combination – Birds migration.

Comparison	Vuong	*p*-value
Beta versus UGz	1.297	0.097
Beta versus UGa	−0.194	0.423
Beta versus Kum	−1.145	0.126
UGz versus UGa	−1.174	0.120
UGz versus Kum	−1.911	0.028
UGa versus Kum	−0.256	0.398

distribution. Also, the estimated coefficients are very different according to distribution specification, particularly the Mass (β_1) and Wingspan (β_2) effects.

The half-normal plot with simulated envelope presented in Figure 7 evidencing that the assumption of Kumaraswamy distribution is more appropriate for this data set.

6. Discussion

The present paper has contributed with a discussion and proposition of parametric modal regression for bounded data. Three new models are proposed, based on the unit-Gamma, Kumaraswamy and unit-Gompertz distributions. The proposed models along with the Beta distribution are applied in modeling three different data sets. Maximum likelihood estimation for model fitting and model selection performed using AIC criterion and Voung test, while model adequacy is checked by residuals plots with simulated envelope. All the four models investigated presents different characteristics and proved to be useful in different context as potential modal regression models for bounded response. Our findings for the modeling highlight these clearly. Furthermore, for reproducible research and practical utility the R package unitModalReg is freely available at https://github.com/AndrMenezes/unitModalReg. As such we envisage that the proposed models will be highly utilized across all relevant fields of science.

Acknowledgments

The authors are thankful to the referees for many valuable suggestions. Josmar Mazucheli gratefully acknowledge the partial financial support from Fundação Araucária (Grant 064/2019 - UEM/Fundação Araucária).

ORCID

André F. B. Menezes http://orcid.org/0000-0002-3320-9834
Josmar Mazucheli http://orcid.org/0000-0001-6740-0445
Subrata Chakraborty http://orcid.org/0000-0002-6405-1486

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6):716–723. doi:10.1109/TAC.1974.1100705.

Aristodemou, K. 2014. New regression methods for measures of central tendency. Ph.D. thesis. In *School of Information Systems*. Brunel University: Computing and Mathematics.

Atkinson, A. C. 1981. Two graphical displays for outlying and influential observations in regression. *Biometrika* 68 (1):13–20. doi:10.1093/biomet/68.1.13.

Chen, Y.-C. 2018. Modal regression using kernel density estimation: A review. *Wiley Interdisciplinary Reviews: Computational Statistics* 10 (4):e1431. doi:10.1002/wics.1431.

Chen, Y., Genovese, C. R., Tibshirani, R. J. and Wasserman, L. et al. 2016. Nonparametric modal regression. *The Annals of Statistics*. 44(2):489–514. doi:10.1214/15-AOS1373.

Core Team, R. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/ .

