

# Multivariate Beta Regression with Application in Small Area Estimation

Debora F. Souza<sup>1</sup> and Fernando A. S. Moura<sup>2</sup>

Multivariate beta regression models for jointly modelling two or more variables whose values belong in the  $(0,1)$  interval, such as indexes, rates or proportions, are proposed for making small area predictions. The multivariate model can help the estimation process by borrowing strength between units and obtaining more precise estimates, especially for small samples. Each response variable is assumed to have a beta distribution so the models could accommodate multivariate asymmetric data. Copula functions are used to construct the joint distribution of the dependent variables; all the marginal distributions are fixed as beta. A hierarchical beta regression model is additionally proposed with correlated random effects. We present an illustration of the proposed approach by estimating two indexes of educational attainment at school level in a Brazilian state. Our predictions are compared with separate univariate beta regressions. The inference process was conducted using a full Bayesian approach.

*Key words:* Bayesian inference; copula function; small domain; education evaluation.

## 1. Introduction

In recent years, numerous applications of the beta distribution have been developed due to the distribution's suitability for modelling rates or proportions. Its properties include being defined on the range  $(0,1)$ , allowing for asymmetry present in these types of variables, and assuming different forms depending on its parameters. The beta regression additionally allows heteroscedastic observations.

Ferrari and Cribari-Neto (2004) proposed a univariate beta regression for modelling rates or proportions and used a classic approach to estimate the model parameters. A Bayesian version of the static beta regression was proposed by Branscum et al. (2007). More recently, Da-Silva et al. (2011) proposed a method for beta time series data, in which the model parameters that are related to the means follow a dynamic model. However, the most frequently proposed use of the beta distribution in the context of regression has been restricted to cases where there is only one dependent variable.

<sup>1</sup> Coordenação de Métodos e Qualidade, Instituto Brasileiro de Geografia e Estatística (IBGE). Rio de Janeiro, Brazil. Email: debora.souza@ibge.gov.br

<sup>2</sup> IM-UFRJ – Statistics Department, Rio de Janeiro, Rio de Janeiro, Brazil. Email: fmoura@im.ufrj.br

**Acknowledgments:** This work is part of the PhD dissertation of Debora F. S., under the supervision of Fernando Moura, in the Graduate Program of Statistics of Universidade Federal do Rio de Janeiro (UFRJ). The authors would like to thank the associate editor and the anonymous referees for their very thoughtful and constructive comments.

We propose a new approach to jointly modelling indexes, rates or proportions, commonly estimated with low accuracy in small samples. Examples of variables that are measured in the range  $(0,1)$  and are related to each other are the proportion of poor people, the mortality rate and the ratio of food expenditures to total expenditures. The models proposed in this article can also be employed in estimating correlated poverty indexes for small domains. While the motivation for this work has been the estimation of rates or proportions in small areas (or domains), the strategy used to achieve this goal can be applied to a more general context. Multivariate models are developed for modelling rates or proportions, offering the possibility of jointly managing related quantities in one single model and enjoying the benefits that this joint approach offers. Borrowing strength across the response variables in the multivariate models proposed here can provide more precise estimates of the quantities of interest.

Cepeda-Cuervo et al. (2014) apply a bivariate strategy using the Farlie-Gumbel-Morgenstern (FGM) copula, modeling the dispersion parameter of the beta regression as proposed in Smithson and Verkuilen (2006) and Simas et al. (2010). However, their approach does not account for any hierarchical structure of the population and no extension to the multivariate case is discussed. Melo et al. (2009), Fabrizi et al. (2011) and Murteira and Ramalho (2014) propose and apply multivariate models for dealing with fractional data.

This article develops multivariate regression models where the dependent variables marginally follow a beta distribution. These models address data fitting in general contexts, and the models are especially advantageous for small area estimation. The beta marginal distributions were reparametrised by the mean and the dispersion, as in Ferrari and Cribari-Neto (2004). The associations between the response variables are considered as a copula function applied to the marginal densities. Copulas are useful tools for building multivariate distributions where the marginal distributions are given or known, allowing individual models be analysed together. Additionally, copula functions allow the representation of various types of dependence between variables. The use of copulas allows flexibility in handling nonlinear relationships between the response variables and is therefore a more general setup than the multivariate normal distribution, which allows only linear relationships. For a complete study on the copula function and its utilities in statistics, see Nelsen (2006).

Two types of multivariate models with beta responses are proposed: a beta regression model, where the marginal densities are connected by a copula function, and a hierarchical beta model with correlation between their means. In a small area estimation context where auxiliary variables and data from multiple characteristics are available, these models can improve the prediction of observations and target-population parameters. Several authors argue that this approach provides better estimates than fitting separate univariate models, because a multivariate model considers the correlations between the response variables after conditioning on the auxiliary variables. Fay (1987) modelled the joint behaviour of the median income in households of three, four and five dwellers. Datta et al. (1999) applied a multivariate mixed linear model and concluded from a simulation study that the multivariate approach provides better results than setting a separate model for each variable. The methods most commonly employed are based on borrowing information from neighbouring or related areas. The models proposed in this article have a direct

application to the small area estimation problem by additionally allowing strength to be borrowed between the response variables.

The article is organised as follows. In Section 2, we propose a multivariate beta regression model by employing copula functions. In Section 3, we apply our proposed models to the small area estimation problems, presenting an illustration with Brazilian education data. Section 4 offers some conclusions and suggestions for further research.

## 2. Multivariate Beta Regression Model Based on Copulas

The structure of dependence between two or more related response variables can be defined in terms of their joint distribution. One way of obtaining a multivariate beta distribution is to join the univariate beta using copula functions, which is one of the most useful tools when the marginal distributions are given or known. The use of copula functions enables the representation of various types of dependence between variables. In practice, this function implies a more flexible assumption about the form of the joint distribution than that given in Olkin and Liu (2003), which assumes that the marginal distributions have the same parameter. Nelsen (2006) defines a copula as a joint distribution function

$$C(u_1, \dots, u_K) = P(U_1 \leq u_1, \dots, U_K \leq u_K), \quad 0 \leq u_j \leq 1,$$

where  $U_j, j = 1, \dots, K$  are uniformly distributed on the interval  $(0,1)$ .

Sklar's theorem, stated here in Theorem 1, shows how to obtain a joint distribution using a copula.

**Theorem 1** *Let  $H$  be a  $K$ -dimensional distribution function with marginal distribution functions  $F_1, \dots, F_K$ . Then, there is a  $K$ -dimensional copula  $C$  such that for all  $(y_1, \dots, y_K) \in [-\infty, \infty]^K$ ,*

$$H(y_1, \dots, y_K) = C(F_1(y_1), \dots, F_K(y_K)). \tag{1}$$

*Conversely, if  $C$  is an  $n$ -dimensional copula and  $F_1, \dots, F_K$  are cumulative distribution functions, then the function  $H$  defined by (1) is a distribution function with marginal distributions  $F_1, \dots, F_K$ . Moreover, if all marginal distributions are continuous,  $C$  is unique. Otherwise, the copula  $C$  is uniquely determined in  $Im(F_1) \times \dots \times Im(F_K)$ , where  $Im(\cdot)$  represents the image of  $(\cdot)$ .*

Let  $\mathbf{y} = ((y_{11}, \dots, y_{1K}), \dots, (y_{n1}, \dots, y_{nK}))$  be a random sample of size  $n$  from a continuous joint distribution with marginal densities  $f_1, \dots, f_K$ . Thus, the likelihood function is given by:

$$L(\Psi) = \prod_{i=1}^n c(F_1(y_{i1}|\Psi), \dots, F_K(y_{iK}|\Psi))f_1(y_{i1}|\Psi) \dots f_K(y_{iK}|\Psi) \tag{2}$$

where  $\Psi$  denotes the set of parameters that define the distribution functions  $F_k$ , the densities  $f_k$ , and the copula-density function  $c(\cdot), k = 1, \dots, K$ .

In (2), we assume that each response variable  $k$  is beta distributed, such that:

$$Y_{ik} | \mu_{ik}, \phi_k \sim Beta(\mu_{ik}, \phi_k), \quad i = 1, \dots, n, \quad k = 1, \dots, K$$

$$g(\mu_{ik}) = \eta_{ik} = \sum_{j=1}^{p_k} x_{ij} \beta_{jk}$$

where  $Beta(\mu_{ik}, \phi_k)$  denotes that  $Y_{ik}$  is beta distributed with mean  $\mu_{ik}$  and variance  $\frac{\mu_{ik}(1-\mu_{ik})}{1+\phi_k}$ ,  $g(\cdot)$  is the link function and  $p_k$  is the number of covariates for the response variable  $k$ .

