SPATIAL LIMITED DEPENDENT VARIABLE MODELS: A REVIEW FOCUSED ON SPECIFICATION, ESTIMATION, AND HEALTH ECONOMICS APPLICATIONS

Published Version:
Anna Gloria Billé, Giuseppe Arbia (2019). SPATIAL LIMITED DEPENDENT VARIABLE MODELS: A REVIEW FOCUSED ON SPECIFICATION, ESTIMATION, AND HEALTH ECONOMICS APPLICATIONS. JOURNAL OF ECONOMIC SURVEYS, 33(5), 1531-1554 [10.1111/joes.12333].

Availability:
This version is available at: https://hdl.handle.net/11585/853783 since: 2022-02-07

Published:
DOI: http://doi.org/10.1111/joes.12333

Spatial Limited Dependent Variable Models: A Review focused on Specification, Estimation and Health Economics applications

Anna Gloria Billé[a,*], Giuseppe Arbia[b]

[a]Faculty of Economics, Free University of Bolzano–Bozen, Bolzano, Italy
[b]Department of Statistical Science, Catholic University of the Sacred Heart, Rome, Italy

Abstract

Modeling individual choices is one of the main aim in microeconometrics. Discrete choice models have been widely used to describe economic agents' utility functions and most of them play a paramount role in applied health economics. On the other hand, spatial econometrics collects a series of econometric tools which are particularly useful when we deal with spatially–distributed data sets. Accounting for spatial dependence can avoid inconsistency problems of the commonly used statistical estimators. However, the complex structure of spatial dependence in most of the nonlinear models still precludes a large diffusion of these spatial techniques. The purpose of this paper is then twofold. The former is to review the main methodological problems and their different solutions in spatial nonlinear modeling. The latter is to review their applications to health issues, especially those appeared in the last few years, by highlighting the main reasons why spatial discrete neighboring effects should be considered and suggesting possible future lines of development in this emerging field. Particular attention has been paid to cross–sectional spatial discrete choice modeling. However, discussions on the main methodological advancements in other spatial limited dependent variable models and spatial panel data models are also included.

Keywords: Spatial Econometrics; Nonlinear Modeling; Limited Dependent Variable Models; Peer Effects; Health Economics.

1. Introduction

Discrete choice models with an explicit consideration of spatial neighboring effects have received less attention in the econometrics literature. Nevertheless, the role of space is becoming paramount in health economics as it is witnessed by the large and increasing amount of publications found in the literature on the subject in recent years, see e.g. Baltagi et al. (2012b), Gravelle et al. (2014), Arbia et al. (2014), and Atella

*Faculty of Economics, Free University of Bolzano–Bozen, Bolzano, Italy. Email: annagloria.bille@unibz.it, web page: https://www.unibz.it/it/faculties/economics-management/academic-staff/person/38038-anna-gloria-bille.

August 1, 2019

et al. (2014).

A plausible reason for the relatively scarce diffusion of spatial discrete choice (SDC) models in health economics and in all the other fields is certainly connected to their complexity, see Fleming (2004). Indeed, there are still a number of methodological problems to be solved connected with the computational burden and accuracy of the various techniques when dealing with large datasets which is often the case in health economics and when using microeconomic datasets. The computational problems can be very serious so that these models have experienced increasing attention in recent years by spatial econometricians, see Smirnov (2010) for a review, and the methodologies proposed are still largely unexplored. The computational problem is relevant because a–spatial discrete choices (DCs) and limited dependent variables (LDVs) are widely used to solve health problems, see Jones (2000) and Jones (2007). The estimation problems also preclude an easy extension to panel data applications, whose diffusion is experiencing a massive increase.

Modeling economic agent–based spatial relationships is a challenging problem to be solved since that individual decisions usually depend upon neighboring agents' decisions. As a matter of fact, DC models nowadays are widely used in health econometrics, see Baltagi et al. (2012a), Berta et al. (2016), Chevalier et al. (2016), and Baltagi et al. (2018). In this literature the methodological problems raised by spatial dependence are often overlooked if not totally neglected. Quite often in health applications the models are estimated in order to evaluate the impact of policies (LeSage et al., 2011) and to evaluate health programs. In this respect it is of crucial importance to estimate accurately the model parameters and to utilize sound methodologies in the associated hypothesis testing procedures. However, in the presence of spatial dependence, the parameter estimators are no more fully efficient producing a bias towards the null hypothesis that can lead, e.g. to discard health policies as ineffective in situations where they could indeed produce relevant benefits. Furthermore, when discrete data are spatially correlated the estimator can also become inconsistent. These are the major reasons why it is of paramount importance to introduce the necessary spatial econometrics corrections in the inferential phases of estimation and hypothesis testing to reduce the probability of type II errors in empirical circumstances.

The present paper is mainly focused on synchronic cross–sectional data analysis. Nevertheless, a brief review on spatial discrete choice panel data models is included. For a review of spatial linear econometric models with panel health data sets the reader is referred to Moscone and Tosetti (2014). The purpose of this paper is twofold. The former is to review the main methodological problems and their different solutions in spatial discrete choice modeling as they have appeared in the econometric literature. The latter is to review their applications to health issues, especially those appeared in the last few years, by highlighting the main reasons why spatial discrete neighboring effects should be considered and then suggesting possible future paths of the development of this emerging field.

The paper is structured in the following way. Section 2 explains spatial limited dependent variable models focusing on discrete choices. In particular, Section 2.1 describes model specifications and identification issues, Section 2.2 reports the game theoretical foundation of the these type of models, Section 2.3 provides the proper definition of the marginal impacts for spatial binary probit models, Section 2.4 briefly reviews the methodological papers which make use of endogenous weighting matrices and/or endogenous agent locations, while Section 2.5 briefly discusses the information added by spatial autoregressive coefficients. Section 3 reports an overview of spatial limited dependent variable models with their main estimation problems and related solutions. It contains also an explanation of the inconsistency problem in Section 3.3 and a brief discussion on the packages in R and Matlab in Section 3.4. Section 4 reviews the empirical applications in health economics which make use of SDC models and tries to stimulate the readers to adopt spatial econometric techniques at least as an alternative in comparison with a–spatial econometric approaches. Finally, Section 5 concludes.

2. Spatial discrete choice model specifications

Spatial econometricians are usually interested in extending standard econometric models by assuming that it is likely the presence of some form of regional/social dependence when we deal with sample units observed over a space. In the following five subsections we present the main discrete choice model specifications, the economic theoretical foundation of these spatial econometric models, the correct specifications of the marginal effects, the case when the weighting matrix is endogenous and the substantive correlation information inside the autoregessive parameter.

2.1. Models' specifications

Spatial binary and multinomial models have received a greater attention in the literature. Therefore, in the following subsections we focus the attention on these two specifications, referring to Section 3 for a brief discussion on other types of limited dependent variable models.

2.1.1. Spatial binary probit models

In this subsection we start with the definition of a general nonlinear nesting model (GNNM) and we progressively define its quite general nested–specifications by introducing different constrains. Let $\mathbf{y}_n$ be a n–dimensional stochastic vector of spatial binary variables, i.e. $\mathbf{y}_n \in \{0,1\}^n$. A GNNM with binary outcomes can be specified as follows

$$\begin{aligned} &\mathbf{y}_n^* = \rho \mathbf{W}_{1,n} \mathbf{y}_n^* + \mathbf{X}_n \boldsymbol{\beta} + \mathbf{W}_{2,n} \tilde{\mathbf{X}}_n \boldsymbol{\theta} + \mathbf{u}_n, \quad \mathbf{u}_n = \lambda \mathbf{M}_n \mathbf{u}_n + \boldsymbol{\varepsilon}_n, \quad \boldsymbol{\varepsilon}_n \sim \mathcal{N}_n\left(\mathbf{0}_n, \boldsymbol{\Sigma}_\varepsilon\right) \\ &\mathbf{y}_n = \mathbb{I}_n\left(\mathbf{y}_n^* > \mathbf{0}_n\right) \end{aligned} \tag{1}$$

where $\mathbf{y}_n^*$ is the n–dimensional vector of latent continuous dependent variables, $\mathbf{y}_n$ is the n–dimensional vector of observed binary dependent variables defined by the n–dimensional indicator function $\mathbb{I}_n\left(\mathbf{y}_n^* > \mathbf{0}_n\right) = \left(\mathbb{I}(y_1^* > 0), \ldots, \mathbb{I}(y_n^* > 0)\right)'$, $\mathbf{X}_n$ is the n by k matrix of exogenous variables including the constant term, $\tilde{\mathbf{X}}_n$ is the same matrix of regressors excluding the constant term, $\mathbf{W}_{1,n}$, $\mathbf{W}_{2,n}$ and $\mathbf{M}_n$ are n–dimensional spatial weighting matrices of known constants, $\boldsymbol{\theta}$ is a $(k-1)$–dimensional vector of parameters that capture local spatial correlation effects, $\boldsymbol{\gamma} = \left(\boldsymbol{\beta}', \rho, \lambda\right)'$ is a $(k+2)$–dimensional vector of parameters with autoregressive coefficients ρ and λ that capture global spillover effects and with $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_j, \ldots, \beta_k)'$, and $\boldsymbol{\varepsilon}_n$ is a multivariate normal vector of innovations with zero mean and finite variance $\sigma_\varepsilon^2 < \infty$, such that $\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \mathbf{I}_n$. Latent variables are then assumed to be linear functions of the regressors, but they are observed through the use of a binary variable that makes the overall model nonlinear in parameters. In the nonlinear case, σ_ε^2 is usually set to 1 for identification.

The inclusion of spatially–lagged dependent variables $\mathbf{W}_{1,n}\mathbf{y}_n^*$ typically causes an endogeneity problem (unobserved in this case), which in turn produces inconsistency of least squares estimators. In the linear regression case, Lee (2002) provided the consistency and efficiency properties of the least squares estimators for specific *mixed–regressive* (i.e. with exogenous regressors or covariates) spatial autoregressive models. The discussion will be referred to the end of Section 3.3. The problem of inconsistency is connected to the multi–directionality nature of spatial dependence in which each site, say i, is a second-order neighbor of itself, implying that spatial spillover effects have the important meaning of feedback/indirect effects also on the site where the shock may have had origin. The problem also makes the overall model a system of n *simultaneous* equations (one for each random variable observed in space), with the consequence that spatial autoregressive models cannot be viewed as simple extensions of natural *recursive* time–series econometric models, see Hamilton (1994). These types of spatial models are then multivariate by definition, with the peculiarity of having statistical information coming from one observation for each random variable in space in a cross–sectional framework.

