Spatial Overdispersed models: a new proposal

Edilberto Cepeda-Cuervo[1], *Michel Córdoba*[2], *Vicente Núñez-Antón*[3]

[1]ecepedac@unal.edu.co, Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia

[2] mfcordobap@unal.edu.co, Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia

[3]vicente.nunezanton@ehu.eus, Departamento de Econometría y Estadística (E.A. III), Universidad del País Vasco UPV/EHU, Bilbao, Spain

In this work, a new class of overdispersed models, the spatial overdispersed model, to analyze counting area data is proposed. The innovative proposed models allow us to obtain information and be able to model the possible existence of spatial dependence for a given variable under study, which is mainly related to the response variable spatial neighbor structures, to the level of the overdispersion related to heteroscedasticity of the data, and to additional spatial structures not being considered in the model itself.

The proposed models assume that the spatial variable under study Y_i, i=1,2,….,n, given the values in all neighborhoods of the i-th region, but not including the i-th region itself (i.e., Y_i), follows a specific overdispersed model that depends on the nature of the response variable, such as, for example, the beta binomial model or negative binomial model, where both the mean and overdispersed models follow given regression structures, including spatial-lag variables. The aforementioned lag-variables are defined as the product of an nxn weight matrix, W, and the n-dimensional vector, $Y=(Y_1,\ldots,Y_n)'$, where W is defined in such a way that it takes into account the desired neighbors' structure.

The spatial overdispersed models are considered for the analysis of specific count data in Colombia, disaggregated by departments (similar to states), and assuming a spatial structure that is defined by the geographical borders of the different departments under study. That is, it is assumed that department i is a neighbor of department j, if they have a common border. The advantages of the proposed models are illustrated with the analysis of two real data sets related to infant mortality and postnatal period of cribbed in Colombia, disaggregated by department, and also considering several independent variables, such as *unsatisfied basic needs, intrafamiliar violence, and fathers educational levels*. The performance of the models is assessed by comparing their performance with that of the most commonly and previously used overdispersion models in the literature.

Models are fitted by adopting a robust Bayesian approach that assumes flat normal prior distributions for the regression parameters, and gamma prior distributions for the dispersion

parameters in the model, whenever no specific model is assumed for it. The Bayesian Information Criteria, BIC, is used as goodness-of-fit criterion in the model selection processes.

Keywords: Spatial regression, Overdispersed models, Bayessian approach.
AMS: 62J12.