Denote by  $BetaM(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\theta})$  the multivariate beta distribution obtained by using  $K$  marginally beta-distributed variables with parameters  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)^T$  and  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)^T$  and a copula function with a vector of parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_L)^T$ . Thus, the structure of dependence between the  $K$  beta responses is defined by their joint distribution, which is obtained by applying a copula function, resulting in the likelihood function given by (2). Under the Bayesian approach, the specification of the model is completed by assigning a prior distribution to  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)$ , the parameter  $\boldsymbol{\beta} = \{\beta_{jk} : j = 1, \dots, p_k; k = 1, \dots, K\}$  and the parameters that define the copula family. Souza (2011) developed and fitted Model (2) using different copulas to predict missing response values. It was also carried out a simulation study to compare bivariate and univariate beta models under different scenarios.

### 2.1. Multivariate Hierarchical Beta Regression Model

In the multivariate beta regression model presented in the previous section, the marginal beta regression coefficients were fixed. However, there are situations in which some or all of the coefficients are assumed to be random. In these cases, the coefficients of each observation have a common average, suffering from the influence of nonobservable effects. Such models are often called mixed-effects models and have applications in several areas. Jiang (2007) discusses linear mixed models and some inference procedures for estimating their parameters. Rao and Molina (2015) shows some use of mixed-effects models in small area estimation.

In this section, we propose a generalisation of the multivariate regression model presented in Section 2 by assuming that some or all of the coefficients associated with the linear predictor of each response variable can be random and correlated.

Let  $y_{idk}$  be the observed value of the  $i^{th}$  microunit within the  $d^{th}$  macrounit for the  $k^{th}$  response variable,  $i = 1, \dots, n_d$ ,  $d = 1, \dots, D$  and  $k = 1, \dots, K$ . Furthermore, let us assume that  $y_{idk}$  and  $y_{i'dk}$  are conditionally independent,  $\forall i \neq i'$ . The multivariate hierarchical beta regression model is defined as

$$y_{id} | \boldsymbol{\mu}_{id}, \boldsymbol{\phi}_{id}, \boldsymbol{\theta} \sim BetaM(\boldsymbol{\mu}_{id}, \boldsymbol{\phi}_{id}, \boldsymbol{\theta}), \quad i = 1, \dots, n_d, \quad d = 1, \dots, D \quad (3)$$

$$y_{idk} | \boldsymbol{\mu}_{idk}, \boldsymbol{\phi}_{idk} \sim Beta(\boldsymbol{\mu}_{idk}, \boldsymbol{\phi}_{idk}), \quad k = 1, \dots, K \quad (4)$$

$$g(\boldsymbol{\mu}_{idk}) = \sum_{j=1}^{p_k} x_{idjk} (\boldsymbol{\beta}_{jk} + \boldsymbol{v}_{dj k}) \quad (5)$$

$$\boldsymbol{v}_{dj k} \sim N\left(0, \boldsymbol{\sigma}_{jk}^2\right), \quad j = 1, \dots, p_k \quad \text{and} \quad k = 1, \dots, K \quad (6)$$

where:  $p_k$  is the number of covariates for the response variable  $k$ ;  $BetaM(\boldsymbol{\mu}_{id}, \boldsymbol{\phi}_{id}, \boldsymbol{\theta})$  denotes a multivariate beta distribution using a copula function with parameter  $\boldsymbol{\theta}$  and the beta marginal distributions for the  $i^{th}$  microunit belonging to the macrounit  $d$ ;  $\mathbf{y}_{id} = (y_{id1}, \dots, y_{idK})^T$ ;  $\boldsymbol{\mu}_{id} = (\mu_{id1}, \dots, \mu_{idK})$ ;  $\boldsymbol{\phi}_{id} = (\phi_{id1}, \dots, \phi_{idK})$ ;  $\boldsymbol{\beta}_k = (\beta_{1k}, \dots, \beta_{p_kk})$ ;  $\mathbf{x}_{idk} = (x_{id1k}, \dots, x_{idp_kk})^T$  and

$$\mathbf{x}_{dk}^T = \begin{pmatrix} x_{1d1k} & \cdots & x_{1dp_kk} \\ x_{2d1k} & \cdots & x_{2dp_kk} \\ \vdots & \cdots & \vdots \\ x_{N_d d1k} & \cdots & x_{N_d dp_kk} \end{pmatrix}.$$

Thus, microunits belonging to the same macrounit have the same coefficient and the coefficients are different between macrounits. Each response variable can have its own set of regressors and these are not necessarily the same.

As generally described in Equations (3) and (5), the model allows all regression coefficients to be random. However, in many applications of hierarchical models, only some coefficients are assumed to be random, specifically the intercept term. To allow fixed and random coefficients, Equation (5) can be changed to

$$g(\mu_{idk}) = \sum_{j=1}^{p_k} x_{idjk} \beta_{jk} + \sum_{j=1}^{p_k} z_{idjk} \nu_{dj} = \mathbf{x}_{idk}^T \boldsymbol{\beta}_k + \mathbf{z}_{idk}^T \boldsymbol{\nu}_{dk},$$

with  $\mathbf{z}_{idk} = (z_{id1k}, \dots, z_{idp_kk})^T$  and  $\boldsymbol{\nu}_{dk} = (\nu_{d1k}, \dots, \nu_{dp_kk})^T$ . If  $z_{idjk} = x_{idjk}$ , the  $j^{th}$  coefficient is random and if  $z_{idjk} = 0$ , the correspondent coefficient is fixed.

In the model described in Equations (3)–(6) all random effects in  $\boldsymbol{\nu}$  could be considered independent, and only the correlations across the response variables would be modelled. However, to allow the averages of the responses to borrow strength across themselves for a given macrolevel  $d$ , all random coefficients for a same covariate  $j$  can be assumed to be correlated. For example, if all covariates are the same for all response models, we have  $\boldsymbol{\nu}_{dj} = (\nu_{dj1}, \dots, \nu_{djK})^T \sim N_K(\mathbf{0}, \boldsymbol{\Sigma}_j)$ ,  $j = 1, \dots, p$  where

$$\boldsymbol{\Sigma}_j = \begin{pmatrix} \sigma_{j1}^2 & \sigma_{j12} & \cdots & \sigma_{j1K} \\ \sigma_{j12} & \sigma_{j2}^2 & \cdots & \sigma_{j2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{j1K} & \sigma_{j2K} & \cdots & \sigma_{jK}^2 \end{pmatrix}.$$

A special case very often used in practice is to assume that only the intercepts are correlated, i.e.,  $\boldsymbol{\nu}_{d1} = (\nu_{d11}, \dots, \nu_{d1K})^T \sim N_K(\mathbf{0}, \boldsymbol{\Sigma}_1)$ .

The dependence of the response variables is evident on two levels: the observations and the linear predictors. This dependence can be favourable for this model with respect to the small area estimation problem because it allows strength to be borrowed across the means, which are interpreted as the true values of indexes, rates or proportions of interest. The logistic link function was used in all applications. The model stated in Equations (3)–(6)

assumes that information about  $K$  response variables and  $D$  macrounits with  $n_d$  microunits  $d = 1, \dots, D$  are available.

Equation (5) relates the averages of the response variables in each  $d^{\text{th}}$  macrounit, and considers specific macrounits' effects. Thus, the mean  $\mu_{idk}$  and  $\mu_{idk'}$  additionally borrow strength among themselves because they are correlated. This is particularly important in the small area estimation problem, in which  $\mu_{idk}$  is interpreted as the true value of the rate or proportion of interest and information from related quantities can produce more accurate estimators.

The vector parameter  $\phi_{id}$  is modelled as presented in the next section. The way it is modelled depends on the specific application considered and it might be subject to restrictions.

### 3. An Example of Small Area Estimation

The models defined in Section 2 were developed for general applications where there are  $K$  related variables, measured in the range (0,1), which can be explained by covariates. Here, we present an example of small area estimation.

The researcher may be interested in estimating functions of the response variables for small domains or for some domains with no sample at all. The multivariate models proposed in the previous section can be applied to make predictions on the nonsampled domains and to produce more accurate estimates for the small domains. Auxiliary information (covariates) must be known for all units at the level being predicted. The information can be obtained from a census or administrative records. We have not considered the case of missing values in explanatory variables.

#### 3.1. Brazilian Educational Data

The Brazilian evaluation of basic education is conducted by the Brazilian National Institute of Education Research (INEP). The evaluation measures the performance of students of the 4<sup>th</sup> and the 8<sup>th</sup> series of elementary school. The tests are performed every two years in urban state schools with more than 20 students. The evaluation of Brazilian education combines performances in the Portuguese language and mathematics tests with socioeconomic information.

The hierarchical structure of the data, organised into municipalities and schools, suggested the use of hierarchical modelling. Only schools with students in the 4<sup>th</sup> series in Rio de Janeiro State were considered in our application.

We considered the whole data of Rio de Janeiro State as our population and in each municipality selected a two-stage simple random sample of schools and students. In fact, we know the score values of all students for all schools. However, we pretend that we only know the sample-school means and sample-school variances for the selected schools and their respective sample sizes. This is not a unrealistic illustration, because information at individual level is not usually available due to issues of confidentiality.

