JTLU

A brief discussion on the treatment of spatial correlation in multinomial discrete models

Francisco J. Bahamonde-Birke
Sociale Geografie en Planologie
Universiteit Utrecht
bahamondebirke@gmail.com

Abstract: Spatial dependence plays a key role in all phenomena involving the geographic space, such as the social processes associated with transport and land use. Nevertheless, spatial dependence in multinomial discrete models has not received the same level of attention as have the other kinds of correlations in the discrete modeling literature, mainly due to the complexity of its treatment. This paper aims at offering a brief discussion on the different kinds of spatial correlation affecting multinomial discrete models and the different ways in which spatial correlation has been addressed in the discrete modeling literature. Furthermore, the paper offers a discussion on the advantages and limitations of the different approaches to treat spatial correlation and it also proposes a compromise solution among complexity, computational costs, and realism that can be useful in some specific situations.

Keywords: Multinomial discrete models, spatial correlation

1   Introduction

Multinomial discrete models play a key role in transport and land use modeling. Almost every analyst aiming at disentangling the relationships between transport and land use from a disaggregated perspective will be confronted with multinomial discrete choice models at some point of their research, as most transport (e.g., destination choices, modal choices, acquisition of mobility tools, etc.) and land use decisions (e.g., residential location, allocation of space, etc.) are of discrete nature. Consequentially, almost all transport and land use modelers may have faced issues regarding the inclusion of spatial correlation into multinomial discrete models. However, and despite its importance, within the discrete modeling literature, spatial correlation among different outcomes[1] and observations has not nearly received a similar level of attention as correlation among possible outcomes given their own characteristics (e.g., red bus / blue bus problem), correlation among different decision levels (e.g., simultaneous mode and destination choice), or correlation among the answers provided by the same individual (e.g., panel or pseudo-panel data).

---

[1] In the discrete modeling literature is customary to refer to the different possible outcomes of a model as alternatives. However, not all possible outcomes of a discrete model are necessarily alternatives. As spatial correlation is a phenomenon affecting a wide range of applications, the more general wording outcome will be preferred in this work.

The consensus seems to be that treating spatial correlation is possible and straightforward from a theoretical viewpoint (as, for instance, using a fully specified Probit model would completely capture the spatial correlation among the different outcomes), but not easy to address given data and computational limitations. Consequentially, in practice, as stated by Sener et al. (2011) spatial correlation ends up being completely ignored in many applications (with destination choice models being a paradigmatic example; Ortúzar & Willumsen, 2011) or treated in an overly simplified fashion. Certainly, it does not help that many approaches that have been proposed to deal with spatial correlation in multinomial discrete models rely upon specifications of the error terms (and consequentially upon kernels) that are not easy to deal with, be it because of the estimation time and complexity, such as the fully (or densely) specified Probit model (Schnier & Felthoven, 2011) or because of limitations in software availability, as in the case of the generalized extreme value model (GEV) (Bhat & Guo, 2004; Sener, Pendyala, & Bhat, 2011).

Consequentially, the main goal of this paper is to offer a brief discussion on the different kinds of spatial correlation affecting multinomial discrete models and on different ways in which spatial correlation has been addressed in the discrete modeling literature, highlighting its particularities which characteristically differentiate spatial correlation from other correlation problems. Furthermore, the paper offers a discussion on the advantages and limitations of the different approaches to treat spatial correlation and it also proposes a compromise solution between complexity, computational costs, and realism that can be useful in some specific situations.

2   Spatial correlation in multinomial discrete models

Tobler's First Law of Geography (Tobler, 1970) postulates that "everything is related to everything else, but near things are more related than distant things." Basically, it postulates that from a geographic perspective, spatial units located closer to each other exhibit a larger degree of interdependence that spatial units located farther away. Furthermore, it states that all spatial units exhibit some degree of spatial interdependence. When dealing with geographic data, as in the case of transport and land use models, and attempting to estimate statistical models, this spatial interdependence takes the form of spatial correlation among the data.

