Bayesian Analysis (2018) **13**, Number 3, pp. 845–872

# A New Regression Model for Bounded Responses

Sonia Migliorati*, Agnese Maria Di Brisco†, and Andrea Ongaro‡

**Abstract.** Aim of this contribution is to propose a new regression model for continuous variables bounded to the unit interval (e.g. proportions) based on the flexible beta (FB) distribution. The latter is a special mixture of two betas, which greatly extends the shapes of the beta distribution mainly in terms of asymmetry, bimodality and heavy tail behaviour. Its special mixture structure ensures good theoretical properties, such as strong identifiability and likelihood boundedness, quite uncommon for mixture models. Moreover, it makes the model computationally very tractable also within the Bayesian framework here adopted.

At the same time, the FB regression model displays easiness of interpretation as well as remarkable fitting capacity for a variety of data patterns, including unimodal and bimodal ones, heavy tails and presence of outliers. Indeed, simulation studies and applications to real datasets show a general better performance of the FB regression model with respect to competing ones, namely the beta (Ferrari and Cribari-Neto, 2004) and the beta rectangular (Bayes et al., 2012), in terms of precision of estimates, goodness of fit and posterior predictive intervals.

**Keywords:** proportions, beta regression, flexible beta, mixture models, MCMC, outliers, heavy tails.

## 1 Introduction

A relevant problem in applied statistics concerns modeling rates, proportions or, more generally, continuous variables restricted to the interval $(0, 1)$ (Kieschnick and McCullough, 2003). The standard linear regression model is unsuitable, since it may fit values outside the support of the variable of interest. As a consequence, in order to still take advantage of linear models, transforming the response variable so that its support becomes the real line has been the preferred method for a long time. However, such an approach has two relevant drawbacks: first, the difficulty encountered when interpreting the estimated parameters with respect to the original response variable (Ferrari and Cribari-Neto, 2004); and second, the failure of the assumptions of normality (proportions typically show asymmetric distributions) and homoscedasticity (Paolino, 2001).

The current branch of research favors staying on the original restricted space, and modeling the response variable of interest according to a beta distribution. Ferrari and

*Department of Economics, Management and Statistics – University of Milano-Bicocca, sonia.migliorati@unimib.it

†Department of Economics, Management and Statistics – University of Milano-Bicocca, agnese.dibrisco@unimib.it

‡Department of Economics, Management and Statistics – University of Milano-Bicocca, andrea.ongaro@unimib.it

 https://doi.org/10.1214/17-BA1079

Cribari-Neto (2004) defined a regression model for the mean of a beta response variable, and proposed a maximum likelihood estimation approach similar to the one for generalized linear models (GLM). Indeed, standard GLM theory is not directly applicable to beta regression models, since the beta distribution is not a dispersion-exponential family (McCullagh and Nelder, 1989). Moreover, Paolino (2001), Smithson and Verkuilen (2006) and Ferrari et al. (2011) extended such a model by considering variable dispersion. As an alternative to maximum likelihood, a Bayesian inferential approach can be adopted (see e.g. Branscum et al. (2007)).

The beta distribution of the response variable allows for good flexibility since it can show very different shapes. However, it fails to model a wide range of phenomena, including heavy tailed responses with a bounded support (Bayes et al., 2012; García et al., 2011) and bimodality. As a first proposal to handle greater flexibility, Hahn (2008) introduced the beta rectangular (BR) distribution, defined as a mixture of a uniform and a beta distribution, and explored its main properties. Moreover, Bayes et al. (2012) defined a BR regression model for both mean and dispersion parameters by considering a Bayesian approach. This model enables heavier tails, and it has been shown to be robust in the presence of outliers.

In order to achieve even greater flexibility, a generic mixture of beta distributions could be taken into account. Indeed, it is well-known that mixture distributions permit more accurate data fitting and robustness (Markatou, 2000; Gelman et al., 2014). As a counterpart, a generic beta mixture (BM) is substantially less tractable. Indeed, it leads to non identifiability, likelihood unboundedness and invariance under relabeling of the mixture components. This generates the well-known label switching problem and undesirable effects on posterior distributions, especially in case of overlapping components. Moreover, some simulation studies confirmed that convergence to the posterior distribution may be hard to achieve in this case and, even considering long chains, the simulated distribution is sensible to initial values.

The aim of this contribution is to provide (adopting a Bayesian approach) a new regression model, that can handle the trade-off between flexibility and tractability. To this end, we introduce the flexible beta (FB) distribution (univariate version of the flexible Dirichlet distribution (Ongaro and Migliorati, 2013)), which is a special mixture of two beta distributions with arbitrary means and common variance. This enables a greater variety of density shapes in terms of tail behavior, asymmetry and multimodality too. Nevertheless, the FB peculiar structure makes it identifiable in a strong sense, and guarantees a bounded likelihood, as well as a finite global likelihood maximum. These theoretical properties make the FB very tractable from a computational perspective, for example with respect to posterior computation. A suitable reparametrization of the FB is proposed, specifically designed for the regression context, which, at the same time, enables a very clear interpretation of the new parameters. In particular, the new parameters comprise the overall regression mean (as an arbitrary suitably chosen function of covariates), the distance between the two group regression means (whose weighted average gives the general mean), a precision parameter, and the mixing weight.

The increased model flexibility will be shown (by means of simulations and applications to real datasets) to successfully handle various relevant data patterns, e.g. presence

of outliers (with robustness properties), unimodality as well as bimodality, displaying a better fit than the beta and BR regression models.

Although in the paper we shall confine ourselves to the analysis of response variables with values on the interval $(0, 1)$, it seems also worthwhile to point out that the proposed model can be easily extended to deal with variables taking values on a generic bounded interval. This can be achieved by an obvious linear transformation of the response, which will not substantially modify the properties and interpretation of the model.

The paper is organized as follows. In Section 2 we introduce the FB distribution, analyzing some important properties, and we propose the new parametrization. In Section 3.1 we define the FB regression model, proving identifiability and likelihood boundedness under very general conditions. Then, in Section 3.2, we discuss the interpretation of the FB regression (and of the BR regression) as a special case of mixture of regression models (Frühwirth-Schnatter, 2006). A regression model with general BM response is also defined. All details concerning Bayesian inference are provided in Section 4; in particular, a Gibbs sampling specifically designed for mixture models is adopted. In order to evaluate the performance of the FB regression model and compare it with the BR and beta regression ones, we set up some simulation studies (Section 5), and perform applications to two real datasets available in the literature (Section 6). In the latter section a comparison with the general BM regression model is provided as well.

# 2 The Flexible Beta Distribution

## 2.1 The Beta Distribution

The beta distribution usefully describes continuous responses bounded on $(0, 1)$. In the standard parametrization of the beta distribution, $Y \sim Beta(\alpha_1, \alpha_2)$, the mean depends on both the parameters, being $\mathbb{E}[Y] = \frac{\alpha_1}{\alpha_1+\alpha_2}$.

In a beta regression model (Paolino, 2001; Ferrari and Cribari-Neto, 2004) the focus is on modeling the mean as a function of some explanatory variables. To such an end, a useful reparametrization of the beta distribution is the following:

$$\begin{cases} \bar{\alpha} = \frac{\alpha_1}{\alpha_1+\alpha_2}, \\ \alpha^+ = \alpha_1 + \alpha_2. \end{cases} \tag{1}$$

The probability density function (pdf) of the beta in the new parametrization $Y \sim Beta(\bar{\alpha}\alpha^+, (1-\bar{\alpha})\alpha^+)$ can be written as:

$$f_B^*(y; \bar{\alpha}, \alpha^+) = \frac{\Gamma(\alpha^+)}{\Gamma(\bar{\alpha}\alpha^+)\Gamma((1-\bar{\alpha})\alpha^+)} y^{\bar{\alpha}\alpha^+-1}(1-y)^{(1-\bar{\alpha})\alpha^+-1} \qquad 0 < y < 1 \tag{2}$$

with $0 < \bar{\alpha} < 1$ and $\alpha^+ > 0$. By construction, the parameter $\bar{\alpha}$ identifies the mean of $Y$, while the parameter $\alpha^+$ is a *precision* parameter such that, for given $\bar{\alpha}$, the variance of the beta distribution decreases as $\alpha^+$ increases being:

$$Var[Y] = \frac{\bar{\alpha}(1-\bar{\alpha})}{\alpha^+ + 1}.$$

## 2.2 The Flexible Beta Distribution

The beta distribution can be viewed as the univariate case of the Dirichlet distribution, which is a continuous multivariate distribution for compositional data, i.e. positive data subject to a unit-sum constraint (typically proportions). In spite of its many properties, the Dirichlet distribution has been shown to be inadequate to model compositional data essentially because of its rigid structure. The flexible Dirichlet distribution (Ongaro and Migliorati, 2013) is a generalization of the Dirichlet distribution, which preserves to a large extent the Dirichlet remarkable tractability, while also allowing for considerably greater flexibility (see Ongaro and Migliorati (2013) and Migliorati et al. (2016)).

As the univariate case of the flexible Dirichlet, the FB distribution can be defined as a special mixture of two beta distributions, $Y \sim pBeta(\alpha_1+\tau, \alpha_2)+(1-p)Beta(\alpha_1, \alpha_2+\tau)$, depending on four parameters $(\alpha_1, \alpha_2, \tau, p)$ such that $\alpha_1 > 0$, $\alpha_2 > 0$, $\tau > 0$ and $0 < p < 1$. If $Y$ is a FB, then its pdf can be written for $0 < y < 1$ as:

$$f_{FB}(y; \alpha_1, \alpha_2, \tau, p) = \\ \frac{\Gamma(\alpha_1+\alpha_2+\tau)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1-1}(1-y)^{\alpha_2-1}\left[p\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1+\tau)}y^{\tau} + (1-p)\frac{\Gamma(\alpha_2)}{\Gamma(\alpha_2+\tau)}(1-y)^{\tau}\right]. \quad (3)$$

The first two moments of the FB are equal to:

$$\mathbb{E}(Y) = \frac{\alpha_1+\tau p}{\phi}, \\ Var(Y) = \frac{\mathbb{E}(Y)(1-\mathbb{E}(Y)) + \tau^2 p(1-p)/\phi}{\phi+1}, \quad (4)$$

where $\phi = \alpha_1 + \alpha_2 + \tau$.

