

Spatial Beta Regression Model with Random Effect

Lida Kalhori and Mohsen Mohammadzadeh*

Tarbiat Modares University

Received: 9/16/2016 Approved: 4/8/2017

Abstract. In many applications we have to encountered with bounded dependent variables. Beta regression model can be used to deal with these kinds of response variables. In this paper we aim to study spatially correlated responses in the unit interval. Initially we introduce spatial beta generalized linear mixed model in which the spatial correlation is captured through a random effect. Then the performances of the proposed model is evaluated via a simulation study, implementing Bayesian approach for parameter estimation. Finally the application of this model on two real data sets about migration rate and divorce rate in Iran are presented.

Keywords. Bayesian estimation; beta regression model; spatial correlation; random effect.

MSC 2010: 91B79; 62P25; 62F15.

1 Introduction

In practice, bounded continuous response variables are very common, specially rates and proportions which are limited in $(0, 1)$ interval. Examples of these variables in official statistics include unemployment rate, illiteracy rate, fertility rate, migration rate, response rate, proportion of time devoted

* Corresponding author

to an activity, proportion of expenses devoted to food. Regression models are widely used to explore the relationship between dependent variables and covariates. The basic assumptions to utilize regression models are independence and normality of the response variable, which may not be realistic for responses in $(0, 1)$ interval. Paolino (2001) assumed that such response variables follow beta distribution due to the support of this distribution. Ferrari and Cribari (2004) proposed the reparametrized beta distribution and applied the maximum likelihood method for parameter estimation, assuming the precision parameter is constant. Cepeda and Gamerman (2005), and Smithson and Verkuilen (2006) applied Bayesian approach to model both the mean and the precision parameters of beta distribution through logit and logarithm link functions respectively. Branscum et al. (2007) studied semi parametric beta regression model using penalized splines. Bayesian approach had been also used by Simas et al. (2010) to study nonlinear beta regression model. Zimprich (2010) added a random effect in the mean model to study reaction time of old people in a longitudinal survey. Further development of mixed beta regression models has been done by Figueroa et al. (2013). They proposed mixed models for both mean and precision parameters and implemented Bayesian approach for parameter estimation. Ferreira et al. (2015) used Bayesian approach to study time series data in $(0, 1)$ interval with correlated errors, they applied their proposed model to study Brazilian and Chilean monthly unemployment rate.

Cepeda et al. (2012) extended the beta regression model for spatial data analysis, in which the spatial effect was included in the model through the distance points as a fixed explanatory variable. This model was applied to study the quality of education in Columbia by Cepeda and Nunez (2013). Spatial analysis of structured additive regression model was done using Integrated Nested Laplace Approximation (INLA) by Gholizadeh et al. (2013). They modeled crime rate data in Tehran as an application of their model. In this paper we focus on mixed effect models to study spatially correlated response variable in $(0, 1)$ interval. We develop beta regression model by adding a random effect to the beta regression model which enables us to include the spatial correlation structure of data in the model.

This paper is arranged in the following order: Section 2 presents a review on beta regression and spatial beta regression models. Section 3 contains Bayesian inference for spatial beta regression model, Section 4 allocated to assess three alternative mixed models to study spatially correlated bounded response variables. These models include regression model for response vari-

able, transformed response variable, and beta regression model. We implemented an intensive simulation study on regular grids for different sample sizes. Section 5 devoted to utilization of the under studied models on two real data sets. The first application is about immigration rate, the data source is the Population and Household Census (PHC) conducted by Statistical Center of Iran. In the second application we investigate divorce rates, information are provided by National Organization for Civil Registration. This paper will be ended by conclusions and further remarks.

2 Regression for Beta Distributed Response

When a variable Y takes values in (a, b) interval, then the realizations of the transformed variable $Y' = \frac{Y-a}{b-a}$ are in $(0, 1)$. If $Y \sim \text{Beta}(a, b)$ then the density function of Y is given by

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}, \quad 0 < y < 1,$$

where $a > 0$, $b > 0$, $\Gamma(\cdot)$ denotes the gamma function and the mean and variance of Y are given by $E(Y) = \frac{a}{a+b}$ and $Var(Y) = \frac{ab}{(a+b)^2(1+a+b)}$, respectively. Beta distribution is very flexible and takes different shapes including uniform, skewed and symmetric in terms of various values of its parameters, (Figure 1).

