Multivariate beta regression

Debora Ferreira de Souza

debora@dme.ufrj.br

Fernando Antônio da Silva Moura

fmoura@im.ufrj.br

Departamento de Métodos Estatísticos - UFRJ

Abstract

Multivariate beta regression models for jointly modeling two or more variables whose values belong to the interval (0,1), such as indexes, rates and proportions are proposed. The multivariate model can help the estimation process, borrowing information between units and obtaining more precise estimates, especially for small samples. Each response variable is assumed to be beta distributed, allowing to deal with multivariate asymmetric data. Copula functions are used to construct the joint distribution of the dependent variables. A simulation study for comparing our approach with independent beta regressions is also presented. An extension to two-level beta regression model is provided. The hierarchical beta regression model assumes fixed and correlated random effects. We present two real applications of our proposed approach. The first application aims to jointly regressing two poverty indexes measured at municipality level of a Brazilian state. The second one applies the approach to modeling two educational attainment indexes. The inference process was conducted under a Bayesian approach.

Key-words:Univariate beta regression, Copula, MCMC

1 Introduction

There are many practical situations involving multivariate regression analysis where the response variables are restricted to the interval $(0, 1)$, such as rates or ratios. Furthermore, these response variables might be correlated, even after conditioning on a set of explanatory variables. The main aim of this work is to propose multivariate regression models for these kind of response variables, taking into account possible correlation among them. As we show in our simulation study this is particular useful for prediction purposes.

For the case of a single response variable, Ferrari and Cribari-Neto (2004) propose a beta regression model, where the dependent variable is continuously measured in the interval (0.1). The density function of the response variable in their model can be written as:

$$f(y|\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \tag{1}$$

where $0 < \mu < 1$ and $\phi > 0$, with $a = \mu\phi$ and $b = (1-\mu)\phi$ are the usual parametrization of the beta density function and $\mu = E(Y)$. The parameter ϕ is related to the variance of the beta distribution, since $Var(Y) = \mu(1-\mu)/(1+\phi)$. This parametrization allows to associate a regression structure to the mean of the beta distribution. Their univariate beta model could be summarized as:

$$\begin{aligned} y_i|\mu_i,\phi &\sim Be(\mu_i(\beta),\phi), \quad i = 1,...,n \\ g(\mu_i) &= \eta_i = \sum_{l=1}^{p} x_{il}\beta_l, \end{aligned} \tag{2}$$

where: $g(\cdot)$ is a strictly monotonic function and twice differentiable which maps the interval $(0,1)$ in $\mathbb{R}$; $\boldsymbol{\beta}^T = (\beta_1,...,\beta_p)$ is a vector of regression coefficients and $x_{i1},...,x_{ip}$, for $i = 1,...,n$ are the observations of the p covariates.

The link function chosen for our examples was the logistic, $g(w) = \log(\frac{1}{1-w})$, although there are other possibilities, such as probit functions and complementary log-log. The parametrization used in (1) allows the data to be analyzed in its original scale without the need of transformation, which makes easier the interpretation of the results. This work follows the Bayesian paradigm, which requires to assign prior distributions to $\boldsymbol{\beta}$ and ϕ. The following independent prior distributions are used for ϕ and $\boldsymbol{\beta}$:

$$\phi \sim Gama(a,b) \quad \text{and} \quad \beta_l \sim N(m_l,\sigma_l^2), \quad l = 1,...,p.$$

For the application described in Section 2.1, we considered relative vague priors by setting $a = b = 0.001$ and $m_l = 0$ and $\sigma_l^2 = 10^6$ for $l = 1,2$.

In the multivariate case, it is desirable to model the dependence between the response variables $(Y_1,...,Y_K)$. The random variables here modeled follow marginal beta distributions and its joint distribution can be obtained using different approaches. In this work, the joint distribution of $(Y_1, ..., Y_K)$ is obtained from the application of a copula function to the marginal distributions of the response variables. In addition to the regression coefficients and precision parameters of the marginal distributions, we estimated the parameters that define the copula family and the ones related to the dependence between the response variables. The results obtained for the multivariate model are compared to those provided by separate beta regressions.

The paper is organized as follows. Section 2 describes the proposed multivariate beta model and provides some important properties of the copula function. An application to two poverty indexes is presented in Section 2.1. Section 2.2 introduces a simulation exercise in which data are generated under the multivariate beta regression and fitted under the univariate beta regression and vice versa. An extension to random coefficient model is presented in Section 3, as well as an application with hierarchical educational data in Section 3.1. Section 4 offers some conclusions and suggestions for further research.

2 Multivariate beta regression

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 joining the univariate beta through copulation, which is one of the most useful tools for working 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 implies a more flexible assumptions 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 function as a joint distribution function

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

where U_j, $j = 1, ..., K$ are uniform distributed in the interval $(0, 1)$.

The Sklar's theorem, stated 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 $F_1, ..., F_K$. Then, there is a unique K-dimensional copula C such that for all $(y_1, ..., y_K) \in [-\infty, \infty]^K$,*

$$H(y_1, ..., y_K) = C(F_1(y_1), ..., F_k(y_K)). \qquad (3)$$

Conversely, if C is a n-dimensional copula and $F_1, ..., F_K$ are distribution functions, then the function H defined by (3) is a distribution function with marginal distributions $F_1, ..., F_K$. Moreover if all marginal are continuous, C is unique. Otherwise, the copula C is unique determined in $Im(F_1) \times ... \times Im(F_K)$, where $Im(\cdot)$ represents the image of $(\cdot)$.

Let $Y_1, ..., Y_K$ be k random variables with marginal distributions $F_1, ..., F_K$, respectively, and joint distribution function $H(y_1, ..., y_K) = C(F_1(y_1), ..., F_K(y_K))$, where $F_j \sim U(0, 1)$, $j = 1, ..., K$ and $C(\cdot)$ is a copula function. Then the density function of $(Y_1, ..., Y_K)$ is given by:

$$
\begin{aligned}
h(y_1, ..., y_K) &= \frac{\partial^n H(y_1, ..., y_K)}{\partial y_1, ..., \partial y_K} \\
&= \frac{\partial^n C(F_1(y_1), ..., F_K(y_K))}{\partial F_1(y_1), ..., \partial F_K(y_K)} \times \frac{\partial F_1(y_1)}{\partial y_1} \times \cdots \times \frac{\partial F_K(y_K)}{\partial y_K} \\
&= c(F_1(y_1), ..., F_K(y_K)) \prod_{j=1}^{K} f_j(y_j) \qquad (4)
\end{aligned}
$$

where

$$
c(F_1(y_1), ..., F_K(y_K)) = \frac{\partial^n C(F_1(y_1), ..., F_K(y_K))}{\partial F_1(y_1), ..., \partial F_K(y_K)} \quad \text{and} \quad f_j(y_j) = \frac{\partial F_j(y_j)}{\partial y_j}, j = 1, ..., K.
$$

Let $\mathbf{y} = ((y_{11}, ..., y_{K1}), ..., (y_{1n}, ..., y_{Kn}))$ be a random sample of size n from the density in (4). Thus, the likelihood function is given by:

$$
L(\mathbf{\Psi}) = \prod_{i=1}^{n} c(F_1(y_{1i}|\mathbf{\Psi}), ..., F_K(y_{Ki}|\mathbf{\Psi})) f_1(y_{1i}|\mathbf{\Psi}) ... f_K(y_{Ki}|\mathbf{\Psi})
$$

where $\mathbf{\Psi}$ denotes the set of parameters that define the distribution functions and the density, as well as the copula function.