The response variables are respectively the averages of proportions of correct answers in Portuguese and mathematics estimated at school level. In this application, these averages of proportions in both disciplines for each selected school are direct estimates based on a sample of students in each selected school.

It is important to note that although the number of correct answers for each student can be assumed to be binomially distributed, the school total cannot. Therefore a logistic model is not feasible here, since we are supposing that data at student level are not available. We further assume that the proportion in each school can be approximated by a beta distribution. This is not a strong assumption because the number of students in each school is not too small.

The main aim is to estimate these indexes for the nonsampled schools and to reduce the errors for the sampled schools. A two-part, multivariate hierarchical beta model was applied. One part relates the direct estimates of schools' proficiency to model parameters, and the other part relates these parameters to the auxiliary variables. The schools' indexes are in the (0,1) interval because the school averages are neither zero nor one for both tests. It is assumed that there is information for all schools, selected or not, on the following chosen covariates: existence of a program to avoid school dropout ( $x_2$ ); lack of books for students ( $x_3$ ); the percentage of teachers who teach less than 60% of the program of their disciplines ( $x_4$ ); proportion of teachers in the school with lower wages ( $x_5$ ); and lack of a library in the school ( $x_6$ ). The variable ( $x_1$ ) refers to the intercept.

The information available about the characteristics of schools is provided by the questionnaires given to school directors and teachers. Schools where there were no answers for at least one of these questionnaires were excluded from the analysis. Municipalities where there was only one state school, after the first mentioned dropouts were additionally eliminated, leaving 82 municipalities. For each one of these 82 municipalities, a random sample of 20% of the schools was selected. In eleven municipalities, all schools were selected. From the total 1,787 schools in the 82 municipalities, only 421 were selected. Within each selected school, a sample of 20% of the students was selected.

The response variables contain sampling error that may be related to the school sample size. To consider this feature, a modification in the multivariate hierarchical model is proposed in the equation of the observations. This modification was proposed by Liu et al. (2014) for a univariate beta model. Because it is natural to assume that the variance of the estimate increases when the sample size decreases, the following two-level model (3)–(6) is proposed:

$$y_{idk} \sim \text{Beta}(\mu_{idk}, \phi_{idk}),$$

where  $y_{idk}$  is the direct estimate (based on the sampling design) of the expected index of proficiency of the discipline  $k$ , of the  $i^{\text{th}}$  school in the  $d^{\text{th}}$  municipality for  $i = 1, \dots, n_d$ ,  $d = 1, \dots, D$ , where  $n_d$  is the number of selected schools for the  $d^{\text{th}}$  municipality.

We assume that the parameter  $\phi_{idk}$  can be different for each sampled school, and its value depends on the sample size through the following function:  $\phi_{idk} = \gamma_k n_{id} - 1$ , where  $\gamma_k$  is a unknown fixed parameter which may vary with the  $k^{\text{th}}$  component of the response vector,  $k = 1, \dots, K$  and  $n_{id}$  is the sample size of the  $i^{\text{th}}$  school in the  $d^{\text{th}}$  municipality. This assumption for  $\phi_{idk}$  is valid only for the sampled schools. For the nonsampled ones we constructed the estimator after inferring about the parameter; see Subsubsection 3.1.2 for details.

For the condition  $\phi_{idk} > 0$  to be satisfied, we must have

$$\gamma_k > \max\{1/n_{id}, \forall(i, d) \in s\} \tag{7}$$

where  $\max\{1/n_{id}, \mathbf{V}(i, d) \in s\}$  denotes the maximum of the inverses of all school sample sizes. Note that  $\gamma_k^{-1}$  can be interpreted as the design effect (*deff*) with respect to the variance of the sample proportion obtained in a simple random sampling with negligible sampling fraction. Therefore, if we have a previous estimate or guess of the *deff* for each response  $k = 1, \dots, K$ , we can use it to set the  $\gamma_k$ s. However, even if this information is not available, we can still obtain estimates of  $\gamma_k$ s through the model.

Taking into account the inequality (7), one should impose the following constraints on the range of  $\gamma_k$ s prior, based on one's prior knowledge about the signal of the intraclass correlation  $\rho_k$  for each variable of interest  $k = 1, \dots, K$ :

- a) if  $\rho_k > 0 \rightarrow \max\{1/n_{id}, \mathbf{V}(i, d) \in s\} < \gamma_k < 1$ ;
- b) if  $\rho_k < 0 \rightarrow \gamma_k > 1$ ;
- c) if one is not sure about the sign of  $\rho_k$  then  $\gamma_k > \max\{1/n_{id}, \mathbf{V}(i, d) \in s\}$ .

A simple type of prior that can be assigned to the  $\gamma_k$ ,  $k = 1, \dots, K$  are independent uniform priors, with ranges obtained as advised above.

The following models were considered in our analysis of the school data:

#### Model A

$$\mathbf{y}_{idk} | \mu_{idk}, \phi_{idk} \sim \text{Beta}(\mu_{idk}, \phi_{idk}), \quad i = 1, \dots, n_d, \quad d = 1, \dots, D$$

$$g(\mu_{id1}) = \beta_{11} + x_{id2}\beta_{21} + x_{id3}\beta_{31} + x_{id4}\beta_{41} + x_{id5}\beta_{51} + x_{id6}\beta_{61} + \nu_{d11}$$

$$g(\mu_{id2}) = \beta_{12} + x_{id2}\beta_{22} + x_{id3}\beta_{32} + x_{id4}\beta_{42} + x_{id5}\beta_{52} + \nu_{d12}$$

$$\nu_{d1} = (\nu_{d11}, \nu_{d12})^T \sim N_2(\mathbf{0}, \mathbf{\Sigma}),$$

#### Model B

$$\mathbf{y}_{id} | \mu_{id}, \phi_{id}, \theta \sim \text{BetaM}(\mu_{id}, \phi_{id}, \theta), \quad i = 1, \dots, n_d, \quad d = 1, \dots, D$$

$$\mathbf{y}_{idk} | \mu_{idk}, \phi_{idk} \sim \text{Beta}(\mu_{idk}, \phi_{idk}),$$

$$g(\mu_{id1}) = \beta_{11} + x_{id2}\beta_{21} + x_{id3}\beta_{31} + x_{id4}\beta_{41} + x_{id5}\beta_{51} + x_{id6}\beta_{61} + \nu_{d11}$$

$$g(\mu_{id2}) = \beta_{12} + x_{id2}\beta_{22} + x_{id3}\beta_{32} + x_{id4}\beta_{42} + x_{id5}\beta_{52} + \nu_{d12}$$

$$\nu_{d1} = (\nu_{d11}, \nu_{d12})^T \sim N_2(\mathbf{0}, \mathbf{\Sigma}),$$

where only the intercepts are assumed to be random.

A preliminary analysis showed that the covariate  $x_6$  ("lack of library in the school") is not statistically significant as a predictor of the index of proficiency in mathematics in the presence of the other covariates. Therefore, we did not use it as a predictor of the second response variable in all models.



Note that Model A generates conditional correlation between the dependent variables given municipality  $d$ , as long as  $\Sigma$  is not diagonal. However, Model B is much more general and useful for small area estimation purposes than Model A, since it allows the dependent variables to be correlated, conditional on the true small area parameters  $\mu_{id}$ ,  $\phi_{id}$  and  $\theta$ . This is equivalent to assuming that the sampling errors of the respective direct estimators are correlated. Furthermore, at first we would think that the use of a suitable copula function makes it possible to assume  $\Sigma$  diagonal in Model B; however, a drawback of adopting this strategy is that this does not allow the municipality random effects to be correlated across the dependent variables.

In the small area context, Models A and B can be regarded neither as a unit-level model, because the response variables are direct estimators, nor as an area-level model, because the municipality random effects are not of the same level as the domains of interest (schools). Since our model can be considered a two-level generalised hierarchical model, the only input response variables required to estimate its model parameters are the design-based direct estimates. Nevertheless, an extension of the model proposed here should include the designed-based variance-covariance matrix as additional information.

Because the scores of all the students are available in the Brazilian microdata test, it is possible to calculate the true observed proportions of the selected schools and to compare them with the direct estimates and the estimates provided by the models.

It is possible to obtain various types of dependence with copula functions. However, there is a wide variety of copula functions. The question thus arises of which copula to use. It makes sense to use the copula that is most appropriate for the data. Silva and Lopes (2008) and Huard et al. (2006) presented proposals for the selection of copulas and models. The criterion proposed by Huard et al. (2006) seeks the most appropriate copula for the data within a previously established set of copulas. Silva and Lopes (2008) implemented the Deviance Information Criterion (DIC) found in Spiegelhalter et al. (2002) and others. This criterion examines the model globally, providing not only the choice of the copula, but also the regressors and the marginal distributions of the response variables. The criteria Akaike Information Criterion (AIC), in Akaike (1973) and Bayesian Information Criterion (BIC), in Schwarz (1978), play a similar role.