Cox, D. R., and D. V. Hinkley. 1974. *Theoretical Statistics*. London: Chapman & Hall/CRC.
Dunn, P. K., and G. K. Smyth. 1996. Randomized quantile residuals. *Journal of Computational and Graphical Statistics* 5 (3):236–244.
Ghitany, M. E., J. Mazucheli, A. F. B. Menezes, and F. Alqallaf. 2019. The unit-Inverse-Gaussian distribution: A new alternative to two-parameter distributions on the unit interval. *Communications in Statistics - Theory and Methods* 48 (14):3423–3438. doi:10.1080/03610926.2018.1476717.
Grassia, A. 1977. On a family of distributions with argument between 0 and 1 obtained by transformation of the Gamma distribution and derived compound distributions. *Australian Journal of Statistics* 19 (2):108–114. doi:10.1111/j.1467-842X.1977.tb01277.x.
Gupta, R. D., and D. Kundu. 1999. Generalized exponential distributions. *Australian New Zealand Journal of Statistics* 41 (2):173–188. doi:10.1111/1467-842X.00072.
Hartigan, P. M. 1985. Algorithm as 217: Computation of the dip statistic to test for unimodality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 34 (3):320–325.
Jones, M. 2009. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology* 6 (1):70–81. doi:10.1016/j.stamet.2008.04.001.
Korosteleva, O. 2019. *Advanced Regression Models with SAS and R*. Taylor & Francis Group: CRC Press.
Kumaraswamy, P. 1980. A generalized probability density function for double-bounded random processes. *Journal of Hydrology* 46 (1–2):79–88. doi:10.1016/0022-1694(80)90036-0.
Lee, M. 1989. Mode regression. *Journal of Econometrics* 42 (3):337–349. doi:10.1016/0304-4076(89)90057-2.
Mazucheli, J., A. F. Menezes, and S. Dey. 2019a. Unit-gompertz distribution with applications. *Statistica* 79 (1):25–43.
Mazucheli, J., A. F. B. Menezes, L. B. Fernandes, R. P. Oliveira, and M. E. Ghitany. 2020b. The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. *Journal of Applied Statistics* 47 (6):954–974. doi:10.1080/02664763.2019.1657813.
Mazucheli, J., A. F. B. Menezes, and M. E. Ghitany. 2018c. The unit-Weibull distribution and associated inference. *Journal of Applied Probability and Statistics* 13 (2):1–22.
Mazucheli, J., A. F. B. Menezes, and S. Chakraborty. 2019b. On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics* 46 (4):700–714. doi:10.1080/02664763.2018.1511774.
Mazucheli, J., A. F. B. Menezes, and S. Dey. 2018a. Improved maximum-likelihood estimators for the parameters of the unit-gamma distribution. *Communications in Statistics - Theory and Methods* 47 (15):3767–3778. doi:10.1080/03610926.2017.1361993.
Mazucheli, J., A. F. B. Menezes, and S. Dey. 2018b. The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics* 9 (1):47–57.
Mazucheli, J., S. R. Bapat, and A. F. B. Menezes. 2020a. A new one-parameter unit-Lindley distribution. *Chilean Journal of Statistics* 11 (1):53–67.
Oliveira, R. P., A. F. B. Menezes, J. Mazucheli, and J. A. Achcar. 2019. Mixture and nonmixture cure fraction models assuming discrete lifetimes: Application to a pelvic sarcoma dataset. *Biometrical Journal* 61 (4):813–826. doi:10.1002/bimj.201800030.
Pan, Y., E. Imani, M. White, and A.-M. Farahmand. 2020. An implicit function learning approach for parametric modal regression. *arXiv Preprint arXiv* 2002:06195.
Pereira, T. L., and F. Cribari-Neto. 2014. Detecting model misspecification in inflated beta regressions. *Communications in Statistics - Simulation and Computation* 43 (3):631–656. doi:10.1080/03610918.2012.712183.
Ramsey, J. B. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)* 31 (2):350–371. doi:10.1111/j.2517-6161.1969.tb00796.x.
Sager, T. W., and R. A. Thisted. 1982. Maximum likelihood estimation of isotonic modal regression. *The Annals of Statistics* 10 (3):690–707. doi:10.1214/aos/1176345865.
Smithson, M., and J. Verkuilen. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11 (1):54–71. doi:10.1037/1082-989X.11.1.54.
Vuong, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57 (2):307–333. doi:10.2307/1912557.
Wang, X., H. Chen, W. Cai, D. Shen, and H. Huang, 2017. Regularized modal regression with applications in cognitive impairment prediction. In: Advances in neural information processing systems.
Yao, W., and L. Li. 2014. A new regression model: Modal linear regression. *Scandinavian Journal of Statistics* 41 (3):656–671. doi:10.1111/sjos.12054.
Zhou, H., and X. Huang; for the Alzheimer's Disease Neuroimaging Initiative. 2020. Parametric mode regression for bounded responses. *Biometrical Journal* 62 (7):1791–1809. doi:10.1002/bimj.202000039.