We assume that each response variable is beta distributed and the structure of dependence between them is defined by their joint distribution which is obtained by applying a copula function. Thus, the multivariate regression model proposed is represented by:

$$
\begin{aligned}
y_{ij}|\mu_{ij}, \phi_j &\sim Be(\mu_{ij}, \phi_j), \quad i = 1, ..., n, \quad j = 1, ..., K \\
g(\mu_{ij}) &= \eta_{ij} = \sum_{l=1}^{p} x_{il} \beta_{lj} \\
(y_{i1}, ..., y_{iK}) &\sim BetaM(\boldsymbol{\mu}_i, \boldsymbol{\phi}, \boldsymbol{\theta}) \qquad (5)
\end{aligned}
$$

where $g(\cdot)$ is the link function and $BetaM(\boldsymbol{\mu}_i, \boldsymbol{\phi}, \boldsymbol{\theta})$ denotes a beta multivariate distribution built by using a copula function with parameter $\boldsymbol{\theta}$ and the beta marginal distributions with their respective vector parameters given by $\boldsymbol{\mu}_i$ and $\boldsymbol{\phi}$, $i = 1, .., n$. Under the Bayesian approach, the specification of the model is completed by assigning a prior distribution to $\boldsymbol{\phi} = (\phi_1, ..., \phi_K)$,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1K} \\ \vdots & \vdots & \vdots \\ \beta_{p1} & \cdots & \beta_{pK} \end{pmatrix}$$

and to the parameters that define the copula family. Table 1 presents the copulas used in this work.

The linear correlation coefficient is not suitable to measure the dependence between variables in a model involving copulation. One most appropriate measure, which can be found in Nelsen (2006), is the statistic τ of Kendall, given by

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

The measure τ of Kendall is related to the parameter θ and can be used to assign a prior to θ.

It is possible to obtain various types of dependence with the use of copula function. However, there is a wide variety of copula functions. The question posed is: what copula should be used ? It makes sense to use the one that is most appropriate for the data under study. Silva and Lopes (2008) and Huard et al. (2006) present proposals for selection of copulas and models. The criterion proposed by Huard et al. (2006) look for the most appropriate copula to the data under analysis within a set of copulas previously established. Silva and Lopes (2008) implemented the DIC criterion (Spiegelhalter et al., 2002), among others, combining the choice of a copula with the marginal distributions. Let $L(\mathbf{y}|\Psi_i, M_i)$ be the likelihood function for the model M_i, where Ψ_i contains the copula parameters and the ones related to the marginal distributions. Define $D(\Psi_i) = -2 \log L(\mathbf{y}|\Psi_i, M_i)$. The criteria AIC, BIC and DIC are given by:

$$\begin{aligned} AIC(M_i) &= D(E[\Psi_i|\mathbf{y}, M_i]) + 2d_i; \\ BIC(M_i) &= D(E[\Psi_i|\mathbf{y}, M_i]) + \log(n)d_i; \\ DIC(M_i) &= 2E[D(\Psi_i)|\mathbf{y}, M_i] - D(E[\Psi_i|\mathbf{y}, M_i]). \end{aligned}$$

where d_i denotes the number of parameters of the model M_i.

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

$$E[D(\Psi_i)|\mathbf{y}, M_i] \approx L^{-1} \sum_l^L D(\Psi_i^{(l)}) \quad \text{and} \quad E[\Psi_i|\mathbf{y}, M_i] \approx L^{-1} \sum_l^L \Psi_i^{(l)}.$$

In what follows, we focus on the bivariate case. The copula functions used in this article are presented in Table 1, as well as the ranges of variation of parameters θ and the measures of dependence τ of Kendall.

Table 1: Copula functions employed

Copula	$C(u,v\|\theta)$	θ	τ
Clayton	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$	$(0,\infty)$	$[0,1]\backslash\{0\}$
FGM	$uv[1+\theta(1-u)(1-v)]$	$[-1,1]$	$[-2/9,2/9]$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right)$	$[-1,1]\backslash\{0\}$	$[-1,1]\backslash\{0\}$
Gaussiana	$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\exp\left\{\frac{2\theta st-s^2-t^2}{2(1-\theta^2)}dsdt\right\}$	$[-1,1]$	$\frac{2}{\pi}\text{arcsen }\theta$

2.1 Application to poverty indexes regression

The data used in our application were obtained from the Brazilian database of the Institute of Applied Economic Research and are available in the site www.ipeadata.gov.br. The response variables are the proportion of poor persons (Y_1) and the infant mortality rate (Y_2) in 168 municipalities in the states of Espírito Santo and Rio de Janeiro for the year 2000. The variables human development index (X) was the explanatory variable used. For the set of data employed, the association dependence measure is 0.42, which implies that the copula to be fitted to the data should allow positive dependence. We considered relative vague priors for the parameter θ related to the four copulas fitted. We respectively set $\theta \sim Gamma(0.001, 0.001)$ and $\theta \sim Unif(-1,1)$ for the Clayton and FGM copulas and $\theta \sim N(0, 10^6)$ and $\theta \sim Unif(-1,1)$ for the Frank and Gaussian copulas.

Because the posterior densities of $\boldsymbol{\beta}$, ϕ and θ as well as their full conditional distributions have not closed form, we use the Metropolis-Hastings algorithm for sampling from these parameters. For all the models with copulas, the convergence of parameters ϕ and θ is quickly reached, showing low autocorrelation, while for the parameter $\boldsymbol{\beta}$, the convergence is slow. In all cases, it was generated two parallels chains with 300000 iterations each and a burn-in of 150000.

Table 2 shows some descriptive statistics of the samples from the posterior of the parameters for the models that used Clayton and FGM copulas. The values of θ for these two copulas can not be directly compared. It should be observed the value of the statistic τ of Kendall provided for each θ to evaluate the degree of dependence created by the copulas. For the Clayton copula, the posterior mean of θ is 0.05, which implies that $\tau = 0.02$, while for the FGM copula, $\theta = -0.45$ yields to $\tau = -0.10$. The 95 %

credible intervals for τ are respectively $[0, 0.08]$ and $[-0.19, 0.00]$, for the Clayton and FGM copulas.

