Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

A new mixed-effects mixture model for constrained longitudinal data

Agnese Maria Di Brisco | Sonia Migliorati

[1]Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy

Correspondence
*Sonia Migliorati, Department of Economics, Management and Statistics, University of Milano-Bicocca - Italy. Email: sonia.migliorati@unimib.it

Summary

Biomedical research often features continuous responses bounded by the interval [0,1]. The well-known beta regression model addresses the constrained nature of these data, while its augmented and mixed-effects variants can address the presence of zeros and/or ones and longitudinal or clustered response values, respectively. However, these models are not robust to the presence of outliers and/or excessive number of observations near the tails. We propose a new augmented mixed-effects regression model based on a special beta mixture distribution that is capable of handling these issues. Extensive simulation studies show the superiority of the proposed model to the models most often used in the literature. The proposed model is applied to two real datasets: one taken from a long-term study of Parkinson's disease (PD) and the other taken from a study on reading accuracy (RA).

KEYWORDS:
proportion, flexible beta distribution, random effect, Bayesian inference, outlier

1 | INTRODUCTION

Bounded continuous response variables (e.g., rates or proportions) are often encountered in many disciplines.[1] The question of modeling these types of constrained outcomes in biomedical trials has been investigated in an analysis of health assessment questionnaires that measure health conditions in terms of percentages,[2] studies on Alzheimer's disease[3] and periodontal disease,[4] and in a microbiome study.[5] In addition to being limited to a restricted interval, bounded responses often show heterogeneity in variance and asymmetry, making standard models unsuitable.[6]

Among parametric approaches, the beta regression model,[6–9] which is based on the assumption of a beta-distributed response variable, has been gaining popularity. The beta distribution allows many diverse shapes (U-shaped, J-shaped, L-shaped, unimodal symmetric, unimodal right- or left-skewed, and constant phenomena), but it fails to represent a wide range of phenomena such as outliers[10] and bimodality.[11]

Given that the presence of outliers often leads to heavy-tailed distributions in the case of bounded responses, a beta-uniform mixture distribution (referred to as a beta rectangular (BR) distribution) has been designed to assign additional weight to the tails of the support.[12] One proposed BR regression model has been found to be robust to the presence of many outlier patterns.[13] Moreover, a regression model for proportions based on the flexible beta (FB) distribution has recently been introduced to achieve even greater flexibility.[11] The FB distribution is defined as a special mixture of two beta distributions. It exhibits good properties, including strong identifiability and likelihood boundedness, which make the model tractable from computational and inferential points of view. The FB regression model has been shown to improve the fit and predictive ability with respect to the beta and the BR regression models in many simulations and real-data contexts.

The excessiveness of 0 and 1 outcome values and longitudinal or clustered measurements are also fundamental issues in the literature, particularly in medical and public health data. The first issue can be addressed through a linear transformation of the original response variable from $[0, 1]$ to the open interval $(0, 1)$,[14] which represents the support of all the cited distributions. Despite its simplicity, this approach is not recommended when a nonnegligible proportion of 0s/ and 1s is present.[9,15] An alternative solution is to augment the probability density function (pdf) of the distribution defined on $(0, 1)$ by adding positive probabilities of the occurrence of the 0s/1s.[15–17] This approach leads to zero-/one-augmented models. The second issue can be addressed by adding random effects to the model, thus obtaining a mixed-effects regression model that handle within-subject correlations. Both of these issues has been addressed in all of the beta and BR regression models.[4,15–18] However, to the best of our knowledge, no model exists that is capable of simultaneously handling an excessive number of boundary values, longitudinal measurements, outliers/heavy tails, and bimodality.

This study generalizes the FB model, which naturally addresses outliers, heavy tails, and bimodality, by introducing an augmented version of the FB distribution, the augmented FB (AFB) distribution, and by defining a new mixed regression model that is appropriate for responses with a hierarchical structure. We adopt a Bayesian approach to inference that is more suitable for complex models such as mixtures, as it avoids the computational and analytical problems of likelihood-based inference and its small-sample limitations. These extensions are motivated by two examples that have been extensively studied in the literature. First, a reading accuracy (RA) dataset, which allows us to investigate the impact of dyslexia and general cognitive ability on reading skills,[19] led us to explore an augmented version of the FB distribution. Second, the Parkinson's Disease (PD) Long-term Study-1 (LS-1) study carried out by the National Institute of Neurological Disorders and Stroke (NINDS)[20,21] led us to extend our mixture model to handle within-subject correlations. For both examples, the new AFB mixed regression model is confirmed preferable to competing models in terms of fit via several diagnostics designed to detect discrepancies between the observed and predicted data.

The remainder of this paper is organized as follows. Section 2 describes the motivating examples, explaining how each inspired our methodological developments. Section 3 briefly describes some useful distributions for proportions and their related regression models, and it introduces a novel regression model with random effects based on the new AFB distribution. Section 4 describes the Bayesian approach to inference as well as several model comparison criteria and model checks based on posterior predictive distributions and cross-validated (leave-one-out) approaches. Section 5 outlines the simulation studies used to evaluate the performance of the AFB model and compare it with the augmented beta rectangular (ABR) and augmented beta (AB) models. Section 6 discusses the results of and main findings from the application of our new regression model to the motivating examples. Finally, Section 7 offers concluding remarks.

2 | MOTIVATING EXAMPLES

The first motivating example, hereafter referred to as "RA", investigates the impact of dyslexia and general cognitive ability on reading skills.[19] In total, 44 children were included in the study, 19 of whom were diagnosed with dyslexia. Their RA level was quantified through psychometric measures as a proportion in $(0, 1]$, where 1 indicates a perfect reading score. In addition to RA, a quantitative covariate for the children's nonverbal abilities (i.e. IQ) was also measured. The RA experiment is well-known among beta regression scholars.[9,14,22–24] A recent study showed the benefit of modeling the constrained outcome, which is affected by a weak bimodality, through an FB regression model rather than through a beta or BR model.[11] It should be noted that the dataset includes a nonnegligible group of children (13 out of 44) who achieved a perfect reading score (i.e., an outcome exactly equal to 1). The simple solution of reducing the 1s by $\epsilon = 0.01$[11,13,14,22,23] avoids the problem but is not recommended when a high proportion of boundary observations is involved.[16] Alternatively, a three component finite mixture has been adopted[24], where one component follows a completely specified distribution, namely a uniform distribution on $(0.98, 1)$, which is used to model the 1s. A criticism of this approach is that it makes it impossible to model the probability of 1s as a function of covariates. A one-AB regression model has also been applied to the problem[9] with promising results, thus motivating us to propose a new augmented mixture regression model.

The second study that inspired the extension of our mixture model concerning random effects is the PD LS-1 study carried out by NINDS for the NINDS Exploratory Trials of Parkinson's Disease (NET-PD) program. The study is a multicenter, double-blind, placebo-controlled, randomized efficacy trial designed to determine the effectiveness of creatine in slowing the clinical decline of PD patients compared to placebo,. Among eligible participants who were treated for and diagnosed with PD within five years, 1,741 participants were enrolled and randomly assigned to creatine (874) or placebo (867) treatment groups. The

participants were evaluated through a questionnaire (at baseline and then annual follow-ups) for a minimum of five years until the end of the study in 2013 for futility based on the results of a planned interim analysis.[20,21] The level of perceived health, our outcome of interest, is measured through the standardized EuroQol vertical visual analog scale (EQ-VAS), a well-established instrument in PD studies,[21] which takes values between 0 (worst imaginable health) and 1 (best imaginable health). The response variable has an asymmetric distribution and a considerable number of observations (169) at the upper boundary of the support. In addition, the presence of two clusters of observations (i.e., a larger one with the trajectories lying near the upper bound of the support, and a smaller one taking values near the lower bound of the support) suggests the need for a flexible model to address constrained longitudinal data with augmentation, such as the AFB model, properly extended to handle within-subject correlation. The proposed model is compared with the beta and BR models applied in previous research.[17]

3 | DISTRIBUTIONS AND REGRESSION MODELS

The pdf of a beta-distributed random variable (rv) $Y \sim \text{Beta}(\bar{\alpha}, \phi)$ can be written as

$$f_{\text{B}}(y; \bar{\alpha}, \phi) = \frac{y^{\bar{\alpha}\phi-1}(1-y)^{(1-\bar{\alpha})\phi-1}}{B(\bar{\alpha}\phi, (1-\bar{\alpha})\phi)}, \quad 0 < y < 1, \tag{1}$$

where $B(\cdot,\cdot)$ is the beta function, $0 < \bar{\alpha} < 1$ and $\phi > 0$. This is known as mean-precision parameterization, since $\bar{\alpha}$ coincides with the mean $\mathbb{E}(Y)$, and the variance $\text{Var}(Y) = \frac{\bar{\alpha}(1-\bar{\alpha})}{\phi+1}$ is a decreasing function of the precision parameter ϕ for given $\bar{\alpha}$. This parameterization is preferred in a regression context because it makes it easier to link the mean (as well as the precision parameter) to the covariates.

A robust alternative to the beta distribution is to specify a mixture of a uniform and a beta distribution (i.e., the BR distribution[12]), which can accommodate outliers by making the tails of the resulting marginal distribution heavy. The pdf of the BR distributed rv $Y \sim \text{BR}(\bar{\alpha}, \phi, p)$ is equal to

$$f_{\text{BR}}(y; \bar{\alpha}, \phi, p) = p + (1-p) f_B(y; \bar{\alpha}, \phi), \quad 0 < y < 1, \tag{2}$$

with $0 \le p \le 1$, $0 < \bar{\alpha} < 1$, and $\phi > 0$. An alternative parameterization of the BR distribution has been devised for regression purposes[13] and depends on the mean of the response, $\gamma = \mathbb{E}(Y) = \frac{p}{2} + (1-p)\bar{\alpha}$, on a parameter $\theta = \frac{p}{1-(1-p)|2\bar{\alpha}-1|}$ that measures the thickness of the tails, and on a precision parameter $\phi > 0$.