Let  $L(\mathbf{y}|\Psi_j, M_j)$  be the likelihood function for the model  $M_j$ , where  $\Psi_j$  contains the copula parameters and those related to the marginal distributions. Define  $D(\Psi_j) = -2\log L(\mathbf{y}|\Psi_j, M_j)$ . The AIC, BIC and DIC are given by:

$$AIC(M_j) = D(E[\Psi_j|\mathbf{y}, M_j]) + 2q_j;$$

$$BIC(M_j) = D(E[\Psi_j|\mathbf{y}, M_j]) + \log(n)q_j;$$

$$DIC(M_j) = 2E[D(\Psi_j)|\mathbf{y}, M_j] - D(E[\Psi_j|\mathbf{y}, M_j])$$

where  $q_j$  denotes the number of parameters of the model  $M_j$ .

Let  $\{\Psi_j^{(1)}, \dots, \Psi_j^{(T)}\}$  be a sample from the posterior distribution obtained via MCMC. Then we have the following Monte Carlo approximations:

$$E[D(\Psi_j)|\mathbf{y}, M_j] \approx T^{-1} \sum_{t=1}^S D(\Psi_j^{(t)}) \quad \text{and} \quad E[\Psi_j|\mathbf{y}, M_j] \approx T^{-1} \sum_{t=1}^T \Psi_j^{(t)}.$$

The linear correlation coefficient is not suitable for measuring the dependence between variables in a model involving copulas since it is not invariant under monotone nonlinear transformation. A further appropriate measure, which can be found in [Nelsen \(2006\)](#), is the Kendall’s  $\tau$  statistic, given by

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

The parameter  $\theta$  has different interpretations, as well as different ranges depending on the copula, as can be seen in [Table 1](#). As the parameter  $\theta$  can be written in terms of the Kendall’s  $\tau$ , it is possible to compare the correlations given by different copulas. Thus, the FGM copula is useful for weak association levels between  $-2/9$  and  $2/9$ . On the other hand, the Clayton copula considers only positive correlations. Another advantage of considering  $\tau$  is that this facilitates the task of assigning a prior to  $\theta$ .

In the following, we focus on the bivariate case. We use the copulas described in [Table 1](#), where the ranges of variation of copula parameters  $\theta$  and the measures of dependence Kendall’s  $\tau$  are presented.

In the following section, the inference process on the parameters of Model B and the indirect estimators of the sampled and nonsampled areas are presented. The inference process and the estimators are analogous for Model A.

### 3.1.1. Inference

We assume that sample selection bias is absent from both models, that is, the sampling scheme is noninformative, see [Pfeffermann et al. \(2006\)](#) for further details. Let  $\mathbf{y}_s$  be the matrix of the response variables for the sampled schools and  $\mathbf{W} = \Sigma^{-1}$ . The posterior density for Model B of all unknown quantities is given by:

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \boldsymbol{\nu}, \mathbf{W} | \mathbf{y}_s) \propto p(\mathbf{y}_s | \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \boldsymbol{\nu}, \mathbf{W}) \times p(\boldsymbol{\nu} | \mathbf{W}) p(\boldsymbol{\beta}) p(\boldsymbol{\gamma}) p(\theta) p(\mathbf{W}).$$

Assuming independent priors for  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$ ,  $\theta$  and  $\mathbf{W}$ , we have:

$$p(\mathbf{y}_s | \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \boldsymbol{\nu}, \mathbf{W}) = \prod_{d=1}^D \prod_{i=1}^{n_d} c(F_1(y_{id1}), \dots, F_K(y_{idK}) | \boldsymbol{\nu}, \boldsymbol{\beta}, \theta, \boldsymbol{\gamma}) \times \prod_{k=1}^K f_k(y_{idk} | \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \nu_{d1k})$$

Table 1. Copula Functions used in this article.

Copula	$C(u, v   \theta)$	$\theta$	$\tau$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(0, \infty)$	$[0, 1] \setminus \{0\}$
FGM	$uv[1 + \theta(1 - u)(1 - v)]$	$[-1, 1]$	$[-2/9, 2/9]$
Frank	$-\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$	$(-\infty, \infty) \setminus \{0\}$	$[-1, 1] \setminus \{0\}$
Gaussian	$\int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left\{\frac{2\theta st - s^2 - t^2}{2(1-\theta^2)}\right\} ds dt$	$[-1, 1]$	$\frac{2}{\pi} \arcsen \theta$
Gumbel	$\exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\}$	$[1, \infty)$	$[0, 1]$

and

$$\begin{aligned}
 p(\boldsymbol{\nu}|\mathbf{W}) &= \prod_{d=1}^D p(\boldsymbol{\nu}_{d1}|\mathbf{W}) \propto \prod_{d=1}^D |\mathbf{W}|^{1/2} \exp\left\{-\frac{1}{2} \boldsymbol{\nu}_{d1}^T \mathbf{W} \boldsymbol{\nu}_{d1}\right\} \\
 &\propto |\mathbf{W}|^{D/2} \exp\left\{-\frac{1}{2} \sum_{d=1}^D \text{tr}(\boldsymbol{\nu}_{d1}^T \boldsymbol{\nu}_{d1} \mathbf{W})\right\} \\
 &\propto |\mathbf{W}|^{D/2} \exp\left\{-\frac{1}{2} \text{tr}\left[\left(\sum_{d=1}^D \boldsymbol{\nu}_{d1}^T \boldsymbol{\nu}_{d1}\right) \mathbf{W}\right]\right\},
 \end{aligned}$$

with  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ ,  $\boldsymbol{\nu} = (\nu_{111}, \nu_{112}, \dots, \nu_{D11}, \nu_{D12})$ ,  $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{21}, \beta_{31}, \beta_{41}, \beta_{51}, \beta_{61})$  and  $\boldsymbol{\beta}_2 = (\beta_{12}, \beta_{22}, \beta_{32}, \beta_{42}, \beta_{52})$ . The cumulative distribution function and the density of the beta distribution for the response variable  $k$  is represented by  $F_k$  and  $f_k$ , respectively. In addition,  $c(\cdot)$  is the density of the copula function.

The posterior distribution of all unknown parameters has no closed form, and thus a Monte Carlo Markov Chain (MCMC) simulation can be applied. Assigning a Wishart prior to  $\mathbf{W}$  and a normal one to the intercepts  $(\beta_{11}, \beta_{12})$  provides a full conditional with known forms for them. Therefore, we can use Gibbs to sample from these parameters. The other parameters are sampled via the Metropolis-Hastings algorithm (Gamerman and Lopes 2006). Samples of the posterior distribution of  $\tau$  are obtained directly from samples of the posterior of  $\theta$ , since  $\tau$  is a function of  $\theta$ .

Souza (2011) discussed different strategies for sampling from the posterior when the random-effect model described in (3)-(6) is fitted, including slice sampling (Neal, 2003). The importance of the posterior parametrisation to the convergence of MCMC algorithm when this model is fitted is shown.

To illustrate the convergence process, Souza (2011) simulated data from the model

$$\begin{aligned}
 y_{id} &\sim \text{BetaM}(\boldsymbol{\mu}_{id}, \boldsymbol{\phi}, \theta), \quad i = 1, \dots, n_d, \quad d = 1, \dots, D \\
 y_{idk} &\sim \text{Beta}(\mu_{idk}, \phi_k) \\
 g(\boldsymbol{\mu}_{idk}) &= \beta_{1k} + x_{id2k} \beta_{2k} + \nu_{d1k} \\
 \boldsymbol{\nu}_{d1} &= (\nu_{d11}, \nu_{d12}) \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})
 \end{aligned}$$

where *BetaM* represents the distribution generated by the Farlie-Gumbel-Morgenstern (FGM) copula with beta marginals. Souza (2011) fixed ( $K = 2$ ) response variables, with  $D = 100$  domains and  $n_d = 20$  units in each one. The following priors were considered:  $\theta \sim U(-1, 1)$ ;  $\beta_{jk} \sim N(0, 10^{-6})$ ,  $j = 1, 2$ ;  $\phi_k \sim \text{Gamma}(0.001, 0.001)$ ,  $k = 1, 2$ ; and  $\mathbf{W} = \boldsymbol{\Sigma}^{-1} \sim \text{Wishart}(2, \mathbf{I}_2)$ , where  $\mathbf{I}_2$  is the identity matrix of order 2. Souza (2011) used the same covariate for both responses. Note that  $\lambda_{dj k} = \beta_{1k} + \nu_{d1k}$ , which is equivalent to  $\lambda_{dj k} \sim N(\beta_{1k}, \sigma_{1k}^2)$ . A simulation study carried out under both ways of parametrisation showed that for the same number of iterations, the convergence is reached faster when the centre parametrisation is considered, that is  $\lambda_{dj k} \sim N(\beta_{1k}, \sigma_{1k}^2)$ . For further theoretical

discussion of how to create strategies for improving MCMC convergence, see [Gilks and Roberts \(1996\)](#).