From the model specification in equation (1), at least three nested–model specifications can be derived. By letting $\boldsymbol{\theta} = \mathbf{0}$ and for notational convenience $\mathbf{W}_{1,n} = \mathbf{W}_n$, a spatial (first–order) autoregressive–regressive probit model with (first–order) autoregressive disturbances (SARAR(1,1)–probit) can be defined, see e.g. Billé and Leorato (2017) and Martinetti and Geniaux (2017), in the following way

$$\begin{aligned} &\mathbf{y}_n^* = \rho \mathbf{W}_n \mathbf{y}_n^* + \mathbf{X}_n \boldsymbol{\beta} + \mathbf{u}_n, \quad \mathbf{u}_n = \lambda \mathbf{M}_n \mathbf{u}_n + \boldsymbol{\varepsilon}_n, \quad \boldsymbol{\varepsilon}_n \sim \mathcal{N}_n\left(\mathbf{0}_n, \boldsymbol{\Sigma}_\varepsilon\right) \\ &\mathbf{y}_n = \mathbb{I}_n\left(\mathbf{y}_n^* > \mathbf{0}_n\right). \end{aligned} \tag{2}$$

Additional conditions are needed for the identification of (ρ, λ) in a SARAR(1,1)–probit model. Specifically, $\mathbf{M}_n$ and $\mathbf{W}_n$ are assumed to be different thus allowing for different mechanisms to govern spatial correlation between shocks affecting the latent model and spatial dependence of the latent variables themselves. Then, the entire spatial dependence can be easily disentangled. It is noticetable that, when $\mathbf{W}_n = \mathbf{M}_n$, then distinguishing among the two spatial effects may become difficult, with possible identification problems of the autoregressive parameters. In this particular case, a sufficient condition to ensure identifiability of the linear model is that the covariates make a material contribution towards explaining variation in the dependent variable, i.e. at least one coefficient β_j $j = 2, \ldots, k$ is statistically significant. From an empirical point of view, disentangling the entire global impact due to (ρ, λ) is important in order to know at least how much of the total spatial spillover effects is due to the direct correlation among the dependent variables and how much is due to unobserved shocks.

The combination of different values of the two spatial autocorrelation parameters in a SARAR framework deserves a certain attention in that it may lead to a re–specification of the model. This effect is similar to the presence of common factors in autoregressive moving average (ARMA) time series models (Harvey, 1990), where under some conditions, for instance, an autoregressive moving average model of orders $p = 1$ and $q = 1$, i.e. ARMA(1,1), may reduce to an autoregressive model of order $p = 1$, i.e. AR(1). In the case of linear spatial models, this effect has been observed e.g. by Arbia (2014) who showed that when the two spatial autocorrelation parameters differ only by the sign a SARAR(1,1) model reduces to a spatial (first–order) autoregressive (SAR(1)) specification. Similar effects maybe observed in spatial discrete choice models.

By letting $\lambda = 0$ or $\rho = 0$ in equation (1), a spatial Durbin probit model or a spatial Durbin error probit model, also capable of informing about local correlation effects, can be defined, respectively. See e.g. Elhorst (2010) and LeSage (2014) for details in the linear case. The spatial Durbin probit model and the spatial Durbin error probit model can be respectively written in the following way

$$\begin{aligned}
&\mathbf{y}_n^* = \rho \mathbf{W}_{1,n} \mathbf{y}_n^* + \mathbf{X}_n \boldsymbol{\beta} + \mathbf{W}_{2,n} \tilde{\mathbf{X}}_n \boldsymbol{\theta} + \boldsymbol{\varepsilon}_n, \quad \boldsymbol{\varepsilon}_n \sim \mathcal{N}_n\left(\mathbf{0}_n, \boldsymbol{\Sigma}_\varepsilon\right) \\
&\mathbf{y}_n = \mathbb{I}_n\left(\mathbf{y}_n^* > \mathbf{0}_n\right),
\end{aligned} \tag{3}$$

$$\begin{aligned}
&\mathbf{y}_n^* = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{W}_{2,n} \tilde{\mathbf{X}}_n \boldsymbol{\theta} + \mathbf{u}_n, \quad \mathbf{u}_n = \lambda \mathbf{M}_n \mathbf{u}_n + \boldsymbol{\varepsilon}_n, \quad \boldsymbol{\varepsilon}_n \sim \mathcal{N}_n\left(\mathbf{0}_n, \boldsymbol{\Sigma}_\varepsilon\right) \\
&\mathbf{y}_n = \mathbb{I}_n\left(\mathbf{y}_n^* > \mathbf{0}_n\right).
\end{aligned} \tag{4}$$

Note that, differently from the case reported in model (2), the models in equations (3) and (4) allow for $\mathbf{W}_{1,n} = \mathbf{W}_{2,n}$ and $\mathbf{W}_{2,n} = \mathbf{M}_n$. Moreover, the spatial interaction effects among the exogenous regressors $\mathbf{W}_{2,n}\tilde{\mathbf{X}}_n$ can be specified for all the regressors in $\tilde{\mathbf{X}}_n$ but also for a subset of them. Finally, other nested spatial models can be specified by letting $\boldsymbol{\theta} = \mathbf{0}$ and $\lambda = 0$ or by letting $\boldsymbol{\theta} = \mathbf{0}$ and $\rho = 0$, identifying a spatial (first–order) autoregressive probit (SAR(1)–probit) model, see e.g. LeSage et al. (2011), and a spatial (first–order) autoregressive error probit (SAE(1)–probit) model, see e.g. McMillen (1992) and Pinkse and Slade (1998), respectively.

The above models are also useful to describe *social interaction effects* or *peer effects*. The notion that an individuals choice is affected by the behavior and/or attributes of her peers is a natural one. In the linear case, the focus is on the so–called *linear–in–means* model of social interactions. Formal econometric analysis of the linear–in–means model begins with the well–known *reflection problem* paper by Manski (1993), i.e. the analysis of the linear version of model in equation (1). As Manski (1993) states *"The reflection problem arises when a researcher observing the distribution of behaviour in a population tries to infer whether the average behaviour in some group influences the behaviour of the individuals that comprise the group"*. He notes that some interaction effects must be excluded to model identifiability since otherwise the endogenous and exogenous effects cannot be distinguished from each other (see also Elhorst, 2010 for a Monte carlo simulation on this issue). This means that some restrictions on the (spatial) weighting matrices are needed to obtain linear–in–means models: indeed, the weighting matrices are typically assumed to be row–stochastic, i.e. the weights are row–normalized. Extensions of the Manski's analysis are e.g. the papers by Bramoullé et al. (2009) and Lee (2007), with a particular focus of the SAR(1) specification with network data.

Fortunately, the reflection problem does not likely arise in binary choice models, see Brock and Durlauf (2007). From a theoretical point of view on the causal interpretation on social interactions in (spatial) discrete choice models see e.g. Brock and Durlauf (2007), Lee et al. (2014), Liu (2018) and references therein. In particular, Brock and Durlauf (2007) show the conditions under which binary choice models with social interactions can be identified or at least partially identified. In addition to the standard assumptions, like e.g. iid distributions on the random payoff terms or linear independence among the observable individual and group–specific regressors, the authors assume (i) random assignment of economic agents to groups and (ii) no unobservable (to the econometrician) group–specific characteristics. In this particular case, the identification of parameters is achieved up to scale. If only assumption (i) is relaxed, and so there is no random assignment of the economic agents to groups, then two possible cases are considered: (a) the assignment is a function of observable variables, in which case the parameters are identified and (b) the assignment is a function of unobservable variables for which the parameters are only partially identified. In contrast, if only the assumption (ii) is relaxed, then the binary model is not identified and the authors suggest two possible classes of strategies to obtain at least a partial identification of the parameters. Finally, a brief discussion of the case of heteroskedasticity is also included. From an empirical point of view, the discussion on social effects is referred to the end of Section 4.

2.1.2. Spatial multinomial probit models

A second case is when we have more than two unordered modalities. A GNNM with unordered multiple outcomes can be defined as follows

$$
\begin{aligned}
&U^*_{ij} = \rho \sum_{h=1}^{n} w_{1,ih} U^*_{hj} + \mathbf{x}'_{ij}\boldsymbol{\beta} + \sum_{h=1}^{n} w_{2,ih} \tilde{\mathbf{x}}'_{hj}\boldsymbol{\theta} + u_{ij}, \quad u_{ij} = \lambda \sum_{k=1}^{J} m_{jk} u_{ik} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_j})\\
&y_{ij} = 1 \quad iff \quad U^*_{ij} > U^*_{ik} \quad for \quad k \neq j\\
&y_{ij} = 0 \quad otherwise
\end{aligned} \tag{5}
$$

where $i, h = 1, \ldots, n$ are individuals, $j, k = 1, \ldots, J$ are alternatives (choice set) and the j–th alternative is chosen if its utility is a maximum respect to all the other alternatives, U^*_{ij} is the utility of individual i associated to alternative (preference) j, $\mathbf{x}_{ij}$ are tipically individual–specific exogenous regressors with $\boldsymbol{\beta}$ coefficients, y_{ij} is the corresponding observed multinomial variable, $\rho \sum_{h=1}^{n} w_{1,ih} U^*_{hj}$ summarizes the dependence structure between individuals' preferences, $\lambda \sum_{k=1}^{J} m_{jk} u_{ik}$ summarizes the dependence structure between unobserved attributes or selected alternatives, $\sum_{h=1}^{n} w_{2,ih}\tilde{\mathbf{x}}'_{hj}\boldsymbol{\theta}$ summarizes the individual–specific local spillover effects, and usually $\sigma_{\varepsilon_j} = 1$ for model identification. Moreover, the alternative–specific error terms σ_{ε_j} remain constant through individuals. It is typically assumed that each individual i faces the same universal choice set $C = \{1, ..., j, ..., J\}$. Finally, note that this model accounts for spatially correlated unobserved alternatives, but it should be also interesting to check if individual–specific unobserved attributes could be spatially correlated.

As in subsection 2.1.1, we can obtain nested–model specifications with some constraints on model in equation (5). By letting $\boldsymbol{\theta} = \mathbf{0}$ a SARAR(1,1)–MNP (multinomial probit) model can be defined. To the best of our knowledge, no one in the literature has yet used this type of model. By letting $\boldsymbol{\theta} = \mathbf{0}$ and $\lambda = 0$, a SAR(1)–MNP model is specified, see Smirnov and Egan (2012), and it can be written as follows

$$\begin{aligned} &U^*_{ij} = \rho \sum_{h=1}^{n} w_{1,ih} U^*_{hj} + \mathbf{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_j}) \\ &y_{ij} = 1 \quad iff \quad U^*_{ij} > U^*_{ik} \quad for \quad k \neq j \\ &y_{ij} = 0 \quad otherwise \end{aligned} \tag{6}$$

where we assume that unobserved utility functions are autocorrelated through $\rho \sum_{h=1}^{n} w_{1,ih} U^*_{hj}$, revealing that individuals' preferences depend also on the preferences of "neighboring" people (the problem in this case is to identify a reasonable topology in order to define individuals' interactions). Smirnov and Egan (2012) called this model *spatial random utility maximization (SRUM)* model. They proposed this type of model in order to measure unobserved spatial interdependencies between households and establish if these interdependencies have a significant effect on the recreational travel choices. Unfortunately, the way in which they capture these unobserved spatial/social effects is based on an aggregation of the neighboring spatial units at a county level, loosing the advantage of considering information at agent–based microeconomic data. The problem of a knowledge diffusion of these type of models into the health field seems to be caused by the insufficient information on the individuals' spatial locations.

