International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:10, No:1, 2016

Spatial Econometric Approaches for Count Data: An Overview and New Directions

Paula Simões, Isabel Natário

Abstract—This paper reviews a number of theoretical aspects for implementing an explicit spatial perspective in econometrics for modelling non-continuous data, in general, and count data, in particular. It provides an overview of the several spatial econometric approaches that are available to model data that are collected with reference to location in space, from the classical spatial econometrics approaches to the recent developments on spatial econometrics to model count data, in a Bayesian hierarchical setting. Considerable attention is paid to the inferential framework, necessary for structural consistent spatial econometric count models, incorporating spatial lag autocorrelation, to the corresponding estimation and testing procedures for different assumptions, to the constrains and implications embedded in the various specifications in the literature.

This review combines insights from the classical spatial econometrics literature as well as from hierarchical modeling and analysis of spatial data, in order to look for new possible directions on the processing of count data, in a spatial hierarchical Bayesian econometric context.

Keywords—Spatial data analysis, spatial econometrics, Bayesian hierarchical models, count data.

I. Introduction

NOWADAYS, the consideration of spatial effects in econometrics modelling evolved to form one of the branches of econometrics [1]. The definition and scope of spatial econometrics has expanded substantially over the last three decades, moving from the "margins of urban and regional modeling" to the mainstream of econometrics methodology [1]. When sample data includes a location component, two scenarios have to be addressed, spatial autocorrelation between observations, and spatial heterogeneity in relations. Under these, fundamental assumptions of traditional statistical methods, that data values are derived from independent observations or that a single relationship with constant variance exists across the sample data, are no longer guaranteed [2]. Traditional econometrics has largely ignore this violation of the Gauss-Markov assumptions used in regression modeling [3]. An adequate alternative is to implement spatial econometrics models that allow to assess the magnitude of the space influences, by introducing a specific weighting scheme, in which relationships among spatial areas are specified [4]. The topology or spatial pattern of the data is carried out by the choice of a spatial weights or contiguity matrix, commonly denoted by the letter W, and represents our comprehension of the spatial association among spatial units [5]. A complete treatment of many aspects of spatial econometrics, including the application of Bayesian estimation methods, is aimed.

P. Simões is with CMA and Área Departamental de Matemática, ISEL - Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Portugal (corresponding author; e-mail: paulasimoes@adm.isel.pt).

I. Natário is with CMA and Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Portugal (e-mail: icn@fct.unl.pt).

This work was funded by national funds through FCT - Fundação para a Ciência e a Tecnologia - in the scope of the project UID/MAT/00297/2013.

In this paper, two main methodologies in spatial econometrics are presented, more addressed for area aggregated spatial data, typically count data. First the employment of the traditional econometric models. These models were designed for continuous data, demanding count data transformation to meet the model's assumptions. Second, an alternative for count data, the use of hierarchical Bayesian models, where data can be modeled as having any distribution (Poisson is the usual choice here). Inference for this Bayesian paradigm needs to be based on simulation methods, namely the Markov Chain Monte Carlo (MCMC) method [6]. However, recently, another spatial econometric approach, that incorporates spatial lag autocorrelation in modelling counts is also available, the spatial autoregressive lag model of counts, recently developed by Lambert, Brown and Florax in 2010 [7]. It showed that some other possible directions for count data are being investigated and, at the same time, that an alternative way of doing Bayesian inference for spatial econometrics models is explored [8].

II. Spatial Perspective in Econometrics

Spatial econometrics is an appropriate area when dealing with data reflecting geographical events, which can accommodate the magnitude of the spatial influences, while maintaining other factors or variables considered important to explain of the phenomenons of interest. In the next sections several theoretical aspects for implementing an explicit spatial perspective model in econometrics are displayed.

A. Econometrics

Econometrics is based upon the development and application of statistical methods for studying and understanding economic phenomenons. It combines statistics with economic theory to analyze and test economic relationships. It is used in various fields of applied economics for estimating economic relationships, testing economic theories, evaluating and implementing government and business policy as well as predicting future behaviors. Its main application is the forecasting of important macroeconomic variables, such as interest rates, inflation rates and gross domestic product. However, currently, the use of econometric methods goes beyond the study of economics, and is also used in areas such as meteorology, biology, political science and education [9].

The kind of analysis that can be performed in econometrics is conditioned by the nature of the data. The data sets can be classified, essentially, into three types, cross-sectional data, time series and panel data.

Cross-Sectional Data A cross-sectional data set contains multiple observations of multiple phenomenons taken at a given point in time. Sometimes the data on all units do not correspond to precisely the same time period. In cross-sectional data analysis is given importance to their values but not to their ordination. The data sets are classified as **pooled cross-sectional data**, when we observe one or more phenomenons in two or more different moments in time, and then join the observations.

Time Series Data A time series data set consists of observations on a variable or several variables over time. Unlike cross-sectional data, the chronological ordering of observations in a time series gives important information, once past events can influence future events, time is an important dimension.

Panel Data A panel data, also known as longitudinal data set, consists of a time series for each cross-sectional member in the data set. For example, we can collect information such as investment or financial data, about the same set of firms over a six-year time period. This information can also be collected on geographical units. The panel data differs from pooled cross-sectional data by the fact that the same cross-sectional units, individuals, firms, countries, are followed over a given time period.

The study and understanding of economics phenomenons in econometrics is carried out resorting to different models and statistical techniques. **Simple** and **multiple regression** is a main tool in econometrics [10]. For example, consider the following scenarios: an economist may be interested in study the dependence of personal consumption expenditure on after-tax or disposal real personal income. This kind of analysis may be important in estimating the marginal propensity to consume, that is, average change in consumption expenditure for a monetary unit of change in real income, [10]. For dealing with statistical relationships among variables in econometrics, essentially with random or stochastic variables, which have probability distributions, the **correlation analysis** is used [10]. The principal aim is to measure the strength or degree of association between two variables. For example, an economist can be interested in studying the correlation of personal consumption with after taxes the real personal income.