Assigning a Wishart prior to  $\Sigma$  is convenient because the full conditional distribution of  $\Sigma$  is known and has close form, which allows the Gibbs sampling algorithm to be employed to sample from it. However, other parameterisations of the matrix  $\Sigma$  can be considered. One simple way of decomposing  $\Sigma$  is as follows:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$$

Another well-known parameterisation is the spectral decomposition. These decompositions of  $\Sigma$  facilitate the elicitation of the prior. The disadvantage is that the full conditional of the parameters no longer have any close form. [Souza \(2011\)](#) analysed two different ways of assigning vague prior distributions to the variances parameters:  $\sigma_k^{-2} \sim \text{Gamma}(\epsilon, \epsilon)$ , for  $k = 1, \dots, K$ , setting  $\epsilon$  small; and one of the approaches proposed by [Gelman \(2006\)](#),  $\sigma_k \sim U(0, M)$  for  $k = 1, \dots, K$ , fixing  $M$  large. For  $\rho_{12}$ , a uniform prior on the interval  $(-1, 1)$  is assigned. [Souza \(2011\)](#) showed that the slice-sampling algorithm is efficient for sampling from the full conditional of the  $\sigma_k$ s, but very slow for sampling from the full conditional of the random effects when a uniform distribution is assigned to the  $\sigma_k$ s. The Metropolis-Hastings algorithm was employed when the gamma prior was adopted for the variances. The simulation studies showed that the posterior distribution of all model parameters does not depend much on the three different ways of assigning prior distributions to  $\Sigma$ .

### 3.1.2. Small Area Estimation

The posterior mean of  $\mu_{idk}$  and its posterior variance can be empirically evaluated by calculating the mean and the variance of the  $T$  iterations of the MCMC algorithm. These values are obtained by jointly simulating the pairs  $(\mu_{id1}^{(t)}, \mu_{id2}^{(t)})$  using the fitted model, where:  $\mu_{idk}^{(t)} = g^{-1}(\beta_{1k}^{(t)} + \nu_{d1k}^{(t)} + \sum_{j=2}^{p_k} \beta_{jk}^{(t)} x_{idjk})$  for  $t = 1, \dots, T$ ,  $k = 1, 2$ ,  $i = 1, \dots, n_d$  and  $d = 1, \dots, D$ .

In addition, it is necessary to define the estimators for the nonsampled schools. Because there is information on the auxiliary variables for these schools, the estimate of the expected index in each nonselected school at each  $(t)$  sample point of the posterior distribution is given by  $\mu_{idk}^{(t)} = g^{-1}(\beta_{1k}^{(t)} + \nu_{d1k}^{(t)} + \sum_{j=2}^{p_k} \beta_{jk}^{(t)} x_{idjk})$ ; for  $t = 1, \dots, T$ ,  $k = 1, 2$ ,  $i = n_d + 1, \dots, N_d$  and  $d = 1, \dots, D$ , where  $N_d$  is the population size in  $d^{\text{th}}$  domain. Because we have  $T$  sample points from the posterior distribution of  $\mu_{idk}$ , we can obtain credibility intervals for the quantities of interest  $\mu_{idk}$  for sampled and non-sampled schools.

It is also possible to make inference about the students' score means at municipality level. Let us assume that  $N_{id}$ , the number of students in each school  $i = 1, \dots, N_d$  for each municipality  $d = 1, \dots, D$ , is known. Then we can generate MCMC samples from

the posterior distribution of the municipality score means,  $(\mu_{d1}, \mu_{d2})$ , as

$$\mu_{dk}^{(t)} = \frac{\sum_{i=1}^{N_d} N_{id} \mu_{idk}^{(t)}}{\sum_{i=1}^{N_d} N_{id}}; \quad t = 1, \dots, T, \quad k = 1, 2, \quad \text{and} \quad d = 1, \dots, D \quad (8)$$

Posterior means and their respective posterior variances for each municipality can be obtained easily by calculating the means and the variances of  $T$  MCMC samples.

### 3.1.3. Some Results: Study 1

We fitted Models A and B, using the copulas listed in Table 1, to the data. We assigned relatively vague priors to all parameters of the two fitted models. In particular, with respect to the  $\gamma_k$ s parameters, we assigned independent uniform priors with the lower limit of their intervals given by inequality (7) and the upper limit equal to  $10^5$ , i.e  $\gamma_k \sim U(0.5, 10^5)$ ,  $k = 1, 2$ . For the other parameters, we set  $\beta_{jk} \sim N(0, 10^6)$ ,  $j = 1, \dots, 6$ ,  $\mathbf{W} \sim \text{Wishart}(2, \mathbf{I}_2)$ , where  $\mathbf{I}_2$  is the identity matrix of order 2. For the parameter  $\theta$  the respective prior was set according to the copula as follows: *Gamma*(0.001, 0.001) for Clayton;  $U(-1, 1)$  for FGM;  $N(0, 10^4)$  for Frank;  $U(-1, 1)$  for Gaussian; and  $U(1, 10000)$  for Gumbel.

In all cases, two parallel chains were generated, each one with 200,000 iterations and a burn-in of 100,000. For all models, the chains were obtained from developing a special code in *Ox* version 5.0 (Doornik 2007). The corresponding two-univariate hierarchical beta model, simply denoted as ‘‘Separated’’ in Table 2, was also adjusted to investigate the benefits of the multivariate framework. Table 2 presents the model selection criteria results for all fitted models. The lower the DIC, AIC and BIC values, the better is the model. As noted in Table 2, the lowest values of the criteria are obtained when Model B is fitted using the Frank and Gaussian copulas. The values are slightly lower for the Frank copula. It should be noted that these copulas allow the widest range for the correlation between the indexes.

Table 3 shows the model parameter estimates for the Frank and Gaussian copulas. The posterior mean for the parameter  $\rho$ , which represents the correlation between the random effects, is approximately 0.60 for both models, as well as the Kendall’s  $\tau$  coefficient. These values indicate that a multivariate approach should be considered in the analysis of the students’ performances.

Table 2. Model selection criteria.

Models	pD	DIC	AIC	BIC	Log-likelihood
Model A	104.03	− 1809.37	− 1989.42	− 1932.83	956.70
Model B1 – Clayton	117.78	− 2137.79	− 2345.34	− 2288.74	1127.78
Model B2 – Fgm	108.85	− 2041.43	− 2231.14	− 2174.54	1075.14
Model B3 – Frank	119.48	− 2237.28	− 2448.24	− 2391.65	1178.38
Model B4 – Gaussian	112.94	− 2239.19	− 2437.07	− 2380.47	1176.06
Model B5 – Gumbel	117.60	− 2213.52	− 2420.73	− 2364.13	1165.56
Separated	69.08	− 1802.51	− 1910.68	− 1850.04	935.80

Table 3. Summary of the model parameters' posterior distribution for the Gaussian and Frank copulas.

Par.	Frank						Gaussian					
	2.50%	50%	97.50%	Mean	Dev.		2.50%	50%	97.50%	Mean	Dev.	
$\beta_{11}$	-0.098	-0.016	0.080	-0.013	0.046		-0.153	-0.064	0.023	-0.064	0.046	
$\beta_{21}$	-0.313	-0.254	-0.191	-0.254	0.031		-0.298	-0.236	-0.171	-0.236	0.032	
$\beta_{31}$	0.000	0.057	0.111	0.056	0.028		0.025	0.085	0.145	0.085	0.030	
$\beta_{41}$	0.137	0.222	0.302	0.221	0.043		0.134	0.215	0.299	0.215	0.042	
$\beta_{51}$	-0.242	-0.161	-0.079	-0.161	0.042		-0.240	-0.147	-0.056	-0.147	0.048	
$\beta_{61}$	-0.002	0.016	0.033	0.017	0.008		-0.001	0.022	0.045	0.023	0.011	
$\gamma_1$	3.091	3.576	4.113	3.581	0.256		3.359	3.857	4.389	3.858	0.267	
$\sigma_1^2$	0.037	0.063	0.101	0.065	0.017		0.037	0.055	0.083	0.056	0.012	
$\sigma_1$	0.194	0.252	0.317	0.253	0.032		0.192	0.234	0.289	0.236	0.025	
$\beta_{12}$	-0.256	-0.150	-0.059	-0.152	0.051		-0.305	-0.223	-0.122	-0.221	0.045	
$\beta_{22}$	-0.276	-0.221	-0.167	-0.221	0.029		-0.268	-0.208	-0.148	-0.208	0.031	
$\beta_{32}$	-0.008	0.048	0.103	0.048	0.028		0.026	0.082	0.138	0.082	0.028	
$\beta_{42}$	0.120	0.198	0.276	0.197	0.040		0.125	0.206	0.286	0.205	0.041	
$\beta_{52}$	-0.219	-0.142	-0.061	-0.141	0.041		-0.209	-0.113	-0.034	-0.115	0.044	
$\gamma_2$	3.424	3.955	4.551	3.965	0.284		3.665	4.230	4.850	4.232	0.297	
$\sigma_2^2$	0.048	0.076	0.121	0.079	0.019		0.051	0.077	0.117	0.079	0.017	
$\sigma_2$	0.220	0.276	0.349	0.278	0.034		0.225	0.278	0.343	0.279	0.030	
$\sigma_{12}$	0.018	0.042	0.079	0.044	0.016		0.020	0.039	0.067	0.040	0.012	
$\rho$	0.383	0.619	0.773	0.608	0.102		0.393	0.605	0.748	0.596	0.091	
$\theta$	6.805	7.815	8.892	7.827	0.544		0.710	0.757	0.797	0.756	0.022	
$\tau$	0.554	0.596	0.633	0.595	0.021		0.502	0.546	0.587	0.546	0.022	

According to the model comparisons used, the Frank copula seems to fit the Brazilian educational data somewhat better than the others. Therefore, we applied the Frank copula to compare the performance of the small area estimates obtained from Model B with its competitors.