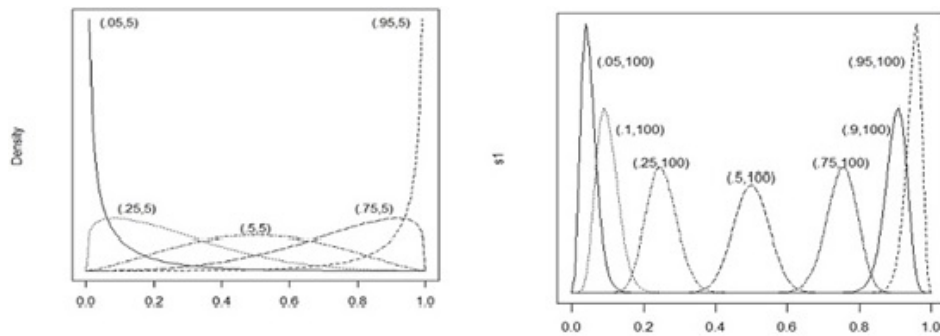


Figure 1. Different shapes of beta distribution for different values of its parameters

For modeling response variable with distribution $\text{Beta}(a, b)$, Paolino (2001) implemented exponential link functions $\alpha = \exp(a)$ and $\gamma = \exp(b)$ for

both parameters a and b . To make a connection between covariates and mean of beta distribution, Ferrari and Cribari (2004) proposed a different parametrization of beta distribution assuming $\mu = \frac{a}{a+b}$, $0 < \mu < 1$, and $\phi = a + b$, $\phi > 0$, therefore $E(Y) = \mu$ and $Var(Y) = \frac{\mu(1-\mu)}{\phi+1}$. Since $Var(Y)$ and ϕ are inversely related to each other, ϕ is usually called precision parameter. Upon this reparametrization, such a distribution is denoted by $Beta(\mu\phi, (1-\mu)\phi)$ and its density function is given by

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

Let $Y_i, i = 1, \dots, n$ be independent random variables with distribution $Beta(\mu_i\phi, (1-\mu_i)\phi)$. Ferrari and Cribari (2004) defined the beta regression model by

$$g(\mu_i) = \sum_{j=0}^{k-1} x_{ij}\beta_j, \quad i = 1, \dots, n,$$

where $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{k-1}) \in R^k$ is the vector of regression coefficients, $x_i = (1, x_{i1}, \dots, x_{i(k-1)})$, are covariates and $g(\cdot)$ is a real valued continuous twice differentiable link function.

In situations, where the precision parameter is not constant over all sampling units, the mean and the precision parameter can be modeled as distinct functions of the explanatory variables as follows:

$$g(\mu_i) = \sum_{j=0}^{k-1} x_{ij}\beta_j, \quad h(\phi_i) = \sum_{\ell=0}^{m-1} w_{i\ell}\vartheta_{\ell}, \quad i = 1, \dots, n, \quad (2)$$

where $x_i = (1, x_{i1}, \dots, x_{i(k-1)})$ and $w_i = (1, w_{i1}, \dots, w_{i(m-1)})$ are two sets of covariates related to the mean and precision parameters, respectively. The link functions $g(\cdot)$ and $h(\cdot)$ are strictly monotonic and twice differentiable, that relate the linear predictors $\sum_{j=0}^{k-1} x_{ij}\beta_j$ and $\sum_{\ell=0}^{m-1} w_{i\ell}\vartheta_{\ell}$ to the mean and the precision parameters of beta distribution, respectively. When we are dealing with rates and proportions we have $0 < \mu < 1$, so a link function needs to map the $(0, 1)$ interval to the entire real line. Different choices are available for the link function $g(\cdot)$. For instance the logit function $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$, the probit function $g(\mu) = \Phi^{-1}(\mu)$ where $\Phi(\cdot)$ is the Normal cumulative distribution function, the complimentary log-log function

$g(\mu) = \log(-\log(1 - \mu))$. The link function for the precision parameter need to satisfy the condition that maps the nonnegative values to the whole real line. One possible choice is the logarithm function $h(\phi) = \log(\phi)$, (McCullagh and Nelder, 2004). Logit and logarithm link functions are applied in literature for modeling beta distributed response variables.