Appendix A

Link function misspecification. To check the specification of logit link function in the applications we provide (i) empirical plots of logit of response versus each covariate along with the locally estimated scatterplot smoothing (LOESS) to indicate the linear relationship between logit of response and covariate and (ii) conducted a formal test inspired by the RESET (regression specification error test) introduced by Ramsey (1969) for linear regression models as a general misspecification test. It should be mention that misspecification means that the estimated model differs from the true data generating process in a way that the former does not provide an accurate description of the latter. Incorrectly specified link function is a way of misspecification (Pereira and Cribari-Neto 2014).

Figures 8 and 9, 10 show that the logit of response variables versus the covariates shows an approximately linear relationship. For the Birds migration application discussed in subsection 4.3, only the variable distance traveled was statistically significant (see the estimates and standard errors in Table 7), the scatterplot shows that the logit of proportion of birds that successfully reach the winter grounds have approximately negative linear relationship between the distance traveled (in km).

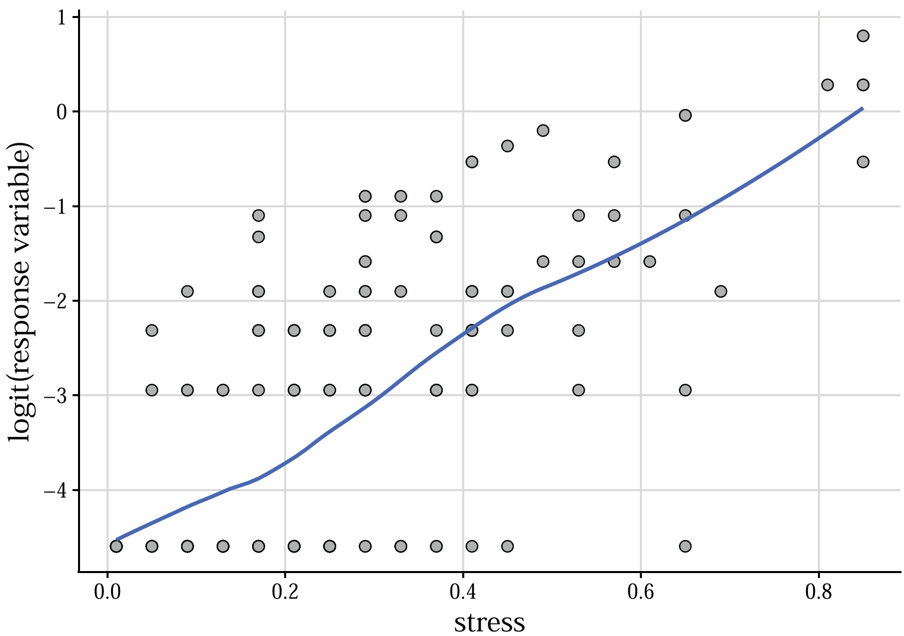

Figure 8. Logit of response variable (anxiety) versus covariate stress. The blue line represents the LOESS – Application 4.1.

The formal misspecification test is conduct adding powers of the fitted linear predictors as new covariates and testing the significance of estimated parameters. In a correct specified link function, the estimated parameters could not be statistically significant. Here, we include the square of fitted linear predictor and use Wald test to check the significance. The results in Table 9 indicate a possible misspecification of link function for the UGz in applications 4.1 and 4.2, for Beta in application 4.2 and UGa in application 4.3.

Table 9. P-values of Wald tests for the covariate square linear predictor adding in the model to check the assumption of link function.