To understand the potential of the FB distribution, it is enlightening to know that it is a mixture of two beta distributions with a common precision parameter $\phi = \alpha_1+\alpha_2+\tau$ and arbitrary (but distinct) means $\lambda_1 > \lambda_2$:

$$f^*_{FB}(y; \lambda_1, \lambda_2, \phi, p) = pf^*_B(y; \lambda_1, \phi) + (1-p)f^*_B(y; \lambda_2, \phi), \quad (5)$$

where $f^*_B$ is the mean-precision parametrized beta (2) and

$$\lambda_1 = \frac{\alpha_1+\tau}{\phi}, \qquad \lambda_2 = \frac{\alpha_1}{\phi}.$$

Therefore, the special mixture structure which defines the FB is simple when looked at from interpretative viewpoint. Moreover, it is able to encompass the main features of a variety of datasets of practical interest. Indeed, the FB distribution greatly extends the variety of shapes of the beta (unimodal, monotone and U-shaped) mainly in terms of bimodality (Figure 1, left panel), asymmetry and tail behavior (Figure 1, right panel). In particular, a comparison of the tail flexibility of the beta and FB models can be performed analytically by examining their density behavior when $y$ tends to 0 or to 1. It is immediate to see that any unimodal beta distribution has zero limit at both bounds

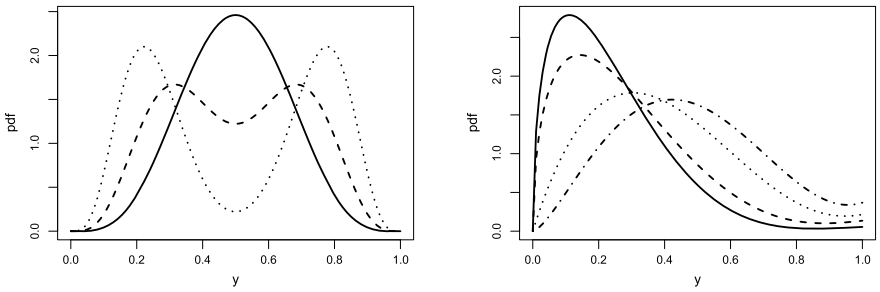


Figure 1: Some examples of FB distributions. Left panel: $\alpha_1 = 5$, $\alpha_2 = 5$, $p = 0.5$ and $\tau$={1 (solid line), 5 (dashed line), 10 (dotted line)}. Right panel: $\alpha_2$ is fixed at 1, while $\alpha_1 = 1.5$, $\tau = 4$, $p = 0.01$ (solid line), $\alpha_1 = 1.5$, $\tau = 3$, $p = 0.03$ (dashed line), $\alpha_1 = 2$, $\tau = 2.3$, $p = 0.05$ (dotted line) and $\alpha_1 = 2.5$, $\tau = 2.1$, $p = 0.08$ (dashed-dotted line).

(0 and 1). Strictly positive finite values can be obtained only by setting at least one of the two $\alpha_i$'s equal to 1, in which case a monotone (or uniform) density is obtained.

Conversely, the FB can give rise to unimodal distributions with heavy tails. Indeed, it can exhibit one positive tail limit when one of the $\alpha_i$'s is equal to 1 and the other is bigger than 1. Thus, the FB includes densities with one heavy tail and possibly large asymmetry (see Figure 1, right panel). Furthremore, by setting $\alpha_1 = \alpha_2 = 1$, both limits are positive and not necessarily equal.

Finally, its peculiar structure makes it theoretically and computationally very tractable. Specifically, the FB can be shown to be identifiable in the following strong sense (typically satisfied only by non-mixture models): two elements of the FB parametric family are equal if and only if the corresponding parameters are the same. This implies, in particular, that there is no invariance under permutations of the components (and therefore no labeling problems). In addition, under an i.i.d. sample with at least three observations, the FB likelihood is a.s. bounded from above and admits a finite global maximum on the assumed parameter space (see Migliorati et al. (2016)). This has positive implications from a computational perspective as well, both in the maximization steps of classical estimation procedures and in posterior computation of Bayesian ones. Note that an arbitrary mixture of two beta distributions does not share the above properties, thus being substantially less tractable than the FB.

A different mixture distribution on (0,1) proposed in the literature in a regression context is the BR. This is defined as the mixture of a uniform and an arbitrary $Beta(\bar{\alpha}, \alpha^+)$ (in the mean-precision parametrization (2)) with probability $p$ and $1 - p$ respectively (Hahn, 2008; Bayes et al., 2012). The BR, despite being somewhat less flexible than the FB (having only three instead of four parameters) is not identifiable. For example, when $\bar{\alpha} = 1/2$ and $\alpha^+ = 2$ (corresponding to the uniform) any value of the mixing proportion $p$ generates the same density. Moreover, the BR model has an unbounded likelihood. Indeed, if one takes $\bar{\alpha}$ equal to one of the observations and lets $\alpha^+$ tend to infinity, then the likelihood tends to infinity for any $p > 0$.

Note that the BR family is not contained in the FB one. Their intersection, in the mean-precision parametrization (5), is obtained by choosing $\phi = 2$ and either $\lambda_1 = 1/2$ or $\lambda_2 = 1/2$. With regard to tail behavior, also the BR, by adding a uniform component to the beta, can give rise to unimodal distributions with strictly positive limits. But it is constrained to have both limits positive (or both zero if the uniform component is dropped), and taking the same values, which might not be suitable for many datasets.

## 2.3 Reparametrization of the Flexible Beta

In order to define a regression model with a FB response, it is convenient to introduce a reparametrization which explicitly includes the mean, possibly complemented with clearly interpretable parameters. In the mean-precision parametrization (5), a sensible proposal is:

$$\begin{cases} \mu = \mathbb{E}(Y) = p\lambda_1 + (1-p)\lambda_2, \\ \phi = \phi, \\ \tilde{w} = \lambda_1 - \lambda_2, \\ p = p, \end{cases} \tag{6}$$

where $\mu$ is the mean, $\tilde{w}$ is a measure of distance between the two mixture components, $p$ is the mixing proportion, and $\phi$ plays the role of a precision parameter. Indeed it can be proved that $Var(Y)$ is a decreasing function of $\phi$.

If we analyze the parametric space, we get that while $\phi$ is free to move in $\mathbb{R}^+$, $\mu$, $p$ and $\tilde{w}$ are linked by some constraints. On the other hand, a variation-independent parametric space may be more appropriate for Bayesian inference through Gibbs sampling (Albert, 2009), which is the approach we shall implement in Section 4. This also allows us to separately model any parameter as a function of the covariates.

To this purpose, we chose to leave $\mu$ and $p$ free to assume values in (0,1), and to properly normalize $\tilde{w}$ to make it free to move on the range (0,1) as well. One can see that, for a given $\mu$ and $p$, the constraints $0 < \lambda_2 < \lambda_1 < 1$ imply that $\tilde{w}$ takes values between 0 and $\min\{\frac{\mu}{p}, \frac{1-\mu}{1-p}\}$. Therefore, we replace $\tilde{w}$ with:

$$w = \frac{\tilde{w}}{\min\left\{\frac{\mu}{p}, \frac{1-\mu}{1-p}\right\}}. \tag{7}$$

The chosen reparametrization guarantees a variation independent parameter space where $p$, $\mu$ and $w$ vary in $(0,1)$ and $\phi > 0$. One might also achieve a variation independent parameter space by normalizing the parameter $p$ instead of $\tilde{w}$. If no covariates are present and $\mu$ is fixed, this choice would lead to a perfectly equivalent model. However, when $\mu$ depends on covariates the two choices give rise to two different models, as we shall discuss in Section 3.2, where the motivation for normalizing $\tilde{w}$ will emerge.

# 3 The Flexible Beta Regression Model

## 3.1 Definition and Some Properties of the Model

Let us consider a vector of independent responses $\mathbf{Y}^T = (Y_1, \ldots, Y_i, \ldots, Y_n)$ which assume values in the unit interval $(0, 1)$. Following the GLM methodology (McCullagh and Nelder, 1989), a regression model can be defined as

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \qquad i = 1, \ldots, n, \tag{8}$$

where $\mu_i$ is the mean of $Y_i$, $\mathbf{x}_i^T = (x_{i0}, x_{i1}, \ldots, x_{ik})$ is a vector of covariates, $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \ldots, \beta_k)$ is a vector of regression parameters and $g(\cdot)$ is an adequate link function, strictly monotone and twice differentiable. The most popular link function is the logit $logit(\mu_i) = \log(\mu_i/(1-\mu_i))$ mainly because of the inherent simple interpretation of the regression coefficients in terms of odds ratios.

If $Y_i$ follows a beta distribution, then the beta regression model is obtained (Ferrari and Cribari-Neto, 2004), while if follows a BR distribution (see Section 2.2), then the BR regression (BRR) model is achieved (Bayes et al., 2012). More precisely, the parametrization used in Bayes et al. (2012) is $Y_i \sim BR(\mu_i, \alpha^+, \nu)$ where $\mu_i = E(Y_i)$, $\nu = p/(1- \mid 2\mu_i - 1 \mid)$ and $\alpha^+$ is the precision parameter of the beta component.

Here we define the FB regression (FBR) model by assuming that each $Y_i$ is independently distributed as a flexible beta: $Y_i \sim FB(\mu_i, \phi, w, p)$ where the parametrization given in Section 2.3 is used. Note that none of the above models is of the GLM type, as the involved distributions do not belong to the dispersion-exponential family (McCullagh and Nelder, 1989).

The regression models so far described only focus on modeling the mean parameter. The response variances, being functions of the corresponding means (see formula (4)), will also vary with the covariates, thus inducing a form of heteroscedasticity. This induced form is a rather natural one, as it takes into account the fact that the maximum value of the variance of a random variable on (0,1) depends on its mean $\mu$, being equal to $\mu(1-\mu)$. However, in some cases, it may be desirable to independently modeling the variance as a function of the covariates. Many authors have proposed extensions in this direction (Paolino, 2001; Smithson and Verkuilen, 2006; Ferrari et al., 2011). This can be easily achieved in the FBR too, as the parameters $\mu$ and $\phi$ do not share any constraint. The regression for the dispersion can be defined as $h(\phi_i) = \mathbf{z}_i^T \boldsymbol{\delta}$, where $h(\cdot)$ is an appropriate link function, $\mathbf{z}_i^T = (z_{i0}, z_{i1}, \ldots, z_{il})$ is a vector of covariates and $\boldsymbol{\delta}^T = (\delta_0, \delta_1, \ldots, \delta_l)$ is a vector of regression parameters.