By letting $\boldsymbol{\theta} = \mathbf{0}$ and $\rho = 0$, a spatial (first–order) autoregressive error multinomial probit (SAE(1)–MNP) model, see e.g. Bolduc et al. (1996a), can be also specified as follows

$$\begin{aligned} &U^*_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_{ij}, \quad u_{ij} = \lambda \sum_{k=1}^{J} m_{jk} u_{ik} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_j}) \\ &y_{ij} = 1 \quad iff \quad U^*_{ij} > U^*_{ik} \quad for \quad k \neq j \\ &y_{ij} = 0 \quad otherwise. \end{aligned} \tag{7}$$

Some models of this type can be found in land–use applications, see e.g. Sidharthan and Bhat (2012) and Chakir and Parent (2009), in which individuals' (land owners') interactive decisions are associated with spatial correlation among the type of use of the land (i.e. parcel units, which corresponds to the alternatives). In particular, both Sidharthan and Bhat (2012) and Ferdous and Bhat (2013) have stressed that spatial dependence among land development intensity levels is justified by the interactions between land owners of the corresponding spatial units. That is, land owners of proximately located spatial units, acting as profit–maximizing economic agents, are likely to be influenced by each other's perceptions of net stream of returns from land use development. This model can be used for example in the context of patient hospital choices, see e.g. Varkevisser et al. (2012), in which individuals maximized their utilities in choosing among different hospitals and the choice cannot depend only on hospital attributes (e.g. hospital's quality) or travel time (which justifies the use of the a–spatial mixed logit model), but also on the presence of spatial autocorrelation between those alternatives. Indeed, the most recent literature is recognizing the importance of spatial competition between alternatives, see Gravelle et al. (2014). Considering a spatial structure, at least in the simplest form of a spatial error, is more important in nonlinear models because in this case we generally have inconsistent estimates rather than a loss of efficiency as in the linear case, see Section 3.3.

2.2. Game theoretical foundations of spatial discrete choice models

The linear SARAR(1,1) model specification can be justified by a Nash equilibrium of a static complete information game with linear–quadratic utilities. The static game implies that there is no time memory among the players' actions, while the complete information game implies that the reaction functions of all the other players are observed and that the players do their actions simultaneously. The same occurs for the SARAR(1,1)–probit specification in equation (2).

Suppose that there are n individuals and they choose their actions to maximaze their utility functions. For the SARAR(1,1)–probit specification in equation (2), $y^*_{i,n}$ for $i = 1, \ldots, n$ are usually interpreted as utility functions. These functions are still linear in parameters but they are unobserved by the researcher. Suppose the researcher observes the binary action $y_{i,n} \in \{0, 1\}$ of individual i, which is directly related to its utility i.e. $y_{i,n} = \mathbb{I}\left(y^*_{i,n} > 0\right)$. In addition, suppose that the individual i's benefit from his/her utility depends on his/her characteristics, on the other's utilities and on the spatially autocorrelated unobserved characteristics: $\rho \sum_{j=1}^{n} w_{ij,n} y^*_{j,n} + x'_{i,n}\boldsymbol{\beta} + \lambda \sum_{j=1}^{n} m_{ij,n} u_{j,n} + \varepsilon_{i,n}$, which can be *strategic substitute* or *strategic complementary* if $\rho < 0$ or if $\rho > 0$, respectively. The same distintion occurs if we consider the sign of λ. Then the utility function of individual i can be written as

$$U_{i,n}\left(y_{i,n} = 1\right) = U_{i,n}\left(y^*_{i,n} > 0\right) = \rho \sum_{j=1}^{n} w_{ij,n} y^*_{j,n} + x'_{i,n}\boldsymbol{\beta} + \lambda \sum_{j=1}^{n} m_{ij,n} u_{j,n} + \varepsilon_{i,n} \tag{8}$$

where the Nash equilibrium is then $y_{i,n} = \mathbb{I}\left(\rho \sum_{j=1}^{n} w_{ij,n} y^*_{j,n} + x'_{i,n}\boldsymbol{\beta} + \lambda \sum_{j=1}^{n} m_{ij,n} u_{j,n} + \varepsilon_{i,n} > 0\right)$ and there is an implicit normalization to zero of the cost function of doing an action. Note that Xu and Lee (2018) and Xu and Lee (2015a) consider a type of SAR(1)–probit model not directly based on latent equations, i.e. $y^*_{i,n} = \rho \sum_{j=1}^{n} w_{ij,n} y_{j,n} + x'_{i,n}\boldsymbol{\beta} + \varepsilon_{i,n}$. However, this model maybe not algebraically consistent, see Klier and McMillen (2008).

For nonlinear spatial econometric models, a dependent variable cannot be expressed as a linear function of the disturbances. As Xu and Lee (2018) stressed "*... one has to investigate whether a model is well specified so that corresponding nonlinear reaction functions would have a solution. ... If a solution would exist, the next issue is to investigate whether the model would have a unique Nash equilibrium or multiple ones*". In order to guarantee at least one solution for model $y^*_{i,n} = \rho\sum_{j=1}^{n} w_{ij,n}y_{j,n} + x'_{i,n}\boldsymbol{\beta} + \varepsilon_{i,n}$, Xu and Lee (2015a) impose that $\rho \geq 0$, therefore defining a strategically complementary game called *supermodular game*, see also Xu and Lee (2018) for details. In case of the model in equation (2) and utility function in (8) no further restrictions have to be imposed. They also raise the issue of equilibrium selection. For details the reader can see the paper by Xu and Lee (2018) and references therein.

Assumptions needed to guarantee a unique Nash equilibrium would probably affect the parameter space of the autoregressive coefficients in equations (8). If the model in equation (2) is an equilibrium relationship, then Kelejian and Prucha (2010) found a proper parameter space that rules out unstable Nash equilibria. A *stable Nash equilibrium* is such that, when there is a small infinitesimal change in probabilities for one player (economic agent), this leads to a situation where two conditions hold: (i) the players (economic agents) who did not change has no better strategy in the new circumstance, (ii) the player (economic agent) who did change is now playing with a strictly worse strategy. Then, a player with the small change in their mixed strategy will return immediately to the Nash equilibrium. The parameter space they defined is based on the definition of the spectral radius of a matrix, so that for model in equation (2) we have:

Lemma 2.1. *Let $\overline{\tau}_{\mathbf{W}_n}$ and $\overline{\tau}_{\mathbf{M}_n}$ denote the spectral radius of the square n–dimensional $\mathbf{W}_n$ and $\mathbf{M}_n$ matrices, i.e.:*
$\overline{\tau}_{\mathbf{W}_n} = max\{|\omega_1|, ..., |\omega_n|\}$ and $\overline{\tau}_{\mathbf{M}_n} = max\{|m_1|, ..., |m_n|\}$, where $\omega_1, ..., \omega_n$ and $m_1, ..., m_n$ are the eigenvalues of $\mathbf{W}_n$ and $\mathbf{M}_n$, respectively. Then, $(\mathbf{I}_n - \rho\mathbf{W}_n)$ and $(\mathbf{I}_n - \lambda\mathbf{M}_n)$ are non singular for all values of ρ in the interval $(-1/\overline{\tau}_{\mathbf{W}_n}, 1/\overline{\tau}_{\mathbf{W}_n})$ and λ in the interval $(-1/\overline{\tau}_{\mathbf{M}_n}, 1/\overline{\tau}_{\mathbf{M}_n})$.

Then, the parameter spaces $(-1/\overline{\tau}_{\mathbf{W}_n}, 1/\overline{\tau}_{\mathbf{W}_n})$ and $(-1/\overline{\tau}_{\mathbf{M}_n}, 1/\overline{\tau}_{\mathbf{M}_n})$ for ρ and λ, respectively, ensure stable Nash equilibria for model in equation (2). So the model in equation (2) can be uniquely defined by 2.1. Note that if all the eigenvalues of $\mathbf{W}_n$ (but the same occurs for $\mathbf{M}_n$) are real and ($\underline{\omega} < 0, \overline{\omega} > 0$), where $\underline{\omega} = min\{\omega_1, ..., \omega_N\}$ and $\overline{\omega} = max\{\omega_1, ..., \omega_N\}$, we are in the particular case in which ρ lies in the interval $(1/\underline{\omega}, 1/\overline{\omega})$. This means that the weighting matrices in 2.1 can be also asymmetric before normalization, so that the eigeinvalues can be also not real.

2.3. Marginal effects

Billé and Leorato (2017) suggest the correct specifications of the marginal effects for spatial nonlinear autoregressive models. In the following we briefly explain how this marginal effects are defined.

In nonlinear regressions, the interpretation of the marginal effects in terms of the change in the conditional mean of $\mathbf{y}$ when regressors $\mathbf{X}$ change by one unit is no longer possible. The effects arising from changes in the explanatory variables depend in a nonlinear fashion on the levels of these variables, that is changes in the explanatory variable near the mean have a very different impact on decision probabilities than changes in very low or high values. For spatial autoregressive probit models, the nonlinearity increases in the evaluation of the marginal effects, see Beron and Vijverberg (2004), LeSage et al. (2011). Recently, Billé (2014) has also pointed out the main consequences in evaluating marginal effects with and without the consideration of heteroskedasticity implied by the spatial autocorrelation coefficient.

Let $\mathbf{x}_{.h} = (x_{1h}, x_{2h}, ..., x_{ih}, ..., x_{nh})'$ an n–dimensional vector of units referred to the h–th regressor, $h = 1, \ldots, k$, and $\mathbf{x}_{i.} = (x_{i1}, x_{i2}, ..., x_{ih}, ..., x_{ik})'$ a k–dimensional vector of regressors referred to unit i. By correctly specifying the conditional expected value and the robust conditional variances of model in (2), the following specifications of the marginal effects has been proposed

$$\begin{aligned} \frac{\partial \mathrm{P}\left(y_i = 1 \mid \mathbf{X}_n\right)}{\partial \mathbf{x}'_{.h}} |_{\bar{\mathbf{x}}} &= \phi\left(\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}^{-1/2}\left\{\mathbf{A}_\rho^{-1}\bar{\mathbf{X}}\right\}_{i.}\boldsymbol{\beta}\right)\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}^{-1/2}\{\mathbf{A}_\rho^{-1}\}_{i.}\beta_h \\ \frac{\partial \mathrm{P}\left(y_i = 1 \mid \mathbf{X}_n\right)}{\partial \mathbf{x}'_{.h}} |_{\mathbf{x}} &= \phi\left(\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}^{-1/2}\left\{\mathbf{A}_\rho^{-1}\mathbf{X}\right\}_{i.}\boldsymbol{\beta}\right)\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}^{-1/2}\{\mathbf{A}_\rho^{-1}\}_{i.}\beta_h \end{aligned} \tag{9}$$

where $\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}$ is the variance–covariance matrix implied by the reduced form of a SARAR(1,1)–probit model in equation (2), $\mathbf{A}_\rho^{-1} = (\mathbf{I} - \rho\mathbf{W})^{-1}$, $\bar{\mathbf{X}}$ is an n by k matrix of regressor–means, $(\cdot)_{i.}$ considers the i–th row of the matrix inside, and $(\cdot)_{ii}$ the i–th diagonal element of a square matrix. Note that $\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}$ reduces to $\boldsymbol{\Sigma}_{\mathbf{u}(\rho)}$ for a SAR(1)–probit specification with $\mathbf{u} = \mathbf{A}_\rho^{-1}\boldsymbol{\varepsilon}$.

The first specification of equations (9) explains the impact of a marginal change in the mean of the h–th regressor, i.e. $\bar{\mathbf{x}}_{.h}$, on the conditional probability of $\{y_i = 1\}$, i.e. $\mathrm{P}\left(y_i = 1 \mid \mathbf{X}_n\right)$, setting $\bar{\mathbf{x}}_{.h'}$ for all the remaining regressors, $h' = 1, \ldots, k-1$. The second specification of equations (9) considers, instead, the marginal impact evaluated at each single value of $\mathbf{x}_{.h}$. This is particularly informative in space since the possibility of evaluating a marginal impact with respect to a particular value $\mathbf{x}_{ih}$ have the same meaning of considering a marginal impact in a particular region/site for regressor h. The results are two n–dimensional square matrices for $\{y_1, y_2, \ldots, y_n\}$. Both the specifications should be evaluated with consistent estimates of the spatial autocorrelation coefficients $\left(\hat{\rho}, \hat{\lambda}\right)$.