The term autocorrelation may be defined as “correlation between members of series of observations ordered in time, as in time series data, or in space, as in cross-sectional data ” [10]. For example, in a time series regression of consumption expenditures the current period depends, among other things, on the consumption expenditure of the previous period. In this situation **autoregressive models** are used. There are also models that incorporate qualitative explanatory variables, called **dummy variables**. These variables that cannot be readily quantified, such as gender, religion, and yet influence the behavior of the dependent variable. It is also possible that the dependent variable in a regression model be qualitative itself. In situations the variable is a ”yes ” or ”no” type, like ownership of a house, possession of an attribute. Some approaches to estimate such models are the **linear model**, the **logit model** or the **probit model**.

Until now, we just refer to models with a single equation, for which there was a single dependent variable y and one or more explanatory variables X’s, with a cause-effect relationship. However, in many situations, this kind of relationship is not meaningful. This happens if y is determined by the X’s, and some of the X’s are, on the other hand, determined by y. In this case, the methodology of **simultaneous equation models** is employed. For time series data, which are frequently and intensively used on empirical research in econometrics, assuming that a time series is stationary, the **ARIMA modeling** can be used for forecasting, however another forecasting method, known as **vector autoregression** (VAR) is also an alternative method for this purpose. (For more details of the enounced methods see, for example, [10]).

The traditional econometric methodologies assume a particular econometric model and try to see if it fits a given set of data. However, there are another econometrics approaches, for example, the ones that use **Bayesian statistics**, which can improve some econometric techniques [11] [12].

B. *Spatial Data*

The ready availability of increasing volumes of geo-referenced data and a user friendly technology to manipulate these in geographic information systems, with the growing attention to a spatial perspective, is stimulating an increasing interest in spatial analysis [5] [1]. Data for which location attributes are an important source of information, when taken into account, yield a spatial modelling approach. The recognition of the spatial dimension can give more meaningful results than an analysis that ignores it [13]. Observations for which the spatial arrangement, i.e., absolute location and/or relative position, is explicitly taken into account, are termed **spatial data**. Such data are the subject of study in many research fields, such as climatology, ecology, epidemiology, econometrics, sociology, among others.

Spatial data analysis focuses on detecting patterns, exploring and modelling relationships between such patterns in order to understand processes responsible for the observed patterns. The spatial data analysis can be applied, for example, to emphasises the role of space as a potentially important explanator of socioeconomic systems, taking spatial patterns, into account. Three main classes of spatial data can be distinguished: **Geostatistical or spatially continuous data**, that is, observations associated with a continuous variation measure over space, given observed values at fixed sampling points, **areal or lattice data**, related to of the discrete measured attribute over space, and **spatial point patterns**, where the objects are point locations at which events of interest have occurred [13].

C. Spatial Econometrics

The definition and scope of spatial econometrics evolved in the literature over the last thirty years [1]. The definition provided in Anselin (1988) [4], states that the domain of spatial econometrics is delineated as "the collection of spatial techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models". Contrasting spatial econometrics to standard econometrics, a straight definition is given as handling with "the specific spatial aspects of data and models in regional science that precludes a straightforward application of standard econometric methods" [4]. This spatial aspects are classified into two main spatial effects, spatial dependence and spatial heterogeneity [1] [13]. Some twenty years later, this definition, whose subject and range are constrained to urban and regional modelling, changed. The enormous growth in the importance and application of spatial techniques in economics as well as in other mainstream sciences, led to the extension of the context, no longer restricted to urban and regional modelling. According to Anselin, to 2006, the subject of spatial econometrics is defined as a "subset of econometric methods that is concerned with spatial aspects present in cross-sectional and space time observations" [1].

When sample data have a location component, two scenarios have to be considered: spatial autocorrelation between observations and spatial heterogeneity in relations. Under these scenarios, fundamental assumptions in traditional statistic methods, namely, that data values are derived from independent observations or that exists a single relationship with constant variance across the sample data, are no longer guaranteed [2]. Spatial econometrics is an adequate alternative that can be used when dealing with observations that describe geographic phenomenons or events [5]. Variables related to location, distance and arrangement (topology) are treated explicitly in model specification, estimation, diagnostic checking and prediction [2]. Similarly to what happens in any statistical modelling, four important tasks can be identified that define the modern spatial econometric methodology: model specification (which deals with the formal mathematical expression for spatial dependence and spatial heterogeneity in econometric models), estimation methods, a specific testing and spatial prediction [1].

Geostatistical data, also termed field data, play an important role in environmental sciences (see, for example, [13] and references there in), but less important in spatial econometrics [5]. The most appropriate perspectives for spatial analysis applications in spatial econometrics are areal data and spatial point process. In this work we will focus on areal data.

D. Spatial Dependence

Spatial association, also referred to as spatial autocorrelation, corresponds to situations where observations or spatial units are non-independent over space, that is, nearby spatial units are associated in some way [13]. Such association can be identified in a number of ways, using a scatter-plot where each value is plotted against the mean of neighbouring areas - the **Moran's scatter plot**, or using a spatial autocorrelation statistic such as **Moran's I** or **Geary's C**. Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation [14]. Both of these statistics require the choice of a spatial weights or contiguity matrix, usually denoted by the letter W, that represents the topology or spatial arrangement of the data and represents our understanding of spatial association among all areas units [5]. Usually, $w_{ii} = 0$, $i = 1, ..., n$, where n is the number of spatial units, but for $i \neq j$, the association measure between area i and area j, w_{ij}, can be defined in many different ways, being the most usual the minimum distance between areas [14].

When spatial autocorrelation is identified, due to its distinct nature, a specialized set of methods is needed. In order to capture dependencies across spatial units, spatially correlated variables are introduced in the model specification. These variables are weighted averages of the neighbours, where the definition of neighbours is carried out through the specification of the spatial weights matrix W. These variables, depending on the problem, can constitute the dependent response as well as the explanatory variables or the error terms [1], [12], [2].