Table 2: 95% Credible intervals, posterior means and posterior standard deviations for $\boldsymbol{\beta}$, ϕ and θ obtained for the models which used Clayton and FGM copulas

Parameter	Clayton				FGM			
	2.5%	97.5%	Mean	Std.	2.5%	97.5%	Mean	Std.
β_{11}	6.60	7.97	7.28	0.35	6.56	7.95	7.26	0.35
β_{21}	-11.82	-9.97	-10.89	0.47	-11.79	-9.92	-10.87	0.48
β_{12}	4.08	5.84	4.98	0.44	4.10	5.84	4.96	0.45
β_{22}	-9.37	-7.00	-8.20	0.60	-9.37	-7.02	-8.18	0.61
ϕ_1	77.28	118.65	96.83	10.58	78.11	119.54	97.73	10.68
ϕ_2	55.45	85.39	69.65	7.62	56.22	86.24	70.15	7.77
θ	0.00	0.18	0.05	0.05	-0.84	-0.01	-0.45	0.21
τ	0.00	0.08	0.02	0.02	-0.19	0.00	-0.10	0.05

Table 3: 95% Credible intervals, median, mean and standard deviation of the posterior for the parameters $\boldsymbol{\beta}$, ϕ and θ using the FGM and Gaussian copulas

Parameter	Frank				Gaussian			
	2.50%	97.50%	Mean	Std.	2.50%	97.50%	Mean	Std.
β_{11}	6.59	7.97	7.27	0.35	6.57	7.95	7.27	0.35
β_{21}	-11.81	-9.96	-10.87	0.48	-11.79	-9.94	-10.87	0.48
β_{12}	4.08	5.81	4.96	0.44	4.06	5.85	4.95	0.45
β_{22}	-9.32	-7.00	-8.18	0.59	-9.38	-6.96	-8.16	0.61
ϕ_1	78.33	120.31	97.91	10.66	77.19	119.41	97.27	10.77
ϕ_2	56.14	85.73	70.08	7.55	55.91	85.70	69.77	7.73
θ	-1.84	0.10	-0.85	0.49	-0.24	0.06	-0.09	0.08
τ	-0.20	0.01	-0.09	0.00	-0.15	0.04	-0.05	0.05

The results obtained for the Frank and Gaussian copulas can be seen in Table 3. In the case of the Frank copula, we have $\theta = -0.85$ corresponding to $\tau = -0.09$, with credible interval $[-0.20, 0.01]\backslash 0$. For the model with Gaussian copula, $\theta = -0.09$ witch results in $\tau = -0.05$. The credible interval to τ is $[-0.15, 0.04]$.

The measure of association τ estimated by each copula is lower than that found before adjustments of the models. Moreover, its value changes sign. This is because, in a multivariate regression analysis, measures of dependence between the response variables are affected by the explanatory variables. In a linear regression analysis, the partial correlation coefficient measures the association between two response variables Y_1 and Y_2

after eliminating the effects of the explanatory variables X_1,...,X_p. Because the response variables follow a beta distribution, a transformation should be applied to use the partial correlation measure. Applying a logit transformation to the response variables, we obtain partial correlation of -0.097, which is consistent with the estimated values of the statistic τ for the models that allow negative values of this measure.

Regarding to the beta regressions parameters, we find that the regression coefficients have the same sign for all copulas, which means that the relationship between response variables and explanatory ones were captured in the same way by all models. The comparison with the results of the separate regressions shows that the use of the copula function did not affect the sign of the regression coefficients. Moreover, the credible intervals of the regression coefficients have interception, showing that their magnitudes are also similar for all models.

We carried out an analysis of the residuals. We define the standardized residual as:

$$r_{ij}^{(l)} = \frac{y_{ij} - \mu_{ij}^{(l)}}{\sqrt{Var(y_{ij})^{(l)}}}$$

where i denotes the i^{th} observation of the j^{th} variable in the l^{th} sample of the posterior distribution obtained by the MCMC method after convergence, with

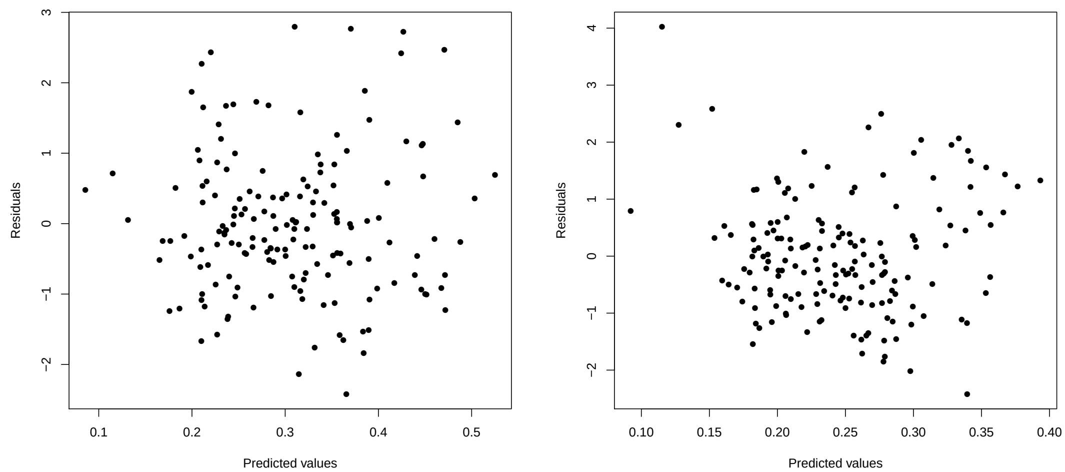
$$\mu_{ij}^{(l)} = g^{-1}(x_{i1}\beta_{1j}^{(l)} + ... + x_{ip}\beta_{pj}^{(l)})$$

and

$$Var(y_{ij})^{(l)} = \frac{\mu_{ij}^{(l)}(1 - \mu_{ij}^{(l)})}{1 + \phi_j^{(l)}}.$$

Figure 1 shows the distribution of standardized residuals against the predicted values for the two response variables when uses the Frank copula. Figure 1 do not show any systematic pattern. The residuals obtained by others copulas show similar behavior and they are not displayed.

Ferrari and Cribari-Neto (2004) define an overall measure of variation explained by univariate regression beta, called pseudo R^2, defined by the root of the correlation coefficient between sample $g(y)$ and $\hat{\eta}$. Thus, $0 \leq R^2 \leq 1$ and the more it is closed to 1, the better the fit is considered. One way of adapting the measure R^2 for the multivariate case is to separately calculate it for each variable. In order to do this, we need to find an estimate of the linear predictor η_{ij} associated to i^{th} observation of the j^{th} variable. Since we have a sample from the posterior distribution of the β, we could obtain the following



(a) Proportion of poor persons (b) Infant mortality rate

Figure 1: Residuals against the predicted values for the variables (a) proportion of poor persons and (b) infant mortality rate in the model that uses Frank copula.

estimate:

$$\hat{\eta}_{ij} = \frac{1}{M}\sum_{l=1}^{M} x_{i1}\beta_{1j}^{(l)} + ... + x_{ip}\beta_{pj}^{(l)}.$$

Thus, for each response variable, the R^2 adapted to the Bayesian context is the root of the sample correlation between the vector $\hat{\boldsymbol{\eta}}_j$ and the vector corresponding to the values of the link function $g(.)$ evaluated at the observed points. Table 4 presents the statistical values of R^2 for the employed models. All models have values of R^2 close to 1, indicating that the models successfully explain much of the total variation and there is no difference between them with respect to their power explanation.

Table 4: R^2 for the fitted models.

Variable	Clay	FGM	Frank	Gaussian	Individual
Proportion of poor	0.9216006	0.9215963	0.9215837	0.9215823	0.9215903
Infant mortality rate	0.9320647	0.9320788	0.9320779	0.9320771	0.9320935

It can be seen in Table 5 that all selection criteria proposed in Silva and Lopes (2008) point to the choice of the Frank copula. This copula is used in the simulation section below. It should be noted the considerable difference between the EPD values for the two separate regressions. The variable proportion of poor has great contribution in the total EPD amount and practically decides what is the best model. Ways to avoid contamination of the criterion caused by the use of variables measured in different scales are still under study.