While spatial correlation has been extensively studied in the context of the estimation of continuous models (see, for instance, kriging processes, Cressie, 1993), the treatment of discrete outcomes poses different challenges. Furthermore, the treatment of multinomial discrete outcomes substantially differs from the treatment of binary discrete outcomes (e.g., LeSage, 2000; Ward & Gleditsch, 2002) as in multinomial discrete models each possible outcome or alternative is associated with an independent latent variable. Consequentially, within a given observation, there exist several functions characterizing the aforementioned latent variables, each with a deterministic component and an error term. Thus, the treatment of spatial correlation substantially differs from the treatment in models for which there is only one error component per observation (such as linear or binary discrete models).

For modeling purposes, the existence of spatial correlation (or correlation in general) is not necessarily a problem as long as the correlation is associated with the deterministic component of the latent variables. However, if the error terms are subject to (spatial) correlation, the usual modeling hypotheses of stochastic independence are violated, leading to biased estimators. Hence, it is necessary to explicitly address correlation in the specification of the model. In certain occasions, it would be possible to reduce the (spatial) correlation affecting the error terms by improving the specification of the deterministic component of the latent variables, or in other words, by including additional variables into the model with the goal of explicitly capturing the causes of the (spatial) correlation, and therefore disentangling (spatial) correlation from unexplained variability (e.g., in some transport and land-use applications,

spatial correlation may be somehow reduced by considering build environment attributes in the deterministic component of latent variable). However, in many cases, such variables may not exist, may not be available to the modeler, or, even if they exist and are available, they may not completely/majorly capture the underlying (spatial) correlation at the level of the error terms.

Given the particularities of spatial correlation, it is important, as a first step, to establish the kind of outcomes being modeled and the level at which spatial correlation arises. In the context of multinomial discrete models, spatial correlation may arise at three different levels, namely:

a) spatial correlation among dependent variables (among possible outcomes or alternatives), within a given observation.

b) spatial correlation among observations.

c) spatial correlation among independent variables.

It is important to note that all three kinds of spatial correlation may coexist, which further complicates their treatment. In the following, I will briefly discuss the three types of spatial correlation.

a) Spatial correlation among dependent variables (among possible outcomes)

This kind of correlation implies that the different possible discrete outcomes (or alternatives when framing a discrete model as a discrete choice model[2]) within each observation are spatially correlated. Hence, this form of correlation necessarily suggests that there must exist a spatial dependence among the different possible outcomes being modeled. Consequentially, this form of correlation is likely to appear when the outcomes are spatial units, i.e., the model aims at establishing the probability that a given spatial unit is observed as dependent variable.

This kind of spatial correlation is fairly common in transport and land use models, as it characterizes processes such as travel destination or residential/work locations choices, in which different spatial units are precisely the possible outcomes of the model (i.e., in residential/work locations the modeler would expect a larger degree of correlation among nearly located spatial units). In the context of the former examples, the existence of spatial correlation would imply that the error terms associated with the underlying utility of nearly located spatial units are similar; thus, if a given spatial unit is characterized by a larger probability than predicted by the model, nearly located spatial units are likely to exhibit a positive deviation as well.

Note that this kind of spatial correlation notably differs from the kind of correlation usually considered in continuous models, as in this case every observation could be considered to be spatially independent (provided no spatial correlation among observations coexists, as discussed at the end of the section).

b) Spatial correlation among observations

This kind of correlation most closely resembles a kriging process of continuous models, in which are not the possible outcome (or alternatives) that exhibits a pattern of spatial correlation, but the different observations. Hence, in this particular case, the spatial dependence is associated with the different observations and not with the different possible outcomes of the model. A good example of the former, in transport and land-use modeling, would be given by an urban space distribution model, in which

---

[2] While the wording "alternative" is customarily used in the literature to refer to the possible outcomes of multinomial discrete models, this wording is only accurate when dealing with choice models. In the literature on transport and land-use, there also exist models, whose possible outcomes can hardly be considered as alternatives, as they are not the result of a decision-making process (e.g., the use given to a space unit as defined by regulation or the availability of different services in a given area). Consequentially, and for the sake of completeness, throughout this paper the wording "possible outcomes" is preferred, but the reader is instructed to note that the intended meaning is equivalent to the wording "alternative" in the choice modeling literature.

the outcome is the use given to spatial unit (for instance, whether a given patch of land is dedicated to residential, industrial or commercial purposes, among others). In this particular case, no (spatial) correlation would be expected among different outcomes (the possible uses given to the spatial units), but, obviously, a modeler would expect that units that are nearly located would have a large probability of exhibiting the same outcome.