E. Spatial Heterogeneity

The term spatial heterogeneity refers to variation in relations over space, due to structural instability or nonstationarity of relationships [2]. Heterogeneity can be related to the spatial structure or to the spatial process generating data. In contrast to spatial dependence, this issue does not always require a separate set of methods. The spatial aspect of heterogeneity is the additional information that may be provided by the spatial structure, for example, this may inform models for heteroscedasticity, spatially varying coefficients, random coefficients and spatial structural change [1]. The specification of spatial heterogeneity in models, for structural instability, can be classified into discrete heterogeneity and continuous heterogeneity. The former consists of a pre-specified set of spatially distinct units, or spatial regimes [3], between which model coefficients and other parameters are allowed to vary. In discrete heterogeneity, specifically, a dummy variable is created for each regime, (i.e. taking a value of one for observation in the regime and zero for all others), and interact with each explanatory variable with each dummy [15]. Continuous heterogeneity specifies how the regression coefficients change over space, either following a predetermined functional form, requiring a **spatial expansion method**, or determined by the data through a local estimation process, as in **geographically weighted regression**(GWR) [1] [15].

Spatial heterogeneity can be identified by some tests, for structural instability with the **Chow Test**, for heteroscedasticity, with **Breusch-Pagan test**. Both tests are used ignoring the presence of spatial dependence, otherwise, the **spatial Breusch-Pagan Test** and the **Spatial Chow Test** should be used instead [15].

F. The W Matrix

The definition of the spatial weights matrix W, where the spatial relationships among spatial units are specified, is very important since estimation results critically depend on the choice of this matrix. There are several approaches to define spatial relations between two locations or spatial units, but they can essentially be classified into two main groups: spatial contiguity approach and the distance based approach. Typical types of neighbouring matrices for spatial contiguity approach are: the linear, the rook, the bishop, the queen contiguity matrices W, and for the distance approach, we have, for example, the k-nearest neighbours or the critical cut-off neighborhood matrices [2].

- **Contiguity Matrix:** Represents a $n \times n$ symmetric matrix, where $w_{ij} = 1$, when i and j are neighbours and 0 when they are not. By convention, the diagonal elements are set to zero. W is usually standardized so that all rows sum to one, $\widetilde{w}_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$, and operations with the W matrix are an average over neighbouring values [2].
  - **Linear contiguity:** Define $w_{ij} = 1$ for regions that share a common edge to the immediate right or left to the region of interest;
  - **Rook contiguity:** Two regions are considered neighbours if they share a common border, and for these $w_{ij} = 1$.
  - **Bishop contiguity:** Define $w_{ij} = 1$ for regions that share a common vertex;
  - **Queen contiguity:** Regions that share a common border or a vertex are considered neighbours, and for these $w_{ij} = 1$.
- **Distance Approach:** makes direct use of the latitude-longitude coordinates associated with spatial data observations [12].
  - **Critical Cut-off Neighborhood:** Two regions i and j are considered neighbours if $0 \leq d_{ij} < d^*$, with d_{ij} the appropriate distance adopted between regions, and d^* representing the critical cut-off or threshold distance, beyond which no direct spatial influence between spatial units is considered.
  - **k-Nearest Neighbor:** Given the centroid distances from each spatial unit i to all units $j \neq i$ ranked as, $d_{ij}(1) < d_{ij}(2) < ... < d_{ij}(n-1)$, for each $k = 1, 2, ..., n-1$, the set $N_k(i) = \{j(1), j(2), ..., j(k)\}$ contains the k closest units to i, and for each given k, the k-nearest neighbor matrix, has the form: $w_{ij} = 1,\ j(i) \in N_k(i),\ i \in 1, ..., n$, and is zero otherwise.

III. Spatial Econometrics Models for Continuous Data

Spatial autoregressive econometrics models are used to model spatial data and provide a relatively complete treatment from a classic perspective [2].

A. Spatial Autoregressive Model-SAR

A first-order spatial autoregressive model, in its simplest form, is given by,

$$\begin{aligned} y &= \rho W y + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{1}$$

where y and ε are $n \times 1$ random vectors, the components of the ε vector are independent identically distributed (i.i.d.), y corresponds to the spatially autocorrelated dependent variable, W is an $n \times n$ spatial contiguity matrix, and ρ represents the autoregressive parameter. This model tries to explain the variation in y only as a linear combination of neighbouring units with no other explanatory variables. It is frequently used for checking residuals for spatial autocorrelation, without the interference of any other [2].

The ordinary least squares estimation is inappropriate method for a model that includes spatial effects. Applying least squares to this model results on a biased estimator for the spatial autoregressive parameter ρ, which leads to inconsistency estimates. With,

$$\widehat{\rho} = (y'W'Wy)^{-1} y'W'y$$

one have,

$$\begin{aligned} E(\widehat{\rho}) &= E[(y'W'Wy)^[-1 y'W'(\rho W y + \varepsilon)] = \\ &= \rho + E[(y'W'Wy)^{-1} y'W'\varepsilon] \neq \rho. \end{aligned}$$

In this model to estimate ρ we should use the maximum likelihood estimator using a "simplex univariate optimization routine", in order to find a value of ρ that maximizes the likelihood function [2]:

$$\begin{aligned} &L(y|\rho, \sigma^2) = \\ &= \tfrac{1}{2\pi\sigma^{2(n/2)}} |I_n - \rho W| \exp\{-\tfrac{1}{2\sigma^2}(y - \rho W y)'(y - \rho W y)\}. \end{aligned}$$

Using this estimate of ρ, the estimate for the parameter σ^2 is provided by

$$\widehat{\sigma}^2 = \tfrac{1}{n}[(y - \widehat{\rho} W y)'(y - \widehat{\rho} W y)].$$

An extension of this spatial model is,

$$\begin{aligned} y &= \rho W y + X\beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{2}$$

where X is a $n \times k$ matrix of observed values of the explanatory variables, X_j (covariates) and β is a $k \times 1$ vector of parameters that reflects the influence of the covariates on the y variation over the spatial sample. This model is also named "mixed regressive-autoregressive model" because it combines the standard regression model with a spatially dependent variable model [2]. As in the previous model, a maximum likelihood iterative estimation is carried out in order to obtain the autoregressive parameter ρ that maximizes the likelihood function, and consequently allows the computation of $\widehat{\beta}$ and $\widehat{\sigma}^2$.