In some applications the values of the response variable depend on the locations of sampling units, so they are spatially correlated. This correlation should be considered in model fitting. Suppose that the response variable at location s_i , $i = 1, \dots, n$ is beta distributed, that is $Y(s_i) \sim \text{Beta}(\mu_i\phi, (1 - \mu_i)\phi)$. Further, assume that $\mathbf{y} = (y(s_1), \dots, y(s_n))^T$ are realizations of random variables $\mathbf{Y} = (Y(s_1), \dots, Y(s_n))^T$ at n distinct locations s_1, \dots, s_n . For simplicity let y_i and Y_i denote the $y(s_i)$ and $Y(s_i)$ respectively, in the following parts. Cepeda et al. (2012) include the spatial neighbors interaction effects in the model, using the spatial lags as explanatory variable in the model as follow

$$g(\mu_i) = \sum_{j=0}^{k-1} x_{ij}\beta_j + \eta(\mathbf{W}\mathbf{y})_i, \quad i = 1, \dots, n,$$

where \mathbf{W} is a spatial weight matrix that specifies the effects of geographical neighborhoods of sample units and η is a regression coefficient related to the spatial term. In this paper, we aim to incorporate the spatial correlation structure of data into the model by adding a random field, through which the spatial correlation can be considered via a specified spatial covariance function. Models of this kind belong to the class of Spatial Generalized Linear Mixed Models (SGLMM) introduced by Diggle et al. (1998).

For beta distributed spatially correlated response variable, we define the spatial beta regression model with random effect as bellow

$$g(\mu_i) = \sum_{j=0}^{k-1} x_{ij}\beta_j + \tau_i, \quad i = 1, \dots, n, \quad (3)$$

where $\tau_i = \tau(s_i)$ denotes a random effect for unit i . The spatial random effect $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$ is a zero mean Gaussian random process, $\boldsymbol{\tau} \sim N_n(0, \Sigma)$, where the spatial covariance structure determined by Σ . Next section is devoted to implement the Bayesian approach for parameter estimation of this spatial beta regression model.

3 Bayesian Approach for Parameter Estimation

According to properties of the spatial random effect in (3) which are mentioned previously, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$ follows a zero mean multivariate normal distribution with covariance matrix Σ . Assume that Σ_{ij} represents the ij th element of Σ and $\Sigma_{ij} = \sigma_{ij}^2 = \sigma_0^2 \exp(\psi|s_i - s_j|)$, where σ_0^2 is the spatial variance and ψ^{-1} is the range parameter. For Bayesian analysis of the spatial beta regression model defined in (3), we apply the following priors

$$\begin{aligned} \beta_j &\sim N(0, 10^p), \quad j = 0, \dots, k-1, & \phi &\sim \text{Gamma}(\xi_1, \xi_2), \\ \sigma_0^{-2} &\sim \text{Gamma}(\epsilon_1, \epsilon_2), & \psi^{-1} &\sim U(\alpha_1, \alpha_2). \end{aligned}$$

If $\boldsymbol{\theta} = (\sigma_0^{-2}, \psi)$ is the vector of parameters related to the random component, then the joint posterior distribution is given by

$$\pi(\boldsymbol{\beta}, \phi, \boldsymbol{\tau}, \boldsymbol{\theta} \mid \mathbf{y}) \propto \prod_{i=1}^n \pi(y_i \mid \boldsymbol{\beta}, \phi, \tau_i) \prod_{i=1}^n \pi(\tau_i \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\beta}) \pi(\phi).$$

The posterior distribution is complicated, so MCMC method is applied for Bayesian inference. Parameters were estimated by using Gibbs sampling from full conditional distributions via R2WinBugs package (Sturtz et al., 2005).

4 Simulation Study

Using ordinary regression model is a common way of modeling continuous response variables. However, when the response variable is a proportion, the transformed response variable can be modeled. Atkinson (1985) proposed the logit function to transform proportions from the interval $(0, 1)$ to \mathbb{R} . Another way to study the continuous bounded response variables that is introduced and mentioned in the last two decades is beta regression model. Thus, the main aim of this section is to compare the following three models as alternatives, in order to study spatially correlated response variable in $(0, 1)$ interval.