Spatial marginal effects are then split into an *average direct impact* and an *average indirect impact*. The average of the main diagonal elements of the n–dimensional matrix, in both the equations, is the average direct effect (i.e., the impact from their own regions). The average of the cumulated off–diagonal elements is the average indirect effect – due to spatial spillover effects (i.e., the impact from other regions). Finally, the average total effects is the sum of them (LeSage and Pace, 2009). Changes in the value of an explanatory variable in a single observation (i.e. a spatial unit) i may influence all the $n-1$ other observations. The scalar summary measure of indirect effects cumulates the spatial spillovers falling on all other observations, but the magnitude of impact will be greatest for nearby neighbors and declines in magnitude for higher–order neighbors. LeSage et al. (2011) pointed out the need to calculate measures of dispersion for these estimates.

Billé and Leorato (2017) provided results on the evaluation of the marginal effects and their measures of dispersion through Monte Carlo simulations. In their paper, they showed that if the estimates of the spatial autocorrelation ρ in equation (2) are slightly biased for different positive/negative true values, then the direct marginal effects remain robust, as expected, but the higher is the true value of the autocorrelation coefficient, i.e. $|\rho| \to 1$, the bigger is the difference between the estimated indirect marginal effects and the true ones. The same conclusion can be made if we only consider the spatial autocorrelation coefficient λ in the error terms, since the diagonal elements of the variance–covariance matrix implied by the reduced form model, i.e. $\{\mathbf{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}$, depend on both the spatial autocorrelation coefficients and enter in the calculation of the marginal impacts in equation (9). Moreover, the above considerations are true if the weighting matrices are correctly specified. Details on model misspecification due to wrongly assumed weighting matrices can be also found in Billé and Leorato (2017). Finally, while most of the literature relies on Monte Carlo methods, some recent papers provided a series of alternatives in the case of linear spatial models (Arbia et al., 2019a).

Observation–level total effects estimates, sorted from low–to–high values of each regressors, can be also viewed as an important measure of spatial variation in the impacts (Lacombe and LeSage, 2013). This kind of interpretation permits also to account for *spatial heterogeneity* due to the variation over space of the marginal impacts with respect to the spatial distribution of the regressors. See Billé et al. (2017) for a two–step approach specifically thought to account for unobserved discrete spatial heterogeneity in the beta's coefficients via iterated local estimation procedures. Finally, note that the specification of these marginal effects (Billé and Leorato, 2017) are different from those proposed by LeSage et al. (2011) and Beron and Vijverberg (2004).

2.4. Endogeneity of the spatial weighting matrix and of the agent locations

To the best of our knowledge, the problem of endogeneity of the spatial weighting matrix and of the agent locations has not been yet considered into nonlinear regression model specifications, i.e. the properties of all the estimators in spatial nonlinear regression models are derived by assuming a fixed spatial weight matrix and exogenous agent locations. Therefore, in the following of this section we provide the state of the art on the endogeneity problem in the linear case, hoping that this will be useful to extending these main contributions to the nonlinear case.

Nowadays, part of the relevant literature is questioning how to correctly specify spatial weighting matrices in parametric spatial models to avoid possible estimation and inference problems. This issue has led to an interesting debate and to several "schools of thought" that span from using spatial semiparametric approaches (Pinkse et al., 2002) to graphical/network theory or endogenous weighting matrices, see e.g. Bhattacharjee and Jensen-Butler (2013), Ahrens and Bhattacharjee (2015). An ideal or "optimal" spatial weight matrix for analyzing all spatial phenomena is surely an unrealistic goal (Bavaud, 1998). Typically, exogenous $\mathbf{W}$ matrices are specified assuming that researchers have at least some prior knowledge of the underlying spatial structure. However, researchers are typically not sure about the form of the $\mathbf{W}$ matrix so that they take into account the possibility of estimating it through the use of some exogenous variables. The problems of endogeneity raised by such an approach are considered in e.g. Kelejian and Piras (2014) and Qu and Lee (2015). Alternatively, one can take advantage of time information, see e.g. Billé and Catania (2018).

Another set of problems emerge in those situations where the choice of agent locations are also endogenous or when agents choose voluntarily their peers thus determining the topology of the system incorporated in the **W** matrix. Under this respect Pinkse and Slade (2010) criticized the current developments of spatial econometrics by observing that "*Economists have studied the locational choices of individuals and of firms, but generally treat the characteristics of locals as given. The purpose of much spatial work, however, is to uncover the interaction among (authorities of) geographic units, who choose, e.g., tax rates to attract firms or social services to attract households. .... An ideal model would marry the two; it would provide a model explaining both individuals location decisions and the action of, say, local authorities*".

The modelling strategy which treats location as endogenous by taking care simultaneously of both locational choices and economic decision in the chosen location is one of the scope of the growing field termed spatial microeconometrics, see Dubé and go Legros (2014) and Arbia et al. (2019c). This field is rapidly emerging building upon the results of various branches of spatial statistics and on the earlier contributions of Arbia and Espa (1996), Duranton and Overman (2005), Arbia et al. (2008), Marcon and Puech (2009), Arbia et al. (2010), Espa et al. (2013), among many others. In the field of linear spatial models the problem has been discussed at a certain length by Arbia (2016) who discussed the impact of endogenous location on the **W** matrix definition. Comparatively less results are available in the area of discrete choices, see e.g. Arbia et al. (2019b).

2.5. A substantive correlation information

Apart from an omitted variable problem whose solution is a purely statistical task, it should be emphasized that the additional information deriving from the geographical location of data is of paramount importance in health economics for a number of reasons. For instance, Atella et al. (2014) developed a spatial Durbin model (SDM) by partitioning the $\mathbf{W}_n$ contiguity weighting matrix into two sub–matrices in order to take into account institutional constraints in a study of per–capita public health expenditure, finding that spatial effects plays a role mainly within entities belonging to the same institutional setting while the between effect is quite negligible. Other examples in health are those of Bolduc et al. (1996a) and Gravelle et al. (2014), which discussions are referred to Section 4.

From a substantive point of view, spatial parameters usually bear an important information content in a way that they cannot be thought as simply nuisance parameters[1]. Indeed, spatial dependence not only means lack of independence between observations, but also an underlying spatial structure, so that the autoregressive coefficient ρ should be interpreted in terms of causal relationship information parameter between y_n^* and its

[1] See Anselin (2002) for a brief discussion on differences between substantive and nuisance correlation parameters.

neighboring values in a discrete context. This should be particularly relevant in all those cases in which we need to describe social interaction/dependence effects between economic agents over space. For instance, it might be interesting to evaluate the probability that a single person takes the decision of choosing a particular health facility that has been affected by the decisions of neighboring economic agents. Moscone et al. (2012) have recently modeled peer effects between economic agents' hospital choices, but their interpretation is more related to a temporal dimension rather than a proximity in space.

3. Estimation

3.1. Generalities

Traditionally, spatial regression models are estimated by maximum likelihood (ML) method. However, this approach can often become computationally unfeasible in the presence of large samples (Arbia et al., 2019d), especially when dealing with discrete dependent variables. In order to solve this issue, some methodological and computational solutions have been recently proposed. Furthermore, in view of the possible computational advantages, many researchers seem to be increasingly incline to use Bayesian inference with the well–known MCMC and Gibbs sampling approaches (LeSage, 2000). At the same time, an emerging literature is seeing the development of semi– and non–parametric techniques (McMillen and McDonald, 2004). In the following of this section we provide a brief review of the main methodological innovations in the econometric subfield of discrete choice and limited dependent variable spatial modeling by distinguishing them according to the nature of the dependent variable, with the purpose to highlight the potential of the proposed solutions.

3.1.1. Binary variables

It is well known from the econometric literature that discrete choice models can be distinguished according to the number of outcomes of the dependent variable. Nonlinear models like *binary probit/logit models* are useful to describe binary dependent variables and both of them have received particular attention in order to introduce spatial spillover effects, see McMillen (1992), Pinkse and Slade (1998), Fleming (2004) and Beron and Vijverberg (2004). However, the spatial dependence structure adds complexity in the estimation of parameters, at least because of the implied heteroskedasticity. Solutions for inconsistency due to heteroskedastic variances in spatial probit/logit models have been proposed by e.g. Case (1992) and Pinkse and Slade (1998). However, there is no consideration in these cases on the information deriving from the off–diagonal elements of the variance–covariance matrix.

Due to the easier accessibility to computer–based solutions, a class of maximum simulated likelihood (MSL) estimators has been proposed to deal with both inconsistency and loss of efficiency, see McMillen (1992) and Beron et al. (2003). Nowadays, a major problem in maximizing this log–likelihood function with MSL approaches is represented by fact that it repeatedly involves the calculation of the determinant of n by n

matrices whose dimension depends on the sample size, in which cases the use of sparse matrices is generally recommended, see also Pace and Barry (1997). The generalized method of moments (GMM) is also affected by this problem in the nonlinear context. For this reason MSL/GMM approaches are still computationally unfeasible. Important contributions are those of Klier and McMillen (2008) in a GMM environment, of Bhat (2011) and Mozharovskyi and Vogler (2016) in the realm of the composite ML estimation, and of Martinetti and Geniaux (2017) for approximate ML estimation. However, the estimator proposed by Klier and McMillen (2008) (a linearization of the GMM proposed by Pinkse and Slade, 1998) has good properties only if the true autocorrelation coefficient is small. The other solutions are, instead, only approximations.

Differently from numerical approximation solutions, Bhat and Sener (2009) copula–based approach does not require simulation machinery and provides a simple closed–form solution, which is computationally feasible even with very large sample sizes. However, when dealing with discrete data there is no unique copula that can be defined and the interpretation of the correlation coefficient is different with respect to the autocorrelation coefficient ρ in model (2). Moreover, in Bhat and Sener (2009) there is no mention of the spillover effects and the estimation of the copula coefficient is based on a parametrization through the use of external exogenous regressors. Wang et al. (2013) proposed a partial maximum likelihood (PML) approach which is based on a trade–off solution between statistical efficiency and computational burden. Their limits are mainly relative to the model specification (a spatial error process which is less attractive for empirical applications), and the absence of a criterion for the partition of the spatial data into groups of pairs of random variables.

Billé and Leorato (2017) overcome these limits. In particular, they proposed a SARAR(1,1)–probit model specification which is more attractive also from an empirical prospective since they accounted for the (unobserved) endogenity problem implied by the inclusion of the term $\rho\mathbf{W}_{1,n}\mathbf{y}_n^*$ (i.e., the correlation among the latent dependent variables) in equation (2). A Kullback–Leibler divergence approach between the continuos Gaussian distributions of the latent variables is also included to provide a criterion for choosing the best partition, i.e. the one that minimizes the loss of statistical information, of the spatial data in terms of couples. They provided proper definitions of the marginal effects for this type of model. The paper also included extensive Monte Carlo simulations to evaluate the finite sample properties of the PML estimator and the marginal effects, also in the case when the spatial model is misspecified due to a wrongly assumed weighting matrix. Finally, the asymptotic analysis of the PML estimator is also derived, providing two direct estimators of the asymptotic variance–covariance matrix and two bootstrap approaches to obtain the estimated standard deviations. Further details are given in Section 2.3. Recently, Lu et al. (2018) have proposed a two–step generalized estimating equation approach in the quasi–maximum likelihood (QML) framework to estimating spatial nonlinear models. They have focused their analysis on spatial binary and count data variables by accounting for potential spatial processes in the error terms. To improve efficiency of the two–step approach they have also proposed a grouping estimator.