The main goals of modelling the indexes of proficiency are to reduce the variability of the direct estimates derived from the sampling design and to obtain accurate estimates for nonsampled schools since the direct estimators can only be obtained for the selected ones. The multivariate model provides estimates for all schools, but we need to evaluate its adequacy. The 95% credible intervals of the predictive proportions by the replica  $y_{idk}^{(t)}$ , for  $i \in s$ , contain 98.1% and 97.8%, respectively, of the observed values for the disciplines of Portuguese and mathematics.

The reduction of the variability of the direct estimates by the application of the model can be assessed by estimating the coefficients of variation (CV) of the direct estimators obtained under design-based approach, that is,  $CV_{\hat{D}}(y_{idk}) = \sqrt{(n_{id} - 1)^{-1} y_{idk}(1 - y_{idk})(1 - n_{id}N_{id}^{-1})} / y_{idk}$  and by calculating the coefficients of variation obtained by employing the models, that is,  $CV_p(\mu_{idk}) = \sqrt{V_p(\mu_{idk})} / E_p(\mu_{idk})$ , where the symbols  $E_p(\cdot)$  and  $V_p(\cdot)$  denote the posterior mean and the posterior variance under the assumed model, respectively. Figure 1 summarises the distribution of the CVs obtained from the estimators. Clearly, the CVs generated by the models are much lower than those obtained through the direct estimation.

We also assessed the relative differences between the small area prediction for each subject provided by the approach employed and the respective true value. The same measure was calculated for the predictions obtained when the two univariate independent hierarchical beta regressions are fitted. Figure 2 shows the box plots for the approaches. It can be seen from Figure 2 that there is some gain in using either Model A or Model B compared to the univariate separated model for both subjects. Model B performs a bit better than Model A in all scenarios. However, Figure 2 shows that model-based estimates

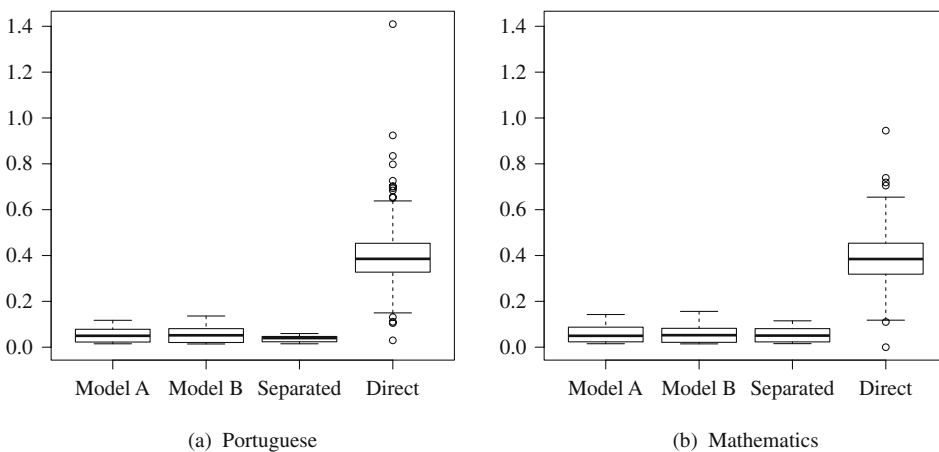
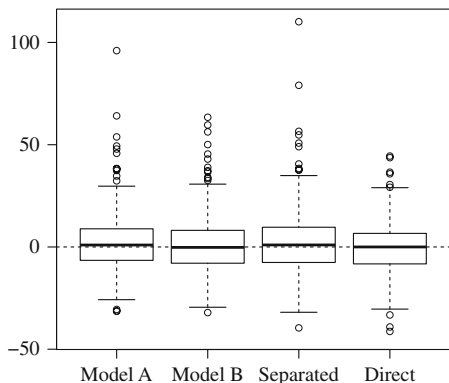
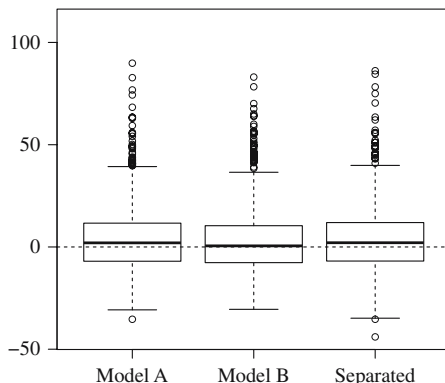


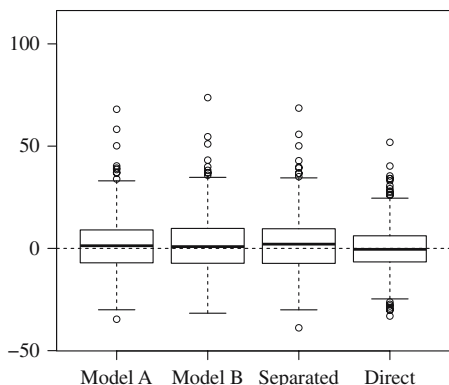
Fig. 1. Box plots of the coefficients of variation of the model-based estimators and the direct estimator for the sampled schools: Portuguese (a) and Mathematics (b).



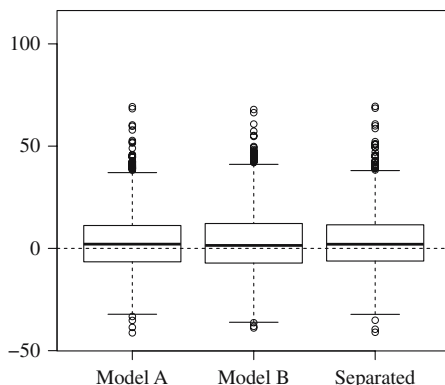
(a) Portuguese test: sampled schools



(b) Portuguese test: nonsampled schools



(c) Mathematics test: sampled schools



(d) Mathematics test: nonsampled schools

Fig. 2. Box plots of the relative differences in (%) between the small area prediction for each subject provided by the approach employed and the respective true value, carried out separately for schools in and out of the sample.

do not seem to substantially reduce the true relative errors of the direct estimates. This might be due to the fact that both sample sizes of schools and students for many municipalities and schools are not small enough to achieve considerable improvement of model-based estimates over designed-based estimates. This issue is investigated further in Subsubsection 3.1.4.

### 3.1.4. Some Results: Study 2

We conducted a second study to investigate the effect of reducing school and student sample sizes on the model-based estimates' improvement over the designed-based estimates. In this second study, the population consists of schools with at least 50 students who had taken both tests. Municipalities that had only one school after these exclusions were also discarded from the population, leaving 96,941 students. Then a simple random sample without replacement of ten percent of schools was selected, ensuring that at least two schools and at most seven schools would be selected per municipality. In the second



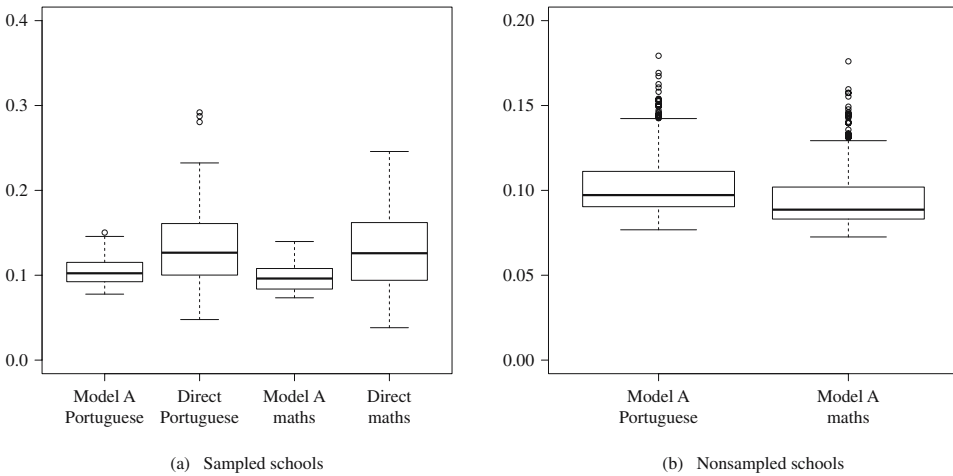


Fig. 3. Coefficients of variation of the model-based A estimator and the direct estimator for the schools in sample (a) and the coefficients of variation of the model-based A estimator for the schools out of sample (b).

stage, a simple random sample without replacement of ten percent of students in each school were selected, imposing a restriction of a maximum of five students per school. The population consists of 32 municipalities in which 87 schools have been selected out of 1,062 schools, making a total of 719 students in the sample. In this study, we only compared the Model A estimates to the direct estimates.