The FB distribution[11] is a special mixture of two beta distributions that has been proposed as a way to handle bimodality, heavy tails, and potentially a large amount of asymmetry. The pdf of $Y \sim \text{FB}(\lambda_1, \lambda_2, \phi, p)$ is equal to

$$f_{\text{FB}}(y; \lambda_1, \lambda_2, \phi, p) = p f_{\text{B}}(y; \lambda_1, \phi) + (1-p) f_{\text{B}}(y; \lambda_2, \phi), \quad 0 < y < 1, \tag{3}$$

where $0 < \lambda_2 < \lambda_1 < 1$, $0 \le p \le 1$, and $\phi > 0$. The parameters λ_1 and λ_2 represent the means of the first and second mixture components, respectively, ϕ is the precision parameter, while p is the usual mixing proportion. A reparameterization of the FB distribution, which is useful for regression purposes, includes the overall mean of the mixture distribution as follows

$$\mu = \mathbb{E}(Y) = p\lambda_1 + (1-p)\lambda_2. \tag{4}$$

Thus, this new parameterization can be completed with $w = (\lambda_1 - \lambda_2)/\min\left\{\frac{\mu}{p}, \frac{1-\mu}{1-p}\right\}$, which is a standardized measure of the distance between the two mixture components, p and ϕ. The parametric space referred to as $f_{\text{FB}}(y; \mu, \phi, w, p)$ is variation independent since $0 < \mu < 1$, $0 < w < 1$, $0 \le p \le 1$, and $\phi > 0$.

We now briefly compare the beta, BR, and FB distributions. The former is a special case of the latter two distributions. The BR and FB distributions display increasing complexity: three and four parameters respectively instead of two. This complexity has immediate consequences for the possible shapes. As is well known, the beta distribution allows many diverse shapes. The BR distribution can have heavier tails for increasing values of θ.[17,18] The FB distribution can accommodate more shapes than the BR distribution, mainly in terms of bimodality, asymmetry, and tails. In particular, the FB distribution can have either heavy (and asymmetric) tails or only one heavy tail, unlike the BR distribution whose heavy tails are completely symmetric. Interestingly, despite being the most complex in terms of number of parameters, the FB distribution has good inferential properties such as strong identifiability and a.s. likelihood boundedness from above.[11] These properties facilitate the computation of estimates and, in a Bayesian context, solve the label switching problem. Note that the BR distribution, despite being less complex than the FB distribution, is not identifiable.

Let $Y_i \in (0,1)$ be the response for subject i ($i = 1, \dots, n$) and $\mathbf{x}_i^{'}$ be the corresponding covariate vector. Then, for all three distributions (i.e. the beta, BR, and FB distributions), a regression model can be defined by linking the mean and the covariates of interest within a generalized linear model (GLM) framework[25]

$$g(\nu_i) = \mathbf{x}_i^{'}\boldsymbol{\beta} \tag{5}$$

where $\nu_i = \mathbf{E}(Y_i)$ is the mean of the response, $g(\cdot)$ is an appropriate link function (typically the logit function $\text{logit}(\nu_i) = \log(\nu_i/(1-\nu_i))$ for interpretability), and $\boldsymbol{\beta}$ is a vector of regression coefficients. If $Y_i \sim \text{Beta}(\bar{\alpha}_i, \phi)$, then (5) is a beta regression model[7] with $\nu_i = \bar{\alpha}_i$; if $Y_i \sim \text{BR}(\gamma_i, \phi, \theta)$ then a BR regression model[13] is obtained with $\nu_i = \gamma_i$; finally, when $Y_i \sim \text{FB}(\mu_i, \phi, w, p)$ we obtain an FB regression model[11] where $\nu_i = \mu_i$. In each model the precision parameter ϕ can either be assumed constant or regressed onto the covariates after a log-transformation[11,13–15].

3.1 | Augmented FB distribution and mixed regression model

The above-described regression models fail to take into account responses lying in the closed interval $[0,\ 1]$ and responses that are measured longitudinally or are clustered. Both issues with the beta[9,15,16,26] and the BR models[4,17,18] have been addressed in the literature. The purpose of this paper is to propose a mixed-effects model based on a novel augmented FB distribution as a way to handle values at the boundaries and within-subject correlation.

The AFB distribution is a three-part mixture distribution that assigns positive probabilities to 0 and 1 and a (continuous) FB density to the interval $(0,\ 1)$. Therefore, its pdf is equal to

$$f_{\text{AFB}}(y; q_0, q_1, \mu, \phi, w, p) = \begin{cases} q_0 & \text{if} \quad y = 0 \\ q_1 & \text{if} \quad y = 1 \\ q_2\ f_{\text{FB}}(y; \mu, \phi, w, p) & \text{if} \quad 0 < y < 1 \end{cases} \tag{6}$$

where the vector (q_0, q_1, q_2) belongs to the simplex; thus, $0 < q_0 < 1$, $0 < q_1 < 1$, $0 < q_2 < 1$, and $q_0 + q_1 + q_2 = 1$. For ease of notation, $q_2 = 1 - q_0 - q_1$. A rv Y following an AFB distribution is denoted by $Y \sim \text{AFB}(q_0, q_1, \mu, \phi, w, p)$. Whenever there are neither 0s nor 1s observed in the response variable, it is possible to simplify (6) by setting one or both probabilities q_0 and q_1 to zero. The marginal mean of the AFB is $\mathbb{E}(Y) = q_1 + q_2\mu$ and its variance is $\text{Var}(Y) = q_2\text{Var}(Y|0 < Y < 1) + q_1 + q_2\mu^2 - \left[q_1 + q_2\mu\right]^2$, where $\mu = \mathbb{E}(Y|0 < Y < 1)$ is given by equation (4) and $\text{Var}(Y|0 < Y < 1) = \frac{\mu(1-\mu)+w^2\phi p(1-p)}{\phi+1}$ is the variance of an FB distributed rv.

Whenever responses show a hierarchical structure (i.e., are measured longitudinally or are clustered) one strategy is to include random effects in the regression model.[27] To define an AFB regression model with random effects, let $y_{ij} \in [0,\ 1]$ be the observed response from subject i ($i = 1, \dots, n$) at visit/cluster j ($j = 1, \dots, n_i$, where n_i is the number of visits or cluster size), and let $\mathbf{y}_i = (y_{i1}, \dots, y_{ij}, \dots, y_{in_i})^{'}$ be the observed response vector for subject i across visits/cluster i, and $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^{'}$ be the observed response vector for all the subjects. The covariates can be regressed onto suitable transformations of the conditional mean μ_{ij} and of the probabilities q_{0ij} and q_{1ij} as follows

$$g_1(\mu_{ij}) = \mathbf{x}_{1ij}^{'}\boldsymbol{\beta}_1 + u_{1j}, \quad g_2(q_{1ij}) = \mathbf{x}_{2ij}^{'}\boldsymbol{\beta}_2 + u_{2j}, \quad g_3(q_{0ij}) = \mathbf{x}_{3ij}^{'}\boldsymbol{\beta}_3 + u_{3j}, \tag{7}$$

where $\mathbf{x}_{1ij}, \mathbf{x}_{2ij}$, and $\mathbf{x}_{3ij}$ ($i = 1, \dots, n$ and $j = 1, \dots, n_i$) are vectors of possibly overlapping covariates, which include the covariates of interest as well as potential confounding variables. Moreover, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\beta_3}$ are the vectors for fixed effects associated with the covariates, and g_1, g_2, and g_3 are the (strictly monotone) link functions from $(0,1)$ to $\mathbb{R}$. A common choice for g_1 is the logit function because of its inherently simple interpretation of the regression coefficients in terms of odds ratios. However, for probabilities q_{1ij} and q_{0ij} we focus on a bivariate logit link: $g_2(q_{1ij}) = \log(q_{1ij}/(1-q_{0ij}-q_{1ij}))$ and $g_3(q_{0ij}) = \log(q_{0ij}/(1-q_{0ij}-q_{1ij}))$. Note that this choice still takes advantage of the ease of interpretation of the logit link: the parameters represent log odds ratios with respect to the baseline category. Moreover, it allows us to handle the constraint between the probabilities $0 < q_{0ij} + q_{1ij} < 1$, unlike the conventional univariate (separate) logits used in many augmented regression models.[4,15,17,18]

Random effects u_{hj} are typically assumed to be independent and normally distributed with a mean of zero and a variance equal to σ_h^2, $h = 1, 2, 3$. Clearly, the model can be easily generalized by assuming that the random effects are not independent and follow a multivariate normal distribution with a nondiagonal covariance matrix

$$\Sigma = diag(\sigma_1, \sigma_2, \sigma_3)\, R\, diag(\sigma_1, \sigma_2, \sigma_3) \tag{8}$$

where $diag(\sigma_1, \sigma_2, \sigma_3)$ is the diagonal matrix of the standard deviations (SDs) of the random effects, and R is the corresponding correlation matrix. Random slopes can also be easily included in the model.

In model (7) the dispersion parameter ϕ, the mixing proportion p, and the distance between the mixture components w are chosen as constants. However, they can be regressed onto the covariates as follows

$$g_4(\phi_{ij}) = \mathbf{x}_{4ij}^{'}\boldsymbol{\beta}_4 + u_{4j}, \quad g_5(w_{ij}) = \mathbf{x}_{5ij}^{'}\boldsymbol{\beta}_5 + u_{5j}, \quad g_6(p_{ij}) = \mathbf{x}_{6ij}^{'}\boldsymbol{\beta}_6 + u_{6j}, \tag{9}$$

where g_5 and g_6 are strictly monotone link functions from $(0, 1)$ to $\mathbb{R}$ (e.g., logit, probit, or complementary log-log), while g_4 is a strictly monotone link function from $\mathbb{R}^+$ to $\mathbb{R}$, the log function being an appropriate choice for this link function.

