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.


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...
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-
variable, 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](image)

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 $\phi = a + b$, $\phi > 0$, therefor $E(Y) = \mu$ and $Var(Y) = \frac{\mu(1-\mu)}{\phi + 1}$. Since $Var(Y)$ and $\phi$ are inversely related to each other, $\phi$ 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 $Y_i, i = 1, \ldots, n$ be independent random variables with distribution Beta($\mu_i\phi, (1-\mu_i)\phi$). Ferrari and Cribari (2004) defined the beta regression model by

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

where $\beta = (\beta_0, \ldots, \beta_{k-1}) \in R^k$ is the vector of regression coefficients, $x_i = (1, x_{i1}, \ldots, x_{i(k-1)})$, 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:

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

where $x_i = (1, x_{i1}, \ldots, x_{i(k-1)})$ and $w_i = (1, w_{i1}, \ldots, w_{i(m-1)})$ 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 $\sum_{\ell=0}^{m-1} w_{i\ell} \theta_\ell$ 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 $s_i$, $i = 1, \ldots, n$ is beta distributed, that is $Y(s_i) \sim \text{Beta}(\mu_i, \phi, (1-\mu_i)\phi)$. Further, assume that $y = (y(s_1), \ldots, y(s_n))^\top$ are realizations of random variables $Y = (Y(s_1), \ldots, Y(s_n))^\top$ at $n$ distinct locations $s_1, \ldots, s_n$. 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

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

where $W$ is a spatial weight matrix that specifies the effects of geographical neighborhoods of sample units and $\eta$ 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

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

where $\tau_i = \tau(s_i)$ denotes a random effect for unit $i$. The spatial random effect $\tau = (\tau_1, \ldots, \tau_n)$ is a zero mean Gaussian random process, $\tau \sim N_n(0, \Sigma)$, where the spatial covariance structure determined by $\Sigma$. 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, \( \tau = (\tau_1, \ldots, \tau_n) \) follows a zero mean multivariate normal distribution with covariance matrix \( \Sigma \). Assume that \( \Sigma_{ij} \) represents the \( ij \)th element of \( \Sigma \) and \( \Sigma_{ij} = \sigma^2_{ij} = \sigma_0^2 \exp(\psi|s_i - s_j|) \), where \( \sigma_0^2 \) is the spatial variance and \( \psi^{-1} \) is the range parameter. For Bayesian analysis of the spatial beta regression model defined in (3), we apply the following priors

\[
\beta_j \sim N(0, 10^0), \quad j = 0, \ldots, k - 1, \quad \phi \sim Gamma(\xi_1, \xi_2), \\
\sigma_0^{-2} \sim Gamma(\epsilon_1, \epsilon_2), \quad \psi^{-1} \sim U(\alpha_1, \alpha_2).
\]

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

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

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.

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

Model 2: \( E(\text{logit}(Y_i) \mid \beta, \tau_i) = \sum_{j=0}^{k-1} x_{ij} \beta_j + \tau_i \),
Model 3: \( \text{logit}(E(Y_i | \beta_i, \tau_i)) = \sum_{j=0}^{k-1} x_{ij} \beta_j + \tau_i \).

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 \( \beta = (\beta_0, \beta_1, \beta_2) = (-1, 2, -1.5) \), \( \phi = 50 \), \( \sigma_0^2 = 0.5 \), \( \psi^{-1} = 0.1 \) 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 (Heidelberger 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 criterions, 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