Despite the above–mentioned estimation limits, an increasing attention has recently been paid to extending the previous model specifications to *panel data*. In this context we recognise the recent work by Pinkse et al. (2006), Arduini (2016) and Baltagi et al. (2016). The first one specifies a dynamic model with a one–step GMM or continuous updating (CU) estimation procedure, whereas the second and the third proposed a semiparametric approach and a Bayesian pairwise approach, respectively. The estimation method proposed by Pinkse et al. (2006) lies within the class of generalized empirical likelihood (GEL) estimators, whose statistical properties tend to be superior with respect to a standard GMM for small/moderate sample sizes. Their dynamic spatial model mainly relies on the assumption that, by dividing the entire sample size into subgroups and groups of units, the spatial dependence between observations into different subgroups that are allocated to the same group must be relatively small. Arduini (2016) proposes a semiparametric apporach in order to relax the assumption on the distribution of the error terms. Although the proposed spatial semiparametric nonlinear least square (SSNLLS) is unfeasible, the author define a two–step procedure to estimate the dynamic spatial model. Finally, in Baltagi et al. (2016) the model assumed to be true is an m–simultaneous equation dynamic spatial model, where $m \in \{1, 2\}$ to define the bivariate case. Then, each of the two simultaneous equations reflect a dynamic SAR(1) model. Therefore, in their paper the bivariate case is relative to the number of (simultaneous) equations, with n observations each. The advantage of the Bayesian procedure over the ML one is that the first can be used with large data set. However, in their Monte Carlo simulations the sample size is equal to $n \in \{100, 500\}$, and there is no a computational comparison with the ML estimator.

3.1.2. Ordered and unordered variables

When we deal with more than two modalities (ordered or unordered), *ordered-response probit/logit models* and *multinomial probit/logit models* are adopted, respectively.

In health economics, ordered response models are usually employed to describe individual inequalities of self–assessed health (SAH) and its reporting heterogeneity (Lindeboom and Van Doorslaer, 2004), state–dependent reporting bias and justification bias (Lindeboom and Kerkhofs, 2009), or scale of reference bias problem (Groot, 2000)[2]. Although many databases require ordered discrete responses in a spatial context, few papers with spatial spillover effects have been found. Among them, two relevant papers are those by Ferdous and Bhat (2013) and Castro et al. (2013). The former developed a spatial panel ordered–response model with spatial dependence introduced in both the exogenous variables and the error terms, while accounting for unobserved spatial heterogeneity and accommodating time–varying dependency effects in a urban land–use application. The latter proposed a spatial random coefficient generalized ordered-response probit (SRC-GORP) model with a spatial intermediate formulation of the dependence structure to analyze injury severity of crashes occurring at urban intersections. Both the estimation procedures rely on the composite marginal likelihood proposed by Bhat (2011).

[2] See Greene et al. (2014) for a recent review of ordered response models for this type of applications.

Multinomial probit/logit (MNP/MNL) models are instead justified by the *random utility theory*, see McFadden (2001) and Manski (1981), and are usually considered in health economics to describe individuals' choices and utilizations of health care services. These models define individual utility functions based on some features that only vary between individuals (i.e. effects through decision-makers) together with some others that only vary among individual choices (i.e. effects among choice alternatives). To ensure a closed–form solution for the individual choice probabilities, multinomial models usually rely on two main assumptions: (i) iid error terms with a type I extreme value (or Gambel) distribution; (ii) unobserved response homogeneity. See Weeks (1997) for a review on specification and estimation of this type of models. Multinomial probit models have been generally preferred because of the independent of irrelevant alternatives (IIA) property which affects the logit (McFadden, 1981). Indeed, this property reflects an individual choice independence which is too restrictive for several applications, so that alternative specifications – *mixed logit models* – were progressively defined. See e.g. McFadden and Train (2000) for an overview of these models.

The IIA property is unlike to hold in spatial autoregressive models. The generalized–extreme value (GEV) class of models (see Hunt et al., 2004, Bhat and Guo, 2004, Bekhor and Prashker, 2008 and Pinjari, 2011) relaxes the iid assumption of the MNL model by allowing the random components of alternatives to be correlated, while maintaining the assumption that they are identically distributed, assuming a Gumbel distribution for the error terms. For instance, Bhat and Guo (2004) proposed a mixed spatially correlated logit (MSCL) model which utilized a GEV structure in order to consider utility correlation between spatial units, and they superimposed a mixed distribution on the GEV structure to capture the unobserved response heterogeneity in a housing choices study. Bekhor and Prashker (2008) examined several GEV models to discuss their adaptability on destination choice situations, with the object to determine the probability that a person from a given origin chooses a particular destination among different available alternatives. Pinjari (2011) has formally obtained the class of multiple discrete–continuous generalized extreme value (MDCGEV) models, and in particular he tested the existence and extracted the general form of the consumption probability in a closed–form, with an application in a household expenditure analysis. Finally, Bhat et al. (2015) developed a spatial multiple discrete-continuous probit (MDCP) model to specify and estimate a model of land–use change that is capable of predicting both the type and the intensity of urban development patterns over large geographic areas. Their formulation also accommodates spatial heterogeneity and heteroskedasticity in the dependent variable, and should be applicable in a wide variety of fields where social and spatial dependencies between decisions' agents lead to spillover effects in multiple discrete–continuous choices (or states). The estimation procedures of the GEV class of models rely on MSL estimation which is time–consuming as mentioned in the previous section, while Bhat et al. (2015) consider the composite marginal likelihood proposed in Bhat (2011).

3.1.3. Count data and limited dependent variables

A different discussion can be made for count data variables. As it is well–known, *count data models* are used when dependent variables consist in a count of positive integers. Due to the nature of these variables, data are usually affected by asymmetric distribution problems and by a high proportions of zero. In health economics these models have been subjected to a wide diffusion in order to analyze the demand for health care and health care utilization.

Negative binomial models (see e.g. Cameron and Trivedi, 1986) have rapidly replaced the Poisson models due to their equidispersion property that did not arise in many health data sets, see e.g. Mullahy (1997) and Gurmu (1997). Moreover, an excess of zeros that is not implied by unobserved heterogeneity requires instead the so–called zero–inflated models and their extensions: finite mixture/latent class models (Deb and Trivedi, 1997) or hurdle/two part models (Mullahy, 1986). A substantial difference between finite mixture models and two part models in health empirical applications is that the former distinguish between users and not users of the health care services (because of the binary process), whereas the latter distinguish among frequently users and the not frequently users[3]. For a discussion see e.g. Deb and Trivedi (2002) and Winkelmann (2004).

Empirical spatial econometric papers with count data dependent variables are still not very popular. Recent promising works are those of Lambert et al. (2010) and Castro et al. (2012). The former developed a two step limited information maximum likelihood (LIML) estimator for a spatial autoregressive Poisson model, with small sample properties evaluated using by Monte Carlo simulations. The latter proposed a spatial lag count model with temporal dependence in a generalized ordered response context, introducing spatial dependences by using a spatial structure on the latent continuous variables and time–varying temporal correlation patterns by means of an appropriate structure for the error term of the latent variable. The estimation procedure is based on the composite marginal likelihood by Bhat (2011).

Finally, some recent important contributions on the estimation of spatial autoregressive tobit models are those by Xu and Lee (2015b), Qu and Lee (2012) and Qu and Lee (2013). In particular, the first paper analysed the asymtpotic properties of the maximum likelihood estimator based on the spatial near–epoch dependence of the dependent variable process, see Jenish (2012), Jenish and Prucha (2009) and Jenish and Prucha (2012). Finite sample properties of the estimator are also included. The second and the third papers focused instead on the asymptotic and finite sample properties of LM test statistics for the spatial simultaneous autoregressive Tobit model.

3.2. Bayesian analysis

Because of the apparent computational advantages, Bayesian techniques have received an increasing attention in several applied research fields, especially those related to agricultural and land use issues. For

[3] Bago d'Uva (2006) proposed a finite mixture hurdle panel (FMH–Pan) model which can accommodate for both mixture and two part models. However, the author stressed that for cross–sectional data the model is characterized by identification problems.

instance, some of them made use of binary variables (Holloway et al., 2002 and Holloway et al., 2007), ordered responses (Wang and Kockelman, 2009b and Wang and Kockelman, 2009a), unordered responses (Chakir and Parent, 2009) and count data (Rathbun and Fei, 2006 and Ver Hoef and Jansen, 2007). Anyway, it should be clarified that the use of Bayesian inference should not be generally preferred to a frequentist approach without a justified statistical reason. In order to briefly explain, in all cases in which a non-correct prior distribution is chosen, the estimates may give misleading results and in most empirical cases we generally do not have sufficient information to define a proper prior. Many uninformative or diffuse priors have been proposed, but for those priors we generally expect that "*the likelihood will dominate the prior*" (i.e. the likelihood function will provide the significant part of information). Being Bayesian inference a different "way" to view the estimation of parameters, a comparison with MSL and other kind of estimators is necessary. For example, in a comparison between MSL estimator and the Gibbs sampling approach Bolduc et al. (1997) found no significant differences. Moreover, recently in LeSage and Pace (2009) it has emerged that the Bayesian MCMC estimation requires extensive simulation, it is time–consuming, not straightforward to implement and it can create converge assessment problems. Therefore, there are no significant advantages over the MSL–based estimators. However, as LeSage (2014) stressed, a Bayesian approach can be used in many situations where a prior knowledge (for example on the well–known W matrix) is required.

3.3. The problem of inconsistency

Although a long list of reasons would justify the use of spatial autoregressive models, the one considered here is the inconsistency of the standard probit estimators.

The error term in a simple probit model summarizes the unknown information coming from other regressors (i.e. omitted variables) which we assume to be uncorrelated with those in $\mathbf{X}_n$. In this case, extremum estimators (such as likelihood–based estimators) are consistent, see Amemiya (1977), Amemiya (1978) and Amemiya (1985). However, unknown forms of misspecification of the functional form (Yatchew and Griliches, 1985), for example when heteroskedastic errors are incorrectly assumed to be homoskedastic, lead to inconsistency of the maximum likelihood estimators in a nonlinear setting (Poirier and Ruud, 1988). Indeed, MLE is consistent if the conditional density of $\mathbf{y}_n|\mathbf{X}_n$ is correctly specified. Misspecification of the functional form in a probit context is equivalent to have a misspecification of the Bernoulli probability for each $y_i, 1 \leq i \leq n$.

In a SAE(1)–probit setting, heteroskedasticity will arise whenever the weights $\mathbf{M}_n$ induce non–constant diagonal terms of the matrix $\mathbf{\Sigma_u} = [\mathbf{B}_\lambda' \mathbf{B}_\lambda]^{-1}$. Indeed, this usually happens even for rather *simple* choices of $\mathbf{M}_n$, such as a k–nearest neighbor matrix. Heteroskedastic probit estimators (Case, 1992) that explicitly consider the diagonal elements of the variance–covariance matrix, i.e. $\operatorname{diag}(\mathbf{\Sigma_u}) = \operatorname{diag}[\mathbf{B}_\lambda' \mathbf{B}_\lambda]^{-1}$, remain consistent. However, the form of heteroskedasticity is generally unknown if it is implied by the spatial autocorrelation coefficient, see McMillen (1995) and Pinkse and Slade (1998).