Model (7), eventually complemented with (9), can be reduced to accommodate various data features. For example, apart from the mean μ_{ij}, some or even all of the remaining parameters can be considered constant or can be free of random effects, which may be necessary both for easy interpretation and to avoid overparameterization. Moreover, if interest lies in directly evaluating the effect of covariates on the marginal mean of the response, an approximate marginal model can be derived.[28] This effect will be illustrated in the analysis of the RA dataset (see Section 6.1).

4 | BAYESIAN INFERENCE: MODEL IMPLEMENTATION AND COMPARISON CRITERIA

Let us assume that, conditional on the random effects, the responses Y_{ij} ($i = 1, \dots, n$ and $j = 1, \dots, n_i$) are independent and have an AFB pdf given by (6), with μ, q_0 and q_1 replaced by μ_{ij}, q_{0ij} and q_{1ij}, as specified by (7). Moreover, let $\boldsymbol{\eta} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \phi, w, p\}^{'}$ be the parameter vector and $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)^{'}$ the random effects vector. Then, the likelihood function, conditional on $\mathbf{u}$, is equal to

$$L(\boldsymbol{\eta}|\mathbf{u}, \mathbf{y}) = \prod_{i=1}^{n}\prod_{j=1}^{n_i} q_{0ij}^{\mathbb{I}(y_{ij}=0)} q_{1ij}^{\mathbb{I}(y_{ij}=1)} \left[(1 - q_{0ij} - q_{1ij}) f_{FB}(y_{ij}, |\mu_{ij}, \phi, w, p)\right]^{\mathbb{I}(0<y_{ij}<1)}, \tag{10}$$

where $\mathbb{I}(\cdot)$ is the indicator function,

$$q_{1ij} = \frac{\exp\left(\mathbf{x}_{2ij}^{'}\boldsymbol{\beta}_2 + u_{2j}\right)}{1 + \exp\left(\mathbf{x}_{2ij}^{'}\boldsymbol{\beta}_2 + u_{2j}\right) + \exp\left(\mathbf{x}_{3ij}^{'}\boldsymbol{\beta}_3 + u_{3j}\right)}, \quad q_{0ij} = \frac{\exp\left(\mathbf{x}_{3ij}^{'}\boldsymbol{\beta}_3 + u_{3j}\right)}{1 + \exp\left(\mathbf{x}_{2ij}^{'}\boldsymbol{\beta}_2 + u_{2j}\right) + \exp\left(\mathbf{x}_{3ij}^{'}\boldsymbol{\beta}_3 + u_{3j}\right)},$$

and

$$\mu_{ij} = \frac{\exp\left(\mathbf{x}_{1ij}^{'}\boldsymbol{\beta}_1 + u_{1j}\right)}{1 + \exp\left(\mathbf{x}_{1ij}^{'}\boldsymbol{\beta}_1 + u_{1j}\right)}.$$

In the current context, likelihood-based inference can be computationally intensive and analytically challenging since it involves both numerical integration and optimization. Indeed, to calculate the maximum likelihood estimates (MLEs) of $\boldsymbol{\eta}$ in the presence of random effects $\mathbf{u}_h$ for any of the link functions in (7) ($h = 1, 2, 3$), one has to integrate out $\mathbf{u}_h$, thus obtaining the marginal distribution of $\mathbf{y}$, and then maximize the likelihood based upon it. However, neither integration nor optimization allows a closed-form solution. Moreover, MLEs are affected by several well-known small-sample problems.

Therefore, we focus on Bayesian inference, in which unobservable random effects are drawn together with the unknown parameters in posterior sampling. Indeed, the AFB distribution is a finite mixture and can thus always be expressed as an incomplete data model where the allocation of each observation to one of the mixture components is unknown.[29] A Bayesian approach to inference based on Markov chain Monte Carlo (MCMC) methods, producing (simulated) posterior distributions for the parameter vector, is particularly suitable for handling incomplete data (and, generally, complex models with many parameters). A recently proposed solution is the Hamiltonian Monte Carlo (HMC)[30,31] algorithm, a generalization of the Metropolis-Hastings algorithm that combines MCMC and deterministic simulation methods. In addition to being more efficient than standard MCMC algorithms, the HMC is easy-to-use because of the easily accessible Stan modeling language[32,33], which requires specifying only the likelihood function (as in (10)) and prior distributions for the unknown parameters.

For the choice of priors, we adopt non- or weakly informative priors to have a minimal impact on the posteriors,[34] and we assume prior independence, which is reasonable because of the variation-independent parametric space of the FB distribution. In particular, the prior for regression parameters $\boldsymbol{\beta}_h$ ($h = 1, 2, 3$) is a diffuse multivariate normal prior with a zero mean vector and a diagonal covariance matrix with "large" values for the variances to induce flatness (i.e., non-informativeness). If the remaining parameters of the continuous part of the model (i.e. ϕ, w, and p) are defined as functions of the covariates (see model (9)), the

same choice holds for additional regression parameters; otherwise a uniform distribution on (0, 1) is selected for w and p, while a gamma(g, g) distribution is chosen for the precision parameter ϕ. Finally, to assign a noninformative prior to the matrix of variances of the random effects, Σ, the latter distribution is decomposed into a vector of SDs and a correlation matrix as in (8). The inverse variances have been given the same vague prior as that for ϕ, while the Lewandowski, Kurowicka, and Joe (LKJ) prior[35] with a shape parameter equal to 1 has been set for the correlation matrix, which is equivalent to giving a uniform prior over all the elements of the correlation matrix.[33]

We diagnose the convergence of the Markov chains to the equilibrium distribution through graphical investigations (trace plots and density plots) and several diagnostics that ascertain stationarity (Geweke and Heidel diagnostics) and mixing (potential scale reduction and effective sample size) of the chains. Initial values for the algorithm are set at random and, to diminish the influence of starting values, the first half of the chains are usually discarded. Moreover, we monitor the level of autocorrelation with the Raftery diagnostic[36] and we set a thinning interval L, where appropriate, such that only the first generated values in every batch of L iterations were kept. A numerical investigation concerning robustness, understood as having a limited impact on inferential conclusions, with respect to different choices of the hyperparameters of the multivariate normal and the gamma priors can be found in Section 1 of the Supporting Information (SI).

We now focus on the model comparison criteria, which typically penalize a model's goodness of fit score for complexity. The latter is a function of the number of parameters, which is less obvious in hierarchical (i.e., longitudinal) models. The widely applicable information criterion (WAIC)[33,37,38] is a fully Bayesian criterion and is also well-defined for non-regular models such as mixtures. It is based on the log pointwise predictive density (LPPD), which evaluates the model's goodness of fit and, having simulated S draws $\boldsymbol{\eta}^s$ $(s = 1, \dots, S)$ from the posterior distribution, can be estimated as follows:

$$\widehat{\text{LPPD}} = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log \left(\frac{1}{S} \sum_{s=1}^{S} f(y_{ij}|\boldsymbol{\eta}^s) \right),$$

where $f(y_{ij}|\boldsymbol{\eta}^s)$ is the density function for subject i at visit/cluster j. Since the LPPD of the observed data overestimates the corresponding value of future data, the quantity needs to be corrected by an estimate of the effective number of parameters[38] to avoid overfitting, i.e.

$$\hat{p}_{\text{WAIC}} = 2 \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \log \left(\frac{1}{S} \sum_{s=1}^{S} f(y_{ij}|\boldsymbol{\eta}^s) \right) - \frac{1}{S} \sum_{s=1}^{S} \log(f(y_{ij}|\boldsymbol{\eta}^s)) \right\}. \tag{11}$$

The difference between the $\widehat{\text{LPPD}}$ and the correction in (11) gives rise to the WAIC once it is multiplied by -2 to convert it to a deviance scale. As a general rule, the smaller the values of the comparison criteria are, the better the model.

In addition to identifying the best model, we also assess the adequacy of the fit and search for discrepancies between the observed and expected data under the regression model. Posterior predictive checks enable us to identify a bad fit by comparing the replicated data from the posterior predictive distribution with the observed data.[39] The graphical representation of the posterior predictive intervals for each marginal prediction $f(\tilde{y}_{ij}|\mathbf{y})$ is persuasive and easy to interpret. An alternative approach to marginal model checks is based on a cross-validation (leave-one-out) approach.[40] The conditional predictive ordinate (CPO) is a widely used diagnostic based on a cross-validation predictive distribution defined as $\text{CPO}_{ij} = f(y_{ij}|\mathbf{y}_{(ij)}) = \int f(y_{ij}|\boldsymbol{\eta}, \mathbf{y}_{(ij)}) \pi(\boldsymbol{\eta}|\mathbf{y}_{(ij)}) d\boldsymbol{\eta}$, where $\mathbf{y}_{(ij)}$ indicates the full data after deleting the observation from subject i at visit/cluster j. To avoid needing to refit the model $\sum_{i=1}^{n} n_i$ times, which would be very computationally intensive, an approximate cross-validation is provided under the assumption that $\pi(\boldsymbol{\eta}|\mathbf{y}_{(ij)}) \approx \pi(\boldsymbol{\eta}|\mathbf{y})$. Then, the estimate of CPO is given by $\widehat{\text{CPO}}_{ij} = S \left(\sum_{s=1}^{S} \frac{1}{f(y_{ij}|\boldsymbol{\eta}^s)} \right)^{-1}$ based only on the draws $\boldsymbol{\eta}^s$ $(s = 1, \dots, S)$ from the posterior distributions.[41] Finally, another well-established measure for detecting outlying observations is the leave-one-out Kullback-Leibler divergence among posteriors, $\text{KL}\left(\pi(\boldsymbol{\eta}|\mathbf{y}), \pi(\boldsymbol{\eta}|\mathbf{y}_{(ij)})\right)$. As a rule of thumb, low values of CPO estimates and high values of KL estimates suggest possible influential observations and outliers.