$$\text{Model 1: } E(Y_i \mid \boldsymbol{\beta}, \tau_i) = \sum_{j=0}^{k-1} x_{ij} \beta_j + \tau_i,$$

$$\text{Model 2: } E(\text{logit}(Y_i) \mid \boldsymbol{\beta}, \tau_i) = \sum_{j=0}^{k-1} x_{ij} \beta_j + \tau_i,$$

$$\text{Model 3: } \text{logit}(E(Y_i | \boldsymbol{\beta}, \tau_i)) = \sum_{j=0}^{k-1} x_{ij}\beta_j + \tau_i.$$

In Models 1 and 2 we assume that Y_i and $\text{logit}(Y_i)$ are normally distributed, so the identity link function is appropriate to link the linear predictors to the mean parameter. In Model 3 that is the spatial beta regression model with random effect, we assume that the response variable is beta distributed, and we use the logit link function.

To evaluate performances of these three models for modeling spatially correlated data in $(0, 1)$ interval, we go through a simulation study. To have both positive and negative regression coefficients and considering the previous studies on beta regression model, the considered values for the model parameters are $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2) = (-1, 2, -1.5)$, $\phi = 50$, $\sigma_0^2 = 0.5$, $\psi^{-1} = 0.1$ and covariates are generated from uniform distribution. Prior distributions are $\beta_j \sim N(0, 10^2)$ for $j = 0, 1, 2$, $\phi \sim \text{Gamma}(0.01, 0.01)$, $\sigma_0^{-2} \sim \text{Gamma}(0.01, 0.01)$, and $\psi^{-1} \sim U(0.001, 20)$. To explore the effect of sample sizes on inference, three different sample sizes of 25, 100 and 225 are considered. For generated data set of size 225 we simulate one chain of size 100,000 for each parameter, and burning the first 50,000 iterations to eliminate the effect of the initial values which were chosen randomly. To avoid correlation problems we considered a spacing of size 100, obtaining an effective sample of size 500 for posterior inference. For samples of size 100 and 25 we need more iterations to reach convergence.

Diagnostic tests of convergence were implemented and there was no evidence of chains divergence. To do the Gelman-Rubin test of convergence (Brooks and Gelman, 1998), two chains with different random initial values were generated simultaneously. The Multivariate Proportional Scale Reduction Factor (MPRF) for all three models are approximately 1.00, which is lower than 1.2, so convergence has not rejected. Geweke's test (Geweke, 1991) and Heidelberg and Welch test (Heidelberg and Welch, 1981) are also shown the convergence of the results of Gibbs samplers for each parameter. The reference diagnostic tests were done using the Coda package in R. Autocorrelation plots of spatial beta regression model parameters in Figure 2, indicate that the chains provided for the parameters are not correlated. The corresponding plots of the other two models are shown the same manner and are not presented for the sake of brevity.

The estimates of the model parameters along with their precision criteria, Bias and RMSE, are calculated based on 100 simulated data sets for each model. The results are summarized in Table 1.

Table 1. Estimates, Bias and RMSE of the model parameters

n	Par.	Value	Model 1			Model 2			Model 3		
			Est.	Bias	RMSE	Est.	Bias	RMSE	Est.	Bias	RMSE
25	β_0	-1	0.286	-1.286	1.286	-1.087	0.087	0.225	-1.071	0.071	0.224
	β_1	2	0.394	-0.803	1.607	2.012	0.006	0.289	2.113	0.056	0.282
	β_2	-1.5	-0.288	-0.808	1.213	-1.388	-0.075	0.270	-1.448	-0.035	0.262
	ϕ	50	124.837	-0.497	159.907	25.983	-0.480	24.363	62.071	0.241	32.304
	σ_0^2	0.5	0.063	-0.874	0.441	0.074	-0.851	0.427	0.270	-0.460	0.504
	ψ^{-1}	0.1	165.692	1655.918	175.078	16.620	165.197	17.917	6.008	59.076	7.686
100	β_0	-1	0.294	-1.294	1.295	-1.041	0.041	0.161	-0.901	-0.099	0.565
	β_1	2	0.421	-0.790	1.580	2.066	0.033	0.178	2.002	0.001	0.108
	β_2	-1.5	-0.299	-0.800	1.201	-1.589	0.059	0.187	-1.564	0.043	0.172
	ϕ	50	129.513	1.590	153.031	31.564	-0.369	42.432	33.724	-0.326	39.947
	σ_0^2	0.5	1.764	0.176	1.494	1.846	2.692	1.563	0.285	-0.431	0.277
	ψ^{-1}	0.1	124.315	1242.146	128.598	3.371	32.708	3.472	1.080	9.799	1.908
225	β_0	-1	0.295	-1.295	1.295	-0.993	-0.007	0.140	-0.954	-0.046	0.374
	β_1	2	0.403	-0.798	1.597	2.052	0.026	0.104	2.046	0.023	0.199
	β_2	-1.5	-0.298	-0.801	1.203	-1.566	0.044	0.123	-1.514	0.009	0.158
	ϕ	50	256.827	4.137	213.868	20.218	-0.596	31.682	67.916	0.358	25.412
	σ_0^2	0.5	1.863	2.727	1.891	3.090	5.180	3.164	0.342	-0.316	0.258
	ψ^{-1}	0.1	130.254	1301.544	135.126	4.990	48.899	5.823	0.301	2.014	0.317