There may be relevant contexts (an example will be shown in Section 6.1) where it is sensible to let the weight $p$, and possibly even the group distance $w$, depend on covariates. This can also be easily accommodated in the FBR model in a similar fashion.

As a consequence of the identifiability of the FB distribution, the FBR model (including all above mentioned extensions with various parameters depending on covariates) is identifiable, under the obviously necessary condition that the corresponding design matrices have full rank (see the Supplementary Material (Migliorati et al., 2017), Section I, for the proof).

**Proposition 1.** *Consider a vector of independent response variables* $Y_i \sim FB(\boldsymbol{\theta}_i)$, $(i = 1, \ldots, n)$ *with* $\boldsymbol{\theta}_i = (\mu_i, \phi_i, w_i, p_i)$. *Suppose that*

$$\mu_i = g_1(\mathbf{x}_i^T \boldsymbol{\beta}_1), \quad \phi_i = g_2(\mathbf{z}_i^T \boldsymbol{\beta}_2), \quad w_i = g_3(\mathbf{t}_i^T \boldsymbol{\beta}_3), \quad p_i = g_4(\mathbf{u}_i^T \boldsymbol{\beta}_4),$$

*where* $\mathbf{x}_i$, $\mathbf{z}_i$, $\mathbf{t}_i$ *and* $\mathbf{u}_i$ $(i = 1, \ldots, n)$ *are vectors of possibly overlapping covariates of dimension* $k_1, k_2, k_3, k_4$ *respectively (with* $k_j \leq n$*),* $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, $\boldsymbol{\beta}_4$ *are vectors of corresponding regression parameters and* $g_1$, $g_3$, $g_4$ *are strictly monotone functions from* $\mathbb{R}$ *to* $(0, 1)$ *while* $g_2$ *is a strictly monotone function from* $\mathbb{R}$ *to* $\mathbb{R}^+$.

*Let us denote by* $FBR(\boldsymbol{\gamma})$ *where* $\boldsymbol{\gamma} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)$ *the corresponding regression model. Let* $\mathbf{Y} \sim FBR(\boldsymbol{\gamma})$ *and* $\mathbf{Y}' \sim FBR(\boldsymbol{\gamma}')$. *Then, if the design matrices* $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{T}$ *and* $\mathbf{U}$ *are of full rank,* $\mathbf{Y} \sim \mathbf{Y}'$ *if and only if* $\boldsymbol{\gamma} = \boldsymbol{\gamma}'$.

Note that the above formulation includes the case where some of the FBR parameters are kept fixed. To make a given parameter not dependent on covariates, it is enough to associate a fictitious constant covariate to it.

It is also possible to show that, under general conditions, the likelihood of the FBR model is bounded from above. Let us first consider, for simplicity, the case where only the mean $\mu_i$ depends on covariates (for the proof see the Supplementary Material, Section I).

**Proposition 2.** *Consider a vector of independent response variables* $Y_i \sim FB(\boldsymbol{\theta}_i)$, $(i = 1, \ldots, n)$ *with* $\boldsymbol{\theta}_i = (\mu_i, \phi, w, p)$ *and* $g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$, $i = 1, \ldots, n$, *as in (8), with* $\boldsymbol{\beta}$ *of dimension* $k + 1$. *Suppose that* $n > k + 2$. *Then, the likelihood* $L(\boldsymbol{\beta}, \phi, w, p|\mathbf{y})$ *is a.s. bounded from above.*

With analogous proof, and suitably increasing the requirement on the sample size $n$, boundedness of the likelihood can be shown even when $p$ and/or $w$ are modeled as functions of covariates. On the contrary, when $\phi$ depends on covariates, one can specify the latter so that even the likelihood of the beta regression model diverges. This may happen because it is possible to make $\phi_i$ diverge only for some $i$'s. A simple example is the following. Suppose to have just one covariate $X$ taking on only two values $x_1 = 1$ and $x_i = 0$ for $i \geq 2$. If $\log \phi_i = \delta_0 + \delta_1 x_i$, then by letting $\delta_1 \to +\infty$ and $\mu_1 = y_1$, we obtain an infinite likelihood for $y_1$ and a strictly positive one for the other observations.

Note that for the BRR model, and therefore for a generic beta mixture one, the likelihood is unbounded whatever parameters are expressed as functions of covariates.

## 3.2 Flexible Beta Regression and Mixture of Regression Models

It is fruitful to understand the relationship between the FBR model and the mixture of regression models (Frühwirth-Schnatter, 2006). In the latter it is assumed that the regression function is not fixed over all realizations, but different groups of observations may display (arbitrarily) different dependencies of the response means on covariates. Their main focus is on separately modeling the group regressions. Conversely, the main inferential objective of the present paper is on assessing the impact of covariates on the

(overall) mean of the response variable. To such an end, a mixture of regression models generally produces not clearly interpretable results.

However, by viewing the FBR as a very special case of mixture of regression models, a different and enlightening perspective on it can be gained. In particular, two important aspects will be clarified, namely the fact that different parametrizations of the FB model imply different forms of group regressions for the FBR one, and the motivation and implications of the chosen parametrization. From (6), the FBR can be written as a mixture of two beta regression models, with common precision $\phi$ and group means given by

$$\begin{cases} \lambda_1 = \mu + (1-p)\tilde{w}, \\ \lambda_2 = \mu - p\tilde{w}. \end{cases} \tag{9}$$

Here $\mu = g^{-1}(\mathbf{x}^T\boldsymbol{\beta})$ has to be understood as a function of covariates. The underlying assumption, inherited from the FB model, is that there are two groups, one of which displays a greater mean ($\lambda_1$) than the other, for any given value of covariates. However, because $\lambda_1 \leq 1$ and $\lambda_2 \geq 0$ for any given $\mu$, the parameters $p$ and $\tilde{w}$ can not freely vary in $(0,1)$. This implies that at least one of the two must depend on $\mu$. A large variety of modeling choices for $p$ and $\tilde{w}$ as functions of $\mu$ unfolds, the simplest two being to fix one of the two parameters and let the other vary with $\mu$.

In our opinion, it is generally more appropriate to fix $p$. Indeed, often there are latent groups of fixed size, but with different behavior with respect to covariates. Furthermore, it seems natural to expect that, when the overall mean proportion $\mu$ tends to 0 (or to 1), the two group means tend to 0 (or to 1) as well. This makes the difference between the two group means $\tilde{w}$ vary, decreasing (and tending to 0) when $\mu$ tends to 0 or to 1. On the other hand, if one fixes the mean difference $\tilde{w}$, when $\mu \to 0$ $\lambda_2$ will tend to 0, whereas $\lambda_1$ will approach the positive value $\tilde{w}$. This is possible only if $p \to 0$ when $\mu \to 0$. Analogously, one can see that when $\mu \to 1$, $p$ is forced to tend to 1. This peculiar behavior of $p$ seems to be well-suited only in very specific situations.

Therefore, our choice is to fix $p$. In this case, we obtain the constraints $0 \leq \tilde{w} \leq \min\{\mu/p, (1-\mu)/(1-p)\}$. The simplest choice for $\tilde{w}$, obtained by normalization, is then $\tilde{w} = w \min\{\mu/p, (1-\mu)/(1-p)\}$. Here the parameter $w$ can vary freely in $(0,1)$, and represents the maximum value of $\tilde{w}$, which is attained at $\mu = p$.

To better grasp the practical implication of this choice, Figure 2 reports a graph of the two group means $\lambda_1$ and $\lambda_2$ as functions of $\mu$. Although at a first glance the specific behavior of the two regression curves might seem somewhat odd, a moment's reflection shows that it is largely explained by boundedness of the response variable and by the requirement that one group regression dominates the other. Indeed, typically in linear regression models the dominance constraint is incorporated by imposing a fixed difference or ratio (for positive variables) of the group regressions. In the present context a fixed ratio is unfeasible, while the difference can be fixed only giving up the condition that both curves tend to zero (one) when $\mu$ tends to zero (one)). Actually, the chosen regression means are, in some sense, as close as possible to the fixed ratio situation: the ratio $\lambda_1/\lambda_2$ is constant for $\mu \leq p$, as it is the ratio $\frac{1-\lambda_1}{1-\lambda_2}$ for $\mu \geq p$. They are piecewise increasing linear functions of $\mu$ and their difference $\tilde{w}$ is (linearly) decreasing

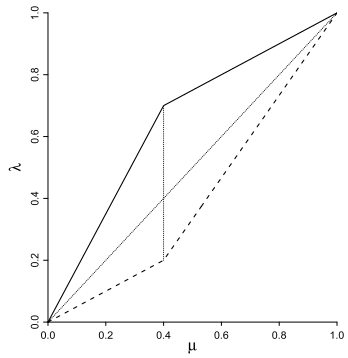


Figure 2: Group means $\lambda_1$ (solid) and $\lambda_2$ (dashed) as functions of $\mu$ with $p = 0.4$ and $w = 0.5$.

to zero when $\mu$ goes to zero or one, as expected. The maximum is attained at $\mu = p$, which is the only point where any acceptable function $\tilde{w}$ can vary in the whole range $(0, 1)$.

Note that the BRR can be viewed as a mixture of regression models as well, remaining valid the general mean structure given in (9). Here one of the two group regressions does not depend on covariates, being identically equal to $\frac{1}{2}$. Furthermore, both the weight $p$ and the mean difference $\tilde{w}$ depend on $\mu$. More specifically, $p = \nu(1 - \mid 2\mu - 1 \mid)$, therefore tending to 0 when $\mu$ tends to 0 or 1, and having a maximum equal to $\nu$ when $\mu = \frac{1}{2}$. In particular, this implies that, under this model, outliers (which are captured by the uniform component) are considered more likely to happen for central mean values than for extreme values of the response variable.