In the general case, let $\mathbf{A}_\rho = (\mathbf{I}_n - \rho\mathbf{W}_n)$ and $\mathbf{B}_\lambda = (\mathbf{I}_n - \lambda\mathbf{M}_n)$. So we get

$$\begin{aligned}\mathbf{y}_n^* &= \rho\mathbf{W}_n\mathbf{y}_n^* + \mathbf{X}_n\boldsymbol{\beta} + \mathbf{u}_n, \quad \mathbf{B}_\lambda\mathbf{u}_n = \boldsymbol{\varepsilon}_n \\ \mathbf{B}_\lambda\mathbf{y}_n^* &= \rho\mathbf{B}_\lambda\mathbf{W}_n\mathbf{y}_n^* + \mathbf{B}_\lambda\mathbf{X}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}_n \\ \mathbf{y}_n^* &= \lambda\mathbf{M}_n\mathbf{y}_n^* + \rho\mathbf{B}_\lambda\mathbf{W}_n\mathbf{y}_n^* + \mathbf{B}_\lambda\mathbf{X}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}_n, \quad \boldsymbol{\varepsilon}_n \sim \mathcal{N}_n\left(\mathbf{0}_n, \boldsymbol{\Sigma}_\varepsilon\right)\end{aligned} \tag{10}$$

which is known as the Cochrane–Orcutt type transformation (Cochrane and Orcutt, 1949), a model in which the resulting disturbances are innovations. Even after the Cochrane–Orcutt transformation, both $\mathbf{W}_n\mathbf{y}_n^*$ and $\mathbf{M}_n\mathbf{y}_n^*$ are correlated with $\boldsymbol{\varepsilon}_n$ because

$$\mathbb{E}\left[\mathbf{y}_n^*\boldsymbol{\varepsilon}_n'\right] = \mathbf{A}_\rho^{-1}\mathbb{E}\left[\mathbf{u}_n\boldsymbol{\varepsilon}_n'\right] = \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1} \tag{11}$$

and these correlations rule out the use of nonlinear least squares methods due to their inconsistency. For the SARAR(1,1)–probit model in equation (2), and its sub–specification SAR(1)–probit by letting $\lambda = 0$, we have $\mathbb{E}\left(\left(\mathbf{W}_n\mathbf{y}_n^*\right)\mathbf{u}_n'\right) \neq \mathbf{0}_n$ where $\mathbf{u}_n = \mathbf{B}_\lambda^{-1}\boldsymbol{\varepsilon}_n$ and $\mathbb{E}\left(\left(\mathbf{W}_n\mathbf{y}_n^*\right)\boldsymbol{\varepsilon}_n'\right) \neq \mathbf{0}_n$, respectively (see Kelejian and Prucha, 1998 and Kelejian and Prucha, 1999 in the linear case). Therefore, consistency can only be achieved by correctly specifying the conditional expected value of model in equation (2).

In the linear case, the paper by Lee (2002) significantly contributed to indentify cases in which the least squares estimation is consistent and efficient. He showed that when the sum of the distances among pairs of random variables in space diverges to infinity, i.e. $\lim_{n\to\infty}\sum_{j=1}^n d_{ij} \to \infty \quad \forall i = 1,\ldots,n$, the ordinary least squares estimator (OLSE) can be consistent. However, two important considerations can be made. First of all, the property that $\lim_{n\to\infty}\sum_{j=1}^n d_{ij} \to \infty \quad \forall i = 1,\ldots,n$ implicitly requires that the weighting matrix $\mathbf{W}_n = \{w_{ij}\}$ is row–normalized, such that each row vector is defined as $w_{i,\cdot} = \frac{d_{i,\cdot}}{\sum_{j=1}^n d_{ij}}$, where $d_{i,\cdot}$ is the i–th row vector of distances. The row–normalization does not ensure the equivalence of the normalized spatial model with the original/unnormalized one (Kelejian and Prucha, 2010). Secondly, if the row–normalization is assumed then the property that $\lim_{n\to\infty}\sum_{j=1}^n d_{ij} \to \infty \quad \forall i = 1,\ldots,n$ precludes the cases of *sparse* weighting matrices, like e.g. based on a first–order contiguity criterion, since in this case each unit has usually a (fixed) finite number of neighbours. The *dense* weighting matrix is instead suitable. However, the property that $\lim_{n\to\infty}\sum_{j=1}^n d_{ij} \to \infty \quad \forall i = 1,\ldots,n$ implies that $\mathbf{W}_n = \{w_{ij}\} \to 0$, so that the OLSE is consistent as long as the spatial connections have relatively small weights: this is something similar to say that there is no spatial autocorrelation. Finally, the spatial autocorrelation coefficent λ e.g. in equation (2) appears linearly in the Cochrane–Orcutt transformed equation (10) only if we consider the latent equations $\mathbf{y}_n^*$, as in the linear case. This is not true in the nonlinear context, so that a more careful attention should be paid on the rate of convergence to establish the asymptotic distribution of least squares estimators.

3.4. Estimation procedures in R and MATLAB

The estimation procedures developed in spatial econometrics are gradually spreading out in the R language, see e.g. Arbia (2014). With respect to discrete choice models, we found the **McSpatial** package to be useful in estimating spatial binary probit models with both the MLE and the Linearized GMM proposed by Klier and McMillen (2008). The **mvProbit** package used the GHK algorithm to numerically approximate the multidimensional integral, which is unfortunately computationally unfeasible. A fast approximated ML procedure is proposed by Martinetti and Geniaux (2017) with their package **ProbitSpatial**. Finally, within the Bayesian estimation, we recognise the **spatialprobit** package. More information are available at `https://cran.r-project.org/web/views/Spatial.html`, especially in the section *spatial regression*.

Further available packages for modeling spatial limited dependent variable models are in the MATLAB software, see e.g. LeSage and Pace (2009). The functions are available at `http://www.spatial-statistics.com/software_index.htm`.

4. Spatial Discrete Choice models in Health

Although a large number of papers dealing with limited dependent variable (LDV) and discrete choice (DC) models with empirical applications can be found in health economics, those models with an explicit reference to space and spatial relationships are still not so common in the literature mainly because of the peculiarities and the micro–scale of health data. Modeling economic agent–based spatial relationships will be instead an approaching problem to be solved since individual decisions usually depend upon peers and neighboring agents' decisions. These dynamics can e.g. occur between economic agents in the demand for health care utilization.

Observational data are though vulnerable to biases in estimating effects due to non–random selection and confounding that are avoided in randomized experimental data. To properly indentify spatial neighbour effects with the use of experimental data the reader is referred e.g. to the papers by Bobonis and Finan (2009) and Lalive and Cattaneo (2009). For a recent review see also Advani and Malde (2018). In most cases, the above–mentioned peculiarities of observational health data, instead, make us unable to correctly use the econometric techniques which differ according to different observed data and they are continuously subject to criticisms and improvements by researchers[4]. The inconsistency due to incorrect functional forms in discrete choice models is particularly important and, moreover, if the researcher is also interested in estimating endogenous effects, then she has also to deal with some other important issues like e.g. *agents' self–selection*, *sorting* and *common shocks*, that are likely to be present in health economic data sets[5].

In the last 20 years, we have had an increased experience in econometric studies as the basis for health policy. As already said, most of them required the use of LDV or DC models to describe health care expenditures, treatment effects analysis, and many others, see e.g. Varin and Czado (2009), Munkin and Trivedi (2008),

[4] See for example Madden (2008) and Norton et al. (2008) for comprehensive debates.

[5] For theoretical contributions in linear spatial models that also account for serial correlation, spatial dependence (also known as *weak dependence*), and common factors (also knwon as *strong dependence*), the reader is referred to Pesaran and Tosetti (2011) and Shi and Lee (2017).

Santos et al. (2017), Varkevisser et al. (2012), Lindeboom and Kerkhofs (2009), Deb et al. (2006) and Basu et al. (2007). However, a very limited number of papers take into account space and spatial structure of discrete health data sets. Some empirical works considered a distance variable or a spatial dummy variable to distinguish between districts/regions (see e.g. Geweke et al., 2003, Wolff et al., 2008 and Nketiah-Amponsah, 2009), but none of them used spatial spillover effects by introducing autocorrelation coefficients.

Bolduc et al. (1996a) proposed the model in equation (7) in order to allow for suspected interdependencies among location choices (so among the alternatives regions or preferences) in a study of the choice of location by general practitioners in their initial work. Indeed, they argued that "*spatial correlation is likely to be present in the data because of the similarity of unobserved attributes in neighboring regions*", and found that the spatial model was to be preferred. The hybrid MNP model approximated the correlation among the utilities of the different locations using a first–order spatial autoregressive process based on a distance decaying relationship. They used a MSL estimation procedure to obtain parameter estimates. Other more recent contributions are those of Bukenya et al. (2003) who proposed a spatial ordered probit model to examine the relationship existing between quality of life (QOL), health and several socioeconomic variables. Bhat and Sener (2009) instead used a binary logit model to study teenagers' physical activity participation levels, a subject of considerable interest in the Public Health as well as in other fields. Instead of cosidering a spatial autoregressive model, they proposed a copula–based apporach and they parametrized the copula coefficient trought a geographical distance measure. Similarly, Sener et al. (2010) proposed a spatial ordered-response model to estimate physical activity participation levels by including unobserved dependences inside clusters of observations (i.e. family units) which affects those participation levels and, in the same way, Sener and Bhat (2012) extended a multinomial model to introduce spatial effects into the error terms with the motivation that it is likely the presence of unobserved residential urban form factors (such as good bicycle and walk path continuity) which may increase participation tendencies in specific activity and unobserved lifestyle perspectives (such as physically active lifestyle attitudes) that affect activity participation decisions based on the proximity of teenagers' residences.

In studies on the demand of health care, health care utilizations, and in all those cases in which we need to describe individual choices of health services among different alternatives, it is generally reasonable to assume that there are unobserved factors, which are correlated among geographical units or among individuals who are proximal in space, as it is frequently the case in health. For instance, in Nketiah-Amponsah (2009) it is likely that unobserved factors (such as tobacco and alcohol consumptions) are correlated over space since it is almost certain that individuals, especially among the youngest, have social interaction effects with those who live in the neighboring areas. This correlation information can be taken into account by specifying a simple spatial error structure of the discrete choice model, which can be used to improve coefficient estimates by avoiding inconsistent estimates[6] and leading to a correct inference approach. In Bolduc et al. (1996b), individuals' utility

[6]The problem of inconsistency in spatial binary nonlinear models is referred to Section 3.3.

functions, which are generally described by individuals' choices of health care services, can be autocorrelated in space due to social interaction effects between those individuals who are proximal in space, that is individuals' choices are also likely to be determined by neighbor individual opinions. Also in child labor (Wolff et al., 2008) and child mortality (Iram and Butt, 2008) studies, it is not to exclude the possibility of specifying a spatial autoregressive discrete choice structure, since it is reasonable to assume the presence of autocorrelation between health status among children who lie within the same neighborhood. In the same way, it could be interesting to see if there are spatial interaction effects among child labor choices taken by individuals (i.e. parents' choices) who are in the closeness. Gravelle et al. (2014) used a spatial autoregressive–regressive (SAR) model in order to detect if a hospital's quality level depends on its rivals' quality levels in a competitive setting. The main finding was that hospitals' quality levels are positively autocorrelated over space and then geographical proximity plays a paramount role in describing hospitals' competition. One way in which hospitals can raise their quality is surely the adoption of advanced technologies. In this context, it is then reasonable to assume that hospitals close in space, which are competitive in terms of technology adoptions, share information about the quantity and the quality of their qualified technologies, which in turn have an impact on the hospitals' attractive potential of patients.