Table 5: Model selection criteria for the copulas analyzed together plus the separated regressions model

Criteria	Clayton	FGM	Frank	Gaussian	Poor	Infant Mortality
p_D	6.13	7	6.98	7.15	2.91	2.76
DIC	-1083.46	-1086.17	-1086.24	-1084.04	-560.11	-525.54
AIC	-1085.73	-1090.17	-1090.19	-1088.34	-555.94	-521.05
BIC	-1070.11	-1074.55	-1074.57	-1072.72	-540.32	-505.43
$EAIC$	-1079.59	-1083.17	-1083.21	-1081.19	-553.03	-518.29
$EBIC$	-1063.98	-1067.55	-1067.59	-1065.57	-537.41	-502.67
EPD	1.61	1.6	1.6	1.61	0.72	0.89
$\log p(\boldsymbol{\Psi})$	544.8	546.59	546.61	545.59	281.51	264.15

2.2 A Simulation Study

The purpose of the simulation study is to evaluate the efficiency of proposed model. Besides, we intend to compare the results of the bivariate model with those provided by fitting separate regression for each response variable, which ignores the correlations structure between them.

Having as motivation the real data analyzed above, we simulate samples from the bivariate model and the univariate model, assuming the existence of a single explanatory variable. The true values of the parameters were set as $\boldsymbol{\beta}_1 = (7.26, -10.86)$ and $\phi_1 = 99.04$, for the first response variable and $\boldsymbol{\beta}_2 = (4.95, -8.16)$ and $\phi_2 = 70.97$, for the second. These values were obtained by fitting univariate beta regression models for the proportion of poor and infant mortality rate and using the human development index (HDI) as the covariate for both models.

We simulated data sets with $n = 50$ and $n = 100$ observations. We use the Frank copula for simulating from the bivariate case. The package "R" was used to generate observations from this copula. We fix the dependence measure tau of Kendall between the response variables at $\tau = 0.1$, $\tau = 0.5$ and $\tau = 0.8$, which correspond to the values of the parameter of the Frank copula $\theta = 0.91$, $\theta = 5.74$ and $\theta = 18.10$, respectively. It is expected that as the correlation between the responses increases, the better is the fit of the bivariate model compared to the separate beta regressions. For each situation considered, we simulated 200 samples. The priors used in this simulation study are the same as the ones described in Section 2.1, for the Frank copula fit.

In order to compare the various models, we used the relative absolute bias, the root of mean square error. The relative absolute bias (RAB) and the root mean square error

Table 6: Relative absolute bias and Mean square error obtained for the bivariate model with $\tau = 0.1$

	Simulated Model: Bivariate, $\tau = 0.1$							
	Fitted Model: Bivariate				Fitted Model:Univariate			
	$n = 50$		$n = 100$		$n = 50$		$n = 100$	
Parameter	RMSE	RAB	RMSE	RAB	RMSE	RAB	RMSE	RAB
β_{11}	0.291	3.159	0.183	2.009	0.307	3.351	0.187	2.058
β_{21}	0.424	3.059	0.261	1.915	0.447	3.228	0.268	1.963
β_{12}	0.294	4.761	0.246	3.893	0.301	4.799	0.254	3.978
β_{22}	0.425	4.130	0.357	3.425	0.435	4.182	0.369	3.498
ϕ_1	20.180	15.921	13.379	10.931	22.364	17.385	14.056	11.340
ϕ_2	15.034	16.675	9.551	10.303	16.531	17.908	9.857	10.509
θ	0.936	82.203	0.630	56.163	-	-	-	-

(RMSE), are respectively defined as:

$$
\begin{aligned}
RAB &= \frac{1}{200} \sum_{r=1}^{200} |\hat{U}^r - U|/U \\
RMSE &= \left[\sum_{r=1}^{200} (\hat{U}^r - U)^2/200 \right]^{1/2}
\end{aligned} \tag{6}
$$

where U denotes the true value of the parameter and U^r its estimate value for the r^{th} simulation. Table 6 compares the fit of the bivariate and univariate models, when the responses exhibit dependence $\tau = 0.1$.

As expected, the bias and the $RMSE$ statistics are lower in samples with size equal to $n = 100$ than those with sample size $n = 50$. The bias and the mean square error are slightly lower than the corresponding values obtained for the univariate model for both sample sizes, ie, the correct model (bivariate) comes close to the true values than the simpler model. When the data were simulated from the bivariate model with $\tau = 0.5$, the conclusions are quite similar.

The comparison between Tables 6 and 8 shows that differences between the bias and the mean square errors are higher for the case that we fit the univariate model when the data has considerable dependence. Thus, the greater the dependence of the response variables, the more serious is the problem of fitting the simpler model to data with complex structure.

In the situation where the responses were generated independently, the bias and mean square errors were smaller for bivariate fit, although this is not the correct model. This suggests that the bivariate model can be fitted even when there is no dependence or very

Table 7: Relative absolute bias and Mean square error obtained for the Bivariate model with $\tau = 0.5$

	Simulated Model: Bivariate, $\tau = 0.5$							
	Fitted Model: Bivariate				Fitted Model: Univariate			
	$n = 50$		$n = 100$		$n = 50$		$n = 100$	
Parameter	RMSE	RAB	RMSE	RAB	RMSE	RAB	RMSE	RAB
β_{11}	0.299	3.269	0.218	2.400	0.300	3.357	0.237	2.629
β_{21}	0.429	3.161	0.311	2.278	0.431	3.221	0.337	2.494
β_{12}	0.336	5.385	0.228	3.659	0.337	5.382	0.253	4.201
β_{22}	0.484	4.713	0.330	3.202	0.489	4.739	0.364	3.644
ϕ_1	19.927	16.277	13.266	10.736	22.858	18.301	14.074	11.211
ϕ_2	12.531	13.877	10.423	11.862	13.288	14.568	11.099	12.325
θ	1.250	16.980	0.771	10.644	-	-	-	-

Table 8: Relative absolute bias and Mean square error obtained for the Bivariate model with $\tau = 0.8$

	Simulated Model: Bivariate, $\tau = 0.8$							
	Fitted Model: Bivariate				Fitted Model: Univariate			
	$n = 50$		$n = 100$		$n = 50$		$n = 100$	
Parameter	RMSE	RAB	RMSE	RAB	RMSE	RAB	RMSE	RAB
β_{11}	0.245	2.707	0.176	1.969	0.301	3.259	0.205	2.237
β_{21}	0.346	2.551	0.255	1.901	0.426	3.087	0.292	2.129
β_{12}	0.278	4.482	0.209	3.363	0.336	5.493	0.225	3.602
β_{22}	0.397	3.878	0.304	2.963	0.480	4.726	0.325	3.185
ϕ_1	18.410	14.810	12.994	10.405	22.319	17.295	15.218	12.095
ϕ_2	12.292	13.490	9.870	10.904	14.888	15.968	10.670	11.705
θ	2.677	11.421	2.062	8.684	-	-	-	-

Table 9: Relative absolute bias and Mean square error obtained for the Univariate model

	Simulated Model: Univariate							
	Fitted Model: Univariate				Fitted Model: Bivariate			
	$n = 50$		$n = 100$		$n = 50$		$n = 100$	
Parameter	RMSE	RAB	RMSE	RAB	RMSE	RAB	RMSE	RAB
β_{11}	0.300	3.416	0.208	2.301	0.293	3.321	0.202	2.232
β_{21}	0.426	3.228	0.301	2.230	0.417	3.142	0.294	2.153
β_{12}	0.290	4.608	0.203	3.232	0.297	4.753	0.204	3.255
β_{22}	0.420	4.031	0.296	2.874	0.430	4.138	0.297	2.898
ϕ_1	20.605	16.511	14.825	11.247	18.562	15.182	14.078	10.804
ϕ_2	16.661	17.041	9.400	10.781	15.255	16.146	9.105	10.648

low one.