In the following, purely for comparison purposes, we shall also consider a regression model (called BMR) where the response variable is a general BM. To our knowledge this kind of model has not yet been proposed in the literature. Besides the difficulties already mentioned in the Introduction (i.e. non identifiability, likelihood unboundedness, convergence of estimation algorithms), the above discussion shows that the implementation of such a model when covariates are present requires an in-depth analysis of the various available possibilities. This is because many and very different regression models can be defined for a general BM distributed response variable, depending on the assumption made on $p$ and $\tilde{w}$ in expression (9). Actually, here the variety of possible choices is even wider, as we are no longer necessarily forced to assume the dominance constraint on the group means, i.e. now $\tilde{w}$ in (9) may take on both positive and negative values.

An appropriate analysis of this aspect goes beyond the scope of this paper. Here we limit ourselves to consider a simple BMR, obtained by fixing $p$ and normalizing $\tilde{w}$, to give a first idea of possible results. In order to mitigate (but not completely eliminate, as we shall see) convergence problems of estimation algorithms, we impose identifiability by setting $p \geq 1/2$, whereas we let the two means $\lambda_1$ and $\lambda_2$ unconstrained.

## 4 Bayesian Inference

The likelihood function for the FBR model (8) based on a sample of $n$ independent observations $\mathbf{y}^T = (y_1, \ldots, y_i, \ldots, y_n)$ is equal to:

$$L(\boldsymbol{\eta}|\mathbf{y}) = \prod_{i=1}^{n} f^*_{FB}(y_i|\mu_i, \phi, w, p), \tag{10}$$

where $\boldsymbol{\eta} = (\boldsymbol{\beta}, \phi, w, p)$, $\mu_i = g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta})$, and $f^*_{FB}(y|\mu, \phi, w, p)$ is given by (5) with $\lambda_1$ and $\lambda_2$ as in (9) and $\tilde{w} = w \min\{\mu/p, (1-\mu)/(1-p)\}$.

Since the allocation of each $i^{th}$ observation to one of the two mixture components is unknown, no explicit solution to the estimation problem exists. Thus, a numerical one is required. Here our preference is for a Bayesian approach, which makes it easy to cope with complex models with many parameters. In particular, we propose to use Makov Chain Monte Carlo (MCMC) techniques such as data augmentation (Tanner and Wong, 1987) and Gibbs sampling (Gelfand and Smith, 1990). A mixture model can be seen as an incomplete data problem (Dempster et al., 1977). Therefore, we can introduce a $n$-dimensional vector of latent variables $\mathbf{v}$, such that $v_i = 1$ if the $i^{th}$ observation belongs to the first mixture component and $v_i = 0$ otherwise. These latent variables are missing data, which can be accommodated within Gibbs sampling by repeating two steps until convergence: one for the parameter simulation conditional on $\mathbf{v}$, and the other for the classification of the observations (i.e. updating $\mathbf{v}$) conditional on knowing the parameter. This enables to compute the "complete-data" posterior distribution $\pi(\boldsymbol{\eta}, \mathbf{v}|\mathbf{y})$ and, finally, the posterior distribution $\pi(\boldsymbol{\eta}|\mathbf{y})$ by marginalization. The computation of the former can be accomplished by resorting to the complete-data likelihood $L_{CD}(\boldsymbol{\eta}|\mathbf{y}, \mathbf{v})$, i.e. the likelihood based on both observed ($\mathbf{y}$) and missing ($\mathbf{v}$) data. More precisely:

$$\pi(\boldsymbol{\eta}, \mathbf{v}|\mathbf{y}) \propto L_{CD}(\boldsymbol{\eta}|\mathbf{y}, \mathbf{v})\pi(\boldsymbol{\eta}) \tag{11}$$

with

$$L_{CD}(\boldsymbol{\eta}|\mathbf{y}, \mathbf{v}) = \prod_{i=1}^{n} [p f^*_B(y_i; \lambda_{1i}, \phi)]^{\{v_i\}} [(1-p) f^*_B(y_i; \lambda_{2i}, \phi)]^{\{1-v_i\}}, \tag{12}$$

where $\lambda_{1i}$ and $\lambda_{2i}$ are given by (9), $f^*_B$ is defined by (2), and $\pi(\boldsymbol{\eta})$ is an appropriate prior distribution.

As for the specification of the prior distribution, we assumed a priori independence, which is a usual choice when no prior information is available. This is feasible in our context since the parametric space is variation independent. Therefore, the joint prior distribution can be factorized as $\pi(\boldsymbol{\eta}) = \pi(\boldsymbol{\beta})\pi(\phi)\pi(w)\pi(p)$.

Moreover, we decided to adopt flat priors, so as to generate the minimum impact on the posteriors (see e.g. Albert (2009)). With respect to the regression parameters, we selected the usual multivariate normal prior $\boldsymbol{\beta} \sim N_{k+1}(\mathbf{a}, \mathbf{B})$ with $\mathbf{a} = \mathbf{0}$ for the mean, and a diagonal covariance matrix with "large" values for the variances in $\mathbf{B}$. For the remaining parameters we chose a gamma distribution $Ga(g, g)$ for $\phi$, which is a

rather standard choice for precision parameters (see e.g. Branscum et al. (2007)), and we selected non-informative uniform priors for the remaining parameters $w \sim U(0,1)$ and $p \sim U(0,1)$. A numerical investigation showed that there is robustness (limited impact on inferential conclusions) with respect to different choices of the hyperparameters $\mathbf{B}$ and $g$, the main consequence being the length of the chains generated to obtain pseudo-independent Monte Carlo samples from the posteriors. Further details on the choice of the prior for $\phi$ can be found in Section II of the Supplementary Material, where a different prior tailored-made for the present context is also considered.

The estimation procedure described above can be easily extended to deal with cases in which the precision parameter is modeled as a function of the covariates. It is enough to replace the prior for $\phi$ with a convenient multivariate normal prior for the regression coefficients $\boldsymbol{\delta} \sim N_{l+1}(\mathbf{c}, \mathbf{D})$. Analogous considerations hold for the parameters $w$ and $p$.

We implemented the Gibbs sampling algorithm through the BUGS software (Thomas, 1994; Lunn et al., 2000) in order to generate a finite set of values from the posterior distribution, and further analyzed the results through the R software (R Core Team, 2016). Both in simulation studies and real-data applications, we chose suitable burn-in periods to avoid the influence of initial values. Furthermore, to properly treat autocorrelations, we also set a thinning interval, say L, such that only the first generated values in every batch of L iterations were kept. Finally, to verify the convergence of the algorithm, several statistical tests were used as convergence diagnostics, with a focus on diagnostic tests for stationarity (Geweke and Heidel diagnostics) and for the level of autocorrelation (Raftery diagnostic) (Mengersen et al., 1999; Ntzoufras, 2011).

In order to compare the models we focused on comparison criteria that take into account the trade-off between the goodness of fit and the complexity of a model. Here, we shall consider some well-known deviance-based criteria, where the deviance is defined as a function of the likelihood $L(\boldsymbol{\eta}|\mathbf{y})$ (see (10) for the FBR model), i.e.: $D(\boldsymbol{\eta}) = -2log[L(\boldsymbol{\eta}|\mathbf{y})]$. Clearly the deviance is a measure of lack of fit, and can simply be estimated from the MCMC output by taking the posterior mean of the deviance $\overline{D}$, i.e. the mean of the deviances of the MC sample.

As for model complexity, different measures can be considered. More precisely, the deviance information criterion ($DIC$) (Spiegelhalter et al., 2002):

$$DIC = \overline{D} + p_D$$

penalizes the complexity of the model via $p_D = \overline{D} - D(\bar{\boldsymbol{\eta}})$, where $\bar{\boldsymbol{\eta}}$ is the vector of posterior means of the parameters. Therefore, $p_D$ can be interpreted as an estimate of the "true" number of parameters, and suggests how many parameters one loses by adopting a Bayesian approach (Brooks, 2002) due to the prior-induced parameter reduction. Note that in some situations, especially in the context of missing data problems like mixture models, the value of $p_D$ can take negative values leading to unreliable criteria values (Celeux et al., 2006). Essentially, this happens because of the lack of identifiability, which is not the case of the FB model.

Alternatively, one can penalize model complexity the same way as is done by the well-known Akaike Information Criterion (AIC, Akaike (1998)) and Bayesian Information Criterion (BIC, Schwarz (1978)), thus obtaining the corresponding Bayesian

counterparts Expected AIC (EAIC) and Expected BIC (EBIC) (Brooks, 2002), i.e.:

$$EAIC = \overline{D} + 2p, \qquad EBIC = \overline{D} + p\log(n),$$

where $p$ is the number of the model parameters and $n$ is the sample size. When dealing with mixture models, the values of such criteria are not implemented by default in BUGS. Nevertheless, they can be easily computed from the MCMC output.

# 5 Simulation Studies

To evaluate the performance of the FBR model we set up some simulation studies. In particular, we evaluated the fitting ability of the FBR model (comparing it with the BRR and beta regression models) under various scenarios, i.e. data coming from beta, BRR, FBR and BMR models (Section 5.1). Moreover, we investigated robustness of the FBR under several patterns of heavy tails (Section 5.2), and of outliers (Section 5.3).

We did not include the BMR model in our comparisons since the estimation algorithm does not appear to be sufficiently reliable when used to perform many replications (unsatisfying convergence diagnostic, bimodal posteriors, etc.) even though identifiability constraints on $p$ have been imposed. In all studies, we set thinning intervals such that the dependence factor of the Raftery–Lewis diagnostic is near 1, thus implying independently generated values (see Section III of the Supplementary Material for more details).

## 5.1 Fitting Study

Here we study the behavior of the FBR when data are generated from the following five scenarios: (i) a FBR with two well separated components; (ii) a FBR with two overlapping components; (iii) a beta regression; (iv) a BRR which is not FBR; (v) a BMR which is neither FBR nor BRR.

For each scenario, we simulated a dataset of length $n = 100$ and replicated it $N = 100$ times. More specifically, we simulated a vector $x$ of length $n$ of i.i.d. observations from a uniform distribution on (-0.5,0.5) and we set $logit(\mu_i) = \beta_0 + \beta_1 x_i$, for each $i = 1, \ldots, n$, for fixed $\beta_0$ and $\beta_1$. Complete details on the five scenarios, together with their graphical illustration by means of the scatter plots of $x$ and $y$, can be found in the Supplementary Material, Section IV.