Spatial dependence is inherent in many aspects to human–decision–making, with the choice decisions of one individual being affected by those of other individuals who are proximal in space. The importance of such spatial dependence has been recognized in a variety of disciplines. In recent years, it has become more common to include social interactions or neighborhood effects (i.e. social network effects) also in discrete choice models, see e.g. Goetzke and Andrade (2010), Li and Lee (2009). In particular, Goetzke and Andrade (2010) stressed the need to include social interactions and correlated effects in mode choice models as one combined spatial spillover variable for two reasons: spatial spillover serves the purpose to avoid a possible omitted variable bias, and, in addition, the spatial spillover variable can be seen as a proxy for the mode–friendliness in the neighborhood. As also Rosenquist and Lehrer (2014) stressed, if such influences are ignored estimates of the impact of policy interventions will, in many cases, be biased because they neglect the indirect pathway that occurs due to spillovers or what is known as the social multiplier effects. This should be a tempting prospect also in applied health economics, where microeconomic geo–referred data will become more and more available in the near future.

5. Conclusions

Accounting for spatial autocorrelation in the discrete or limited dependent variable is a fundamental challenge in the econometrics literature. One of the most important reasons for the relatively scarce diffusion of these models is certainly their complexity, which often require MSL or Bayesian algorithms to estimate them. To this purpose, some methodological and computational solutions have been proposed, but the aim of developing optimal estimators is still unreached. In our literature review, however, we have found that only a small number of papers use the above–mentioned models with the purpose of solving health economics issues. This fact stresses the need to popularize the potential of these models in this applied field: this was one of the aims of the present paper.

Most of the sample sizes used in health empirical applications are of the order of millions of observations because of their micro–scale nature. Indeed, Bell and Dalton (2007) highlighted the problem of specifying a weighting matrix for micro–scale or individual data, in which the difficulties are related to correctly describe all the relationships among economic agents. Data of this kind will become more and more available in the near future with the diffusion of Big Data and with the current state–of–the–art methods we are still largely unprepared to manage them and to correctly use the whole amount of information. Filling this gap in the literature will surely lay the foundations for the development of *Spatial Microeconometrics*, which can provide an unbelievable impact especially in Health Economics.

References

Advani, A. and Malde, B. (2018). Credibly identifying social effects: Accounting for network formation and measurement error. *Journal of Economic Surveys*, 32(4):1016–1044.

Ahrens, A. and Bhattacharjee, A. (2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics*, 3(1):128–155.

Amemiya, T. (1977). The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica: Journal of the Econometric Society*, pages 955–968.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. *Econometrica: Journal of the Econometric Society*, pages 1193–1205.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard university press.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267.

Arbia, G. (2014). *A primer for spatial econometrics: with applications in R*. Springer.

Arbia, G. (2016). Spatial econometrics: A broad view. *Foundations and Trends® in Econometrics*, 8(3–4):145–265.

Arbia, G., Bera, A. K., Doğan, O., and Taşpınar, S. (2019a). Testing impact measures in spatial autoregressive models. *International Regional Science Review*.

Arbia, G., Berta, P., and Dolan, C. (2019b). Measurement error induced by locational uncertainty when estimating discrete choice models with a distance as regressor.

Arbia, G. and Espa, G. (1996). *Statistica economica territoriale*. Cedam.

Arbia, G., Espa, G., and Giuliani, D. (2019c). *Spatial microeconometrics*. Routledge. In preparation.

Arbia, G., Espa, G., Giuliani, D., and Dickson, M. M. (2014). Spatio-temporal clustering in the pharmaceutical and medical device manufacturing industry: A geographical micro-level analysis. *Regional Science and Urban Economics*, 49:298–304.

Arbia, G., Espa, G., Giuliani, D., and Mazzitelli, A. (2010). Detecting the existence of space–time clustering of firms. *Regional Science and Urban Economics*, 40(5):311–323.

Arbia, G., Espa, G., and Quah, D. (2008). A class of spatial econometric methods in the empirical analysis of clusters of firms in the space. *Empirical Economics*, 34(1):81–103.

Arbia, G., Ghiringhelli, C., and Mira, A. (2019d). Estimation of spatial econometric linear models with large datasets: How big can spatial big data be? *Regional Science and Urban Economics*.

Arduini, T. (2016). Distribution free estimation of spatial autoregressive binary choice panel data models.

Atella, V., Belotti, F., Depalo, D., and Mortari, A. P. (2014). Measuring spatial effects in the presence of institutional constraints: The case of italian local health authority expenditure. *Regional Science and Urban Economics*, 49:232–241.

Bago d'Uva, T. (2006). Latent class models for utilisation of health care. *Health economics*, 15(4):329–343.

Baltagi, B. H., Egger, P. H., and Kesina, M. (2016). Bayesian spatial bivariate panel probit estimation. In *Spatial Econometrics: Qualitative and Limited Dependent Variables*, pages 119–144. Emerald Group Publishing Limited.

Baltagi, B. H., Jones, A. M., Moscone, F., and Mullahy, J. (2012a). Special issue on health econometrics: editors introduction. *Empirical Economics*, 42(2):365–368.

Baltagi, B. H., Moscone, F., and Santos, R. (2018). Spatial health econometrics. In *Health Econometrics*, pages 305–326. Emerald Publishing Limited.

Baltagi, B. H., Moscone, F., and Tosetti, E. (2012b). Medical technology and the production of health care. *Empirical Economics*, 42(2):395–411.

Basu, A., Heckman, J. J., Navarro-Lozano, S., and Urzua, S. (2007). Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatments of breast cancer patients. *Health economics*, 16(11):1133–1157.

Bavaud, F. (1998). Models for spatial weights: a systematic look. *Geographical analysis*, 30(2):153–171.

Bekhor, S. and Prashker, J. N. (2008). Gev-based destination choice models that account for unobserved similarities among alternatives. *Transportation Research Part B: Methodological*, 42(3):243–262.

Bell, K. P. and Dalton, T. J. (2007). Spatial economic analysis in data-rich environments. *Journal of Agricultural Economics*, 58(3):487–501.

Beron, K. J., Murdoch, J. C., and Vijverberg, W. P. (2003). Why cooperate? public goods, economic power, and the montreal protocol. *Review of Economics and Statistics*, 85(2):286–297.

Beron, K. J. and Vijverberg, W. P. M. (2004). *Advances in Spatial Econometrics: Methodology, Tools and Applications*, chapter Probit in a Spatial Context: A Monte Carlo Analysis, pages 169–195. Springer Berlin Heidelberg, Berlin, Heidelberg.

Berta, P., Martini, G., Moscone, F., and Vittadini, G. (2016). The association between asymmetric information, hospital competition and quality of healthcare: evidence from italy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):907–926.

Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (macml) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, 45(7):923–939.

Bhat, C. R., Dubey, S. K., Alam, M. J. B., and Khushefati, W. H. (2015). A new spatial multiple discrete-continuous modeling approach to land use change analysis. *Journal of Regional Science*, 55(5):801–841.

Bhat, C. R. and Guo, J. (2004). A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B: Methodological*, 38(2):147–168.

Bhat, C. R. and Sener, I. N. (2009). A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems*, 11(3):243–272.

Bhattacharjee, A. and Jensen-Butler, C. (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics*, 43(4):617–634.

Billé, A. G. (2014). Computational issues in the estimation of the spatial probit model: A comparison of various estimators. *The Review of Regional Studies*, 43(2, 3):131–154.

Billé, A. G., Benedetti, R., and Postiglione, P. (2017). A two–step approach to account for unobserved spatial heterogeneity. *Spatial Economic Analysis*, 0(0):1–20.

Billé, A. G. and Catania, L. (2018). Dynamic spatial autoregressive models with time-varying spatial weighting matrices. *Available at SSRN 3241470*.

Billé, A. G. and Leorato, S. (2017). Quasi-ml estimation, marginal effects and asymptotics for spatial autoregressive nonlinear models. Technical report, Faculty of Economics and Management at the Free University of Bozen.

Bobonis, G. J. and Finan, F. (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics*, 91(4):695–716.

Bolduc, D., Fortin, B., and Fournier, M.-A. (1996a). The effect of incentive policies on the practice location of doctors: a multinomial probit analysis. *Journal of labor economics*, 14(4):703–732.

Bolduc, D., Fortin, B., and Gordon, S. (1997). Multinomial probit estimation of spatially interdependent choices: an empirical comparison of two new techniques. *International Regional Science Review*, 20(1-2):77–101.

Bolduc, D., Lacroix, G., and Muller, C. (1996b). The choice of medical providers in rural benin: a comparison of discrete choice models. *Journal of health economics*, 15(4):477–498.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.

Brock, W. A. and Durlauf, S. N. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics*, 140(1):52–75.

Bukenya, J. O., Gebremedhin, T. G., and Schaeffer, P. V. (2003). Analysis of rural quality of life and health: A spatial approach. *Economic Development Quarterly*, 17(3):280–293.

Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data. comparisons and applications of some estimators and tests. *Journal of applied econometrics*, 1(1):29–53.

Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics*, 22(3):491–508.

Castro, M., Paleti, R., and Bhat, C. R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation research part B: methodological*, 46(1):253–272.

Castro, M., Paleti, R., and Bhat, C. R. (2013). A spatial generalized ordered response model to examine highway crash injury severity. *Accident Analysis & Prevention*, 52:188–203.

Chakir, R. and Parent, O. (2009). Determinants of land use changes: A spatial multinomial probit approach. *Papers in Regional Science*, 88(2):327–344.

Chevalier, A., Moscone, F., and Mullahy, J. (2016). Preface to the papers on health econometrics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):883–884.

Cochrane, D. and Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, 44(245):32–61.

Deb, P., Munkin, M. K., and Trivedi, P. K. (2006). Bayesian analysis of the two-part model with endogeneity: application to health care expenditure. *Journal of Applied Econometrics*, 21(7):1081–1099.

Deb, P. and Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of applied Econometrics*, 12(3):313–336.

Deb, P. and Trivedi, P. K. (2002). The structure of demand for health care: latent class versus two-part models. *Journal of health economics*, 21(4):601–625.

Dubé, J. and go Legros, D. (2014). *Spatial econometrics using microdata.* John Wiley & Sons.

Duranton, G. and Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106.

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, 5(1):9–28.

Espa, G., Arbia, G., and Giuliani, D. (2013). Conditional versus unconditional industrial agglomeration: disentangling spatial dependence and spatial heterogeneity in the analysis of ict firms distribution in milan. *Journal of Geographical Systems*, 15(1):31–50.

Ferdous, N. and Bhat, C. R. (2013). A spatial panel ordered-response model with application to the analysis of urban land-use development intensity patterns. *Journal of Geographical Systems*, 15(1):1–29.