When modeling choices (i.e., considering discrete choice models) this kind of correlation may also appear every time the decision maker is associated with a given spatial distribution, e.g., when assessing the decisions of individuals or companies based on different geographic areas.

c) Spatial correlation among independent variables

This kind of correlation arises when spatially correlated variables are used as explanatory variables in the model. As it was previously discussed, independent variables affect the deterministic component of the model do not violate the hypotheses of independence among error terms. Hence, under perfect circumstances, the existence of this kind of correlation will lead to losses in efficiency and in domain (akin to multicollinearity issues in linear models), but to unbiased estimators. However, in most cases the modeler cannot rely on absolutely accurate information, and, consequentially, the explanatory variables also include a small measurement/imputation error.

In most modeling situations, small measurement/imputation errors are likely to be harmless for modeling purposes; however, if these measurement/imputation errors are also subject to spatial correlation (as they are likely to be; think, for instance, of property prices in a large geographic area), they will induce correlation in the entire model (depending on the characteristics of the model, the correlation will be induced at the level of the possible outcomes, at the level of the observations or at both). Furthermore, as the errors normally depend on the magnitude of the variable being measured, they are also likely to induce (or increase the) heteroscedasticity.

As previously mentioned, these different kinds of correlation may coexist leading to a highly involved modeling situation. A well-known example of a situation, in the field of transport and land-use, in which the three aforementioned types of spatial correlation are likely to coexist are relocation decisions of companies currently based on different locations of a geographic area (spatial correlation among observations: companies based on closely related locations prior to relocation are likely to behave more similarly than companied located farther away), which may or may not decide to relocate to a new location within the same area (spatial correlation among possible outcomes: closely related possible relocation targets are likely to exhibit similar unobserved properties), for which price information is estimated on the basis of aggregated values (correlation among independent variables: closely related locations are likely to exhibit similar prices, and consequentially exhibit a similar level of error if the estimation is not accurate).

3   Econometric frameworks

The customary approach to address multinomial discrete outcomes is based upon the assumption that they are an expression of an underlying latent variable (or utility function, when framing the discrete model as a discrete choice model in accordance with Random Utility Theory, Thurstone, 1927; McFadden, 1974). Hence, it is assumed that a given discrete outcome i, belonging to a set of possible outcomes $A(n)$ will be observed for the observation n if and only if its underlying latent variable U_{in} is larger than the latent variables U_{jn} associated with all remaining outcomes j belonging to $A(n)$. The aforementioned U_{in}, in turn, will be expressed as the sum of a representative component (V_{in}) and an error term (ε_{in}), which leads to the following expression (Train, 2009; Ortúzar & Willumsen, 2011):

$$U_{in} = V_{in} + \varepsilon_{in} \tag{1}$$

Here, the representative component *(V$_{in}$)* represents all attributes that can be quantified by an observer, and it is usually characterized through concrete and measurable properties of the outcomes (alternatives in choice models) and the observations (decision-makers in choice models); the error term *(ε_{in})*, in turn, represents all unknown or abstract elements affecting the decision.

It is customary to assume that the error terms follow an independent and identically distributed Normal distribution or an EV Type 1 distribution, leading either to the homoscedastic Probit model or to the Multinomial Logit model (MNL; Domencich & McFadden, 1975), respectively. The assumption of independent and identically distributed error terms inherent to the homoscedastic Probit and Logit models basically implies that the deviation of the observed values from the values predicted by the model do not relate to the deviations observed for any other observation or outcome. It also implies that the variability of the deviations between the observed and predicted values is constant across observations/possible outcomes. These assumptions, however, are clearly violated by the existence of spatial correlation, which postulates the existence of special dependence between nearly located spatial units.