We estimated the vector of parameters in the FBR model, $\boldsymbol{\eta} = (\beta_0, \beta_1, \phi, p, w)$, as described in Section 4. In addition, for each simulated dataset in each scenario, we fitted a BRR and a beta regression model. In Table 1 we reported the mean values of the estimated DIC, EAIC and EBIC in each scenario. We also added, in parenthesis, the % of times the FBR ended up being the preferred model in relation to the beta and the BRR ones, respectively, for each criteria. We can observe that, on average, the FBR model is preferred in all scenarios and for all criteria. In particular, we can see that in the first two scenarios the FBR model exhibits a significantly better fit. In scenario III, with beta distributed response, the FBR model guarantees the best fit most of the time.

| Scenario | Model | DIC | EAIC | EBIC |
|---|---|---|---|---|
| | FBR | -372.4273 | -366.8956 | -353.8698 |
| I | BRR | -219.0147 (91.4) | -170.1883 (99.0) | -159.7676 (99.0) |
| | Beta | -112.4524 (100.0) | -109.4642 (100.0) | -101.6487 (100.0) |
| | FBR | -241.2042 | -234.2879 | -221.262 |
| II | BRR | -82.2750 (100.0) | -76.70478 (100.0) | -66.2841 (100.0) |
| | Beta | -82.0041 (100.0) | -79.0067 (100.0) | -71.19119 (100.0) |
| | FBR | -258.9465 | -214.5167 | -201.4907 |
| III | BRR | -204.2782 (100.0) | -199.1029 (99.0) | -188.6823 (98.2) |
| | Beta | -203.3696 (100.0) | -200.4014 (100.0) | -192.5859 (97.2) |
| | FBR | -108.0562 | -91.70023 | -78.67438 |
| IV | BRR | -97.07001 (82.0) | -87.68871 (67.0) | -77.26803 (61.0) |
| | Beta | -37.5946 (100.0) | -34.60769 (100.0) | -26.79218 (100.0) |
| | FBR | -252.3715 | -246.3384 | -233.3125 |
| V | BRR | -150.5611 (99.0) | -133.7362 (100.0) | -123.3155 (100.0) |
| | Beta | -18.92058 (100.0) | -15.95819 (100.0) | -8.142679 (100.0) |

Table 1: Mean values of the comparison Criteria DIC, EAIC and EBIC (% of selection of the FBR model versus the BRR or the beta regression models in parenthesis), in the five scenarios.

In scenario IV, the beta model has by far the worst fit. On the contrary, the BRR and FBR models seem to have a good performance, the FBR often guaranteeing a better fit. In the last scenario the FBR model is by far the preferred one.

The bias (and Mean Squared Error (MSE) in parenthesis) for the estimates of the regression parameters $\beta_0$ and $\beta_1$ are shown in Table 2. In the first two scenarios the FBR model guarantees less biased estimates with significantly lower MSEs, especially for the more relevant parameter $\beta_1$. In scenario III (beta distributed response) all models provide accurate estimates. In scenario IV, the beta model performs poorly, unlike the other models. Although $\beta_1$ is better estimated by the BRR model, the FBR displays an overall satisfactory behavior. In the last scenario the FBR shows a better performance especially with respect to $\beta_1$.

## 5.2 Heavy Tails Study

To test the behavior of the FBR in the presence of heavy tails, we considered the following five scenarios. The first two replicate a simulation study proposed by Bayes et al. (2012), where the BR model performs better than the beta. Specifically, we simulated samples of size $n = 100$ from a beta distribution of parameters $\bar{\alpha} = 0.2$ and $\alpha^+$ respectively equal to 10 and 30. Then, we sampled without replacement 5% of observations, and we replaced them by values generated from a uniform distribution in $(q, 1)$, where $q$ is the 0.999 quantile of the simulated beta distribution (right tail contamination).

The third scenario is identical to second one but with a different value of $\bar{\alpha}$, i.e. 0.6. As a fourth scenario, we defined a contaminated beta with a left heavy tail. We choose

| Scenario | Model | $\beta_0$ | $\beta_1$ |
|---|---|---|---|
| I | FBR | -0.0011 (0.0110) | 0.0212 (0.0172) |
| | BRR | -0.1119(0.0455) | -0.4514 (0.4446) |
| | Beta | -0.0934 (0.0209) | -0.2960 (0.2055) |
| II | FBR | -0.0032 (0.0049) | 0.0161 (0.0344) |
| | BRR | -0.0572 (0.0101) | -0.2593 (0.1319) |
| | Beta | -0.0441 (0.0083) | -0.1804 (0.0966) |
| III | FBR | -0.0008 (0.0013) | -0.0044 (0.0177) |
| | BRR | -0.0093 (0.0014) | -0.0160 (0.0173) |
| | Beta | -0.0006 (0.0013) | -0.0005 (0.0176) |
| IV | FBR | -0.1017 (0.0197) | -0.2946 (0.1451) |
| | BRR | -0.0638 (0.0126) | -0.1848 (0.1037) |
| | Beta | -0.1626 (0.0399) | -0.3821 (0.2507) |
| V | FBR | -0.0472 (0.0167) | 0.0320 (0.0243) |
| | BRR | 0.0841 (0.0399) | -0.5769 (0.4132) |
| | Beta | -0.0325 (0.0122) | -0.1138 (0.1105) |

Table 2: Bias (and Mean Square Error (MSE) in parenthesis) of the estimates of the regression parameters in the five scenarios.

| Scenario | Model | DIC | EAIC | EBIC |
|---|---|---|---|---|
| I | FB | -161.4934 | -148.7277 | -138.307 |
| | BR | -149.3861 (86.0) | -144.4894 (79.0) | -136.6738 (61.0) |
| II | FB | -229.523 | -224.0233 | -213.6026 |
| | BR | -231.7928 (43.0) | -227.3423 (41.0) | -219.5268 (31.0) |
| III | FB | -201.9912 | -194.144 | -183.7233 |
| | BR | -185.1471 (99.0) | -180.2845 (97.0) | -172.469 (93.0) |
| IV | FB | -221.451 | -213.9913 | -203.5706 |
| | BR | -185.8266 (100.0) | -180.1325 (100.0) | -172.317 (100.0) |
| V | FB | -114.75 | -96.60665 | -90.16298 |
| | BR | -73.23275 (99.0) | -68.87421 (97.0) | -64.04146 (92.0) |

Table 3: Mean values of the comparison Criteria DIC, EAIC and EBIC (% of selection of the FB model versus the BR model in parenthesis) in the five heavy tail scenarios.

$\bar{\alpha} = 0.3$ and $\alpha^+ = 30$, and a contamination percentage equal to 8%. The contaminating values were generated from a uniform distribution in $(0, q)$, where $q$ is the 0.001 quantile of the simulated beta distribution. In the last scenario we kept the same parameters of scenario IV with the exception of $\alpha^+ = 5$.

Table 3 reports the values of the comparison criteria. The BR performs slightly better in the second scenario, while the FB fits better in the four remaining ones. By looking at the histograms corresponding to the various scenarios, it appears that the BR displays some fitting difficulties when there is only one heavy tail and the remaining one tends to zero. This is due to the fact that BR densities display the same values on the two tails, unlike the FB ones.

## 5.3 Robustness Study with Covariates

We now explore sensitivity to outliers when covariates are taken into account. Following Bayes et al. (2012), we simulated values from a beta regression model $Y_i \sim B(logit(\bar{\alpha}_i), \alpha^+ = 30)$, given $logit(\bar{\alpha}_i) = \beta_0 + \beta_1 x_i$, $\beta_0 = 0.5$, $\beta_1 = 1$ and $x_i$ generated from a Uniform distribution on $(-3; 3)$.

Then, we contaminated the dataset by decreasing/increasing a given percentage of the $y$ values according to the following four patterns: (i) decrease 0.7 units of the response values associated to high $x$ values, (ii) increase 0.7 units of the response values associated to low $x$ values, (iii) decrease 0.5 units of the response values for central $x$ values, (iv) decrease and increase 0.5 units of the response values both for high and low $x$ values. In the first three scenarios we contaminated 5% of the simulated observations, while in the fourth scenario 6%, as shown in the scatterplots reported in Figure 3.

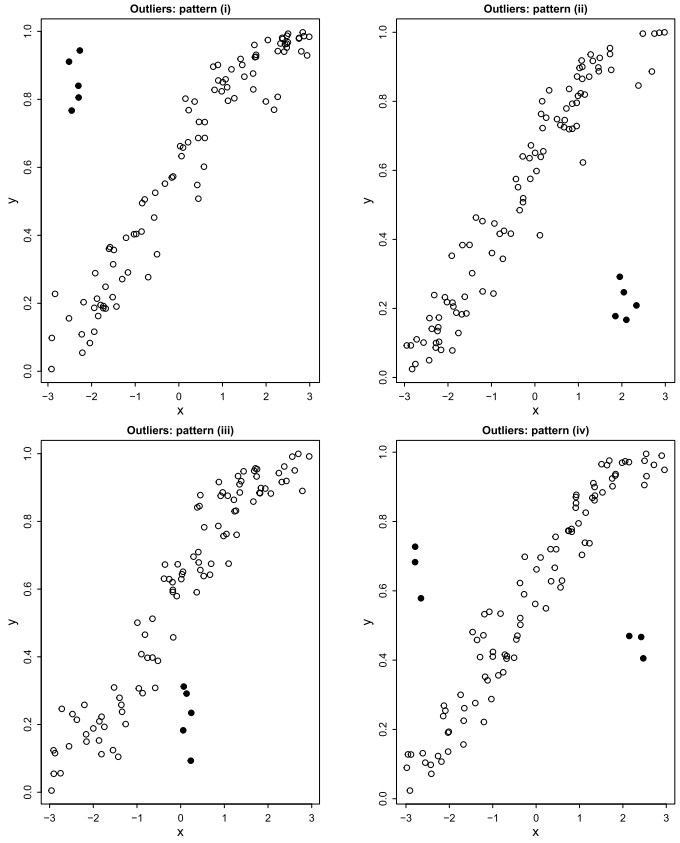


Figure 3: Generation of contaminated datasets under four different patterns.