5 | SIMULATION STUDIES

Some simulation studies have been designed to evaluate the performance of the AFB regression model and to compare it to alternative models. The performance of both nonmixed and mixed augmented models has been evaluated in the presence of outliers. Below, we describe the rationale behind each study, the samples' simulation schemes, and the main inferential results. For each simulated dataset the vector of parameters in the AFB model, $\boldsymbol{\eta} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi, w, p\}'$ was estimated through the HMC algorithm as described in Section 4. In addition, the ABR and AB models have been estimated for comparison purposes.

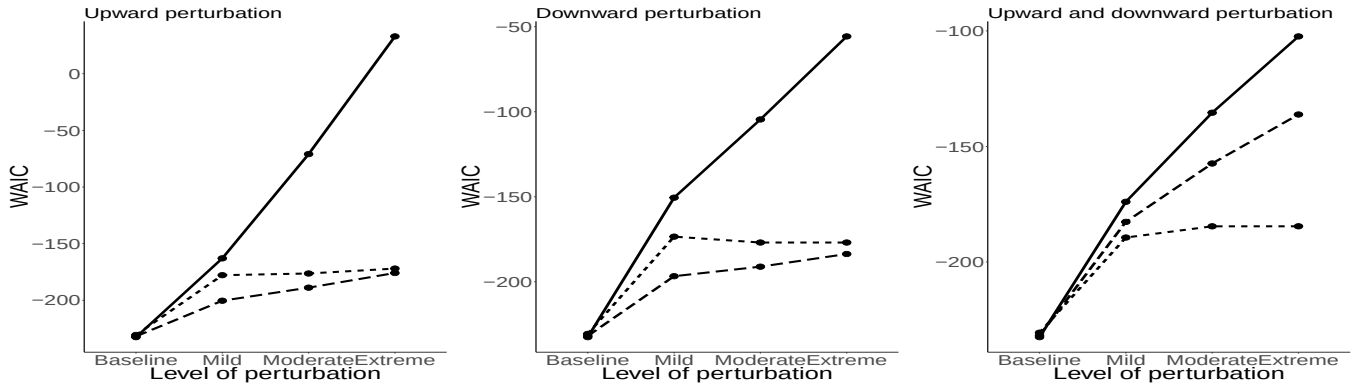

FIGURE 1 Average WAIC of the fitted regression models (the AFB model is the dashed line, the ABR model is the dotted line, and the AB model is the solid line) for upward perturbation (left-hand panel), downward perturbation (center panel) and half-upward/ half-downward perturbation (right-hand panel).

5.1 | Augmented nonmixed models in the presence of outliers

This simulation study evaluates the performance of the AFB model in the presence of outliers under many different scenarios. In total, $B = 100$ datasets, with $n = 200$ observations each, are randomly generated following the AB regression rationale outlined below. Given n i.i.d. observations from a quantitative covariate, x_i, uniformly distributed on $(-1.5, 1.5)$, the continuous component of the model is assumed to be beta-distributed, $Y_i \sim Beta(\bar{\alpha}_i, \phi)$, where the mean is a function of the simulated covariate such that $\text{logit}(\bar{\alpha}_i) = \beta_{1,0} + \beta_{1,1} x_i$ for $i = 1, \dots, n$. The regression parameters have been set to $\beta_{1,0} = 0.5$, $\beta_{1,1} = 1.5$, while $\phi = 50$. In addition, augmentation in both 1 only and in 0 and 1 is taken into account regarding the discrete part of the model. Specifically, for augmentation in 0 and 1, parameters q_{1i} and q_{0i} are regressed onto the quantitative covariate x_i for $i = 1, \dots, n$ as follows

$$\log\left(\frac{q_{1i}}{(1 - q_{0i} - q_{1i})}\right) = \beta_{2,0} + \beta_{2,1} x_i, \quad \log\left(\frac{q_{0i}}{(1 - q_{0i} - q_{1i})}\right) = \beta_{3,0} + \beta_{3,1} x_i, \tag{12}$$

where $\beta_{2,0} = -3$, $\beta_{2,1} = 2$, $\beta_{3,0} = -3$, and $\beta_{3,1} = -2$. For augmentation in only 1, the discrete part of the model is simplified by setting $\text{logit}(q_{1i}) = \beta_{2,0} + \beta_{2,1} x_i$ and $q_{0i} = 0$ for $i = 1, \dots, n$.

For each replication, the vector $(y_1, \dots, y_n)'$ simulated from the AB model constitutes the baseline scenario, which is further contaminated by inducing $r = 6$ outliers following three patterns of perturbation: upward, downward, and half-upward/half-downward perturbations. The outliers are generated by adding (or subtracting) a value Δ to (from) their baseline values. Ultimately, for each of the three scenarios, four levels of perturbation are set up (baseline, mild, moderate, and extreme perturbation) with increasing (decreasing) values of Δ. Only the results for the case of augmentation in 0 and 1 under the three perturbation scenarios are reported given that similar conclusions can be inferred for the case of augmentation in 1 only. A detailed description of all the simulations is presented in Section 2.2 of the SI.

For the case of only upward or only downward perturbation, the analysis of the WAIC values (see the left-hand and center panels of Figure 1) reveals an average fit of the AFB that is better than that of the ABR and AB models, regardless of the intensity of the perturbation factor Δ. For perturbations in both directions, the fit of the AFB model, despite being much better than that of the AB model, is comparable to the fit of the ABR model under mild perturbation but worse under moderate or extreme perturbation (see the right-hand panel of Figure 1). This result occurs because the AFB model fits only one of the two clusters of outliers. In the baseline scenario (i.e., no perturbation), the AFB model performs better than the ABR model in almost all replications, but it does not outperform the AB model. This finding is a foregone conclusion, as the AB model is the model from which the data have been simulated. In contrast, when perturbation is present, the AFB model is better than the AB model in almost all replications regardless the type of perturbation, while it clearly outperforms the ABR model only if perturbation occurs in only one direction (either upward or downward).

Finally, KL divergences are computed to assess the ability of each model to identify outliers in the case of extreme perturbations and the results are reported in Section 2.1 of the SI along with some useful remarks. Overall, we can say that the AB model clearly identifies the outliers, but it has low goodness-of-fit criteria values. The ABR model experiences some difficulties in detecting outlier observations, but it improves the criteria with respect to the AB model, thus showing greater robustness

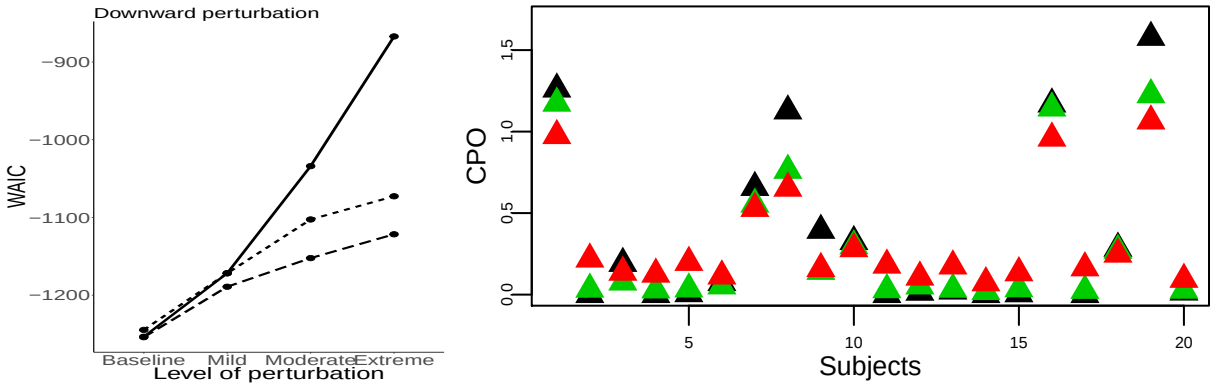

FIGURE 2 Left-hand panel: WAIC of the fitted regression models (the AFB model is the dashed line, the ABR model is the dotted line, and the AB model is the solid line). Right-hand panel: CPO estimates of the perturbed observations in case of extreme perturbation for the AFB (red triangles), ABR (green triangles), and AB (black triangles) models.

than the AB model. The AFB model succeeds in both identifying outliers and predicting future observations, and it fits the data better than the other models. This finding is true especially when outliers exist in a single direction.

5.2 | Augmented mixed-effects models (for longitudinal data) in the presence of outliers

This second simulation framework evaluates the effect of perturbation in the case of longitudinal data. Special attention is paid to analyzing the robustness of the posterior estimates. To this end, $B = 100$ datasets are randomly generated from an AB regression model for longitudinal data with augmentation in 1. Each dataset consists of $n = 200$ subjects observed at four visits. Given a Bernoulli distributed covariate, x_i $(i = 1, \dots, 200)$, assuming a value of 1 with probability 0.6, and given the time-dependent covariate, t_j $(j = 1, \dots, 4)$, the response variable $0 < y_{ij} \leq 1$ is simulated from an AB distribution according to the following regression scheme which includes a regression equation for the precision ϕ:

$$\text{logit}(\bar{\alpha}_{ij}) = \beta_{1,0}+\beta_{1,1}x_i+\beta_{1,2}t_j+\beta_{1,3}x_it_j+u_{1i}, \quad \text{logit}(q_{1ij}) = \beta_{2,0}+\beta_{2,1}x_i+u_{2i}, \quad \log(\phi_{ij}) = \beta_{4,0}+\beta_{4,1}x_i+\beta_{4,2}t_j+\beta_{4,3}x_it_j. \quad (13)$$

Here $u_i = (u_{1i}, u_{2i})'$ is the random effect vector following a bivariate normal distribution $N\left(\mathbf{0}, \Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ & \sigma_2^2 \end{bmatrix}\right)$. The parameters are set as follows: $\boldsymbol{\beta}_1 = (-0.1, 1.5, -0.3, 0.1)'$, $\boldsymbol{\beta}_2 = (-2, 0.5)'$, $\boldsymbol{\beta}_4 = (2.3, 0.5, 0.8, -0.1)'$, $\sigma_1 = 0.3$, $\sigma_2 = 0.6$, $\rho_{12} = 0.65$. Furthermore, 10% of the observations are randomly selected from each baseline visit from the replicated dataset; these observations are decreased by subtracting Δ to generate outliers. In addition to the baseline scenario, three scenarios with different perturbation intensities are established making Δ equal to 0.1 (mild), 0.2 (moderate), or 0.3 (extreme). The perturbation effect emphasizes bimodality, inducing outliers in both covariate categories.