In the following we will consider three econometric approaches (regarding the distribution of the error terms) that has been used to capture spatial correlation in the literature.

a) The fully specified Probit model

The Probit model is based upon the assumption that error terms follow a Normal distribution. Opposite to the homoscedastic Probit, the fully specified Probit model allows for a fully specified correlation pattern among error terms (i.e., a fully specified covariance matrix).

According to the fully specified Probit model, the probability of outcome *i* for the observation *n* is given by (Train, 2009):

$$P_{in} = \int I(V_{ijn} + \varepsilon_{ijn} < 0 \; \forall \; j \neq i) \cdot \frac{e^{-\frac{1}{2} \cdot \tilde{\varepsilon}_{in}^{T} \cdot \Omega_i^{-1} \cdot \tilde{\varepsilon}_{in}}}{(2\pi)^{\frac{1}{2}(J-1)} |\Omega_i|^{1/2}} \cdot d\tilde{\varepsilon}_{in} \tag{2}$$

where *I(%)* is a logic operator indicating whether *%* holds or not, while $\tilde{\varepsilon}_{in}$ is a vector of differences between ε_{in} and ε_{jn} $\forall \; j \neq i$, *Ω* represents the covariance matrix and *J* stands for the cardinality of *A(n)* (i.e., the total number of possible outcomes for observation *n*). This integral has no closed-form and it must be evaluated numerically. The integral must be computed over *J-1* dimensions and the covariance matrix will consist of *J·(J-1)/2* parameters to be estimated.

As the reader may have already noticed, in eq. 2 the covariance matrix *Ω* refers exclusively to the error terms associated with the latent variables of the different outcomes; hence, this formulation only allows capturing spatial correlation among outcomes and not among different observations. Nevertheless, it is straightforward to adapt eq. 2 to consider correlation among observations by defining $\tilde{\varepsilon}_{in}$ as a vector of differences between ε_{in} and $\varepsilon_{im} \; \forall \; m \neq n$ and *Ω* as the covariance matrix among different observations, whose cardinality is given by the number of observations *N*. In this case, the integral must be computed over *N-1* dimensions (and the covariance matrix will consist of *N·(N-1)/2* elements, forcing the modeler to impose some restrictions).

b) The GEV and the GNL models

The Generalized Extreme Value model (GEV) (McFadden, 1978) bases upon the assumption that the error terms follow an EV1 distribution. While still retaining closed-form probability functions, opposite to the MNL, it allows relaxing the assumption of independently distributed (i.i.d.) error terms. Under the GEV model, the probability of a certain outcome is given by (Daly & Bierlaire, 2003):

$$P_{in} = \frac{e^{V_i} \cdot \dfrac{\partial G(e^{V_1}, e^{V_2} \ldots, e^{V_J})}{\partial e^{V_i}}}{\mu \cdot G(e^{V_1}, e^{V_2} \ldots, e^{V_J})} \tag{3}$$

where G may be any function satisfying the following properties:

i) $G(y) \geq 0 \quad \forall\, y \in \widetilde{\,}^{\,J}_{+}$

ii) G is homogeneous of degree $\mu > 0$.

iii) $\lim\limits_{y_i \to +\infty} G(y_1 \ldots y_i \ldots y_J) = +\infty \quad \forall\, i = 1 \ldots J$

iv)   The partial derivatives of G exist and are continuous. The kth derivative is

        negative if k is odd and non-positive if k is even $\forall\, i = 1 \ldots J$.

It can be shown that the Nested Logit models (NL) (Williams, 1977; Daly & Zachary, 1978), the Paired Combinatorial Logit (PCL) (Chu, 1989) or the Cross-Nested Logit (CNL) (Bierlaire, 2001) are particular cases of the GEV. However, for the purpose of this paper we will focus on the Generalize Nested Logit model (GNL) (Wen & Koppelman, 2001), as it seems to be the most suitable model of the GEV family to address spatial correlation (and the one whose use has been reported in the literature to deal with it).