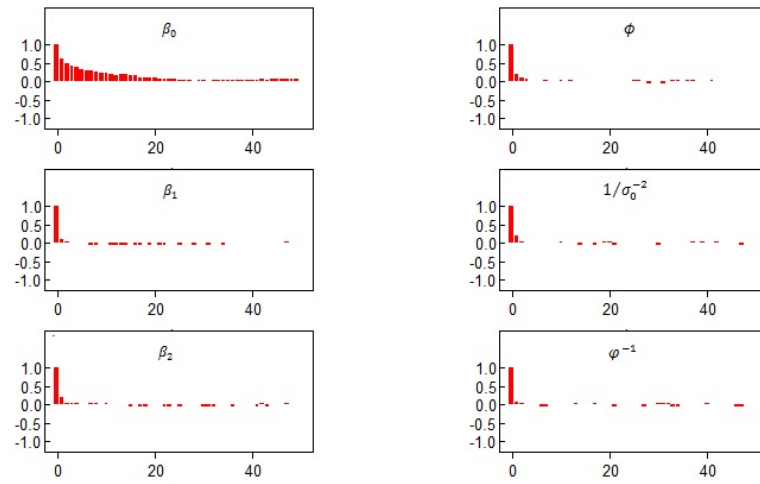


Figure 2. Autocorrelation plots of parameters of spatial beta regression model

Considering the estimated values of the parameters for various sample sizes, it is clear that the Model 1 does not fit properly. The regression parameters β are similarly estimated for two other models. But for the parameters ϕ , σ_0^2 and ψ^{-1} , the Model 3 is more satisfactory, because the Bias and RMSE of the parameter estimates are smaller than the Bias and RMSE of the Model 2, especially for smaller sample sizes. It is also worth to note that Model 2 may be make predictions out of logical range of response value which is $(0, 1)$ interval. So, we recommend to use the spatial beta regression model for spatially structured responses in $(0, 1)$ interval. Prediction surface in a grid of size 100 using Model 3 is presented in Figure 3.

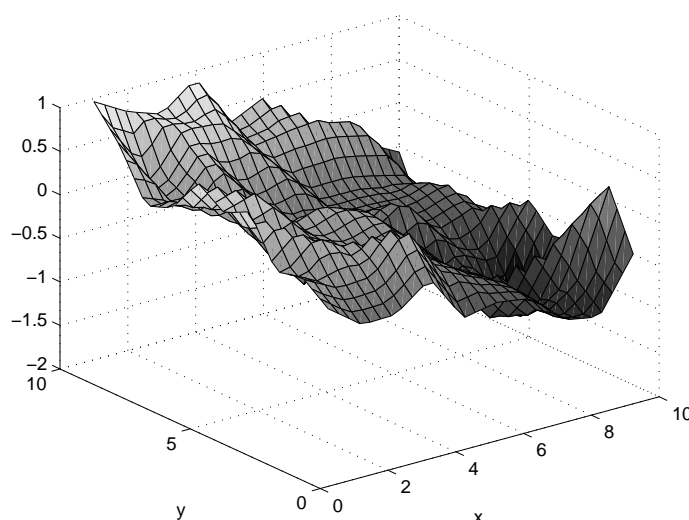


Figure 3. Prediction surface of the spatial random effect

5 Applications

Population and Household Census which is conducted every five years by Statistical Center of Iran provide information in many aspects of population characteristics and composition. One of the main processes that caused changes in demographic characteristics is the migration phenomena. Another issue that seems to be critical is increasing of divorces rates. In this section