The average WAIC values reported in Figure 2 (the left-hand panel) show that the AFB model outperforms the competing models under all the perturbation patterns. At baseline, although this scenario obviously favors the AB model, as the data are simulated from it, our new model outperforms the AB model in 43% of the replications and outperforms the ABR model in 90% of the replications. In the remaining scenarios, the AFB model has a better fit than the AB models in all replications and a better fit than the ABR model in 95 to 99 % of replications. For the extreme perturbation scenario, we analyze the CPO estimates relative to the perturbed observations, recalling that the lower the CPO is, the lower the likelihood of observing the response given the model. The right-hand panel of Figure 2 reveals that most of the strongly influential outliers (i.e., those with CPO estimates less than 0.5) are more likely in the AFB model, which devotes its second mixture component to them, than in the other models. To identify outliers, the AB model shifts its mean downward, thus worsening the overall fit but also better describing the less-influential perturbed values, namely those with CPO estimates greater than 0.5. Conversely, the ABR model, due to its well-known robustness, is less affected by outliers than the AB (the green triangles are always above the black ones when the CPO estimates are less than 0.5), but it barely fits the data (the green triangles are always below the red ones, i.e. the AFB model estimates, when the CPO estimates are less than 0.5) because of the lack of flexibility induced by the uniform mixture component.

Finally, Table 1 reports the posterior estimates and the coverage (i.e., the percentage of times the 95% posterior credible interval contains the true value of the parameter) for each model and each level of perturbation. First, it is worth noting that neither the regression parameters for the discrete part of the model, $\boldsymbol{\beta}_2$, nor the covariance parameters of the random effects are affected by the perturbation or the chosen model, thus leading to similar estimates and a stable coverage percentage for increasing levels of perturbation for all models. In contrast, the perturbation affects the regression parameter vectors $\boldsymbol{\beta}_1$ (the model for the mean) and $\boldsymbol{\beta}_4$ (the model for the precision). Concerning $\boldsymbol{\beta}_1$, the estimates of the intercept $\beta_{1,0}$ and the interaction coefficient $\beta_{1,3}$ decrease as the perturbation becomes more extreme, and this trend is compensated by an increase in the estimate of the single covariate coefficients $\beta_{1,1}$ and $\beta_{1,2}$ in all the models. Nevertheless, the AB model shows low coverages, especially for moderate to extreme perturbations, thus confirming its inability to handle outliers. Regarding the AFB and ABR models, the effect of perturbation manifests in different ways. While the estimate of the intercept $\beta_{1,0}$ is slightly less affected by extreme perturbation under the ABR model, the AFB model provides more robust estimates for all three regression coefficients $\beta_{1,1}$, $\beta_{1,2}$, and $\beta_{1,3}$, with higher coverage for moderate to extreme perturbation. The same applies to $\boldsymbol{\beta}_4$, for which the AFB model displays the best coverage for moderate to extreme perturbation, while the AB shows the worst coverage (i.e., the coverage collapses toward 0).

Pert.		$\beta_{1,0}$=-0.1	$\beta_{1,1}$=1.5	$\beta_{1,2}$=-0.3	$\beta_{1,3}$=0.1	$\beta_{2,0}$=-2	$\beta_{2,1}$=0.5	$\beta_{4,0}$=2.3	$\beta_{4,1}$=0.5	$\beta_{4,2}$=0.8	$\beta_{4,3}$=-0.1	σ_1=0.3	σ_2=0.6	ρ_{12}=0.65
Base	AFB	-0.09 (0.94)	1.50 (0.96)	-0.30 (0.95)	0.10 (0.98)	-2.05 (0.94)	0.52 (0.94)	2.24 (0.95)	0.49 (0.96)	0.86 (0.95)	-0.08 (0.97)	0.30 (0.94)	0.59 (0.93)	0.65 (0.97)
	ABR	-0.09 (0.98)	1.49 (0.94)	-0.30 (0.99)	0.10 (0.94)	-2.05 (0.98)	0.51 (0.98)	2.28 (0.96)	0.49 (0.99)	0.80 (0.98)	-0.09 (0.99)	0.30 (0.94)	0.61 (0.91)	0.66 (0.96)
	AB	-0.09 (0.93)	1.50 (0.94)	-0.30 (0.95)	0.10 (0.96)	-2.05 (0.96)	0.52 (0.95)	2.28 (0.93)	0.50 (0.95)	0.80 (0.93)	-0.10 (0.96)	0.30 (0.95)	0.59 (0.95)	0.65 (0.98)
Mild	AFB	-0.11(0.90)	1.49 (0.91)	-0.30 (0.90)	0.08 (0.86)	-1.97 (0.94)	0.54 (0.94)	2.35 (0.93)	0.43 (0.94)	0.89 (0.91)	-0.16 (0.92)	0.31 (0.92)	0.59 (0.93)	0.64 (0.96)
	ABR	-0.13 (0.91)	1.54 (0.91)	-0.29 (0.88)	0.08 (0.88)	-2.05 (0.96)	0.50 (0.96)	2.39 (0.96)	0.57 (0.95)	0.80 (0.96)	-0.21 (0.93)	0.31 (0.95)	0.60 (0.95)	0.66 (0.98)
	AB	-0.14 (0.90)	1.55 (0.87)	-0.29 (0.90)	0.08 (0.84)	-2.05 (0.94)	0.52 (0.95)	2.37 (0.92)	0.60 (0.95)	0.80 (0.94)	-0.26 (0.85)	0.31 (0.87)	0.59 (0.94)	0.65 (0.97)
Mod	AFB	-0.18 (0.68)	1.54 (0.87)	-0.28 (0.76)	0.08 (0.86)	-2.05 (0.94)	0.52 (0.96)	2.29 (0.92)	0.56 (0.97)	0.82 (0.91)	-0.14 (0.90)	0.29 (0.83)	0.59 (0.94)	0.64 (0.97)
	ABR	-0.17 (0.73)	1.54 (0.87)	-0.27 (0.62)	0.08 (0.82)	-2.05 (0.91)	0.55 (0.95)	2.35 (0.89)	0.50 (0.93)	0.78 (0.85)	-0.08 (0.91)	0.28 (0.82)	0.56 (0.95)	0.65 (0.96)
	AB	-0.18 (0.71)	1.60 (0.70)	-0.28 (0.76)	0.06 (0.63)	-2.05 (0.93)	0.52 (0.96)	2.22 (0.91)	0.99 (0.68)	0.85 (0.89)	-0.54 (0.14)	0.32 (0.83)	0.60 (0.94)	0.64 (0.97)
extr.	AFB	-0.22 (0.52)	1.56 (0.84)	-0.27 (0.69)	0.08 (0.82)	-2.05 (0.96)	0.52 (0.96)	1.97 (0.72)	0.86 (0.78)	0.90 (0.83)	-0.21 (0.89)	0.28 (0.82)	0.58 (0.95)	0.64 (0.97)
	ABR	-0.20 (0.58)	1.57 (0.81)	-0.26 (0.38)	0.07 (0.65)	-2.02 (0.92)	0.49 (0.96)	1.97 (0.71)	0.85 (0.73)	0.90 (0.83)	-0.19 (0.85)	0.28 (0.83)	0.58 (0.92)	0.63 (0.96)
	AB	-0.22 (0.51)	1.65 (0.61)	-0.27 (0.56)	0.04 (0.41)	-2.05 (0.94)	0.52 (0.93)	1.83 (0.56)	1.56 (0.21)	0.95 (0.75)	-0.87 (0.00)	0.32 (0.81)	0.59 (0.90)	0.64 (0.97)

TABLE 1 Posterior means and coverages (in parenthesis) of the AFB, ABR, and AB regression models.

6 | MOTIVATING EXAMPLES: DATA ANALYSES AND FINDINGS

6.1 | Reading accuracy data

The sample consists of a group of 44 children, 19 of whom have been diagnosed with dyslexia.[19] The response variable, y, quantifies the RA performance as a proportion in $(0, 1]$ while a quantitative covariate, x, measures the standardized nonverbal IQ of the children. In the group of non-dyslexic readers, 13 children achieved a perfect score on the RA test ($y = 1$). Research[11] has shown the superiority of the FB regression model without augmentation with respect to the beta and BR models when the response variable is transformed from $(0,\ 1]$ to $(0,\ 1)$.