Basically, the GNL is a particular case of the GEV, which assumes the following generator function (following the formulation by Bhat & Guo, 2004):

$$G(y_1 \ldots y_i \ldots y_J) = \sum_{i=1}^{J-1} \sum_{j=i+1}^{J} \left((\alpha_{i,ij} \cdot y_i)^{1/\mu} + (\alpha_{j,ij} \cdot y_j)^{1/\mu} \right)^{\mu} \tag{4}$$

Here, $0 < \mu \leq 1$ represents a correlation parameter that can be estimated. Correlation increases as μ approaches zero and collapses to the MNL, when μ approaches one. $0 < a_{i,ij} \leq 1$ is known as the allocation parameter and, as $\sum \alpha_{i,ij} = 1$, it roughly represents the proportion to which the error term ε_i is associated with the correlated pair (i.e., nest) $\varepsilon_i, \varepsilon_j$. The correlation between ε_i and ε_j increases as $a_{i,ij}$ and $a_{j,ij}$ get larger. Consequentially, the GNL model allows estimating the degree of correlation between any pair of latent variables in a discrete model, without sacrificing the closed-form probabilities that characterize GEV models.

c) The Mixed Logit (ML) model

Mixed logit models (Cardell & Dunbar, 1980) assume that the stochastic component of the model is given by the sum of an i.i.d. error term (ε_{in}) following an EV1 distribution and another stochastic element vin (known as mixing distribution) that can follow any desired distribution. It allows not only

to account for heteroscedasticity and correlation between different possible outcomes and observations (as the additional stochastic elements are not subject to homoscedasticity and no-autocorrelation restrictions) but to consider any desired functional form. This way, the latent variable of a given outcome would be given by:

$$U_{in} = V_{in} + \upsilon_{in} + \varepsilon_{in} \tag{5}$$

In this case, and opposite to the models of the GEV family, the probability of a given outcome cannot be longer depicted through a closed-form expression and it is necessary to integrate the probabilities over the probability density function of the mixing distribution. Hence, the probability takes the following form (Train, 2009):

$$P_{in} = \int P_{in}^{*}(V_n, \upsilon_n) \cdot f(\upsilon_n) \cdot d\upsilon_n \tag{6}$$

where the first component P^{*}_{in} stands for the usual MNL probabilities and $f(\upsilon_n)$ is the probability density function of the mixing distribution. In this representation, V_n and υ_n represent vectors accounting for the latent variables associated with all possible outcomes of the model and for all mixing distributions included in the model, respectively. As usually this integral cannot be analytically solved, it is necessary to compute it relying on numerical techniques.

Depending on the assumptions regarding the mixing distributions, the ML framework allows addressing correlation among outcomes, observations as well as interactions between error terms and independent variables. However, this high flexibility does not come at no-costs as every mixing error term to be considered implies that the integral depicted in eq. 6 increases in one dimension.

4 Treating spatial correlation in multinomial discrete models

As previously outlined, the main reason why spatial autocorrelation has received much less attention than other correlation problems is its high degree of complexity. Opposite to other problems in which possible outcomes are correlated, problems in which spatial correlation arise, are normally characterized by a very large number of possible outcomes; for instance, a residential location model can easily have a couple hundred possible outcomes. Thus, considering spatial correlation may require the evaluation of integrals in several dimensions (when following the Probit or the ML approaches) and the estimation of an unwieldy large number of parameters (the number of parameters in the covariance matrix increases quadratically with the number of possible outcomes). This *curse of dimensionality* leads to prohibitive computational costs. Similarly, when treating spatial correlation among observations, the size of the problem widely differs from other problems such as panel data, in which only the answers provided by the same individual are assumed to be correlated (and consequentially are assumed to be independent of the answers provided by different individuals). Yet, the dimensionality problem is even more grievous when treating spatial correlation among observations, as the correlation matrix has n more parameters to be estimated ($N\cdot(N-1)/2$ parameters with N representing the total amount of observations) than the number of available observations, meaning that estimating a fully specified correlation matrix is not only prohibitively expensive but also theoretically impossible.

