Spatial Statistics 21 (2017) 279–303

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Geographically Weighted Beta Regression

Alan Ricardo da Silva *, Andreza de Oliveira Lima

Departamento de Estatistica, Universidade de Brasilia, Brazil

ARTICLE INFO

Article history:
Received 18 January 2017
Accepted 28 July 2017
Available online 9 August 2017

Keywords:
Generalized Linear Models
Beta regression
Spatial analysis
Geographically Weighted Regression

ABSTRACT

Linear regression models are often used to describe the relationship between a dependent variable and a set of independent variables. However, these models are based on the assumption that the error (or, in some cases, the response variable) is normally distributed with constant variance and that the relations are equal throughout space. Thus, these models may not be the most appropriate to adjust spatially varying rates and proportions. The Beta Regression model deals with rates and proportions and has been shown to be a good approach to model this type of data, since it naturally adapts to variables constrained to an interval of the real line and exhibiting heteroscedasticity, which is a common characteristic in this type of data. In addition, to deal with spatial non-stationarity, Geographically Weighted Regression (GWR) allows for variability in the parameters by an extension of the linear regression model, providing a better understanding of the spatial phenomenon. Therefore, we propose the Geographically Weighted Beta Regression (GWBR) model which combines the features of the above models such that a better fit is provided in the study of spatially varying continuous variables restricted to an interval of the real line. We applied this model to analyze the proportion of households that have telephones in the state of Sao Paulo, Brazil. The results were more appropriate than those obtained by the global models and the Geographically Weighted Regression model, following statistics such as AICc, pseudo-R^2, log-likelihood and by the reduction of spatial dependence computed by Moran's I.

* Corresponding author.
E-mail address: alansilva@unb.br (A.R. da Silva).

http://dx.doi.org/10.1016/j.spasta.2017.07.011
2211-6753/

1. Introduction

Geographically Weighted Regression (GWR) pertains to a linear model and allows the spatial modeling of non-stationary processes and was defined by Fotheringham et al. (1996), Brunsdon et al. (1996, 1998), and Fotheringham et al. (2002). However, it is used when the distribution of the data is Gaussian. Other distributions were incorporated into the GWR framework such as Logistic (Atkinson et al., 2003), Poisson (Nakaya et al., 2005), Multinomial Logistic (Luo and Kanala, 2008), and Negative Binomial (Silva and Rodrigues, 2014), but in many applications the dependent variable represents a rate or proportion that is usually defined between (0, 1), making the classic GWR (or the others distributions cited before) inappropriate to model this kind of data, because the estimated dependent variable can be outside of the interval (0, 1) (Dyke and Patterson, 1952). The first idea is to apply a transformation to the response variable, but the parameters interpretation would not be straightforward. Another problem occurs when the proportion p is near 0 or 1, showing some asymmetry and hence, violating the normality assumption required by classical GWR.

The use of a beta distribution in a regression model dates back to Falls (1974), Brehm and Gates (1993), McDonald and Xu (1995) and Sulaiman et al. (1999), but the structure of data restricted to the interval (0, 1) was defined by Ferrari and Cribari-Neto (2004) and named Beta Regression. This model uses a reparameterization of the classical beta distribution depending on only 2 parameters and because of its versatility, as shown in Fig. 1, utility and software availability (Cribari-Neto and Zeileis, 2010; Swearingen et al., 2011), it is being used in several areas such as medicine (Peplonska et al., 2012), economics (Castellani et al., 2012), public policy (Pereira et al., 2014), engineering (Eskelson et al., 2011), and in spatial statistics by using remotely sensed spectral data (Korhonen et al., 2015).

Thus, the objective of this paper is to extend beta regression concepts to Geographically Weighted Regression, namely Geographically Weighted Beta Regression (GWBR), in order to model rate or proportion data restricted to the interval (0, 1) in a spatial context, and, in this way, providing to the analyst with another option to model the data. Section 2 presents the specifications of GWBR. Section 3 presents some simulations and a real application to evaluate the potential of GWBR in relation to classical GWR and beta regression. Conclusions are drawn in Section 4.

2. Specifications of geographically weighted beta regression

The beta distribution has density given by

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

where $0 < y < 1$, $a > 0$, $b > 0$ and $\Gamma(.)$ is the gamma function.

Although the beta distribution does not belong to the Generalized Linear Models (GLM), the idea of Ferrari and Cribari-Neto (2004) was to create a similar structure. As is well known, the mean and variance of beta distribution are given by:

$$E(y) = \frac{a}{a+b} \tag{2}$$

$$Var(y) = \frac{ab}{(a+b)^2(a+b+1)}. \tag{3}$$

Letting $\mu = E(y)$ and $\phi = a + b$, then $a = \mu\phi$ and $b = (1-\mu)\phi$. Therefore, from (2) and (3),

$$E(y) = \mu \qquad \text{and} \qquad Var(y) = \frac{V(\mu)}{1+\phi} \tag{4}$$

where $V(\mu) = \mu(1-\mu)$.

The reparameterization of the beta distribution as a function of the mean μ and the precision parameter ϕ is (Ferrari and Cribari-Neto, 2004):

$$f(y) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\, y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \tag{5}$$

where $0 < \mu < 1$ and $\phi > 0$.