we are going to use the spatial beta regression model with random effect to analyze two real data sets for considering these subjects. The first data set contains migration data provided by Population and Household Census (PHC) conducted by Statistical Center of Iran. Proportion of immigrant who comes to each province is our response variable and we will call it migration rate. Second data set contains information about number of divorces in all provinces provided by National Organization for Civil Registration, covariates information are handed by PHC. In this data set we are interested to study a proportion defined by the number of divorces to the number of marriage and consider it as divorce rate. In the following subsections we will distinctly explain covariates for each data set and aim to analyze response variables using Models 1, 2 and 3 introduced in previous sections, and choose the best model among them based on DIC criterion.

5.1 Immigration rate among provinces of Iran

In this section we mention proportion of immigrants who comes to each province from other provinces and refer to it as migration rate. We consider performance of cost credit of executive agencies from public income (PI), value added (VA), unemployment rate (UR), and proportion of population with higher education (HE) as covariates. We initially interested to investigate the following three models

$$\text{Model 1*}: y_i = \beta_0 + \beta_1 PCI_i + \beta_2 VA_i + \beta_3 UR_i + \beta_4 HE_i + \tau_i,$$

$$\text{Model 2*}: \log\left(\frac{y_i}{1-y_i}\right) = \beta_0 + \beta_1 PCI_i + \beta_2 VA_i + \beta_3 UR_i + \beta_4 HE_i + \tau_i,$$

$$\text{Model 3*}: \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 PCI_i + \beta_2 VA_i + \beta_3 UR_i + \beta_4 HE_i + \tau_i,$$

where $i=1,\dots,31$. We fitted the models and estimated the parameters implementing the Bayesian approach using the priors introduced in Section 4, for the fixed precision parameter case. Going through the models step by step, yielded to the following final results for each of these models and model

Table 2. Parameter estimates, SD and 95% CI of immigration rate for all models

Model	Parameters						DIC
	β_0	β_1	ϕ	σ_0^2	ψ^{-1}		
1	Est.	0.031	0.029	744.100	523.500	6.504	
	SD	0.034	0.010	262.400	285.800	6.579	-114.049
	95%CI	(-0.04, 0.10)	(0.01, 0.05)	(346.70, 1341.0)	(57.48, 1087.0)	(0.002, 18.94)	
2	Est.	-3.682	0.646	31.730	35.090	10.280	
	SD	0.066	0.064	42.880	49.740	5.479	11.017
	95%CI	(-3.81, -3.56)	(0.52, 0.77)	(6.13, 171.80)	(6.23, 170.50)	(1.50, 19.48)	
3	Est.	-3.628	0.641	324.200	86.220	9.905	
	SD	0.075	0.055	100.500	72.530	5.486	-204.760
	95%CI	(-3.77, -3.47)	(0.53, 0.74)	(161.0, 541.60)	(12.740, 287.20)	(0.97, 19.51)	

estimates along with their standard deviations are presented in Table 2.

$$\text{Model 1: } y_i = \beta_0 + \beta_1 VA_i + \tau_i,$$

$$\text{Model 2: } \log\left(\frac{y_i}{1-y_i}\right) = \beta_0 + \beta_1 PCC_i + \tau_i,$$

$$\text{Model 3: } \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 PCC_i + \tau_i,$$

According to DIC criteria, Model 3 is the best model among the three models. But, it would be better to be cautious because the SD value for estimation of the precision parameter ϕ in Model 3, is greater than the estimated SD for ϕ in Model 2. On the other hand results of simulation study shown that the precision parameter is underestimated using Model 2. Due to all these reasons we choose Model 3 to study immigration rate. Estimates of the model parameters and 95% confidence intervals presented in Table 2 show that the effective covariate is the performance of cost credit of executive agencies from public income. It is shown that provinces with better welfare and social situations absorbing more immigrants. In the other word, the people tend to migrate to reach higher standards of living.