B. Spatial Error Model-SEM

The spatial error model is a regression model with spatial autocorrelation in the residuals defined by

$$\begin{aligned} y &= X\beta + u \\ u &= \lambda W u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{3}$$

where, y is an $n \times 1$ vector of dependent variable, X is a $n \times k$ matrix of observed explanatory variables and β is a vector of parameters that reflects the influence of this variables on variation of variable y. W is a known $n \times n$ spatial contiguity matrix, the parameter λ is a coefficient on the autocorrelated residuals u and ε is defined as in the previous model. Spatial autocorrelation in the least-squares residuals can be detected by an appropriate statistical test, like the one based on the Moran's I statistics.

C. General Spatial Model

The most general form of a general autoregressive spatial model, that includes both the spatial lag term and a spatially correlated error structure, is

$$\begin{aligned} y &= \rho W_1 y + X\beta + u \\ u &= \lambda W_2 u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{4}$$

where y contains an $n \times 1$ vector of the dependent variable and X represents an $n \times k$ matrix of observed explanatory variables. W_1 e W_2 are known $n \times n$ spatial weight matrices, defining spatial relations between spatial units, using contiguity or the distance based approach, considering ρ, $\beta\,\lambda, u$ and ε as defined for the previous models. The log likelihood function for this model, is given by

$$L = C - \tfrac{n}{2}\ln(|A|) + \ln(|B|) - \tfrac{1}{2\sigma^2}(e'B'Be) \tag{5}$$

where C denotes an inessential constant, $e = (Ay - X\beta)$, $A = (I_n - \rho W_1)$, $B = (I_n - \lambda W_2)$. Using the following expressions for β and σ^2,

$$\begin{aligned} \beta &= (X'A'AX)^{-1}(X'A'ABy) \\ \sigma^2 &= \tfrac{e'e}{n} \end{aligned} \tag{6}$$

with $e = By - X\beta$, the log likelihood for this model can be maximized through an optimization algorithm that allows to obtain the values of ρ and λ. The values of the other parameters β and σ^2 are calculated as a function of the maximum likelihood values of ρ, λ and the sample data in y and X.

D. Spatial Durbin Model

There is a model where "spatial lag" of the dependent variable, as well as a spatial lag of the explanatory variables in matrix X, are added in a traditional linear model, given by

$$\begin{aligned} y &= \rho W y + X\beta_1 + WX\beta_2 + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{7}$$

This model is called the spatial Durbin model, where y contain an $n \times 1$ vector of the dependent variable, X correspond to the $n \times k$ matrix containing the observed explanatory variables with an associated parameters vector, β_1, W is the spatial weight matrix, and ρ represents the parameter of the spatial lag of the dependent variable. The matrix product WX represents a spatial lag of the explanatory variables, with associated $k \times 1$ parameters vector β_2.

The vectors of parameters, β_1 and β_2, can be expressed as

$$\begin{aligned} \beta_1 &= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'y \\ \beta_2 &= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Wy, \end{aligned} \tag{8}$$

with $\widetilde{X} = X\,W\,X$.

The log-likelihood function for this model is given by the following expression,

$$ln(L) = C + \ln|I_n - \rho W| - \tfrac{n}{2}\ln(e_1'e_1 - 2\rho e_2' e_1 + \rho^2 e_2' e_2) \tag{9}$$

where C denotes an inessential constant, $e_1 = y - \widetilde{X}\beta_1$, $e_2 = Wy - \widetilde{X}\beta_2$, $\widetilde{X} = X\,W\,X$.
Given the value of ρ that maximizes the log-likelihood function (9), $\widehat{\rho}$, the estimates for β_1 and β_2 in (7) can be computed using

$$\widehat{\beta} = (\beta_1 - \widehat{\rho}\beta_2) \tag{10}$$

and an estimate for σ^2 is obtained trough

$$\widehat{\sigma}^2 = \tfrac{(y - \widehat{\rho}Wy - \widetilde{X}\widehat{\beta})'(y - \widehat{\rho}Wy - \widetilde{X}\widehat{\beta})}{n}. \tag{11}$$

It should be noted that when this model is used, the explanatory variables matrix $\widetilde{X}$ can suffer from severe collinearity problems in some applications, being necessary take this possibility into account.

E. Lagrange Multipliers Tests

In this section some tests are presented some tests to choose the best spatial autocorrelation model for the specific situation.

The Moran's I test for spatial autocorrelation is not suited to choose the best spatial autocorrelation model for the specific form of spatial dependence in data. For this propose the Lagrange Multiplier (LM) test (based on the residuals e of a gaussian linear regression model) has proven to be more adequate. This is due to the fact that this test statistics takes a different form, whether the alternative hypothesis to the null hypothesis, of non-existence of spatial autocorrelation, is related with a spatial error or a spatial lag model [16] [17] [18].

LM test for spatial error: the test statistics is

$$\text{LM}_{\text{error test}} = \frac{e^T W e}{tr(W^T W + W^2)},$$

where W is the weight matrix and $tr(\cdot)$ is the trace of the matrix. Under the null hypothesis, this statistics is approximately qui-squared distributed with 1 degree of freedom.

LM test for spatial lag dependence: the test statistics is

$$\text{LM}_{\text{lag test}} = \\ = \frac{\left(e^T W y e/\hat{\sigma}^2\right)^2}{(WX\widehat{\beta})^T M (WX\widehat{\beta})/\hat{\sigma}^2 + tr(W^T W + W^2)},$$

where $M = I - X(X^TX)^{-1}X^T$. Under the null hypothesis, this statistics is approximately qui-squared distributed with 1 degree of freedom.

Robust LM test for spatial error: This is a robust version of the LM test for spatial error robust to the presence of spatial lag dependence. The test statistics is

$$\text{LMR}_{\text{error test}} = \text{LM}_{\text{error test}} - \text{LM}_{\text{lag test}}.$$

Under the null hypothesis, this statistics is approximately qui-squared distributed with 1 degree of freedom.