Our initial analysis aims to evaluate the clustering ability of the AFB regression model in presence of latent groups, namely by assuming ignorance about whether a child belongs to the group of dyslexic readers. We fix augmented regression models for the conditional mean, $\text{logit}(\nu_i) = \beta_{1,0} + \beta_{1,1}x_i$, and the probability of values at the upper boundary, $\text{logit}(q_{1i}) = \beta_{2,0} + \beta_{2,1}x_i$. For comparison purposes, we estimate the AFB regression model ($\nu_i = \mu_i$) as well as the ABR ($\nu_i = \gamma_i$) and AB ($\nu_i = \bar{\alpha}_i$) regression models. For each model, we simulate MCMCs of length 10,000 and we assess the convergence to the equilibrium distribution using the analytical and graphical tools described in Section 4. It is worth noting that the estimates of the parameters belonging to the discrete part of the model are nearly identical and highly significant across the three competing models (see Table 2). In particular, the probability of achieving a perfect score ($y = 1$) given $x = 0$ (i.e., the mean of the standardized nonverbal IQ covariate) is equal to $\exp(\beta_{2,0})/(1 + \exp(\beta_{2,0})) \approx 0.23$, meaning that a perfect score in reading is $\exp(\beta_{2,0}) \approx 0.3$ times more likely to occur than a nonperfect score (i.e., $0 < y < 1$) conditional on $x = 0$. The standardized nonverbal IQ covariate has a significant augmenting effect for positive values (i.e., above-average IQ values) and decreasing for negative values (i.e., below-average IQ values). Specifically, we may observe that the estimated odds ratio is approximately $\exp(\beta_{2,1}) \approx 3.97$ for all models, meaning that an increase of 1 in the x score increases the odds that the RA score is 1 by 3.97. Instead, there are differences among the estimates concerning the continuous part of the models. First, it should be noted that the slope parameter $\beta_{1,1}$ is statistically significant in only the AB model and results in an increase in the odds of the target outcome (i.e., the expected reading score ν_i) of $\exp(\beta_{1,1}) \approx 1.5$ times for each unit increase in the x score. Conversely, the CIs for $\beta_{1,1}$ in the AFB and ABR models cross 0, implying a nonsignificant effect on the odds $\nu_i/(1 - \nu_i)$ due to a unit increase in the covariate. The intercept of the regression function for ν_i is significant in all models but different inferential conclusions can be drawn. Given $x = 0$, the target outcome ν_i

Parameters	AB			ABR			AFB		
	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
$\beta_{1,0}$ (intercept)	0.939	0.128	(0.685;1.190)	0.783	0.226	(0.328;1.138)	0.839	0.125	(0.614; 1.107)
$\beta_{1,1}$ (IQ)	0.411	0.129	(0.157;0.663)	0.302	0.197	(-0.078;0.635)	0.131	0.152	(-0.092;0.522)
$\beta_{2,0}$ (intercept)	-1.219	0.427	(-2.107;-0.429)	-1.230	0.422	(-2.101;-0.459)	-1.232	0.433	(-2.140;-0.447)
$\beta_{2,1}$ (IQ)	1.381	0.479	(0.517;2.409)	1.378	0.477	(0.517;2.389)	1.392	0.484	(0.521;2.446)
ϕ	10.783	2.662	(6.303;16.626)	17.153	13.961	(6.389;57.914)	28.471	11.737	(8.699;52.163)
p							0.216	0.173	(0.051;0.890)
w							0.656	0.200	(0.069;0.862)
θ				0.143	0.127	(0.004;0.450)			
WAIC	12.8			21.0			8.8		

TABLE 2 RA data: posterior means, SDs, and CIs for the parameters of the AFB, ABR, and AB regression models together with the WAIC values (models with one covariate).

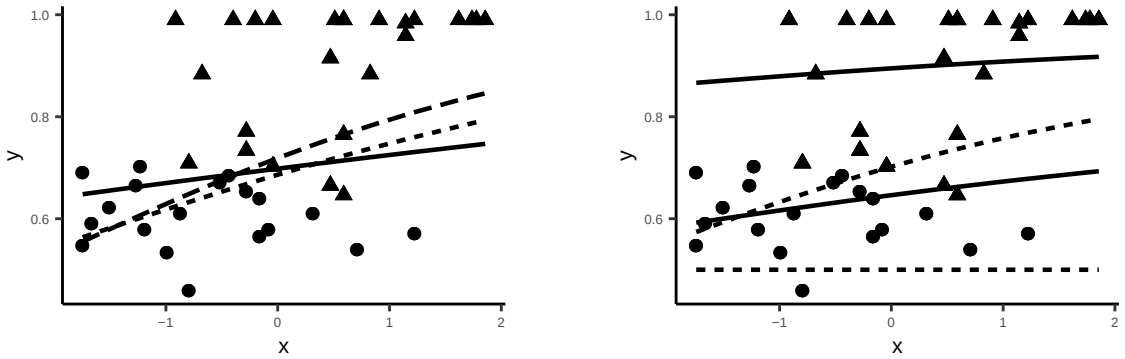

FIGURE 3 Left-hand panel: regression curves for the AFB (solid line), ABR (dotted line), and AB (dashed line) models. Right-hand panel: regression curves for the two component means of the AFB model (λ_{1i} and λ_{2i} in the solid lines) and the ABR model (1/2 and $\bar{\alpha}_i$ in the dotted lines).

is $\exp(\beta_{1,0}) \approx 2.6$ times more likely to occur than the nontarget outcome (i.e., the difference from a perfect reading score $1 - \nu_i$) in the AB model and approximately 2.2 and 2.3 more likely in the ABR and AFB models, respectively. Figure 3 shows that the regression curve for the AFB model is the least-inclined, suggesting a lower impact of the covariate on the mean of the response variable. The mixture structure of the AFB model enables it to correctly identify the presence of two clusters as highlighted by the right-hand panel of Figure 3. Despite also being a mixture model, the ABR model does not provide the same (good) evidence due to its uniform component, the mean of which is constant and equal to 1/2, a value far from the central body of observations. The AFB model, which has the lowest WAIC value, is also the preferable choice in terms of fit.

A second analysis of the RA dataset has the very different aim of investigating the fit of the AFB model by assuming a functional relationship between intelligence (quantified through the standardized IQ), dyslexic status (a dichotomous covariate z reflecting being dyslexic (0) or not (1)) and RA. For this purpose, we define the following regression model that includes regression functions for the conditional mean, the probability of 1s, and the precision parameter:

$$\text{logit}(\nu_i) = \beta_{1,0} + \beta_{1,1} z_i, \quad \text{logit}(q_{1i}) = \beta_{2,0} + \beta_{2,1} x_i + \beta_{2,2} z_i, \quad \log(\phi_i) = \beta_{4,0} + \beta_{4,1} x_i + \beta_{4,2} z_i + \beta_{4,3} x_i z_i. \tag{14}$$

For comparison purposes, we estimate the AFB ($\nu_i = \mu_i$), ABR ($\nu_i = \gamma_i$) and AB ($\nu_i = \bar{\alpha}_i$) regression models. The effects of the quantitative covariate x and of the interaction are removed from the regression model for the conditional mean and the probability of 1s, respectively, due to lack of statistical significance in all competing models. It is noteworthy that, by including the covariate z in the regression function for q_{1i}, a data separation problem arises given that all children with perfect RA score belong to the group of nondyslexic readers. In a Bayesian framework, one possible solution to data separation is adopting weakly informative prior distributions for the regression parameters. We choose a Cauchy prior distribution with a center of 0 and a scale of 2.5[42] for the regression parameter $\boldsymbol{\beta}_2$. The fit of all models improves greatly, and the AFB model remains the preferable choice, providing the lowest WAIC value (see Table 3). Focusing on the estimated parameters for the regression function for ν_i, it emerges that $\exp(\beta_{1,1})$, which can be interpreted as the difference in the odds of the target outcome ν_i for the group of nondyslexic readers compared to the group of dyslexic readers, is approximately equal to 2.2 in the AFB model, while it is

Parameters	AB			ABR			AFB		
	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
$\beta_{1,0}$ (intercept)	0.405	0.066	(0.274;0.531)	0.382	0.067	(0.253;0.515)	0.433	0.085	(0.290;0.650)
$\beta_{1,1}$ (dyslexic)	0.872	0.215	(0.446;1.320)	0.754	0.241	(0.339;1.214)	0.794	0.240	(0.378;1.290)
$\beta_{2,0}$ (intercept)	-4.399	3.537	(-13.735;-1.330)	-4.241	2.611	(-11.270;-1.477)	-4.994	4.929	(-22.705;-1.467)
$\beta_{2,1}$ (IQ)	0.832	0.533	(-0.152;1.926)	0.844	0.535	(-0.189;1.912)	0.807	0.524	(-0.134;1.835)
$\beta_{2,2}$ (dyslexic)	3.990	3.605	(0.735;13.682)	3.846	2.672	(0.883;10.960)	4.623	4.967	(0.676;22.309)
$\beta_{4,0}$ (intercept)	4.299	0.495	(3.244;5.183)	4.311	0.496	(3.282;5.191)	4.425	0.581	(3.239;5.592)
$\beta_{4,1}$ (IQ)	0.436	0.535	(-0.569;1.550)	0.409	0.534	(-0.596;1.474)	0.396	0.607	(-0.819;1.499)
$\beta_{4,2}$ (dyslexic)	-1.753	0.680	(-3.052;-0.312)	-1.374	0.928	(-2.974;0.600)	-1.437	0.993	(-3.113;0.742)
$\beta_{4,3}$ (dyslexic×IQ)	-1.570	0.768	(-3.024;-0.046)	-2.102	1.166	(-4.802;-0.251)	-2.011	1.080	(-4.376;0.002)
p							0.369	0.318	(0.011;0.976)
w							0.267	0.224	(0.016;0.746)
θ				0.080	0.071	(0.003;0.271)			
WAIC	-25.6			-19.5			-25.7		

TABLE 3 RA data: posterior means, SDs and CIs for the parameters of the AFB, ABR, and AB regression models together with the WAIC values (models with two covariates).

equal to 2.1 in the ABR model and 2.4 in the AB model. It is noteworthy that the estimates of the second and third equations in (14) (the models for q_1 and ϕ, respectively) give different values for the parameters of the x covariate under the three models (i.e., a lower impact under the AFB model with a sort of offset in the intercept estimate). From an interpretative viewpoint, the probability of achieving a perfect score ($y = 1$) is governed by belonging to either the dyslexic or nondyslexic group: the odds of a perfect score is more than 50 times greater for the control group than for the dyslexic group, given $x = 0$). Moreover, focusing on the regression model for the precision parameter ϕ, the interaction between the quantitative covariate x and the dichotomous covariate z appears to be significant in all the models. In particular, given $x = 0$, the logarithm of the precision is approximately equal to 4.3 for the dyslexic group in all the models and decreases down to 3 for the group of nondyslexic readers in the AFB and ABR models and down to 2.546 for the AB model. All models estimate a negative linear effect of the covariate x on the logarithm of the precision for the group of nondyslexic readers and a positive linear effect for the dyslexic group. These results indicate behavioral differences between the dyslexic and nondyslexic groups in terms of the location and scale effects. Interestingly, the same regression model for ν_i and ϕ_i but without augmentation can lead to misleading results. The results concerning the non-augmented models are provided in Section 3 of the SI.