5.2 Divorce rate in provinces of Iran

Divorce is a social issue that seems to increase and become more popular among youth population, on the other hand the number of marriages are

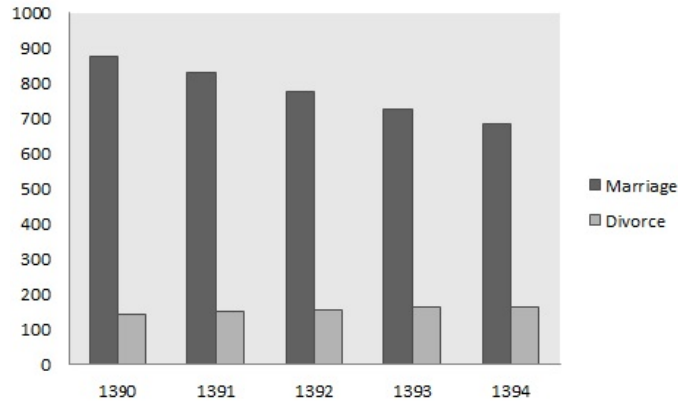


Figure 4. Number of marriages and divorces divided by 1000 in Iran from 1390 to 1394

being reduced. Both of these social problems will have consequences on both families and society. Figure 4 shows the number of marriages and divorces divided by 1000 from year 1390 to 1394. In this subsection we aim to study divorce rate. Covariates are unemployment rate (UR), proportion of people at age of marriage (MAR), and proportion of population with higher education (HE). We initially by following models

$$\text{Model 1*}: y_i = \beta_0 + \beta_1 UR_i + \beta_2 Mar_i + \beta_3 HE_i + \tau_i,$$

$$\text{Model 2*}: \log\left(\frac{y_i}{1-y_i}\right) = \beta_0 + \beta_1 UR_i + \beta_2 Mar_i + \beta_3 HE_i + \tau_i,$$

$$\text{Model 3*}: \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 UR_i + \beta_2 Mar_i + \beta_3 HE_i + \tau_i,$$

where $i = 1, \dots, 31$. The model parameters are estimated using Bayesian approach and following final models are approved

$$\text{Model 1: } y_i = \beta_0 + \tau_i,$$

$$\text{Model 2: } \log\left(\frac{y_i}{1-y_i}\right) = \beta_0 + \beta_1 UR_i + \beta_2 Mar_i + \tau_i,$$

$$\text{Model 3: } \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 UR_i + \beta_2 Mar_i + \tau_i.$$

Table 3. Parameter estimates, SD and 95% CI of divorce rate

Model	Parameters							DIC
	β_0	β_1	β_2	ϕ	σ_0^2	ψ^{-1}		
1	Est.	0.22	-	-	293.7	275.7	8.458	
	SD	0.0504	-	-	120.2	112.7	6.171	-73.072
	95%CI	(0.12, 0.32)	-	-	(125.9, 599.8)	(111.8, 555.3)	(0.21,19.33)	
2	Est.	-6.487	-3.192	9.232	32.51	30.69	10.39	
	S	1.845	1.243	3.076	46.79	40.53	5.497	12.644
	95%CI	(-10.3, -2.95)	(-5.58,-0.83)	(3.31,15.62)	(5.33,150.5)	(5.66,148.9)	(1.47, 19.64)	
3	Est.	-5.753	-2.991	7.998	89.31	39.47	10.06	
	SD	1.874	1.391	3.146	61.28	45.12	5.669	-92.641
	95%CI	(-9.49, -2.32)	(-5.68, -0.21)	(2.24, 14.39)	(31.68, 263.10)	(6.08, 166.0)	(0.92, 19.45)	

No covariate enters to the first model, but unemployment rate and proportion of people at age of marriage come to the two other models. Results are presented in Table 3.

As seen in Table 3, Model 3 in which we assume that the response variable is beta distributed has the smallest DIC among all the fitted models. Model 3 demonstrates that unemployment rate and proportion of people at marriage age have effect on divorce rate.

Upon demographic transition theory, in developing countries moving from a preindustrial to an industrial economic system there are changes in economic and social situations, such as growing unemployment rate. In our study the significant regression coefficient related to unemployment rate is -2.993, that demonstrates increasing unemployment rate results in reduction of divorce rate. Based on demographic transition theory, it can be concluded that unemployment rate entered to the model as an index of development.

As Table 3 shows, divorce rate will rise as the number of people at age of marriage increases. Rising people at age of marriage may be a sign of having more population exposure to marriage. It can be concluded that increasing marriages results in growth of divorce risk. It is a general result for those who are at the age of marriage may be varied among different age groups. Significant spatial parameters could be a consequence of cultural effect on social issues.