3 Multivariate hierarchical beta regression model

In the previous section was presented a multivariate beta regression model in which the marginal beta regression coefficients were fixed. However, there are situations that are reasonable to assume that some or all of the coefficients are random. In these cases, the coefficients of each observation have a common average, suffering the influence of non-observable effects. Such models are often called mixed effects models with response in the exponential family, with applications in several areas. Jiang (2007) discusses linear mixed models and some inference procedures for estimating its parameters.

In this section we propose a generalization 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_{ijk} be the observed value of the k^{th} response variable in the j^{th} first unit level of the i^{th} second unit level, $k = 1, ..., K$, $j = 1, ..., n_i$ and $i = 1, ..., m$. Furthermore, let us assume that y_{ijk} and $y_{i'jk}$ are conditional independents, $\forall\ i \neq i'$. The multivariate hierarchical beta regression model is defined as:

$$\mathbf{y}_{ij} \sim BetaM(\boldsymbol{\mu}_{ij}, \boldsymbol{\phi}, \boldsymbol{\theta}), \quad j = 1, ..., n_i, \quad i = 1, ..., m \tag{7}$$

$$g(\mu_{ijk}) = \mathbf{x}_{ij}^T \boldsymbol{\lambda}_{ik}, \quad k = 1, ..., K \tag{8}$$

$$\lambda_{ilk} = \beta_{lk} + \nu_{ilk}, \tag{9}$$

$$\boldsymbol{\nu}_{il} = (\nu_{il1}, ..., \nu_{ilK})^T \sim N_K(\mathbf{0}, \boldsymbol{\Sigma}_l), \quad l = 1, ..., p \tag{10}$$

where: $BetaM(\boldsymbol{\mu}_{ij}, \boldsymbol{\phi}, \boldsymbol{\theta})$ denotes a beta multivariate distribution built by using a copula function with parameter $\boldsymbol{\theta}$ and the beta marginal distributions; $\mathbf{y}_{ij} = (y_{ij1}, ..., y_{ijK})^T$; $\mathbf{x}_{ij}^T = (x_{ij1}, ..., x_{ijp})^T$; $\boldsymbol{\lambda}_{ik} = (\lambda_{i1k}, ..., \lambda_{ipk})^T$; $\boldsymbol{\phi} = (\phi_1, ..., \phi_K)^T$ and

$$\mathbf{x}_i^T = \begin{pmatrix} x_{i11} & \cdots & x_{i1p} \\ x_{i21} & \cdots & x_{i2p} \\ \vdots & \cdots & \vdots \\ x_{in_i1} & \cdots & x_{in_ip} \end{pmatrix}.$$

From (9) and (10) follows $\lambda_{il} \sim N(\beta_l, \boldsymbol{\Sigma}_l)$, $l = 1, ..., p$. This parametrization was proposed by Gelfand et al. (1995) to improve convergence of mixed linear models. The authors show that this parametrization is able to reduce the autocorrelation of the Gibbs sampling chains, speeding up the convergence of the model parameters. The implementation of their approach for fitting multivariate beta regression also improves convergence when Gibbs sampling and Metropolis-Hastings algorithms are used. See Appendix 1 for details.

The model described from (7) to (10) requires that observations within each i^{th} level are available. This assumption is necessary in order to avoid some difficulties in estimating the matrix $\boldsymbol{\Sigma}$. The Multivariate hierarchical beta regression model allows for interesting particular cases. If we regard the responses to be conditional independents given their means and their precision parameters, we have univariate beta regression with random regression coefficients. Two beta regressions with random intercepts were used in the application described in Section 3.1.

As generally described in equations (7), (8) and (9), the model allows all regression coefficients to be random, however, in many applications of hierarchical models only some coefficients are assumed to be random, specially the intercept term. In the model (7)-(10) all random effects in $\boldsymbol{\nu}$ could be considered independent and only the correlation between the response variables would be contemplated. However, to allow the averages of the responses also exchange information among themselves, it is considered that within each level i, and for each coefficient of the response variable l, the random effects concerning the response variables are correlated, i.e: $\boldsymbol{\nu}_{il} = (\nu_{il1}, ..., \nu_{ilK})^T \sim N_K(\mathbf{0}, \boldsymbol{\Sigma}_l)$ where

$$\boldsymbol{\Sigma}_l = \begin{pmatrix} \sigma_{l1}^2 & \sigma_{l12} & \cdots & \sigma_{l1K} \\ \sigma_{l12} & \sigma_{l2}^2 & \cdots & \sigma_{l2K} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{l1K} & \sigma_{l2K} & \cdots & \sigma_{lK}^2 \end{pmatrix}.$$

In this model, the dependence of the response variables appears at two levels: at the observations and at the linear predictors. This can be a point in favor of it, because it allows the exchange of information between the means, which are interpreted as the true values of indices, rates or proportions of interest. The model (7)-(10) assumes that information about K response variables and m second level units, with n_i first level units, $i = 1, ..., m$ are available.

The equation (8) relates the averages of the response variables in each i^{th} second level units, and considers specific second level unit effects. Thus, the mean μ_{ijk} and $\mu_{ijk'}$ also exchange information among themselves due to the fact that they are correlated.

3.1 Application to educational data

The data used to illustrate the application of the multivariate beta regression with random coefficients were extracted from the Second International Science Survey. This survey was carried out in 1984 by the International Association for the Evaluation of Educational Achievement. The data is described in Goldstein (2003) and it is available on the site of *MLWin* package, version 2.13. The data contain the results of six tests applied to 2439 Hungarian students in 99 schools. The number of students per school varies from 12 to 34, with mean 25. In order to reduce each test score to the same scale, Goldstein (2003) divided each test score by the total number of items in the test. Goldstein (2003) fitted a multivariate hierarchical normal model to the data. Here, we compare the goodness of fit of multivariate hierarchical model with the multivariate hierarchical beta model with and without copula. The two response variables used in all models were the scale scores in Biology and Physics, respectively denoted by Y_1 and Y_2. The variable gender of the student (X) was the single covariate employed. The indexes i, j and k respectively refer to school, student and the response variable. Because some values of the scores were found to be 0 or 1, we modified them by applying the transformation proposed by Smithson and Verkuilen (2006). Another alternative is to assign positive probabilities to 0 and 1, see Ospina and Ferrari (2010) for details.