In Table 4 we reported the mean values of the estimated DIC, EAIC and EBIC in each scenario. We also added, in parenthesis, the % of times the FBR ended up being the

| Scenario | Model | DIC | EAIC | EBIC |
|---|---|---|---|---|
| | FBR | -282.2429 | -276.1177 | -263.0918 |
| I | BRR | -264.4611 (99.0) | -259.154 (99.0) | -248.7333 (99.0) |
| | Beta | -120.3277 (100.0) | -117.3498 (100.0) | -109.5342 (100.0) |
| | FBR | -281.6261 | -275.7628 | -262.737 |
| II | BRR | -265.9097 (100.0) | -260.5964 (100.0) | -250.1757 (98.0) |
| | Beta | -113.0299 (100.0) | -110.0527 (100.0) | -102.2372 (100.0) |
| | FBR | -295.9161 | -282.4695 | -269.4437 |
| III | BRR | -267.9762 (99.0) | -262.1149 (96.0) | -251.6942 (94.0) |
| | Beta | -233.4922 (100.0) | -230.511 (100.0) | -222.6955 (100.0) |
| | FBR | -222.5819 | -202.8648 | -189.839 |
| IV | BRR | -261.5993 (10.0) | -256.2259 (0.0) | -245.8052 (0.0) |
| | Beta | -166.3029 (100.0) | -163.3154 (100.0) | -155.4999 (100.0) |

Table 4: Mean values of the comparison Criteria DIC, EAIC and EBIC (% of selection of the FBR model versus the BRR or the beta regression models in parenthesis), in the four scenarios.

| Scenario | Model | $\beta_0$ | $\beta_1$ |
|---|---|---|---|
| | FBR | 0.0916 (0.0105) | -0.0674 (0.0058) |
| I | BRR | -0.0353 (0.0031) | -0.1020 (0.0115) |
| | Beta | -0.0121 (0.0013) | -0.3363 (0.11389) |
| | FBR | -0.1325 (0.0201) | -0.0868 (0.0089) |
| II | BRR | -0.0532 (0.0047) | -0.1063 (0.0124) |
| | Beta | -0.2138 (0.0473) | -0.3076 (0.0953) |
| | FBR | -0.1404 (0.0220) | -0.0698 (0.0064) |
| III | BRR | -0.0522 (0.0053) | -0.0676 (0.0055) |
| | Beta | -0.1268 (0.0180) | -0.0558 (0.0041) |
| | FBR | -0.0332 (0.0036) | -0.1897 (0.0373) |
| IV | BRR | -0.0453 (0.0033) | -0.1157 (0.0150) |
| | Beta | -0.0252 (0.0018) | -0.2216 (0.0500) |

Table 5: Bias (and MSE in parenthesis) of the estimates of the regression parameters in the four scenarios.

preferred model in relation to the beta and the BRR ones, respectively, for each criteria. The FBR performs better than the competing models in the first three scenarios both in terms of fit of the model and in terms of robustness of the estimates of the regression parameters. Conversely, in the forth scenario the BRR is the preferred model, although the FBR parameter estimates are reasonably accurate (see Table 5).

It is noteworthy that the above beta regression model originally chosen by Bayes et al. (2012) to ascertain robustness gives rise to values of the dependent variable ranging from 0 to 1 and centered in 1/2. This is in agreement with the BRR model, where one of the two mean regression components (i.e. the uniform one) is centered in 1/2. Therefore we further explored a framework withe response values higher than 1/2. This is simply obtained by changing the support of the covariate which is now generated

from a Uniform distribution on $(0, 3)$. In such a case the FBR model displays a better fitting than the BRR one even in the forth contamination scenario.

# 6 Applications

In this section we focus on two well-known datasets, already analyzed in the literature, to show how the FBR can be successfully applied, comparing it with the two competing models, namely the BRR and the beta regression ones. At the end of each subsection a comparison with the BMR model in terms of fitting criteria is also provided.

## 6.1 Reading Accuracy Data

As a first application, we consider a dataset about the reading accuracy performance, quantified as a proportion on $(0, 1)$, of a group of 44 children, 19 of which have been diagnosed with dyslexia (Pammer and Kevan, 2007). As covariate a quantitative measure of the children's non verbal abilities (IQ) is available. Smithson and Verkuilen (2006) have already illustrated how the adoption of standard regression models with response variables restricted on $(0, 1)$ can produce misleading results. In addition, Ferrari et al. (2011) applied the beta regression to this dataset.

In order to convey a more thorough description of the models' performances, we consider two cases, with and without information about the children's dyslexic status.

**Regression model for the mean – one explanatory variable**

At first we estimate the models including only the covariate $X_1$ about the non-verbal IQ converted to $z$-scores. The FBR model for the reading accuracy level, $Y_i \sim FB(\mu_i, \phi, p, w)$ for $i = 1, \ldots, n$ is defined as:

$$logit(\mu_i) = \beta_0 + \beta_1 X_{i1},$$

where $\beta_0$ and $\beta_1$ are the unknown regression parameters.

To compare the FBR model with the beta and with the BRR ones, we simulated MCMCs of length 5000, discarded the first half values, and used a thinning interval set equal to 1 for the beta model, to 3 for the BRR model and to 5 for FBR model. These values satisfy the various diagnostic tests mentioned in Section 4.

The results are shown in Table 6, in Table 7 and in Figure 4.

| Parameter | $\beta_0$ | $\beta_1$ | $\phi$ | $p$ | $w$ | $\alpha$ |
|---|---|---|---|---|---|---|
| FBR | 1.258 | 0.053 | 35.609 | 0.410 | 0.861 | |
| BRR | 1.32 | 0.593 | 4.759 | | | 0.127 |
| Beta | 1.390 | 0.608 | 4.629 | | | |

Table 6: Posterior mean of the parameters under the FBR, the BRR and the beta regression models.

| Criterion | DIC | EAIC | EBIC |
|---|---|---|---|
| FBR | -145.097 | -138.655 | -129.734 |
| BRR | -62.863 | -57.724 | -50.587 |
| Beta | -63.173 | -60.169 | -54.816 |

Table 7: Model comparison criteria for the FBR, the BRR and the beta regression models.

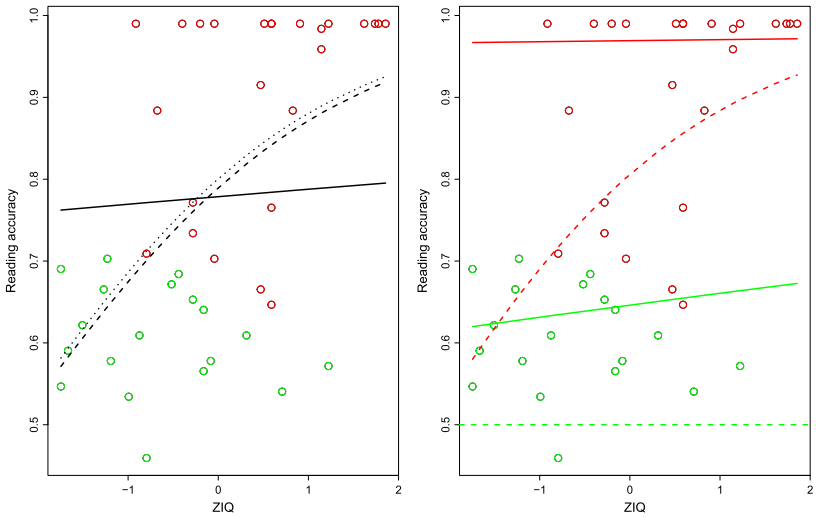


Figure 4: Left panel: fitted regression curves for the mean of the response variable under the FBR (solid), the BRR (dashed) and the beta regression (dotted) models. Right panel: FBR group means $\lambda_{1i}$ and $\lambda_{2i}$ (solid) and BRR group means (dashed) (1/2 for the uniform component). Green refers to dyslexics and red to controls.

The FBR exhibits a striking better fit than the other two models, which have similar performances in this example, including very close estimates of common parameters. Of particular relevance is the difference in the estimates of $\beta_1$: the FBR model suggests a very little influence of the covariate (in agreement with data pattern), whereas the other two models indicate a relatively strong positive dependence. The difference in the precision parameter estimates also provides indication of a better fit of the FBR.

To better grasp the behavior of the models, the right panel in Figure 4 reports the group regression means of the FBR and the BRR models. It is evident the better ability of the FBR to describe the presence of two groups which correspond to the dyslexic/non-dyslexic children, with the exception of seven outliers in the group of the non-dyslexic displaying unusually low values very close to the dyslexics' ones.

Lastly, we simulated draws from the posterior predictive distributions (Albert, 2009) of the three models. The posterior predictive means and the 90% posterior predictive intervals (Figure 5) confirm the better ability of the FBR model in predicting the observed data and in correctly classifying them into two clusters.

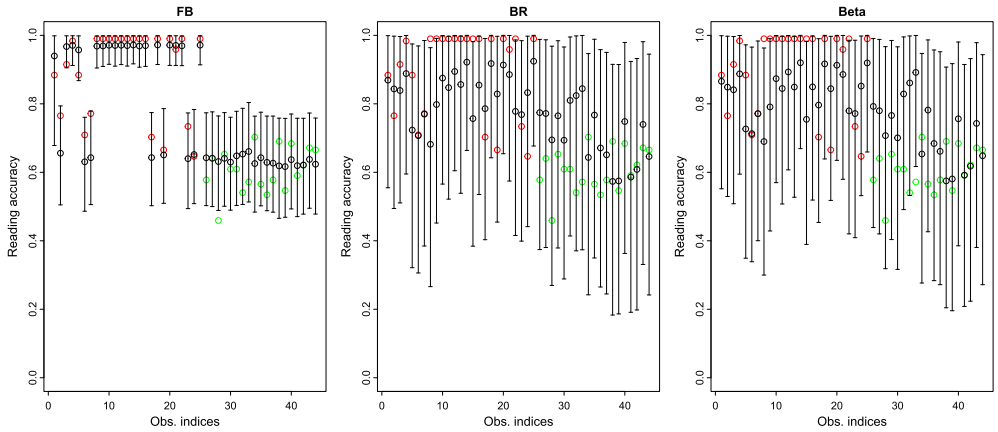


Figure 5: Posterior predictive means and 90% predictive intervals for each subject (the observed values of reading accuracy are reported in green for the group of dyslexic children, in red for the controls).