6 Discussion and Results

In this paper we compared three models for analysis of spatially correlated response variable in (0, 1) interval. These models include linear regression

model with spatial random effect for original data, linear regression model with spatial random effect for transformed data, and the suggested spatial beta regression model with random effect. Simulation over regular grids for different sample sizes shows that modeling spatially correlated bounded response variable utilizing spatial beta regression model with random effect will result in better fit, especially for small sample size.

The outcome obtained from application on two real data sets, respectively about migration and divorce rate, also demonstrate better fitting by assuming the response variable is beta distributed. Such dependent variables are very common in actual data analysis and we encountered considerable number of these response variables in official statistics. Spatial beta regression model with random effect provides a suitable tool to analyze such data that are spatially correlated.

Acknowledgement

The authors are thankful to the referees for their many helpful comments that greatly improved this paper. We also wish to acknowledge for the support from Ordered and Spatial Data Center of Excellence of the Ferdowsi University of Mashhad.

References

- Atkinson, A.C. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. New York: Oxford University Press.
- Branscum, A.J., Johnson, W.O. and Thurmond, M. (2007). Bayesian Beta Regression: Applications to Household Expenditure Data and Genetic Distance Between Food and Mouth Diseases Viruses, *Australian and New Zealand Journal of Statistics*, **49**, 287-301.
- Brooks, S.P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, **7**, 434-455.
- Cepeda, E.D. and Gamerman, D. (2005). Bayesian Methodology for Modeling Parameters in the Two Parameter Exponential Family, *Revista Estadística*, **57**, 168-169.
- Cepeda, E.D. and Nunez, A.V. (2013). Spatial Double Generalized beta Regression Models Extensions and Application to Study Quality of Education in Colombia, *Journal of Educational and Behavioral Statistics*, **38**, 604-628.
- Cepeda, E.D., Urdinola, B.P. and Rodriguez, D. (2012). Double Generalized Spatial Econometric Models, *Communications in Statistics-Simulation and Computation*, **41**, 671-685.

-
- Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1998). Model based Geostatistics, *Journal of the Royal Statistical Society*, **47**, 299-350.
- Ferrari, S. and Cribari, F. (2004). Beta Regression for Modelling Rates and Proportions, *Journal of Applied Statistics*, **31**, 799-815.
- Ferreira, G., Figueroa-Zuniga, J.I. and de Castro, M. (2015). Partially Linear Beta Regression Model with Autoregressive Errors, *TEST*, **24**, 752-775.
- Figueroa-Zúñiga, J.I., Arellano-Valle, R.B. and Ferrari, S.L. (2013). Mixed beta Regression: A Bayesian Perspective, *Computational Statistics and Data Analysis*, **61**, 137-147.
- Geweke, J. (1991). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments (Vol. 196), Minneapolis, MN, USA: Federal Reserve Bank of Minneapolis, Research Department.
- Gholizadeh, K., Mohammadzadeh, M. and Ghayoumi, Z. (2013). Spatial Analysis of Structured Additive Regression and Modeling of Crime Data in Tehran City Using Integrated Nested Laplace Approximation, *Journal of Applied Statistical Science*, **7**, 103-124.
- Heidelberger, P. and Welch, P.D. (1981). A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations, *Communications of the ACM*, **24**, 233-245.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.
- Paolino, P. (2001). Maximum Likelihood Estimation of Models with Beta Distributed Dependent Variables, *Political Analysis*, **9**, 325-346.
- Simas, A.B., Barreto-Souza, W. and Rocha, A.V. (2010). Improved Estimators for a General Class of Beta Regression Models, *Computational Statistics and Data Analysis*, **54**, 348-366.
- Sturtz, S., Ligges, U. and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R, *Journal of Statistical Software*, **12**, 1-16.
- Smithson, M. and Verkuilen, J. (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables, *Psychological Methods*, **11**, 54-71.
- Zimprich, D. (2010). Modeling Change in Skewed Variables Using Mixed Beta Regression Models, *Research in Human Development*, **7**, 9-26.

Lida Kalhori

Department of Statistics,
Tarbiat Modares University,
Tehran, Iran.
email: *lida.kalhori@modares.ac.ir*

Mohsen Mohammadzadeh

Department of Statistics,
Tarbiat Modares University,
Tehran, Iran.
email: *mohsen_m@modares.ac.ir*