Finally, note that if interest lies in modeling the marginal mean of the response, an approximate marginal model[28] can be estimated. In particular, combining the marginal mean $\mathbb{E}(Y) = q_1 + (1 - q_1)\mu$ with model (14), the marginal regression function $\text{logit}(\mathbb{E}(Y_i)) = \omega_0 + \omega_1 x_i + \omega_2 z_i$ is obtained. Although its parameters $(\omega_0, \omega_1, \omega_2)$ are nonlinear functions of the estimated parameters of (14), a first order Taylor expansion leads to the following approximate posterior estimates $(0.670, 0.276, 1.355)$. It is worth noting that $exp(\omega_2)$, which is the difference in the odds of the marginal mean for the group of nondyslexic readers compared to the dyslexic group, is approximately equal to 3.88 (for further discussion on this see Section 3 of the SI).

6.2 | Parkinson's Disease data

The sample consists of $n = 1,741$ subjects measured at up to $n_i = 7$ visits. The response variable $0 < y_{ij} \leq 1$ quantifies the level of perceived health, where a value of 1 indicates the best imaginable health status. Missing values are caused by drop-outs and missed visits; these missing values are treated as missing at random for the purpose of this study. The regression functions for the mean and the probability of augmentation in 1 are defined as follows

$$\text{logit}(\nu_{ij}) = \beta_{1,0} + \beta_{1,1}x_{1ij} + \beta_{1,2}x_{2ij} + \beta_{1,3}x_{1ij}x_{2ij} + u_{1i}, \quad \text{logit}(q_{1ij}) = \beta_{2,0} + \beta_{2,1}x_{1ij} + u_{2i}, \tag{15}$$

where x_{1ij} indicates the randomly assigned treatment, which is equal to 1 if the treatment is creatine and 0 if the treatment is placebo, and x_{2ij} refers to the time of the visit. The mean of the continuous part of the model is a function of the treatment, the time of the visit, and their interaction and the probability q_1 is a function of only the treatment. Given the longitudinal structure of the data, random intercepts are added. In particular, $\mathbf{u_i} = (u_{1i}, u_{2i})'$ follows a bivariate normal distribution $N\left(\mathbf{0}, \Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ & \sigma_2^2 \end{bmatrix}\right)$. To compare the AFB ($\nu_{ij} = \mu_{ij}$) regression model with the ABR ($\nu_{ij} = \gamma_{ij}$) and AB ($\nu_{ij} = \bar{\alpha}_{ij}$) models, we simulate MCMCs following the Bayesian approach to inference described in Section 4, and we check their convergence to the equilibrium distribution.

We first note that the fitting ability of the AFB model (WAIC $= -18,325$) is superior to that of the competing models (the WAIC is equal to $-17,397$ for the AB model and $-17,369$ for the ABR model). Table 4 reports the posterior means, SDs and 95% CIs for the estimated parameters. Although the estimates differ somewhat across the three models, the same parameters

Parameters	AB			ABR			AFB		
	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
$\beta_{1,0}$ (intercept)	1.6038	0.0298	(1.5458; 1.6625)	1.6224	0.0287	(1.5673; 1.6417)	1.6296	0.0282	(1.5748; 1.6847)
$\beta_{1,1}$ (treatment)	-0.0395	0.0417	(-0.1218;0.0420)	-0.0501	0.0394	(-0.1271; 0.0275)	-0.0425	0.0384	(-0.1191; 0.0322)
$\beta_{1,2}$ (time)	-0.0741	0.0058	(-0.0856;-0.0629)	-0.0752	0.0050	(-0.0849;-0.0656)	-0.0775	0.0051	(-0.0874; -0.0676)
$\beta_{1,3}$ (treatment×time)	-0.0010	0.0082	(-0.0171; 0.0151)	0.0008	0.0071	(-0.0131; 0.0146)	0.0019	0.0073	(-0.0125; 0.0160)
$\beta_{2,0}$ (intercept)	-6.5400	0.3520	(-7.2796; -5.8928)	-6.5011	0.3411	(-7.2111; -5.8860)	-6.5090	0.3438	(-7.2272; -5.8779)
$\beta_{2,1}$ (treatment)	-0.4135	0.2637	(-0.9338; 0.0973)	-0.4174	0.2603	(-0.9366; 0.0895)	-0.4178	0.2581	(-0.9236; 0.0941)
ϕ	14.3422	0.2503	(13.8575; 14.8358)	25.0421	0.6287	(23.8384; 26.2901)	19.4845	0.3662	(18.7769; 20.2123)
σ_1	0.5981	0.0133	(0.5718; 0.6241)	0.6172	0.0134	(0.5915; 0.6443)	0.5916	0.0131	(0.5661; 0.6174)
σ_2	2.4836	0.2119	(2.1028;2.9339)	2.4310	0.2054	(2.0580; 2.8634)	2.4518	0.2092	(2.0703; 2.8838)
ρ_{12}	0.8024	0.0556	(0.6916;0.9083)	0.7480	0.0538	(0.6401; 0.8498)	0.7672	0.0539	(0.6556; 0.8686)
θ				0.1194	0.0095	(0.1012; 0.1383)			
p							0.9898	0.0013	(0.9869; 0.9923)
w							0.7546	0.0181	(0.7188; 0.7904)

TABLE 4 PD data: posterior means, SDs and CIs for the parameters of the AFB, ABR, and AB regression models.

are identified as being significant. In particular, all parameters referring to creatine ($\beta_{1,1}$ and $\beta_{1,3}$ in the mean model and $\beta_{2,1}$ for the probability of perfect health) show a CI crossing 0, implying a nonsignificant effect. This finding reflects the fact that the LS-1 study was terminated due to the futility of creatine as a disease-modifying agent in PD. The estimates of the parameters relative to the probability of a response equal to 1 (perfect health) are very similar across the three models and suggest a lower probability for creatine treatment than for the placebo ($\beta_{2,1}$ is negative). The intensity of the effect, which is conditional on other covariates and random effects, is $\exp\left(\beta_{2,1}\right) \approx 0.66$ for all the models. This can be interpreted as an odds ratio, namely the odds of $y_{ij} = 1$ (perfect health) in the creatine group over the odds of perfect health in the placebo group. With respect to the regression parameters for the mean, conditional on time and on the random effects, the odds ratio between the patients receiving creatine or a placebo is equal to $\exp(\beta_{1,1}) \approx 0.96$ in the AFB model. Moreover, conditional on the treatment and on the random effects, for a one-unit year increase in time, the expected decrease in the odds (i.e., the ratio between the expected EQ-VAS μ_{ij} and the difference from perfect health $1 - \mu_{ij}$) is equal to $(1 - \exp(\beta_{1,2})) * 100 = 7,5\%$ for the patients in the placebo group and $(1 - \exp(\beta_{1,2} + \beta_{1,3})) * 100 = 7,3\%$ for the patients in the creatine group.

We can better understand the added value of the AFB model by analyzing the remaining parameter estimates, namely, the mixing proportion p and the normalized distance between the mixture components w. The estimate of p is equal to 0.99, suggesting the presence of two clusters: one cluster including most (99%) of the observations and another cluster containing the other 1% (over 150) of the observations. Moreover, as the estimate of the normalized distance between the two clusters is large (0.75), it follows that the observations in the second cluster can be identified as potential outliers. Interestingly, all the models estimate a positive correlation among the random effects, which implies that the higher the observed individual EQ-VAS score, the higher the probability that the subject will have perfect health. Further insight into the behavior of the three competing models is provided by the graphical representation of the regression curves in Figure 4. Recall that the AB model identifies a single curve, whereas the two mixture models identify both the general curve relative to the overall mean, and two curves concerning each mixture component (λ_{1ij} and λ_{2ij} for the AFB model and $\bar{\alpha}_{ij}$ and 1/2 for the ABR model). To ensure a thorough analysis, two separate graphs are plotted for the creatine and placebo groups, even though they show no relevant differences. First, it is worth noting that the regression curve of the AB model is lowered because of the presence of a small group of outliers near 0. The curves associated with the cluster mean λ_{1ij} of the AFB model and $\bar{\alpha}_{ij}$ of the ABR model almost overlap regardless of the absence (upper panel) or presence (lower panel) of treatment. The evident difference between these two competing models concerns the cluster means of the other component. The second cluster mean λ_{2ij} of the AFB model adapts suitably to the group of observations far from the central body, while the other cluster mean of the ABR model is forced to be identically equal to 1/2, resulting in a poor fit to the observed data. The analyses of the posterior predictive distributions and the posterior predictive bounds (see the Section 4 of the SI) further illustrate the models' behavior and the superior performance of the AFB model.