**Regression model for the mean – two explanatory variables**

As a second step, we considered a more complicated FBR model for the mean by adding the dichotomous covariate $X_2$ about being dyslexic (value 1) or not (value 0), and the interaction between the latter and $X_1$, i.e. $X_3 = X_1 X_2$:

$$logit(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}, \quad i = 1, \ldots, n.$$

For space constraints, here we only report the main results of the analysis, a complete description being available in the Supplementary Material, Section V. In this case too there is a huge difference in the fitting performance of the FBR with respect to the other two models which, again, have a very similar general behavior. In particular, the FBR model provides a better fit of the control group. This is most evident from the right panel of Figure 6 where the FBR model's flexibility is used to accurately distinguish the main body of the control group from the outliers, i.e. the children with response values similar to the dyslexic ones. The BRR model cannot fit these outliers equally well, as it devotes to them a regression curve constantly fixed at 0.5. The overall regression mean relative to the dyslexic group of the beta and BRR models fits better than the FBR one (left panel of Figure 6). However, the right panel of Figure 6 shows that in the FBR model data are well fitted by the regression curve relative to the most relevant sub-group ($p = 0.828$).

Indeed, the 90% posterior predictive intervals of the FBR are the smallest ones (see Figure 3 of the Supplementary Material). In particular, it is noteworthy that the seven outliers in the control group are more precisely predicted displaying shorter and better centered intervals. This confirms the potential of the FBR model when outliers are present.

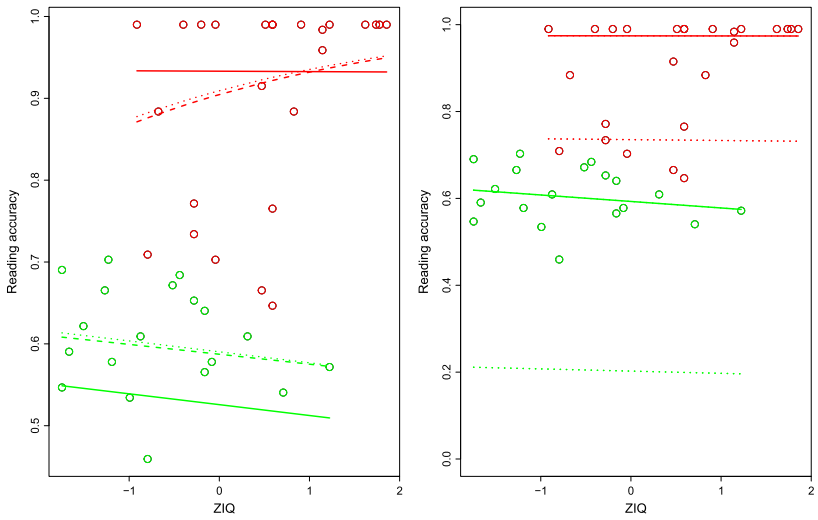


Figure 6: Left panel: regression curves for the group of dyslexics (in green) and for the control group (in red), for the beta regression model (dotted), the BRR model (dashed) and the FBR model (solid). Right panel: FBR curves for the two sub-groups ($\lambda_{1i}$ solid and $\lambda_{2i}$ dotted) relative to the dyslexics (in green) and to the controls (in red).

To further improve the fit of the FBR model one can let the parameter $w$ (or $p$) depend on covariates, namely $X_2$. We therefore implemented the above model, but writing $w$ as:

$$logit(w_i) = \gamma_0 + \gamma_1 X_{i2}, \quad i = 1, \ldots, n.$$

Although the goodness-of-fit criteria do not change significantly, the fit of the dyslexic group greatly improves. Indeed, the corresponding two sub-group regression means are now identical (estimated $w$ almost equal to 0) and well adapting to observed data, whereas the regression curves for the controls are essentially unchanged.

Finally, we fitted the BMR model described in Section 3.2 both for the one and for the two explanatory variable cases. In the former we found a quite better performance of the BMR, whereas in the latter the BMR and of the FBR models are comparable, and clearly outperform the beta regression and the BRR ones. We believe that the superiority of the BMR model in the first case is essentially due to the different variability of the two groups as identified by the models. Indeed, the group with higher values of the response displays an exceptionally low variability.

## 6.2 Australian Institute of Sport Data

The aim of this second application is a comparison between the BRR and the FBR models on the dataset used by Bayes et al. (2012) to demonstrate robustness properties of their BRR model. This dataset (Australian Institute of Sport) is included in the R

library sn. A subset of 37 rowing athletes has been selected, specifying the body fat percentage as the response variable *Bfat* restricted to $(0, 1)$, and the lean body mass (*lbm*) as a quantitative covariate. We observe the presence of two outliers (filled-in circles corresponding to subjects $16^{th}$ and $30^{th}$ in Figure 7).

We provide results both on the entire dataset and after removing the outliers, with the purpose of evaluating the robustness of the FBR model and of comparing it with the two competing models. For the scope of this application, we defined the FBR model for the response variable *Bfat*, $Y_i \sim FB(\mu_i, \phi, p, w)$ for $i = 1, \ldots, n$, by specifying:

$$logit(\mu_i) = \beta_0 + \beta_1 X_{i1},$$

where $X_1$ is the *lbm* covariate, $\beta_0$ and $\beta_1$ are the regression parameters.

To estimate the relevant parameters we performed a Gibbs sampling, as described in Section 4. For all the models we generated 5000 values of the Markov Chain, varying the thinning interval among 1 and 22 to achieve convergence. The parameter posterior means concerning the three models are reported in Table 8. All the models produce

| | Parameters | $\beta_0$ | $\beta_1$ | $\phi$ | $p$ | $w$ | $\alpha$ |
|---|---|---|---|---|---|---|---|
| All obs. | FBR | 0.743 | -0.037 | 227.099 | 0.917 | 0.498 | |
| | BRR | 0.908 | -0.037 | 231.665 | | | 0.192 |
| | Beta | 0.095 | -0.027 | 91.397 | | | |
| Without outliers | FBR | 0.893 | -0.039 | 977.242 | 0.446 | 0.117 | |
| | BRR | 0.855 | -0.037 | 233.088 | | | 0.072 |
| | Beta | 0.832 | -0.038 | 233.053 | | | |

Table 8: Posterior parameter means for the FBR, the BRR and the beta regression models.

similar regression coefficient estimates when outliers are absent, whereas when they are taken into consideration the estimates (especially of the more relevant $\beta_1$) appear to be rather stable only for the FBR and the BRR model.

A much better understanding of the implications of such estimates is obtained by inspection of the graph with the corresponding regression curves (see Figure 7, left panel). When outliers are absent (thick curves) the three models provide almost overlapping curves. When outliers are introduced, as expected, the beta regression curve substantially alters its inclination. Conversely, the BRR and especially the FBR regression curves are subject to little changes, thus demonstrating their robustness. Interestingly, these changes have opposite directions. This behavior can be explained by looking at the group regression curves (right panel of Figure 7). The two models share a nearly identical curve modeling the most relevant group. To such a curve the FBR adds a second curve which accurately fits the two outliers, whereas the BRR model adds a curve constantly equal to 0.5, which is far above the whole dataset.

The result is that the BRR overall regression curve (left panel) appears not to be well centered, lying to the right of most data points. This slightly worse behavior is confirmed by the comparison criteria values displayed in Table 9. Note also the outstanding superior fitting of the FBR in the case without outliers.

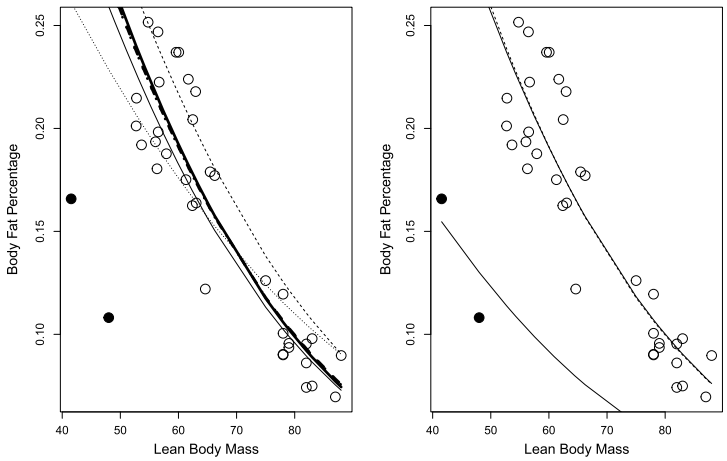


Figure 7: Left panel: fitted regression curves for the mean of *Bfat* under the beta model (dotted), the BRR (dashed) and the FBR (solid); thick curves represent the situation without outliers, thin ones with outliers. Right panel: group means for the FBR (solid) and for the BRR (dashed; the uniform component equal to 1/2 lies outside the scatter plot).

| Obs. | Model | DIC | EAIC | EBIC |
|---|---|---|---|---|
| All obs. | FBR | -169.125 | -162.358 | -154.303 |
| | BRR | -159.292 | -154.116 | -147.672 |
| | Beta | -136.371 | -133.439 | -128.606 |
| Without outliers | FBR | -218.719 | -202.458 | -194.681 |
| | BRR | -160.637 | -155.612 | -149.390 |
| | Beta | -160.997 | -158.064 | -153.398 |

Table 9: Model comparison criteria for the FBR, the BRR and the beta regression models.

With respect to the models fitted on the entire dataset, we analyzed the Bayesian residuals (Gelman et al., 2014). Results are discussed in the Supplementary Material, Section V. It is shown that only the FBR and beta regression models clearly identify the two outliers.

We also sampled from the posterior predictive distribution and we plotted the observed values for *Bfat* against the predicted values for each model (Figure 8). One can see that in the beta regression model the two outliers fall outside 90% predictive interval. The BRR model seems capable of modeling all the sample values. However, the two outliers are modeled with the uniform, thus producing 90% predictive intervals too wide to be informative. Conversely, the FBR provides accurate predictive intervals for all sample values.