Robust LM test for spatial lag dependence: This is a robust version of the LM test for spatial lag dependence robust to the presence of spatial error. The test statistics is

$$\text{LMR}_{\text{lag test}} = \text{LM}_{\text{lag test}} - \text{LM}_{\text{error test}}.$$

Under the null hypothesis, this statistics is approximately qui-squared distributed with 1 degree of freedom.

LM test for spatial error and spatial lag dependence: the test statistics is

$$\text{LM}_{\text{SARMA}} = \text{LMR}_{\text{error test}} + \text{LMR}_{\text{lag test}}.$$

Under the null hypothesis, this statistics is approximately qui-squared distributed with 2 degree of freedom.

IV. Generalized Linear Models with Random Effects

The generalized linear models introduced by Nelder and Wedderburn (1972) have been playing an increasingly important role in statistical analysis, due to the large number of models that they encompass and facility of analysis associated with rapid computer development, in responding to situations which are not adequately explained by the normal linear model [19]. The generalized linear models with random effects are a different way of modelling the outcome (y) and accounting for covariates and random effects, either spatially structured or not, to account for spatial autocorrelation in the analysis of spatial data. The outcome (y) is assumed to come from a distribution of the exponential family with mean parameter λ_i.

> ***Exponential family:*** A family of probability density functions or probability mass functions is said to bellong to the exponential family if it can be expressed as
>
> $$f(y|\boldsymbol{\theta}) = h(y)c(\boldsymbol{\theta})exp(\textstyle\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(y)),$$
>
> here $h(y) \geq 0$, $t_1(y), ..., t_k(y)$ are real-valued functions of the observation y, not dependent on $\boldsymbol{\theta}$, $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-value parameter $\boldsymbol{\theta}$, not dependent on y.

Many common distributions belong to the exponential family, including the discrete distributions, binomial, Poisson, and negative binomial, and continuous distributions, normal, gamma and beta.

The relationship between λ_i and the linear predictor on a vector of covariates X_i is established through a link function $g(.)$:

$$g(\lambda_i) = \eta_i = X_i\beta.$$

Random effects may be included in the model through:

$$\eta_i = X_i\beta + \mu_i + \epsilon_i,$$

where $\mu = (\mu_1, ..., \mu_n)$ is an unstructured random component, multivariate Normal distributed with zero mean and diagonal variance-covariance matrix $\sigma^2 I_n$; $\epsilon = (\epsilon_1, ..., \epsilon_n)$ is a spatially structured random component, multivariate Normal distributed with zero mean and variance-covariance matrix Σ. Different structures for Σ have been proposed to model spatial autocorrelation.

For the particular case of Gaussian models (presented in the previous section), Σ has a SAR specification, with

$$\Sigma = \sigma^2(I_n - \rho W)^{-1}(I_n - \rho W')^{-1}.$$

Another possible specification is the conditional autoregressive (CAR) specification, with

$$\Sigma = \sigma^2(I_n - \rho W)^{-1}.$$

In both cases, W is a symmetric matrix.

The spatial econometrics models, where the variance-covariance matrix has a SAR specification, can be expressed as generalised linear model with random effects.

Although inference is typically carried out with the classical maximum likelihood method, eventually needing some numerical methods, it can also be done under the Bayesian paradigm, with pretty much the same results. On a Bayesian setting, inference for these models often requires the use of numerical techniques, such as Markov Chain Monte Carlo (MCMC) methods [6]. In the next sections the conditional autoregressive specification will be considered to account for the spatial autocorrelation in data, explored in a hierarchical Bayesian framework.

V. Spatial Econometrics Approaches for Count Data

For studying the spatial patterns in count data, several types of spatial models may be employed, namely, trough the classical spatial econometric models [20], employing a Bayesian hierarchical model [21], or considering regression models for count data in a Bayesian framework [22] [23]. Another spatial econometric approach, that incorporates spatial lag autocorrelation in modelling counts is also available, the spatial autoregressive lag model of counts, recently developed by Lambert, Brown and Florax in 2010 [7].

A. Traditional Econometric Methods

For modeling count data using spatial econometric models for continuous data, by means of an spatial autoregressive (SAR) model, or a spatial error (SEM) model, or even with both, spatial lag and spatial error considered simultaneously through the general spatial model, it is necessary to convert the count dependent variable into an approximately continuous variable. These models were designed for continuous data, demanding count data transformation to meet the model's assumptions, as well as for making use of the tests described in Section III-E which rely on a linear model.

1) From a Poisson Log-linear Model to a Linear Log Model: Considering a Poisson distribution for the number of count outcomes observed in each spatial unit the most widely used transformation is the log transformation, taking into account that the dependent variable needs to be converted into a rate variable by inclusion of an offset. A linear model is then fitted to the transformed data, and the conversion into a rate variable allows comparisons between the results obtained trough Poisson log-linear model (fitted before the covariates had also to be transformed) and the spatial econometrics models, as exemplified below.

Consider the Poisson log-linear model

$$Y \sim \text{Poisson}(\text{offset.var} \times \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)),$$

for which

$$E[Y] = \mu_Y = \text{offset.var} \times \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$$
$$\Leftrightarrow \log\left(E\left[\frac{Y}{\text{offset.var}}\right]\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

Consider now the usual log transformation to be performed to the Poisson data in order to achieve symmetry:

$$W = \begin{cases} \log(Y), & Y > 0 \\ 0, & Y = 0 \end{cases}$$

Focus on the first branch of W definition. Here:

$$\begin{aligned} W &= \log(Y) = \log(\mu_Y) + \log\left(1 + \frac{(Y-\mu_Y)}{\mu_Y}\right) \approx \\ & \quad \text{(2nd order Taylor expansion of the log)} \\ &\approx \log(\mu_Y) + \frac{(Y-\mu_Y)}{\mu_Y} - \frac{(Y-\mu_Y)^2}{2\mu_Y^2} \end{aligned}$$