The following three models were fit to the data:

Model 1: Two-level model proposed by Goldstein (2003), which assumes bivariate normal distribution for the response variables. It can be written as:

$$\begin{aligned} y_{ijk} &\sim N(\mu_{ijk}, \sigma_k^2), \quad i = 1, ..., 99, \quad j = 1, ..., n_i, \quad k = 1, 2 \\ \mu_{ijk} &= \beta_{1k} + x_{ij}\beta_{2k} + \nu_{ik} \\ \boldsymbol{\nu}_i &= (\nu_{i1}, \nu_{i2}) \sim N_2(\mathbf{0}, \mathbf{\Sigma}_\nu). \end{aligned}$$

Goldstein (2003) uses classical approach to make inference about the model parameters. Here we employed a Bayesian approach and assigned the following prior distribution to the model parameters: $\beta_{lk} \sim N(0, 10^{-6})$, $\sigma_k^{-2} \sim Gama(0.001; 0.001)$, $l = 1, 2$, $k = 1, 2$, and $\mathbf{\Sigma}_\nu^{-1} \sim Wishart(2, \mathbf{I}_2)$, where $\mathbf{I}_2$ is the 2×2 identity matrix.

Model 2: Multivariate beta hierarchical model without copula:

$$\begin{aligned}
y_{ijk} &\sim Beta(\mu_{ijk}, \phi_k), \quad i = 1, ..., 99, \quad j = 1, ..., n_i, \quad k = 1, 2 \\
g(\mu_{ijk}) &= \beta_{1k} + x_{ij}\beta_{2k} + \nu_{ik} \\
\boldsymbol{\nu}_i &= (\nu_{i1}, \nu_{i2}) \sim N_2(\mathbf{0}, \mathbf{\Sigma}_\nu),
\end{aligned}$$

with $\beta_{lk} \sim N(0, 10^{-6})$, $\phi_k \sim Gama(0.001; 0.001)$, $l = 1, 2$, $k = 1, 2$, and $\mathbf{\Sigma}^{-1} \sim Wishart(2, \mathbf{I}_2)$.

Model 3: Multivariate beta hierarchical model with Gaussian copula

$$\begin{aligned}
\mathbf{y}_{ij} &\sim BetaM(\boldsymbol{\mu}_{ij}, \boldsymbol{\phi}, \theta), \quad i = 1, ..., 99, \quad j = 1, ..., n_i, \quad k = 1, 2 \\
g(\mu_{ijk}) &= \beta_{1k} + x_{ij}\beta_{2k} + \nu_{ik} \\
\boldsymbol{\nu}_{i1} &= (\nu_{i1}, \nu_{i2}) \sim N_2(\mathbf{0}, \mathbf{\Sigma}_\nu),
\end{aligned}$$

with $\theta \sim U(-1, 1)$, $\beta_{lk} \sim N(0, 10^{-6})$, $\phi_k \sim Gama(0.001; 0.001)$, $l = 1, 2$, $k = 1, 2$, and $\mathbf{\Sigma}^{-1} \sim Wishart(2, \mathbf{I}_2)$.

A special program made in *Ox 5.10* was used to fit models 2 and 3, while to fit model 1 we used *Winbugs 1.4.3*. The three models were compared by using the following criteria measure: AIC, BIC, DIC and predictive likelihood $\Big((L(\hat{\Psi})\Big)$.

Table 10 shows the values of AIC, BIC, DIC (with the contribution p_D) and the predictive likelihood $(L(\hat{\Psi}))$, where Ψ denotes the vector of parameters of the corresponding model. For the model 3, the effective number of parameters estimated by p_D is well above others, because it is a more complex model. However, the values of other statistics are quite lower than those of the other two models, indicating that it has the best adjustment. It is worth noting that the DIC for the normal model is much larger than the ones that assume beta distribution. According to DIC criterion, the most appropriate model is the more complex model. The predictive likelihood criterion leads to the same conclusion.

Table 11 shows summary measures of the posterior for the parameters of the three models. It should be noted that no 95% credible interval contains the zero. The sex of the student is an important factor for explaining both responses. The 95% credible interval for the degree of the association τ between the responses at the student level is

Table 10: DIC, AIC, BIC, number of effective parameters and the logarithm of predictive likelihood for the three models

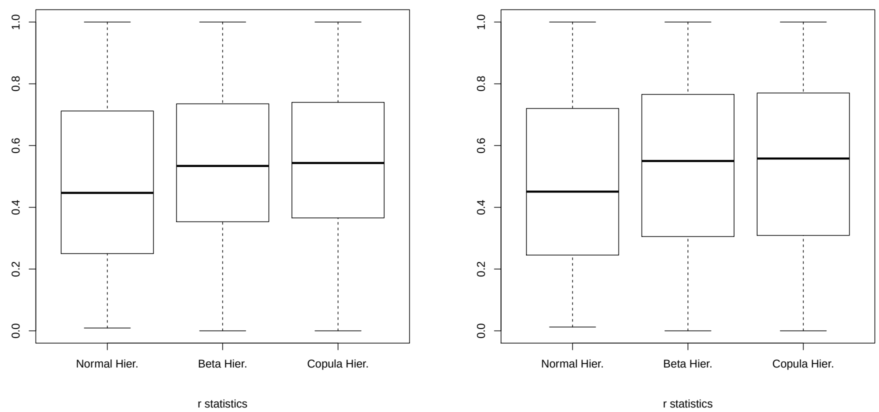
Model	DIC	AIC	BIC	p_D	$\log p(\Psi)$
1	-3314.22	-3682.50	-3614.51	189.14	1982.06
2	-5350.87	-5529.13	-5516.16	188.26	2769.57
3	-5667.59	-5851.15	-5838.18	193.57	2930.58

given by $(0.199, 0.245)$, indicating that there is some association, even being low. For θ, we have $(0.308, 0.376)$. The analysis of DIC and the predicted values showed that it is important to include this parameter in the model. The correlation at the school level is high with an average of 0.756 and 0.77 for the models 2 and 3, respectively.

Table 11: Summary measures of the posterior for models 2 and 3

Parameter	Model 2					Model 3				
	2.5%	50%	97.5%	Mean	Std.	2.5%	50%	97.5%	Mean	Std.
β_{11}	0.896	1.018	1.141	1.018	0.063	0.906	1.031	1.156	1.031	0.064
β_{21}	-0.156	-0.087	-0.018	-0.087	0.035	-0.148	-0.076	-0.003	-0.076	0.038
β_{12}	1.174	1.323	1.469	1.322	0.074	1.204	1.350	1.500	1.350	0.076
β_{22}	-0.500	-0.418	-0.329	-0.418	0.044	-0.505	-0.420	-0.338	-0.420	0.043
ϕ_1	4.169	4.408	4.654	4.408	0.122	4.200	4.439	4.684	4.440	0.123
ϕ_2	2.939	3.107	3.290	3.111	0.089	2.986	3.154	3.333	3.156	0.089
σ_1^2	0.244	0.330	0.454	0.335	0.053	0.237	0.321	0.443	0.326	0.052
σ_1	0.494	0.574	0.673	0.577	0.046	0.486	0.566	0.665	0.569	0.045
σ_2^2	0.337	0.455	0.628	0.462	0.075	0.342	0.459	0.634	0.466	0.075
σ_2	0.580	0.675	0.792	0.678	0.054	0.584	0.677	0.796	0.681	0.054
σ_{12}	0.207	0.292	0.413	0.297	0.052	0.211	0.297	0.422	0.302	0.054
ρ_{12}	0.644	0.759	0.846	0.756	0.052	0.663	0.778	0.859	0.773	0.050
θ	-	-	-	-	-	0.308	0.343	0.376	0.342	0.017
τ	-	-	-	-	-	0.199	0.223	0.245	0.223	0.012

Figure 2 shows the graphics of the statistics $r_{ijk} = P(y_{ijk} < y^*_{ijk})$, where y^*_{ijk} is the random predict value of y_{ijk}, for both response variables and the three models compared. The ideal situation is that the value of r_{ijk} be near to 0.5, indicating that there is neither underestimation or overestimation. It can be seen from Figure 2 that the multivariate beta models have quite similar performance with respect to the r measure and on average perform better than the multivariate normal model.