Application	Beta	UGz	UGa	Kum
4.1	0.1566	0.0750	0.1124	0.2484
4.2	0.0389	0.0017	0.1824	0.3524
4.3	0.4335	0.6082	0.0087	0.8384

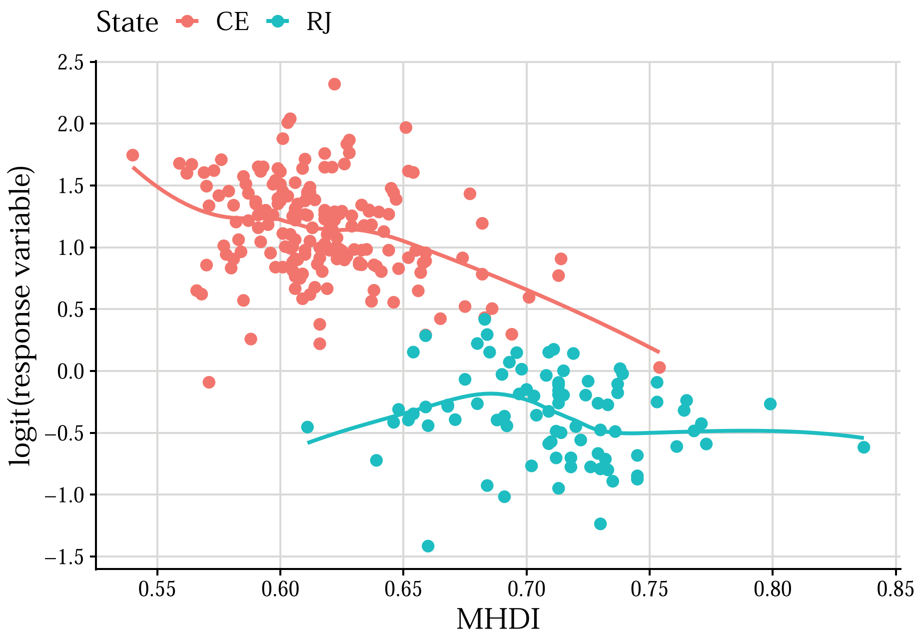

Figure 9. Logit of response variable (proportion of votes) versus covariate MHDI colored by state. The lines represent the LOESS – Application 4.2.

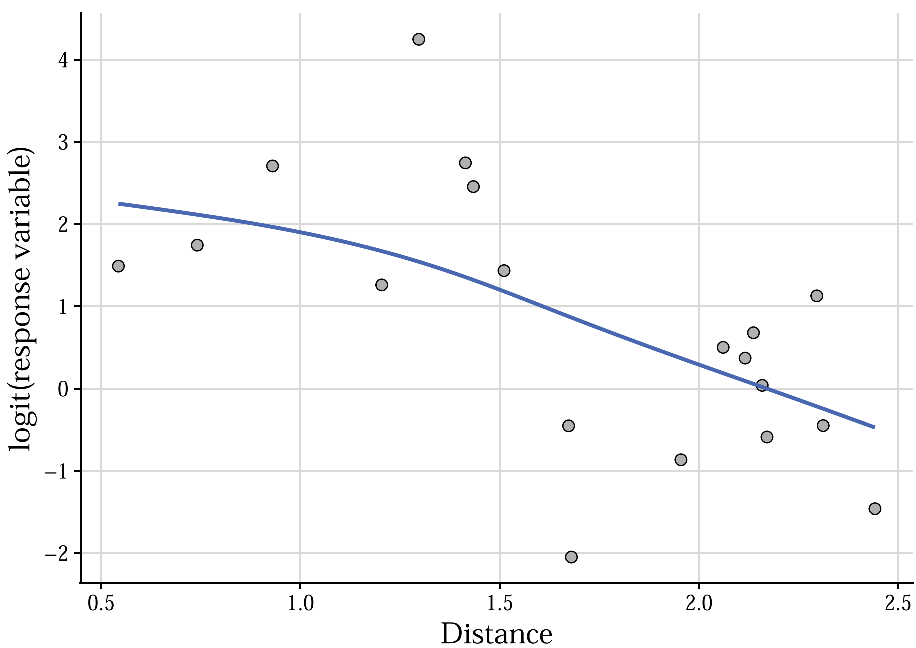

Figure 10. Logit of response variable (proportion of birds that reach the winter) versus covariate distance. The blue line represents the LOESS – Application 4.3.