So,

$$\begin{aligned} E[W] &= E[\log(Y)] \approx \\ &\approx \log(\mu_Y) + E\left[\frac{(Y-\mu_Y)}{\mu_Y}\right] - E\left[\frac{(Y-\mu_Y)^2}{2\mu_Y^2}\right] = \\ &= \log(\mu_Y) - \frac{1}{2\mu_Y} = \\ &= \log(\text{offset.var} \times \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)) - \\ &- \frac{1}{2\text{offset.var} \times \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)} = \\ &= \log(\text{offset.var}) + (\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) - \\ &- \frac{\exp(-\beta_0 - \beta_1 x_1 - \ldots - \beta_k x_k)}{2\text{offset.var}} \approx \\ & \quad \text{(1st order Taylor expansion of the exponencial)} \\ &\approx \log(\text{offset.var}) + (\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) - \\ &- \frac{1 - \beta_0 - \beta_1 x_1 - \ldots - \beta_k x_k}{2\text{offset.var}} \\ &= \log(\text{offset.var}) + \underbrace{\left\{\beta_0\left(1 + \frac{1}{2\text{offset.var}}\right) - 1\right\}}_{\text{Intercept}} + \\ &+ \beta_1 x_1 \left(1 + \frac{1}{2\text{offset.var}}\right) + \\ &+ \ldots + \beta_k x_k \left(1 + \frac{1}{2\text{offset.var}}\right) \end{aligned}$$

Consequently, and because in this particular application $P(Y = 0)$ is very small (there are always accidents!) so that $E[W]$ can be well approximated for the value above, the coefficients of the linear model are approximately the same as the coefficients of the Poisson log-linear model provided that the values of the covariates $x_1, x_2, \ldots, x_n$ are multiplied by $\left(1 + \frac{1}{2\text{offset.var}}\right)$.

B. Bayesian Hierarchical Models for Count Data

Data for which the spatial dimension is relevant, when considered within a regular or irregular lattice, generally would revel spatial autocorrelation, with closer spatial units having similar values. For count data defined into spatial units of the lattice, there are some alternative models which can be applied, wherein the spatial dependency structure is defined conditionally. The large-scale variation data is normally integrated in the model through a regression component which is added to the structure of the mean of observations.

Part of the spatial autocorrelation can be modeled by including known covariate risk factors in a regression model, but it is common for spatial structure to remain in the residuals after accounting these covariates effects. For modeling the residual autocorrelation, the most common feature is to expand the linear predictor with a set of spatially correlated random effects, as a part of a Bayesian hierarchical model [21].

The referred random effects are usually represented by a conditional autoregressive model (CAR) [24], which induces *a priori* spatial autocorrelation through the contiguity structure of the spatial units. Different CAR prior distributions commonly used for modeling spatial autocorrelation have been establish in the statistics literature, from the intrinsic and Besag, York and Mollié (BYM) proposals [24], as well as the alternatives developed by Leroux, Lei and Breslow [25] and Stern and Cressie [26] where each model is a special case of a Gaussian Markov random field. The next subsection describes and explains different Bayesian hierarchical models for Poisson count data.

1) Hierarchical Log-Poisson Model: Considering n spatial units on a spatial domain, let $y = (y_1, ..., y_n)^T$ and $E = (E_1, ..., E_n)^T$ be, respectively, the observed cases and the expected population under risk for the n spatial units, where Y_i come from a Poisson distribution, with expected value, $E(Y_i) = E_i\,\theta_i$. Let $X_i = (X_{i1}, ..., X_{ip})^T$ be a set of p covariates associated with spatial unit i, for $i = 1, ..., n$, the first of which corresponds to the intercept term, $\beta = (\beta_1, ..., \beta_p)^T$ be the corresponding regression coefficients, and $log(E) = (log(E_1), ..., log(E_n))^T$ be a vector of known offsets. The expected values of the responses are related to the linear predictor via an invertible link function, that in this case, is the natural log function, having $\lambda_i = log(E_i\,\theta_i)$, $i = 1, ..., n$ the log relative risks.
The hierarchical Log-Poisson Model is defined as follows,

[27], [28]:

$$Y_i|\lambda_i \sim Poisson(exp(\lambda_i))$$

$$\lambda_i = X_i^T\beta + \epsilon_i + log(E_i) \quad (12)$$

$$\epsilon \sim Normal(\mathbf{0}, \sigma^2\boldsymbol{D}(\rho));$$

In (12), $E_i = exp(O_i)$, $\epsilon = (\epsilon_1, ..., \epsilon_n)^T$ is a set of random effects taken to follow *a priori* a conditional auto-regressive (CAR) model [29], with covariance matrix $\boldsymbol{D}(\rho) = (I - \rho MW)^{-1}M$, where $W = ((w_{ij}))$ is the $n \times n$ contiguity matrix, that defines the neighbouring structure, and $M = ((m_{ij}))$ is a $n \times n$ diagonal matrix whose i-th diagonal entry is given by $\frac{1}{w_{i+}}$, with $w_{i+} = \sum_{j=1}^{n} w_{ij}$.

The parameter ρ defines the extent of spatial dependence and σ^2 is the measure of global variability. The main interest is to infer the true relative risks θ_i's and to estimate the model parameters $\boldsymbol{\delta}^T = (\boldsymbol{\beta}^T, \rho, \sigma^2)$. The parameters of interest are the $\theta_i s$ when the objective is mapping, however for studying spatial association between variables and spatial dependence, the parameters β and ρ are the main interest.

A simpler hierarchical Log-Poisson model only including unstructured random effects, the **independence model**, is first presented. However, when the residuals display spatial autocorrelation, this is insufficient and the spatial structure is considered through a global CAR prior, here considered with two different approaches, the **BYM** and the **Leroux** models. These methods are implemented in a Bayesian setting, where inference is based on Markov chain Monte Carlo (MCMC) simulation.