We fitted Model A to the sample using the same priors described in Study 1. Figure 3 summarises the distribution of the CVs obtained from the Model A estimator and the direct estimator for both subjects. As expected, the CVs obtained by the Model A are still much lower than those obtained by the direct estimation for sampled schools.

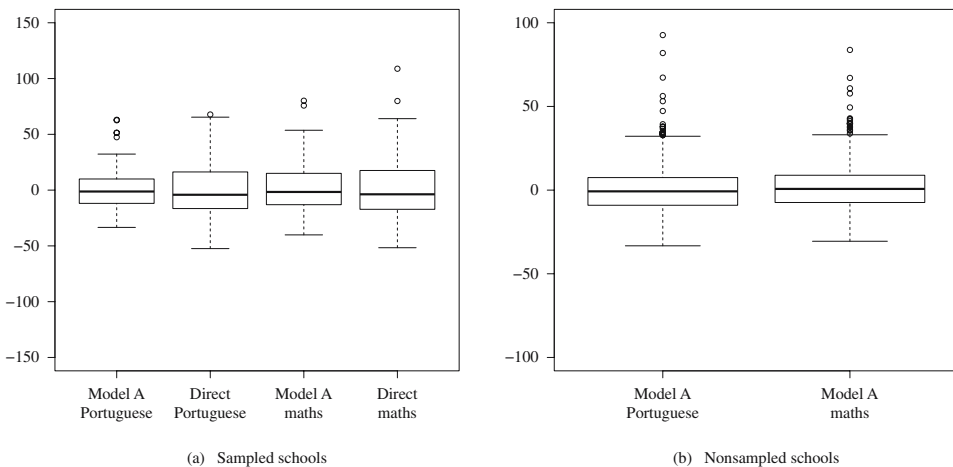


Fig. 4. Box plots of the relative differences (in %) with respect to the true value for the model-based A estimator and the direct estimator for the schools in sample (a); Coefficients of variation of the model-based A estimator for the schools out of sample (b).

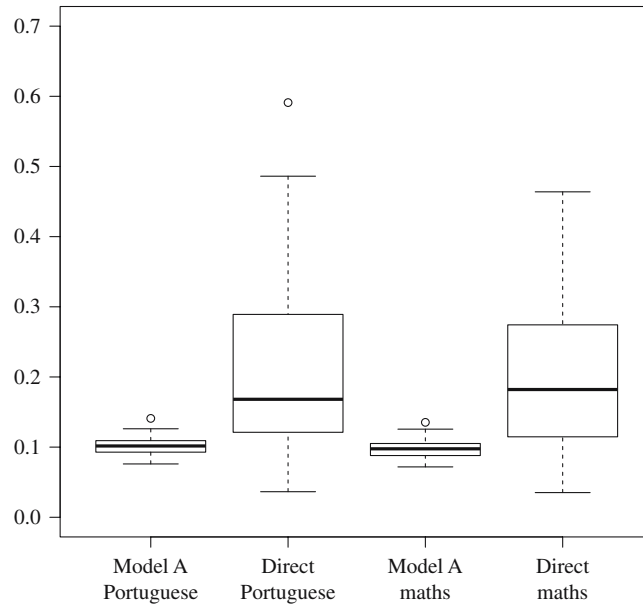


Fig. 5. Coefficients of variation of the model-based A estimator and the direct estimator calculated at municipality level for the subjects of Portuguese and mathematics.

Figure 4 shows the box plots of the relative differences between the school mean prediction for each subject provided by model-based A and designed-based approaches with respect to true value. It can be seen from Figure 4 that the reduction of the prediction errors of the proposed model-based A estimates with respect to the direct

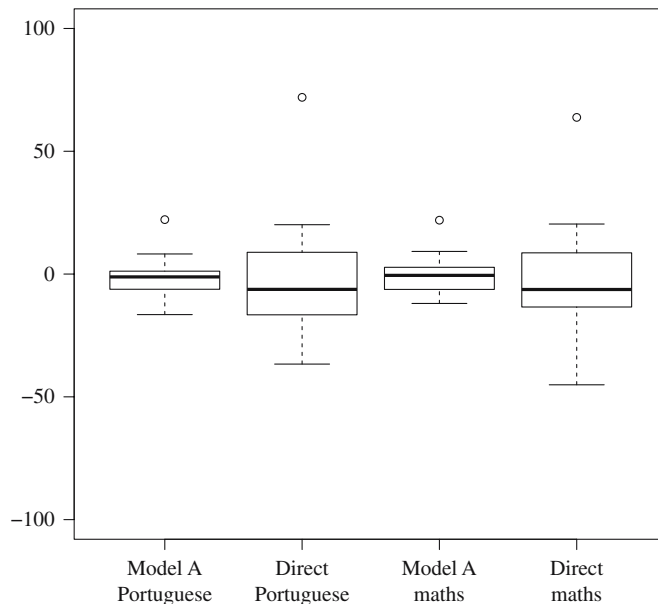


Fig. 6. Box plots of the relative differences in (%) with respect to the true value for the model-based A estimator and the direct estimator calculated at municipality level.

estimator is more appreciable than the respective result obtained from Study 1 for both subjects.

We also obtained the posterior means and the posterior variances, and consequently the coefficients of variation under the assumed model, for all 32 municipalities using formulas in Subsubsection 3.1.2. Estimates of the coefficients of variation of the direct estimators were obtained by using standard formulae to estimate the means and the variances in two-stage cluster sampling without replacement with equal probability of selection in both stages (schools and students), see, for example [Lohr \(1999, 147\)](#) for formulae.

[Figures 5](#) and [Figure 6](#) respectively present box plots of the coefficients of variation and the relative differences with respect to the true value of the Model A estimates and the direct estimates at municipality level for the subjects of Portuguese and mathematics. As expected, the gain in precision of the model-based estimates with respect to the direct estimates at municipality level are much higher than at school level for both subjects.

#### 4. Concluding Remarks and Suggestions for Future Work

The proposed models have the advantage of keeping the response variables at their original scale. Another advantage is the use of copulas, which are marginal free, that is, the degree of the association of the variables is preserved regardless of the marginal distributions. Thus, if two indexes are correlated, whatever marginal is adopted, the measure of dependence is the same. The use of copula functions in beta marginal regressions allows the joint analysis of the response variables by taking advantage of their dependency structure. The application of multivariate models with beta responses is an appealing alternative to models that require transforming the original variables. The choice between the proposed models and their competitors in the literature should be guided by the goals of the researcher, who must observe the models' predictive power and goodness of fit. The disadvantage of models that use copulas is that they are time consuming when simulating samples from the posterior distributions of the model parameters or functions of them.

In Section 2, we propose a multivariate hierarchical model with two levels. The variables are correlated on the first level with the aid of a copula function. Despite being applicable in general situations, this model has been developed especially for the small area estimation problem to allow strength to be borrowed across the areas or small domains of interest. The random effects of the same area are assumed to be correlated, and the random effects of different areas have the same variance-covariance matrix. In the illustration presented, the multivariate hierarchical model estimated the expected indexes of proficiency for nonsampled schools and additionally presented a significant reduction of the coefficients of variation compared to the direct estimates at school as well as at municipality level.

Sample household surveys are important sources of potential applications of the models proposed in this work. Examples of variables measured in the range  $(0,1)$  are the unemployment rate and the poverty gap, the latter of which measures, on average, the distance between the poor and the poverty line. These variables are important measures both for planning and in the knowledge of the population conditions, but are rarely available for small geographic levels or population subgroups for intercensus periods. Prediction of these poverty indexes could be performed using the models proposed in this article.

It is important to note that this work focuses on building multivariate regression models in which the marginal distributions are beta. This work notes the advantages of these models over corresponding univariate models and proposes a strategy for estimating their parameters. However, the theory of copula functions can be applied to any multivariate model that can be built for any known marginal distributions, provided that the distributions of response variables are different. We can even have continuous and discrete variables in the same model. To build a model for other distributions is straightforward, but each model has a peculiar and practical feature. In the specific case of the beta model, the mean and the dispersion have been adopted as the model parameters, where the latter parameter controls the variance. Other parametrisations are possible but could lead to additional difficulties. Various strategies can be defined by the researcher according to the available data. Some important strategies are first to fix the marginal and then obtain the appropriate copulas or to estimate models with different copulas and marginal densities and decide which is the “best” model by applying a model-comparison approach.