<table>
<thead>
<tr>
<th>n</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n Par. Value</td>
<td>Est. Bias RMSE</td>
<td>Est. Bias RMSE</td>
</tr>
<tr>
<td>25</td>
<td>β₀ -1</td>
<td>0.286 -1.286</td>
<td>-1.087 0.087</td>
</tr>
<tr>
<td></td>
<td>β₁ 2</td>
<td>0.394 -0.803</td>
<td>1.607 2.012</td>
</tr>
<tr>
<td></td>
<td>β₂ -1.5</td>
<td>-0.288 -0.808</td>
<td>1.213 -1.388</td>
</tr>
<tr>
<td></td>
<td>ϕ 50</td>
<td>124.837 -0.497</td>
<td>159.907 25.983</td>
</tr>
<tr>
<td></td>
<td>σ² 0.5</td>
<td>0.063 -0.874</td>
<td>0.441 0.074</td>
</tr>
<tr>
<td></td>
<td>ψ⁻¹ 0.1</td>
<td>165.692 1655.918</td>
<td>175.078 16.620</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n Par. Value</td>
<td>Est. Bias RMSE</td>
<td>Est. Bias RMSE</td>
</tr>
<tr>
<td>100</td>
<td>β₀ -1</td>
<td>0.294 -1.294</td>
<td>1.295 -1.041</td>
</tr>
<tr>
<td></td>
<td>β₁ 2</td>
<td>0.421 -0.790</td>
<td>1.580 2.066</td>
</tr>
<tr>
<td></td>
<td>β₂ -1.5</td>
<td>-0.299 -0.800</td>
<td>1.201 -1.589</td>
</tr>
<tr>
<td></td>
<td>ϕ 50</td>
<td>129.513 1.590</td>
<td>153.031 31.564</td>
</tr>
<tr>
<td></td>
<td>σ² 0.5</td>
<td>1.764 0.176</td>
<td>1.491 1.846</td>
</tr>
<tr>
<td></td>
<td>ψ⁻¹ 0.1</td>
<td>124.315 1242.146</td>
<td>128.598 3.371</td>
</tr>
<tr>
<td>225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n Par. Value</td>
<td>Est. Bias RMSE</td>
<td>Est. Bias RMSE</td>
</tr>
<tr>
<td>225</td>
<td>β₀ -1</td>
<td>0.295 -1.295</td>
<td>1.295 -0.993</td>
</tr>
<tr>
<td></td>
<td>β₁ 2</td>
<td>0.403 -0.798</td>
<td>1.597 2.052</td>
</tr>
<tr>
<td></td>
<td>β₂ -1.5</td>
<td>-0.298 -0.801</td>
<td>1.203 -1.566</td>
</tr>
<tr>
<td></td>
<td>ϕ 50</td>
<td>256.827 4.137</td>
<td>213.868 20.218</td>
</tr>
<tr>
<td></td>
<td>σ² 0.5</td>
<td>1.863 2.727</td>
<td>1.891 3.090</td>
</tr>
<tr>
<td></td>
<td>ψ⁻¹ 0.1</td>
<td>130.254 1301.544</td>
<td>135.126 4.990</td>
</tr>
</tbody>
</table>

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

© 2016, SRTC Iran
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 $\beta$ are similarly estimated for two other models. But for the parameters $\phi, \sigma_0^2$ and $\psi^{-1}$, 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.

![Prediction surface of the spatial random effect](image)

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

Model 1*: \[ y_i = \beta_0 + \beta_1 PCI_i + \beta_2 VA_i + \beta_3 UR_i + \beta_4 HE_i + \tau_i, \]

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

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 \( i = 1, \ldots, 31 \). 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \phi )</th>
<th>( \sigma_i^2 )</th>
<th>( \psi^{-1} )</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Est.</td>
<td>0.031</td>
<td>0.029</td>
<td>744.100</td>
<td>523.500</td>
<td>6.504</td>
<td>-114.949</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.034</td>
<td>0.010</td>
<td>262.400</td>
<td>285.800</td>
<td>6.579</td>
<td>-114.949</td>
</tr>
<tr>
<td></td>
<td>95%CI</td>
<td>(-0.04, 0.10)</td>
<td>(0.01, 0.05)</td>
<td>(346.70,1341.0)</td>
<td>(57.48, 1087.0)</td>
<td>(0.002, 18.94)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Est.</td>
<td>-3.682</td>
<td>0.646</td>
<td>31.730</td>
<td>35.090</td>
<td>10.280</td>
<td>11.017</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.066</td>
<td>0.064</td>
<td>42.880</td>
<td>49.740</td>
<td>5.479</td>
<td>11.017</td>
</tr>
<tr>
<td></td>
<td>95%CI</td>
<td>(-3.81, -3.56)</td>
<td>(0.52, 0.77)</td>
<td>(6.13, 171.80)</td>
<td>(6.23, 170.50)</td>
<td>(1.50, 19.48)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Est.</td>
<td>-3.628</td>
<td>0.641</td>
<td>324.200</td>
<td>86.220</td>
<td>9.905</td>
<td>9.905</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.075</td>
<td>0.055</td>
<td>100.500</td>
<td>72.530</td>
<td>5.486</td>
<td>204.760</td>
</tr>
<tr>
<td></td>
<td>95%CI</td>
<td>(-3.77, -3.47)</td>
<td>(0.53, 0.74)</td>
<td>(161.0, 541.60)</td>
<td>(12.740, 287.20)</td>
<td>(0.97, 19.51)</td>
<td></td>
</tr>
</tbody>
</table>

According to DIC criteria, Model 3 is the best model among the three models. But, it would better to be cautious because the SD value for estimation of the precision parameter \( \phi \) in Model 3, is greater than the estimated SD for \( \phi \) 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
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

Model 1*:\[ y_i = \beta_0 + \beta_1 UR_i + \beta_2 Mar_i + \beta_3 HE_i + \tau_i, \]

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

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 \( i = 1, \ldots, 31 \). The model parameters are estimated using Bayesian approach and following final models are approved

Model 1: \[ y_i = \beta_0 + \tau_i, \]

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

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.22</td>
<td>0.0504</td>
<td>(0.12, 0.32)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-</td>
<td>1.845</td>
<td>(-10.3, -2.95)</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-6.487</td>
<td>1.874</td>
<td>(-9.49, -2.32)</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>-5.753</td>
<td>1.874</td>
<td>(-9.89, -2.32)</td>
</tr>
</tbody>
</table>

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


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