Finally, we fitted the BMR model described in Section 3.2 to both situations. When all observations are considered, the values of the criteria were -167.7623 (DIC), -157.1428

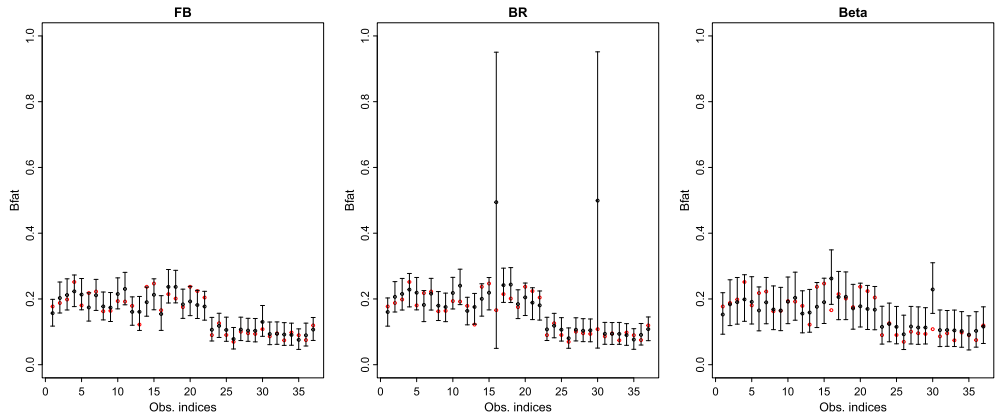


Figure 8: Posterior predictive means and 90% predictive intervals for each subject (in red the observed values of *Bfat*).

(EAIC) and -146.5244 (EBIC), showing a slightly worse performance than the FBR, but better than the beta and the BRR. As for the case without outliers, the estimation algorithm for the BMR model failed to converge.

# 7 Concluding Remarks

When modeling responses bounded to the unit interval, the FBR model turns out to be a good compromise between tractability and flexibility. Indeed, our results show that it greatly expands the modeling potential of the beta, without demanding the theoretical and computational intricacy of a general beta mixture.

In particular, the special mixture structure defining the FBR ensures good theoretical properties (strong identifiability and likelihood boundedness) uncommon for mixture models (not even possessed by the simple mixture giving the BRR). In turn, this leads to computational tractability in terms of convergence rate and stability of the MCMC.

At the same time, the FBR model displays easiness of interpretation as well as a remarkable fitting capacity for a large variety of data patterns. As expected, the FBR model outperforms the beta regression and the BRR models in presence of (even weak) bimodality, as shown both by the simulation studies and by the first real data application. More surprisingly, even when data are simulated from a beta or a BRR model, or more generally are unimodal (as in the second real data application once the outliers have been removed) the FBR model shows a comparable or, more often, better performance, especially in terms of fitting criteria. A similar behavior is observed in presence of outliers, where both the FBR and the BRR are by far preferable to the beta regression model. In addition, the presented real data applications show that the FBR provides a more accurate interpretation of data patterns. This behavior is confirmed by the posterior predictive intervals, which are typically shorter and better centered for the FBR model.

The reasons of the generally better behavior of the FBR seem to lie in its ability to remove two important limitations of the BRR model: (1) the latter requires the density in the two tails to have exactly the same behavior and (2) it devotes to one of the two groups a fixed regression curve identically equal to 1/2, whereas both regression curves can suitably adapt to data in the FBR. Indeed, the only case, among the studied ones, where the BRR performs significantly better is when outliers are introduced both below and above the main regression and their values are not far from 1/2.

For comparison purposes, we also defined a general BMR model and applied it to the two real data sets. From our limited experience the BMR seems to clearly outperform the FBR when two well distinct groups are present with highly different variability, as in the reading accuracy case with one explanatory variable. In the other cases results are comparable or the FBR is preferable. Furthermore, the rate of convergence of the BMR estimation algorithm is often very slow, and sometimes the algorithm fails to converge (as in the sport dataset without outliers) even though identifiability constraints on $p$ have been imposed. This computational difficulty prevented us to conduct a simulation study to systematically compare our model with the BMR.

From a computational perspective, the use of WinBUGS makes the implementation of the model fairly easily accessible for practitioners. However, it may not be the most efficient choice. An interesting alternative to the BUGS software is the more recent STAN software (Stan Development Team, 2016). The latter, in our limited experience, ensures shorter runtimes and, sometimes, better convergence characteristics. We plan to better investigate STAN capability specially in more complicated contexts, such as the general BMR model and its possible extensions to multivariate response variables.

Finally, it seems worthwhile to stress the possibility that some of the parameters $\phi$, $p$, and $w$ of the FBR model may depend on covariates too. This greatly further expands its flexibility potential, enabling it to model more complex data patterns, as also illustrated in the application at the end of Section 6.1.

## Supplementary Material

Supplementary Material for A New Regression Model for Bounded Responses (DOI: 10.1214/17-BA1079SUPP; .pdf). The online supplementary material contains proofs of the Propositions 1 and 2, sensible recommendations about the choice of priors for $\phi$, a list of the thinning intervals adopted as to guarantee Raftery–Lewis diagnostics in most cases around 1. Furthermore, it includes a detailed description of the Simulation scenarios of Section 5.1, further results for the regression model for the mean with two explanatory variables for the reading accuracy dataset (Section 6.1) and a detailed analysis of residuals for sport data of Section 6.2.

## References

Akaike, H. (1998). "Information theory and an extension of the maximum likelihood principle." In *Selected Papers of Hirotugu Akaike*, 199–213. Springer. MR1486823. 856

Albert, J. (2009). *Bayesian computation with R*. Springer Science & Business Media. MR2839312. doi: https://doi.org/10.1007/978-0-387-92298-0. 850, 855, 863

Bayes, C. L., Bazàn, J. L., and Garcìa, C. (2012). "A new robust regression model for proportions." *Bayesian Analysis*, 7(4): 841–866. 845, 846, 849, 851, 858, 860, 861, 865

Branscum, A. J., Johnson, W. O., and Thurmond, M. C. (2007). "Bayesian Beta Regression: Applications to Household Expenditure Data and Genetic Distance between Foot-and-Mouth Disease Viruses." *Australian & New Zealand Journal of Statistics*, 49(3): 287–301. MR2405396. doi: https://doi.org/10.1111/j.1467-842X.2007.00481.x. 846, 856

Brooks, S. (2002). "Discussion on the paper by Spiegelhalter, Best, Carlin and Van Der Linde." 856, 857

Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., et al. (2006). "Deviance information criteria for missing data models." *Bayesian Analysis*, 1(4): 651–673. 856

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society. Series B*, 39: 1–38. 855

Ferrari, S. and Cribari-Neto, F. (2004). "Beta regression for modelling rates and proportions." *Journal of Applied Statistics*, 31(7): 799–815. 845, 847, 851

Ferrari, S. L., Espinheira, P. L., and Cribari-Neto, F. (2011). "Diagnostic tools in beta regression with varying dispersion." *Statistica Neerlandica*, 65(3): 337–351. MR2857878. doi: https://doi.org/10.1111/j.1467-9574.2011.00488.x. 846, 851, 862

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media. 847, 852

García, C., Pérez, J. G., and van Dorp, J. R. (2011). "Modeling heavy-tailed, skewed and peaked uncertainty phenomena with bounded support." *Statistical Methods & Applications*, 20(4): 463–486. 846

Gelfand, A. E. and Smith, A. F. (1990). "Sampling-based approaches to calculating marginal densities." *Journal of the American Statistical Association*, 85(410): 398–409. 855

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis. 846, 867

Hahn, E. D. (2008). "Mixture densities for project management activity times: A robust approach to {PERT}." *European Journal of Operational Research*, 188(2): 450–459. 846, 849

Kieschnick, R. and McCullough, B. D. (2003). "Regression analysis of variates observed on (0, 1): percentages, proportions and fractions." *Statistical Modelling*, 3(3): 193–213. 845

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). "WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility." *Statistics and Computing*, 10(4): 325–337. 856

Markatou, M. (2000). "Mixture models, robustness, and the weighted likelihood methodology." *Biometrics*, 56: 483–486. 846

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press. 846, 851

Mengersen, K. L., Robert, C. P., and Guihenneuc-Jouyaux, C. (1999). "MCMC convergence diagnostics: a review." *Bayesian Statistics*, 6: 415–440. 856

Migliorati,S., Di Brisco, A. M., and Ongaro, A. (2017). "Supplementary Material for A New Regression Model for Bounded Responses". *Bayesian Analysis*. doi: https://doi.org/10.1214/17-BA1079SUPP. 851

Migliorati, S., Ongaro, A., and Monti, G. S. (2016). "A structured Dirichlet mixture model for compositional data: inferential and applicative issues." *Statistics and Computing*, 1–21. 848, 849

Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS*, volume 698. John Wiley & Sons. 856

Ongaro, A. and Migliorati, S. (2013). "A generalization of the Dirichlet distribution." *Journal of Multivariate Analysis*, 114: 412–426. 846, 848

Pammer, K. and Kevan, A. (2007). "The contribution of visual sensitivity, phonological processing, and nonverbal IQ to children's reading." *Scientific Studies of Reading*, 11(1): 33–53. 862

Paolino, P. (2001). "Maximum likelihood estimation of models with beta-distributed dependent variables." *Political Analysis*, 9(4): 325–346. 845, 846, 847, 851

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org/. 856

Schwarz, G. (1978). "Estimating the dimension of a model." *The Annals of Statistics*, 6(2): 461–464. 856

Smithson, M. and Verkuilen, J. (2006). "A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables." *Psychological Methods*, 11(1): 54. 846, 851, 862

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit (with discussion)." *Journal of the Royal Statistical Society: Series B*, 64(4): 583–639. 856

Stan Development Team (2016). *Stan Modeling Language Users Guide and Reference Manual*. URL http://mc-stan.org/. 869

Tanner, M. A. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association*, 82(398): 528–540. 855

Thomas, A. (1994). "BUGS: A statistical modelling package." *RTA/BCS Modular Languages Newsletter*, 2: 36–38. 856

**Acknowledgments**

We are grateful to the referee and to an associate editor for their constructive comments, which greatly helped to improve the paper. Research partially financially supported by the Italian Ministry of University and Research, grants F.A. 2016 from the University of Milano-Bicocca.