Fleming, M. M. (2004). Techniques for estimating spatially dependent discrete choice models. In *Advances in spatial econometrics*, pages 145–168. Springer.

Geweke, J., Gowrisankaran, G., and Town, R. J. (2003). Bayesian inference for hospital quality in a selection model. *Econometrica*, 71(4):1215–1238.

Goetzke, F. and Andrade, P. M. (2010). Walkability as a summary measure in a spatially autoregressive mode choice model: an instrumental variable approach. In *Progress in Spatial Analysis*, pages 217–229. Springer.

Gravelle, H., Santos, R., and Siciliani, L. (2014). Does a hospital's quality depend on the quality of other hospitals? a spatial econometrics approach. *Regional science and urban economics*, 49:203–216.

Greene, W., Harris, M. N., Hollingsworth, B., and Weterings, T. A. (2014). Heterogeneity in ordered choice models: A review with applications to self-assessed health. *Journal of Economic Surveys*, 28(1):109–133.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of health economics*, 19(3):403–420.

Gurmu, S. (1997). Semi-parametric estimation of hurdle regression models with an application to medicaid utilization. *Journal of applied econometrics*, 12(3):225–242.

Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Holloway, G., Lapar, M., and Lucila, A. (2007). How big is your neighbourhood? spatial implications of market participation among filipino smallholders. *Journal of Agricultural Economics*, 58(1):37–60.

Holloway, G., Shankar, B., and Rahman, S. (2002). Bayesian spatial probit estimation: a primer and an application to hyv rice adoption. *Agricultural Economics*, 27(3):383–402.

Hunt, L. M., Boots, B., and Kanaroglou, P. S. (2004). Spatial choice modelling: new opportunities to incorporate space into substitution patterns. *Progress in Human Geography*, 28(6):746–766.

Iram, U. and Butt, M. S. (2008). Socioeconomic determinants of child mortality in pakistan: Evidence from sequential probit model. *International Journal of Social Economics*, 35(1/2):63–76.

Jenish, N. (2012). Nonparametric spatial regression under near-epoch dependence. *Journal of Econometrics*, 167(1):224 – 239.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150(1):86 – 98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, 170(1):178 – 190.

Jones, A. (2007). *Applied econometrics for health economists: A practical guide*, volume 2. Radcliffe Publishing.

Jones, A. M. (2000). Health econometrics. In *Handbook of health economics*, volume 1, pages 265–344. Elsevier.

Kelejian, H. H. and Piras, G. (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, 46:140–149.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two–stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2):509–533.

Kelejian, H. H. and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1):53–67.

Klier, T. and McMillen, D. P. (2008). Clustering of auto supplier plants in the united states: generalized method of moments spatial logit for large samples. *Journal of Business & Economic Statistics*, 26(4):460–471.

Lacombe, D. J. and LeSage, J. P. (2013). Use and interpretation of spatial autoregressive probit models. *The Annals of Regional Science*, pages 1–24.

Lalive, R. and Cattaneo, M. A. (2009). Social interactions and schooling decisions. *The Review of Economics and Statistics*, 91(3):457–477.

Lambert, D. M., Brown, J. P., and Florax, R. J. (2010). A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application. *Regional Science and Urban Economics*, 40(4):241–252.

Lee, L.-F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric theory*, 18(2):252–277.

Lee, L.-F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2):333–374.

Lee, L.-F., Li, J., and Lin, X. (2014). Binary choice models with social network under heterogeneous rational expectations. *Review of Economics and Statistics*, 96(3):402–417.

LeSage, J. (2014). What regional scientists need to know about spatial econometrics. *The Review of Regional Studies*, 44(1):13–32.

LeSage, J. and Pace, R. K. (2009). Introduction to spatial econometrics. *Boca Raton, FL: Chapman & Hall/CRC.*

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32(1):19–35.

LeSage, J. P., Kelley Pace, R., Lam, N., Campanella, R., and Liu, X. (2011). New orleans business recovery in the aftermath of hurricane katrina. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(4):1007–1027.

Li, J. and Lee, L.-F. (2009). Binary choice under social interactions: an empirical study with and without subjective data on expectations. *Journal of Applied Econometrics*, 24(2):257–281.

Lindeboom, M. and Kerkhofs, M. (2009). Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work. *Journal of Applied Econometrics*, 24(6):1024–1046.

Lindeboom, M. and Van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of health economics*, 23(6):1083–1099.

Liu, X. (2018). Simultaneous equations with binary outcomes and social interactions. *Econometric Reviews*, pages 1–17.

Lu, C., Wang, W., and Wooldridge, J. M. (2018). Using generalized estimating equations to estimate nonlinear models with spatial data. *Available at SSRN 3265976*.

Madden, D. (2008). Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of health economics*, 27(2):300–307.

Manski, C. (1981). *Alternative Estimators and Sample Designs for Discrete Choice Analysis*. The MIT Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Marcon, E. and Puech, F. (2009). Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, 10(5):745–762.

Martinetti, D. and Geniaux, G. (2017). Approximate likelihood estimation of spatial probit models. *Regional Science and Urban Economics*, 64:30 – 45.

McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272.

McFadden, D. (2001). Economic choices. *American Economic Review*, pages 351–378.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, 32(3):335–348.

McMillen, D. P. (1995). Selection bias in spatial econometric models. *Journal of Regional Science*, 35(3):417–436.

McMillen, D. P. and McDonald, J. F. (2004). Locally weighted maximum likelihood estimation: Monte carlo evidence and an application. In *Advances in spatial econometrics*, pages 225–239. Springer.

Moscone, F. and Tosetti, E. (2014). Spatial econometrics: Theory and applications in health economics. In *Encyclopedia of Health Economics*, pages 329–334. Elsevier.

Moscone, F., Tosetti, E., and Vittadini, G. (2012). Social interaction in patientsh́ospital choice: evidence from italy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(2):453–472.

Mozharovskyi, P. and Vogler, J. (2016). Composite marginal likelihood estimation of spatial autoregressive probit models feasible in very large samples. *Economics Letters*, 148:87–90.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3):341–365.

Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models. *Journal of Applied Econometrics*, 12(3):337–350.

Munkin, M. K. and Trivedi, P. K. (2008). Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics*, 143(2):334–348.

Nketiah-Amponsah, E. (2009). Demand for health insurance among women in ghana: Cross sectional evidence. *International Research Journal of Finance and Economics*, 33(1).

Norton, E. C., Dow, W. H., and Do, Y. K. (2008). Specification tests for the sample selection and two-part models. *Health Services and Outcomes Research Methodology*, 8(4):201–208.

Pace, R. K. and Barry, R. (1997). Sparse spatial autoregressions. *Statistics & Probability Letters*, 33(3):291–297.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Pinjari, A. R. (2011). Generalized extreme value (gev)-based error structures for multiple discrete-continuous choice models. *Transportation Research Part B: Methodological*, 45(3):474–489.

Pinkse, J., Slade, M., and Shen, L. (2006). Dynamic spatial discrete choice using one-step gmm: an application to mine operating decisions. *Spatial Economic Analysis*, 1(1):53–99.

Pinkse, J. and Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, 85(1):125–154.

Pinkse, J. and Slade, M. E. (2010). The future of spatial econometrics. *Journal of Regional Science*, 50(1):103–117.

Pinkse, J., Slade, M. E., and Brett, C. (2002). Spatial price competition: a semiparametric approach. *Econometrica*, 70(3):1111–1153.

Poirier, D. J. and Ruud, P. A. (1988). Probit with dependent observations. *The Review of Economic Studies*, 55(4):593–614.

Qu, X. and Lee, L.-F. (2012). Lm tests for spatial correlation in spatial models with limited dependent variables. *Regional Science and Urban Economics*, 42(3):430 – 445. Special Section on Asian Real Estate Market.

Qu, X. and Lee, L.-F. (2013). Locally most powerful tests for spatial interactions in the simultaneous sar tobit model. *Regional Science and Urban Economics*, 43(2):307–321.

Qu, X. and Lee, L.-F. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2):209–232.

Rathbun, S. L. and Fei, S. (2006). A spatial zero-inflated poisson regression model for oak regeneration. *Environmental and Ecological Statistics*, 13(4):409.

Rosenquist, J. and Lehrer, S. (2014). Peer effects, social networks, and healthcare demand. In *Encyclopedia of Health Economics*, pages 473–478. Elsevier.

Santos, R., Gravelle, H., and Propper, C. (2017). Does quality affect patients choice of doctor? evidence from england. *The Economic Journal*, 127(600):445–494.

Sener, I. N. and Bhat, C. R. (2012). Flexible spatial dependence structures for unordered multinomial choice models: formulation and application to teenagers activity participation. *Transportation*, 39(3):657–683.

Sener, I. N., Eluru, N., and Bhat, C. R. (2010). On jointly analyzing the physical activity participation levels of individuals in a family unit using a multivariate copula framework. *Journal of Choice Modelling*, 3(3):1–38.

Shi, W. and Lee, L.-F. (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics*.

Sidharthan, R. and Bhat, C. R. (2012). Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis*, 44(4):321–349.

Smirnov, O. A. (2010). Modeling spatial discrete choice. *Regional science and urban economics*, 40(5):292–298.

Smirnov, O. A. and Egan, K. J. (2012). Spatial random utility model with an application to recreation demand. *Economic Modelling*, 1(29):72–78.

Varin, C. and Czado, C. (2009). A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, 11(1):127–138.

Varkevisser, M., van der Geest, S. A., and Schut, F. T. (2012). Do patients choose hospitals with high quality ratings? empirical evidence from the market for angioplasty in the netherlands. *Journal of Health Economics*, 31(2):371–378.

Ver Hoef, J. M. and Jansen, J. K. (2007). Spacetime zero-inflated count models of harbor seals. *Environmetrics*, 18(7):697–712.

Wang, H., Iglesias, E. M., and Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, 172(1):77–89.

Wang, X. and Kockelman, K. M. (2009a). Application of the dynamic spatial ordered probit model: Patterns of land development change in austin, texas. *Papers in Regional Science*, 88(2):345–365.

Wang, X. and Kockelman, K. M. (2009b). Baysian inference for ordered response data with a dynamic spatial-ordered probit model. *Journal of Regional Science*, 49(5):877–913.

Weeks, M. (1997). The multinomial probit model revisited: A discussion of parameter estimability, identification and specification testing. *Journal of Economic Surveys*, 11(3):297–320.

Winkelmann, R. (2004). Health care reform and the number of doctor visitsan econometric analysis. *Journal of Applied Econometrics*, 19(4):455–472.

Wolff, F.-C. et al. (2008). Evidence on the impact of child labor on child health in indonesia, 1993–2000. *Economics & Human Biology*, 6(1):143–169.

Xu, X. and Lee, L.-F. (2015a). Estimation of a binary choice game model with network links. *Working paper available at https://xuxingbai.weebly.com/uploads/5/7/7/4/57743313/sar_binary_11_25_2015.pdf*.

Xu, X. and Lee, L.-F. (2015b). Maximum likelihood estimation of a spatial autoregressive tobit model. *Journal of Econometrics*, 188(1):264 – 280.

Xu, X. and Lee, L.-F. (2018). Theoretical foundations for spatial econometric research. *Regional Science and Urban Economics*.

Yatchew, A. and Griliches, Z. (1985). Specification error in probit models. *The Review of Economics and Statistics*, pages 134–139.