(a) Biology (b) Physics

Figure 2: Boxplots of the r statistics, considering the hierarchical models with normal responses, beta responses and beta responses with Gaussian copula for the (a) Biology and (b) Physics scores.

4 Final Remarks and future work

The models proposed have the advantage of keeping the response variables in their original scale. Another advantage refers to the use of copulas which are marginal-free, i.e, the degree of association of variables is preserved whatever the marginal distributions are. Thus, if two indexes are correlated whatever the marginal adopted, the measure of dependence is the same. The use of copula functions in beta marginal regressions allows to jointly analyze the response variables, by taking advantage of their dependency structure and keeping the variables in their original scale. 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 its competitors in the literature should be guided by the goals of the researcher, who must observe the predictive power and the goodness of fit of them. The disadvantage of models that uses copulas is their time consuming for simulating samples from the posterior distributions of the model parameters or functions of them.

The application with the poverty indexes data show that there is no much difference between the univariate and multivariate models with respect to the estimation of their common parameters. However, the criteria for model selection, pointed to the choice of the model that makes use of Frank copula, suggesting that this copula fit better to the data used in our application. Estimates of the parameters and predictions were similar for all models, which makes us to conclude that the choice of the copula function is not too relevant for this application. The dependence between the response variables in the application data set was low and thus bivariate fit has not get any improvement when compared with the univariate fit. However, as shown in the simulation study, as the measure of dependence between the response variables increases, the greater is the improvement of the bivariate model over the univariate one. In situations where the dependence is high, the use of the bivariate model might be quite worth.

The analysis of the results obtained for the second application shows that the the use of beta distribution for fitting response variables on the interval $(0, 1)$ is likely to yield better fitting than the customary normal model. Moreover, the introduction of the random coefficient model for the beta regression seems to be useful for modeling intra-class correlation within nested level units. However, the parametrization of the random coefficient used for making inference of the hierarchical model parameters, seems to be essential for achieving fast convergence when MCMC is employed.

It is important to note that this work focuses on building multivariate regression models in which the marginal distributions are Beta. It points out its advantages over corresponding univariate models and the difficulties of estimating their parameters. However, the theory of copula functions can be applied to any multivariate models that can be built for any known marginal distributions, allowing that the distributions of response variables involved be different. We can even have continuous and discrete variables in the same model. To build a model for others distributions is straightforward, but each model has a peculiar and practical feature, and the estimation process should always be taken into account when we propose a new model. In the specific case of the Beta model, has been adopted the mean and the dispersion as the model parameters, where the latter parameter controls the variance. Other parameterizations are possible, but could lead to additional difficulties. Various strategies can be defined by the researcher, according to the available database, some important ones are: first fixe the marginal and then obtain the more appropriate copulas; estimate models with different copulas and marginal and decide what is "the best" model by applying a model comparison approach.

We have not considered omission in the explanatory variables in our model formulation, which could be another possible extension of the models proposed here. Further work should also be done for obtaining objective priors for the univariate and bivariate models.

Acknowledgments

This work is part of the Phd dissertation of Debora F. S., under the supervision of Fernando Moura, in the Graduate Program of UFRJ. Fernando Moura receives financial support from the National Council for Scientific and Technological Research (CNPq-Brazil, BPPesq).

Appendix 1: Computational issues

This appendix describes the computational details for sampling from the posterior distributions of the Multivariate hierarchical beta regression model (MHBR) via MCMC. Let denote $\mathbf{W}_l = \boldsymbol{\Sigma}_l^{-1}$ and $\mathbf{W}_{1:p} = \{\mathbf{W}_1, ..., \mathbf{W}_p\}$. Thus, the likelihood of the MHBR model is given by:

$$L(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta}, \mathbf{W}_{1:p}) = p(\boldsymbol{\nu}|\mathbf{W}_{1:p})p(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\theta}).$$

Developing the two terms of it, we have:

$$\begin{aligned}
p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{W}_{1:p}) &= \prod_{i=1}^{m}\prod_{j=1}^{n_i} c\left(F_1(y_{ij1}), ..., F_K(y_{ijK})|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\theta}\right) \prod_{k=1}^{K} p(y_{ijk}|\phi_k, \mu_{ijk}) \\
&= \prod_{i=1}^{m}\prod_{j=1}^{n_i} c\left(F_1(y_{ij1}), ..., F_K(y_{ijK})|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\theta}\right) \\
&\quad \times \prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{k=1}^{K} \Gamma(\phi_k) \frac{y_{ijk}^{\phi_k\mu_{ijk}-1}(1-y_{ijk})^{\phi_k(1-\mu_{ijk})-1}}{\Gamma(\phi_k\mu_{ijk})\Gamma(\phi_k(1-\mu_{ijk}))} \\
&\propto \prod_{i=1}^{m}\prod_{j=1}^{n_i} c\left(F_1(y_{ij1}), ..., F_K(y_{ijK})|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\theta}\right) \\
&\quad \times \Gamma(\phi_k)^{\sum n_i} \prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{k=1}^{K} \frac{y_{ijk}^{\phi_k\mu_{ijk}}(1-y_{ijk})^{\phi_k(1-\mu_{ijk})}}{\Gamma(\phi_k\mu_{ijk})\Gamma(\phi_k(1-\mu_{ijk}))}
\end{aligned}$$

and

$$\begin{aligned} p(\boldsymbol{\lambda}|\boldsymbol{\beta}, \mathbf{W}_{1:p}, \boldsymbol{\phi}, \boldsymbol{\theta}) &= \prod_{l=1}^{p} p(\boldsymbol{\lambda}_l|\mathbf{W}_l, \boldsymbol{\beta}_{l\cdot}) = \prod_{l=1}^{p}\prod_{i=1}^{m} p(\boldsymbol{\lambda}_{il\cdot}|\mathbf{W}_l, \boldsymbol{\beta}_{l\cdot}) \\ &= \propto \prod_{i=1}^{m} |\mathbf{W}_l|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}_{il\cdot} - \boldsymbol{\beta}_{l\cdot})^T \mathbf{W}_l (\boldsymbol{\lambda}_{il\cdot} - \boldsymbol{\beta}_{l\cdot})\right\}, \end{aligned}$$

where $\boldsymbol{\lambda}_{il\cdot} = (\lambda_{il1}, ..., \lambda_{ilK})^T$, $i = 1, ..., M$ and $\boldsymbol{\beta}_{l\cdot} = (\beta_{l1}, ..., \beta_{lK})$ is the l^{th} raw of $\boldsymbol{\beta}$, $l = 1, ..., p$. Thus, the posterior density of all model parameters are:

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{W}_{1:p}|\mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{W}_{1:p}) p(\boldsymbol{\lambda}|\boldsymbol{\beta}, \mathbf{W}_{1:p}, \boldsymbol{\phi}, \boldsymbol{\theta}) p(\boldsymbol{\beta}) p(\boldsymbol{\phi}) p(\boldsymbol{\theta}) p(\mathbf{W}_{1:p}) \\ &\propto p(\mathbf{y}|\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}) p(\boldsymbol{\lambda}|\boldsymbol{\beta}, \mathbf{W}_{1:p}) p(\boldsymbol{\theta}) \left\{\prod_{k=1}^{K} p(\boldsymbol{\phi}_k)\right\} \left\{\prod_{l=1}^{p} p(\boldsymbol{\beta}_{l\cdot}) p(\mathbf{W}_l)\right\}, \end{aligned}$$

The posterior distribution above has no close form. However some of its full conditional have, provided that are assigned independent normal priors to $\boldsymbol{\beta}_{l\cdot}$ and independent Wishart priors to $\mathbf{W}_l = \boldsymbol{\Sigma}_l^{-1}$, $l = 1, ..., p$:

$$\begin{aligned} p(\boldsymbol{\beta}_{l\cdot}|\boldsymbol{\beta}_{(-l)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{W}_{1:p}, \mathbf{y}) &\propto p(\boldsymbol{\beta}_{l\cdot}) \prod_{i=1}^{m} p(\boldsymbol{\lambda}_{il\cdot}|\boldsymbol{\beta}_{l\cdot}, \mathbf{W}_l) \\ &\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_{l\cdot}^T \left(m\mathbf{W}_l + \mathbf{B}_l^{-1}\right) \boldsymbol{\beta}_{l\cdot} \right.\right. \\ &\qquad \left.\left. -2\boldsymbol{\beta}_{l\cdot}^T \left(\mathbf{W}_l \sum_{i=1}^{m} \boldsymbol{\lambda}_{il\cdot} + \mathbf{B}_l^{-1}\mathbf{b}_l\right)\right]\right\}; \\ p(\mathbf{W}_l|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{W}_{(-l)}, \mathbf{y}) &\propto p(\mathbf{W}_l) \prod_{i=1}^{m} p(\boldsymbol{\lambda}_{il\cdot}|\boldsymbol{\beta}_{l\cdot}, \mathbf{W}_l) \\ &\propto |\mathbf{W}_l|^{(d_l-K-1)/2} \exp\left\{-\frac{1}{2} tr(\mathbf{D}_l \mathbf{W}_l)\right\} \times \\ &\qquad \times \prod_{i=1}^{m} |\mathbf{W}_l|^{1/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{\lambda}_{il\cdot} - \boldsymbol{\beta}_{l\cdot})^T \mathbf{W}_l (\boldsymbol{\lambda}_{il\cdot} - \boldsymbol{\beta}_{l\cdot})\right\}. \end{aligned}$$

Thus, the full conditional of $\boldsymbol{\beta}_{l\cdot}$ is bivariate normal distributed with mean $\mathbf{b}_l^*$ and variance-covariance matrix $\mathbf{B}_l^*$ where

$$\mathbf{B}_l^{*-1} = m\mathbf{W}_l + \mathbf{B}_l^{-1} \quad \text{e} \quad \mathbf{b}_l^* = \mathbf{B}_l^* \left(\mathbf{B}_l^{-1}\mathbf{b}_l + \mathbf{W}_l \sum_{i=1}^{m} \boldsymbol{\lambda}_{il\cdot}\right).$$

The full conditional of $\mathbf{W}_l$ is Wishart distributed with parameters $d_l + m$ and $\mathbf{D}_l + \sum_{i=1}^{m}(\boldsymbol{\lambda}_{il\cdot} - \boldsymbol{\beta}_{l\cdot})(\boldsymbol{\lambda}_{il\cdot} - \boldsymbol{\beta}_{l\cdot})^T$.

The remaining conditional distributions have no close forms and the Metropolis-Hastings algorithm is employed to sample from them. The kernels of these distributions are given bellow:

$$
\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{W}_{1:p},\mathbf{y}) &\propto \prod_{i=1}^{m}\prod_{j=1}^{n_i} c\left(F_1(y_{ij1}),...,F_K(y_{ijK})|\boldsymbol{\phi},\boldsymbol{\mu},\boldsymbol{\theta}\right),\\
p(\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\theta},\mathbf{W}_{1:p},\mathbf{y}) &\propto \left\{\prod_{i=1}^{m}\prod_{j=1}^{n_i} c\left(F_1(y_{ij1}),...,F_K(y_{ijK})|\boldsymbol{\phi},\boldsymbol{\mu},\boldsymbol{\theta}\right)\right.\\
&\quad \left.\times\prod_{k=1}^{K} p(y_{ijk}|\phi_k,\mu_{ijk})\right\}\prod_{i=1}^{m}\prod_{l=1}^{p} p(\boldsymbol{\lambda}_{il\cdot}|\boldsymbol{\beta}_{l\cdot},\mathbf{W}_l),\\
p(\boldsymbol{\phi}_k|\boldsymbol{\beta},\boldsymbol{\phi}_{(-k)},\boldsymbol{\theta},\boldsymbol{\lambda},\mathbf{W}_{1:p},\mathbf{y}) &\propto p(\phi_k)\prod_{i=1}^{m}\prod_{j=1}^{n_i} c\left(F_1(y_{ij1}),...,F_K(y_{ijK})|\boldsymbol{\phi},\boldsymbol{\mu},\boldsymbol{\theta}\right)\\
&\quad \times\prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{k=1}^{K} p(y_{ijk}|\phi_k,\mu_{ijk}),
\end{aligned}
$$

$k = 1, ..., K$.

References

Ferrari, S. L. P., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. Journal of Applied Statistics 31 (7), 799–815.

Gelfand, A. E., Sahu, S. K., Carlin, B. P., 1995. Efficient parametrisations for normal linear mixed models. Biometrika 82 (3), 479–488.

Goldstein, H., 2003. Multilevel Statistical Models, 3rd Edition. Edward Arnold, London.

Huard, D., Évin, G., Favre, A.-C., 2006. Bayesian copula selection. Computational Statistics and Data Analysis 51, 809–822.

Jiang, J., 2007. Linear and Generalized Linear Mixed Models and Their Applications. Springer Series in Statistics. Springer, New York.

Nelsen, R. B., 2006. An Introduction to Copulas, 2nd Edition. Springer, New York.

Olkin, I., Liu, R., 2003. A bivariate beta distribution. Statistics and Probability Letters 62, 407–412.

Ospina, R., Ferrari, S. L. P., 2010. Inflated beta distributions. Statistical Papers 51, 111–126.

Silva, R. S., Lopes, H. F., September 2008. Copula, marginal distributions and model selection: a bayesian note. Statistics and Computing 18 (3), 313–320.

Smithson, M., Verkuilen, J., 2006. A better lemon-squeezer? maximum likelihood regression with beta-distributed dependent variables. Psychological Methods 11, 54–71.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society Series B 64 (3), 583–639.