As can be seen in a simulation study in [Souza \(2011\)](#), when responses share exactly the same set of regressors, the results of univariate and multivariate approaches show little difference. In the abovementioned study and applications with real data, the model selection criteria were unable to show which approach was preferable. However, in the application presented in this article where the explanatory variables are not the same for both responses, we could see a better performance of the multivariate model. Similar findings were reported for other models and can be seen in [Bartels and Fiebig \(1991\)](#), [Gueorguieva and Agresti \(2001\)](#) and [Teixeira-Pinto and Normand \(2009\)](#).

It should be noted that the CV of the direct estimates were calculated under a design-based approach and the measure of precision of the proposed model under a model-based (Bayesian model-based) approach. Therefore, although we can interpret  $\gamma_k^{-1}$  as the deff, the calculation is based on model-based premises, and thus the ratio between the model-based variance estimate and its respective estimate of the direct estimator variance could be different from the design-based deff for each school. A possible extension of our proposed approach should allow the values of deff to vary with schools.

Another point is that in practical situations where the response variables can have values of zero or one, the beta distribution will not be adequate. One possible way to circumvent this problem is to use a mixture of distributions so that the zeros and ones can be accommodated. [Ospina and Ferrari \(2012\)](#) propose a general class of inflated regression models to fit data with such features. We have not considered omitted values in the explanatory variables in our model formulation, which could be another possible extension of the models proposed here.

## 5. References

- Akaike, H. 1973. “Information Theory and an Extension of the Maximum Likelihood Principle.” In *Second International Symposium on Information Theory*, edited by B.N. Petrov and F. Csaki, 267–281. Budapest: Akademiai Kiado.
- Bartels, R. and D.G. Fiebig. 1991. “A Simple Characterization of Seemingly Unrelated Regressions Models in Which OLS is Blue.” *The American Statistician* 45: 137–140. Doi: <http://dx.doi.org/10.2307/2684378>.

- Branscum, A.J., W.O. Johnson, and M.C. Thurmond. 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: 287–301. Doi: <http://dx.doi.org/10.1111/j.1467842X.2007.00481.x>.
- Cepeda-Cuervo, E., J.A. Achcar, and L.G. Lopera. 2014. "Bivariate Beta Regression Models: Joint Modeling of the Mean, Dispersion and Association Parameters." *Journal of Applied Statistics* 41: 677–687. Doi: <http://dx.doi.org/10.1080/02664763.2013.847071>.
- Da-Silva, C.Q., H.S. Migon, and L.T. Correia. 2011. "Dynamic Bayesian Beta Models." *Computational Statistics and Data Analysis* 55: 2074–2089. Doi: <http://dx.doi.org/10.1016/j.csda.2010.12.011>.
- Datta, G.S., B. Day, and I. Basawa. 1999. "Empirical Best Linear Unbiased and Empirical Bayes Prediction in Multivariate Small Area Estimation." *Journal of Statistical Planning and Inference* 75: 269–279. Doi: [http://dx.doi.org/10.1016/S0378-3758\(98\)00147-5](http://dx.doi.org/10.1016/S0378-3758(98)00147-5).
- Doornik, J.A. 2007. *Object-Oriented Matrix Programming Using Ox*, 3 ed. London: Timberlake Consultants Press.
- Fabrizi, E., M.R. Ferrante, S. Pacei, and C. Trivisano. 2011. "Hierarchical Bayes Multivariate Estimation of Poverty Rates Based on Increasing Thresholds for Small Domains." *Computational Statistics and Data Analysis* 55: 1736–1747. Doi: <http://dx.doi.org/10.1016/j.csda.2010.11.001>.
- Fay, R.E. 1987. "Application of Multivariate Regression to Small Domain Estimation." In *Small Area Statistics*, edited by R. Platek, J. Rao, C. Särndal, and M. Singh, 91–102. New York: Wiley.
- Ferrari, S.L.P. and F. Cribari-Neto. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31: 799–815. Doi: <http://dx.doi.org/10.1080/0266476042000214501>.
- Gamerman, D. and H.F. Lopes. 2006. *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*, 2 ed. London: Chapman & Hall.
- Gelman, A. (2006). "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis* 1: 515–534. Doi: <http://dx.doi.org/10.1214/06-BA117A>.
- Gilks, W. and G. Roberts. 1996. "Strategies for Improving mcmc." In *Markov Chain Monte Carlo in Practice*, edited by S.R.W. Gilks and D. Spiegelhalter, 89–114. London: Chapman & Hall.
- Gueorguieva, R.V. and A. Agresti. 2001. "A Correlated Probit Model for Joint Modeling of Clustered Binary and Continuous Responses." *Journal of the American Statistical Association* 96: 1102–1112. Doi: <http://dx.doi.org/10.1198/016214501753208762>.
- Huard, D., G. Évin, and A.-C. Favre. 2006. "Bayesian Copula Selection." *Computational Statistics and Data Analysis* 51: 809–822. Doi: <http://dx.doi.org/10.1016/j.csda.2005.08.010>.
- Jiang, J. 2007. *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Series in Statistics. New York: Springer.
- Liu, B., P. Lahiri, and G. Kalton. 2014. "Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions." *Survey Methodology* 40: 1–13. Available at: <http://www.statcan.gc.ca/pub/12-001-x/2014001/article/14030-eng.pdf> (accessed 1 December 2016).

- Lohr, S.L. 1999. *Sampling: design and analysis*, 1st ed. Place of publication: Brooks/Cole Publishing Company, California, USA.
- Melo, T.F., K.L. Vasconcellos, and A.J. Lemonte. 2009. "Some Restriction Tests in a New Class of Regression Models for Proportions." *Computational Statistics and Data Analysis* 53: 3972–3979. Doi: <http://dx.doi.org/10.1016/j.csda.2009.06.005>.
- Murteira, J.M.R. and J.J.S. Ramalho. 2014. "Regression Analysis of Multivariate Fractional Data." *Econometric Reviews* 0: 1–38. Doi: <http://dx.doi.org/10.1080/07474938.2013.806849>.
- Neal, R.M. (2003). "Slice Sampling." *The Annals of Statistics* 31: 705–767. Doi: <http://dx.doi.org/10.1214/aos/1056562461>.
- Nelsen, R.B. 2006. *An Introduction to Copulas*, 2 ed. New York: Springer.
- Olkin, I. and R. Liu. 2003. "A Bivariate Beta Distribution." *Statistics & Probability Letters* 62: 407–412. Doi: [http://dx.doi.org/10.1016/S0167-7152\(03\)00048-8](http://dx.doi.org/10.1016/S0167-7152(03)00048-8).
- Ospina, R. and S.L. Ferrari. 2012. "A General Class of Zero-or-One Inflated Beta Regression Models." *Computational Statistics and Data Analysis* 56: 1609–1623. Doi: <http://dx.doi.org/10.1016/j.csda.2011.10.005>.
- Pfeffermann, D., F.A. da Silva Moura, and P.L. do Nascimento Silva. 2006. "Multi-Level Modelling Under Informative Sampling." *Biometrika* 93: 943–959. Doi: <http://dx.doi.org/10.1093/biomet/93.4.943>.
- Rao, J.N.K. and I. Molina. 2015. *Small area estimation*, 2nd ed. Hoboken, New Jersey: Wiley.
- Schwarz, G. (1978). "Estimating the Dimension of a Model." *The Annals of Statistics* 6: 461–464. Doi: <http://dx.doi.org/10.1214/aos/1176344136>.
- Silva, R.S. and H.F. Lopes. 2008. "Copula, Marginal Distributions and Model Selection: a Bayesian Note." *Statistics and Computing* 18: 313–320. Doi: <http://dx.doi.org/10.1007/s11222-008-9058-y>.
- Simas, A.B., W. Barreto-Souza, and A.V. Rocha. 2010. "Improved Estimators for a General Class of Beta Regression Models." *Computational Statistics and Data Analysis* 54: 348–366. Doi: <http://dx.doi.org/10.1016/j.csda.2009.08.017>.
- Smithson, M. and J. Verkuilen. 2006. "A Better Lemon-Squeezer? Maximum Likelihood Regression With Beta-Distributed Dependent Variables." *Psychological Methods* 11: 54–71. Doi: <http://dx.doi.org/10.1037/1082-989X.11.1.54>.
- Souza, D.F. 2011. "Regressão Beta Multivariada com Aplicações em Pequenas Áreas." Ph.D. thesis, Instituto de Matemática da Universidade Federal do Rio de Janeiro. Available at: <http://www.pg.im.ufrj.br/teses/Estatistica/Doutorado/018.pdf> (accessed 1 December 2016).
- Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A.V.D. Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64: 583–639. Doi: <http://dx.doi.org/10.1111/1467-9868.00353>.
- Teixeira-Pinto, A. and S.-L. T. Normand. 2009. "Correlated Bivariate Continuous and Binary Outcomes: Issues and Applications." *Statistics in Medicine* 28: 1753–1773. Doi: <http://dx.doi.org/10.1002/sim.3588>.

Received February 2015

Revised January 2016

Accepted February 2016