The CAR specification relies on the prior conditional distribution on the spatial error terms, where the distribution of ϵ_i conditioned on ϵ_{-i}, where $\epsilon_{-i} = (\epsilon_1, ..., \epsilon_{i-1}, \epsilon_{i+1}, ..., \epsilon_n)$, is given, and only the neighbours of areas i, according to the chosen spatial structure, are considered. CAR prior are then specified as a set of n univariate full conditional distributions, $f(\epsilon_i|\epsilon_{-i})$, for $i = 1, ..., n$, rather than via the multivariate specification.

1.1) The Independence Model, [28]

The independence model accounts for possible effects of over-dispersion, for Poisson model, if the covariates included in model (12) have removed all of the spatial structure in the response, and can be implemented with the independence prior,

$$\mu_i \sim N(0, \sigma^2), \qquad \sigma^2 \sim U(0, M_\sigma), \quad (13)$$

where μ_i replaces ϵ_i in (12). The the variance parameter is assigned an uniform prior on the interval $(0, M_\sigma)$, with M_σ large (here taken to be 1000).

1.2) Besag-York-Mollié (BYM) Model [24]

The CAR prior proposed by the intrinsic model [24] is given by

$$\epsilon_i|\epsilon_{-i} \sim N\left(\frac{\sum_{j=1}^{n} w_{ij}\epsilon_j}{\sum_{j=1}^{n} w_{ij}}, \frac{\sigma^2}{\sum_{j=1}^{n} w_{ij}}\right). \quad (14)$$

The conditional expectation is the average of the random effects in neighbouring areas, while the conditional variance is inversely proportional to the number of neighbors. τ^2 is assigned an uniform prior on the interval $(0, M_\tau)$, with M_τ large (here taken to be 1000).

The limitation with this model is that it can only represent strong spatial autocorrelation and produces random effects that are excessively smooth. Therefore, an extension of this model that allows for both, weak and strong spatial autocorrelation, is obtained by replacing ϵ_i in (12) by $\mu_i + \epsilon_i$, with μ_i defined by (13) and ϵ_i is defined by (14). This model is known as the **BYM model**.

1.3) Leroux, Lei and Breslow Method [25]

The previous model requires two random effects to be estimated, for each data point, whereas only their sum is identifiable from data. To get through this, [25] proposed an alternative, the CAR prior for modeling spatial autocorrelation using a single set of random effects. The prior is given by

$$\epsilon_i|\epsilon_{-i} \sim N(\frac{\rho\sum_{j=1}^{n} w_{ij}\epsilon_j}{\rho\sum_{j=1}^{n} w_{ij}+1-\rho}, \frac{\sigma^2}{\rho\sum_{j=1}^{n} w_{ij}+1-\rho}). \quad (15)$$

where ρ is a spatial autocorrelation parameter, with $\rho = 0$ corresponding to independence and $\rho = 1$ corresponding to strong spatial autocorrelation. An uniform prior on the unit interval is specified for ρ, $\rho \sim U(0, 1)$, and the uniform prior on the interval $(0, M_\tau)$ is adopted for τ^2.

When $\rho = 1$ the independence model is obtained.

The CAR priors defined by these models enforce a single global level of spatial smoothing for the set of random effects, which, for the Leroux model, is controled by ρ.

C. Spatial Econometrics Models for Count Data - SAR-Poisson Count Model

The methodologies just presented for describing the complex dynamic that result from the existence of spatial interaction in count data do not incorporate a spatial lag process in the spatial autoregressive count models. To fill this gap, recent developments in spatial econometrics approaches for count data have been proposed, suggesting a count estimator that models the response variable as a function of neighbouring counts. According with [7] "A Spatial Autoregressive Poisson model (SAR-Poisson) was investigated in a series of Monte Carlo experiments and estimated using a two-step limited information maximum likelihood approach " [30]. This model includes a spatially lagged dependent variable as a covariate maintaining the required distributions assumptions as well as the consistency of the linear SAR model commonly used in spatial econometric literature.

D. Regression Models for Count Data

Different methods for understanding geographical variation in counts, that evidencing some kind of heterogeneity are also available. Modeling strategies based on the use of spatial random effects models are used in order to capture unobserved spatial heterogeneity in the data, in a Bayesian perspective. Regression models, such as, the negative binomial regression, the generalized Poisson regression or the zero inflated regression models are considered [23], [22].

E. Another Possible Directions for Count Data

An alternative approach is suggested by Bhati [20], [31] to overcome the fundamental assumptions in traditional statistical methods as well as the difficulty, in certain situations, of assuming prior distributions in the Bayesian hierarchical models. He suggests the use of the cross entropy method to avoid parametrical distributional assumptions.

It is still being considered to use spatial point patterns models for these type of data, since these constitute a valid alternative in the aim of understanding temporal and spatial effects [13]. Considering as an object of interest the spatial location of events under study, i.e., the spatial location where the phenomenon of interest occurred, to be modeled. The phenomenon is described by occurrences, identified as points located in space over time, in order to study the spatial and temporal distribution testing hypotheses about the observed pattern.

VI. Alternative Way of Doing Bayesian Inference for Spatial Econometrics Models - INLA

In a Bayesian setting, although the spatial econometric models can be fitted using methods and standard software for generalised linear models with random effects, it is sometimes extremely computationally demanding. As such it might worth explore another way of fitting these models, by using INLA, the Integrated Nested Laplace Approximation, implemented in R-INLA [32]. R-INLA gives an alternative way of accomplishing Bayesian inference to spatial econometrics models, not yet explored and potentially a very interesting approach [8].

VII. Conclusion

Carrying out this research work, it was realized that the currently available models are based on the normality assumption, which is, sometimes, inappropriate. In this context, it is intended to extend some of the existing models, in order to be able to model non-continuous data, within a Bayesian hierarchical framework, with estimation methods and specific tests that will make them useful. It was further realized that the tests available for spatial autocorrelation heavily depend on a linear model assumption, which is frequently not appropriated. It is probably needed a generalization of the tests for a vaster class of models as, for example, the generalized linear models.