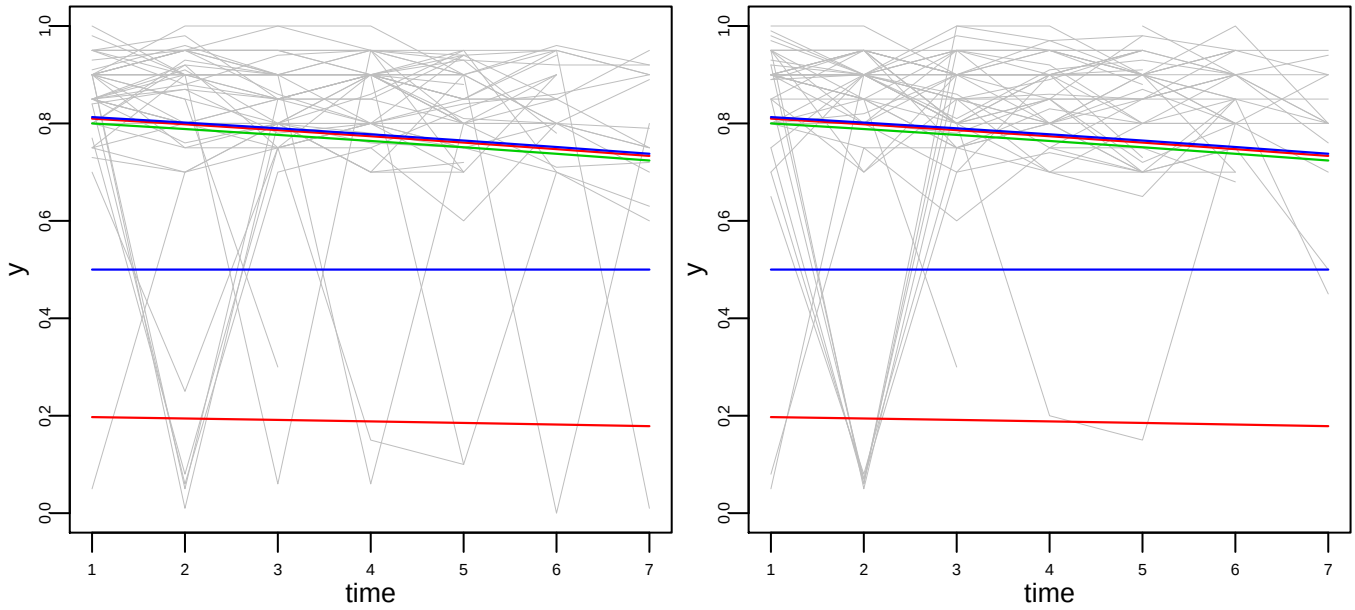

FIGURE 4 Longitudinal profiles of the EQ-VAS scores of 50 randomly selected subjects treated with a placebo (left-hand panel) or creatine (right-hand panel) plus the regression lines of the AFB (red lines), ABR (blue lines), and AB (green line) models.

7 | DISCUSSION

Motivated by two empirical applications extensively studied in the biomedical literature, we proposed a new regression model for bounded responses based on the AFB distribution, together with its extension to handle longitudinal data by means of random effects. Our model allows the mean response, as well as the probabilities of 0s and 1s, to be modeled as functions of covariates, thus enabling the identification of covariates that explain the 0 (e.g., disease-free), progressing, and 1 (e.g., completely diseased) response values. Moreover, the new model can be easily extended by allowing the covariates to be regressed onto the precision parameter, as well as the two remaining parameters of the AFB distribution which deserve a clear and full interpretation (i.e., the distance between the mixture components and the mixing proportion). Although the AFB regression model is not designed as a robust model for proportions, its mixture structure enables one of the two components to be dedicated to outlier or excessive number of tail-area observations, thus producing a very good, sometimes outstanding, performance. The estimation issues have been addressed using a Bayesian approach. It must be noted that the time needed to achieve convergence might be long for large datasets with missing values and highly complex hierarchical structures such as the PD dataset. The runtimes of the AFB and ABR models are broadly similar, while those of the AB model are the shortest due to fewer parameters and the absence of a latent mixture structure. The ABR model sometimes exhibited poor convergence, and it required longer chains to achieve the equilibrium distribution, probably because of the inherent complexity of its mixture structure and the lack of several important theoretical properties. The presence of these properties, particularly strong identifiability and likelihood boundedness, makes the AFB model remarkably tractable from a computational point of view, despite being the most complex (i.e., the model with more parameters) among the three models. The proposed AFB mixed-effects regression model, illustrated for response variables with values in the interval $[0, 1]$, can be easily extended to address variables that take values in a generic bounded interval. This extension is achieved through simple ad hoc rescaling (i.e., a linear transformation) of the response, which does not affect the properties or interpretation of the model. Further interesting extensions include a more complex treatment of missing values (treated here as missing at random) and random effects misspecification issues. These extensions are left for future research.

ACKNOWLEDGEMENTS

We acknowledge the NINDS for providing the PD LS-1 dataset. Anyone interested in the dataset can make a request to NINDS. We thank the Associate Editor and three anonymous referees for their valuable comments.

SUPPLEMENTARY MATERIAL

Supporting Information (including the STAN code) is available for this article online.

References

1. Kieschnick R, McCullough BD. Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. *Stat Modelling.* 2003;3(3):193-213.
2. Ghosh A. Robust inference under the beta regression model with application to health care studies. *Stat Methods Med Res.* 2019;28(3):871-888.
3. Xu XS, Samtani M, Yuan M, Nandy P. Modeling of bounded outcome scores with data on the boundaries: application to disability assessment for dementia scores in Alzheimer's disease. *The AAPS Journal.* 2014;16(6):1271-1281.
4. Bandyopadhyay D, Galvis DM, Lachos VH. Augmented mixed models for clustered proportion data. *Stat Methods Med Res.* 2017;26(2):880-897.
5. Chen K, Cheng Y, Berkout O, Lindhiem O. Analyzing proportion scores as outcomes for prevention trials: a statistical primer. *Prev Sci.* 2017;18(3):312-321.
6. Paolino P. Maximum likelihood estimation of models with beta-distributed dependent variables. *Polit Anal.* 2001;9(4):325-346.
7. Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. *J Appl Stat.* 2004;31(7):799–815.
8. Branscum AJ, Johnson WO, Thurmond MC. Bayesian beta regression: applications to household expenditure data and genetic distance between foot-and-mouth disease viruses. *Aust N Z J Stat.* 2007;49(3):287-301.
9. Liu F, Eugenio EC. A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. *Stat Methods Med Res.* 2016;27(4):1024-1044.
10. García CB, García Pérez J, van Dorp JR. Modeling heavy-tailed, skewed and peaked uncertainty phenomena with bounded support. *Stat Methods Appl.* 2011;20(4):463-486.
11. Migliorati S, Di Brisco AM, Ongaro A. A new regression model for bounded responses. *Bayesian Anal.* 2018;13(3):845-872.
12. Hahn ED. Mixture densities for project management activity times: a robust approach to PERT. *Eur J Oper Res.* 2008;188(2):450-459.
13. Bayes CL, Bazán JL, García C. A new robust regression model for proportions. *Bayesian Anal.* 2012;7(4):841-866.
14. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol Methods.* 2006;11(1):54-71.
15. Galvis DM, Bandyopadhyay D, Lachos VH. Augmented mixed beta regression models for periodontal proportion data. *Stat Med.* 2014;33(21):3759-3771.
16. Ospina R, Ferrari SL. A general class of zero-or-one inflated beta regression models. *Comput Stat Data Anal.* 2012;56(6):1609-1623.
17. Wang J, Luo S. Augmented beta rectangular regression models: a Bayesian perspective. *Biometrical Journal.* 2016;58(1):206-221.
18. Wang J, Luo S. Bayesian multivariate augmented beta rectangular regression models for patient-reported outcomes and survival data. *Stat Methods Med Res.* 2017;26(4):1684-1699.

19. Pammer K, Kevan A. The contribution of visual sensitivity, phonological processing, and nonverbal IQ to children's reading. *Sci Stud Read.* 2007;11(1):33-53.

20. Elm JJ. Design innovations and baseline findings in a long-term Parkinson's trial: the national institute of neurological disorders and stroke exploratory trials in Parkinson's Disease Long-Term Study - 1. *Mov Disord.* 2012;27(12):1513-1521.

21. Kieburtz K, Tilley BC, Elm JJ, et al. Effect of creatine monohydrate on clinical progression in patients with Parkinson disease: a randomized clinical trial. *JAMA.* 2015;313(6):584-593.

22. Cribari-Neto F, Zeileis A. Beta regression in R. *J Stat Softw.* 2010;34(2):1-24.

23. Ferrari SL, Espinheira PL, Cribari-Neto F. Diagnostic tools in beta regression with varying dispersion. *Stat Neerl.* 2011;65(3):337–351.

24. Grün B, Kosmidis I, Zeileis A. Extended beta regression in R: shaken, stirred, mixed, and partitioned. *J Stat Softw.* 2012;48(11):1-25.

25. McCullagh P, Nelder JA. *Generalized Linear Models*. London:Chapman & Hall. 2nd ed. 1989.

26. Figueroa-Zuniga JI, Arellano-Valle R, Ferrari SL. Mixed beta regression: a Bayesian perspective. *Comput Stat Data Anal.* 2013;61(May):137–147.

27. Wu L. *Mixed Effects Models for Complex Data*. New York:Chapman and Hall/CRC. 2009.

28. Todem D, Kim K, Hsu WW. Marginal mean models for zero-inflated count data. *Biometrics.* 2016;72(3):986-994.

29. Frühwirth-Schnatter S. *Finite Mixture and Markov Switching Models*. New York:Springer Science & Business Media. 2006.

30. Duane S, Kennedy A, Pendleton BJ, Roweth D. Hybrid Monte Carlo. *Phys Lett B.* 1987;195(2):216-222.

31. Neal RM. An improved acceptance procedure for the hybrid Monte Carlo algorithm. *J Comput Phys.* 1994;111(1):194-203.

32. Carpenter B, Gelman A, Hoffman MD, et al. Stan : a probabilistic programming language. *J Stat Softw.* 2017;76(1).

33. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. *Bayesian Data Analysis*. New York, NY: CRC Press. 2014.

34. Albert J. *Bayesian Computation with R*. New York:Springer Science. 2nd ed. 2009.

35. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. *J Multivar Anal.* 2009;100(9):1989-2001.

36. Mengersen KL, Robert CP, Guihenneuc-Jouyaux C. MCMC convergence diagnostics: a review. *Bayesian Statistics.* 1999;6:415-440.

37. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res.* 2010;11(Dec):3571-3594.

38. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput.* 2017;27(5):1413-1432.

39. Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Stat Sin.* 1996;6(4):733-807.

40. Carlin BP, Louis TA. *Bayesian Methods for Data Analysis*. Boca Raton:CRC Press. 3rd ed. 2008.

41. Dey DK, Chen MH, Chang H. Bayesian Approach for Nonlinear Random Effects Models. *Biometrics.* 1997;53(4):1239-1252.

42. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat.* 2008;2(4):1360-1383.