For future work, it is also intended to continue studying and applying spatial econometric models, for modelling data with spatial dependencies or spatial heterogeneities, through classical inference methodologies, as well as with Bayesian methods. It is an objective to proceed with the study, development and implementation of Bayesian spatial econometric models, including the Bayesian regression model, the Bayesian first order autoregressive model, as other Bayesian spatial autoregressive models [11], as well as gaining more insight on Bayesian hierarchical models for spatial data [21], in a econometric approach, heading for “spatial Bayesian hierarchical econometric models” for processing of count data.

References

[1] L. Anselin, “Thirty years of spatial econometrics,” *Papers in Regional Science*, vol. 89, pp. 3–25, 2010.
[2] J. LeSage, *The theory and Practice of Spatial Econometrics*. University of Toledo, 1999.
[3] L. Anselin(a), “Spatial dependence and spatial structural instability in applied regression analysis,” *Journal of Regional Science*, vol. 30, pp. 185–207, 1990.
[4] L. Anselin(b), *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.
[5] M. Fischer, *Spatial Analysis and Geocomputation*. Springer, 2006.
[6] G. D. and L. H., *Markov Chain Monte Carlo-Stocastic Simulation for Bayesian Inference*. Chapman and Hall, CRC, 2006.
[7] D. Lambert *et al.*, “A two-step estimator for a spatial lag model of counts-theory, small sample performance and an application,” *Regional Science and Urban Economics*, vol. 40, pp. 241–252, 2010.
[8] R. Bivand *et al.*, “Approximate bayesian inference for spatial econometric models,” *Spatial Statistics*, vol. 9, pp. 146–165, 2014.
[9] J. Wooldrige, *Introductory Econometrics, A Modern Approach*. Thomson South Western, 2006.
[10] D. Gujarati, *Basic Econometrics*. McGrawHill, 1995.
[11] J. LeSage and R. Pace, *Introduction to Spatial Econometrics*. CRC Press, 2009.
[12] G. Arbia, *Spatial Econometrics, Statistical Foundations and Aplications to Regional Convergence*. Springer, 2006.
[13] N. Cressie, *Statistics for Spatial Data*. Jonh Wiley & Sons,Inc., 1993.
[14] M. Carvalho and I. Natário, *Análise de Dados Espaciais*. Sociedade Portuguesa de Estatística, 2008.
[15] L. Anselin, *Spatial Regression Analysis in R - A Workbook*. Center for Spatially Integrated Social Sciences, 2007.
[16] L. Anselin and S. Rey, “Properties of tests for spatial dependency in linear regression models,” *Geographical Analysis*, vol. 23, pp. 112–131, 1991.
[17] L. Anselin(a) *et al.*, *Small sample properties of tests for spatial dependence in regression models: some further results. In:Luc anselin and Raymond Florax (eds.), New Directions in Spatial Econometrics*. Geographical Analysis. Berlin:Springer-Verlag, 1995.
[18] L. Anselin(b) *et al.*, “Simple diagnostic tests for spatial dependence,” *Regional Science and Urban Economics*, vol. 26, pp. 77–104, 1996.
[19] M. Turkman and G. Silva, *Modelos Lineares Generalizados- da teoria á prática*. Sociedade Portuguesa de Estatística, 2000.
[20] M. Quddus, “Modelling area-wide count outcomes with spatial correlation and heterogeneity -an analysis of london crash data,” *Accident Analysis and Prevention*, vol. 40, pp. 1486–1497, 2008.
[21] S. Banerjee *et al.*, *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2004.
[22] S. GschlöSSl and C. Czado, “Modelling count data with overdispersion and spatial effects,” *Statistical Papers*, vol. 49, pp. 531–552, 2008.
[23] R. Haining *et al.*, “Modelling small area counts in the presence of overdispersion and spatial autocorrelation,” *Computational Statistics and Data Analysis*, vol. 53, pp. 2923–2937, 2009.
[24] J. Besag *et al.*, “Bayesian image restoration with two applications in spatial statistics (with discussion),” 2004.
[25] B. Leroux *et al.*, “Estimation of disease rates in small areas: A new mixed model for spatial dependence. in me halloran, d berry (eds.),” *Statistical Models in Epidemiology, the Environment, and Clinical Trials, Springer-Verlag*, 1999.

[26] H. Stern and N. Cressie, "Inference for extremes in disease mapping, in ab. lawson, a. biggeri, d. böhning, e. lesaffre, jf. viel, r. bertollini (eds.)," *Disease Mapping and Risk Assessment for Public Health,John Wiley & Sons*, 1999.
[27] S. Dass *et al.*, "Experiences with aproximate bayes inference for the poisson-car model." Techinal Report RM679 Department of Statistics and Probability, Michigan State University, Tech. Rep., 2010.
[28] D. Lee, "Carbayes: an r package for bayesian spatial modeling with conditional autoregressive priors," *Journal of Statistical Software*.
[29] J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society B*, vol. 36, No.2, pp. 192–236, 1974.
[30] A. Doucet *et al.*, *Sequential Monte Carlo Methods in Practice*. Springer, 2001.
[31] A. Bhati, "A generalized cross-entropy approach for modeling spatially correlated counts," *Econometric Reviews*, vol. 27, pp. 574–595, 2008.
[32] R. H., "Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion)," *Journal of the Royal Statistical Society, Series B*, vol. 71, pp. 319–392, 2009.

Paula Simões is an assistant at the mathematics area departmental of Higher Institute of Engineering of Lisbon and a researcher of the Mathematics and Applications Center, of the Universidade NOVA de Lisboa, as PhD student in Statistics and Risk Management. She holds a master degree in Probability and Statistics from University of Lisbon. Her researcher area is spatial statistics and spatial econometrics.

Isabel Natário is an Assistant Professor in the Mathematics Department of the Faculty of Science and Technology of the Universidade NOVA de Lisboa and an investigator of the Statistics and Application Center of the University of Lisbon. She holds a PhD in Probability and Statistics from University of Lisbon. Her research area is spatial statistics and hierarchical Bayesian models, having published a number of peered reviewed papers and book in these subjects.