Given the aforementioned problems, spatial correlation is considered making use of spatial weighting matrices. Instead of estimating all elements of the correlation matrix (without considering addi-

tional information), spatial weighting matrices *(W)* are used to establish the degree of spatial correlation among different outcomes/observations. Then, *W* is used to estimate the correlation matrix taking advantage of relationships established in *W* (i.e., without the necessity of estimating all elements of a fully-specified covariance matrix). Spatial weighting matrices are normally defined exogenously (Bhat & Guo, 2004; Schnier & Felthoven, 2011; Weiss & Habib, 2017; among many others), but they can also be defined parametrically, imposing a functional form (Sener et al., 2011; Zhou, Wang, & Holguín-Veras, 2016). The latter approach still allows for estimating the degree of interdependence among outcomes/observation but it substantially reduces the amount of parameters to be estimated (compared with the estimation of a fully-specified covariance matrix).

The spatial weighting matrix is defined in such a way, that the degree of spatial dependence between two possible outcomes/observations diminishes (i.e., approaches zero) as the distance between them increases, hence, satisfying Tobler's First Law of Geography. However, as Zhou et al. (2016) extensively discuss, the concept of distance between two spatial units (especially in social sciences) may go beyond Euclidean distance, encompassing concepts such as economic or demographic distance, among others. The extent to which the spatial dependence is affected by different distances between spatial units depends on the model itself. Consequentially, it seems appropriate to allow for the model to endogenously determine the best way to establish the distance between spatial units and to utilize a parametric depiction of the spatial weighting matrix.

As previously mentioned, different approaches have been utilized to capture different kinds of spatial correlation in the past. In the following, and without claiming completeness, I will discuss some of them, in light of their advantages and limitations.

a) Probit model to capture spatial correlation among possible outcomes

As previously mentioned, a fully specified Probit model would require the estimation of *J·(J-1)/2* parameters, where *J* represents the total number of possible outcomes. As the former would require an implausible large number of observations, the covariance matrix Ω (see eq. 2) is usually specified making use of the aforementioned spatial weighting matrix *W*, so that:

$$\Omega = (I - \rho \cdot W)^{-1} \cdot \sigma^2 \cdot (I - \rho \cdot W)^{-T} \tag{7}$$

where ρ is a parameter to be estimated and σ^2 a vector of variances of independently distributed disturbances ε following a Normal distribution. The elements of vector σ^2 may be both homoscedastic (Garrido & Mahmassani, 2000) or heteroscedastic (Schnier & Felthoven, 2011). While most applications define the structure of *W* exogenously (e.g., Schnier & Felthoven, 2011, define the elements of *W* as the inverse of the squared distance between two possible outcomes or Garrido & Mahmassani, 2000, define the elements of *W* on the basis of adjacency), there are no theoretical restrictions preventing the analyst from using a parametrical definition of the elements, to be estimated endogenously (although empirical restrictions may indeed exist).

While the specification of the model is not complex, from a theoretical perspective, the same cannot be said for the estimation, which will require solving a multidimensional integral in *J-1* dimensions. As this integral must be solved numerically, it would lead to prohibitive costs as the number of possible outcomes increases. Hence, most applications relying on the Probit model to consider spatial correlation among outcomes limit the number of possible spatial units: e.g., Schnier and Felthoven (2011) consider up to 24 possible outcomes, while Garrido and Mahmassani (2000) considered up to 41.

b) Probit model to capture spatial correlation among observations

Basically, the same considerations mentioned in the previous point also apply when the Probit model is used to capture spatial correlation among observations. However, the main difference, in this case, is that the size of the covariance matrix and the degree of the multidimensional integral cannot be limited, as it is directly given by the number of observations N (note that as a corollary, in this case, the error terms cannot be considered to be heteroscedastic, as it would require the estimation of $N-1$ parameters, completely exhausting the degrees of freedom of the model).

To tackle this problem, the authors relying upon this approach have focused their effort on the computation of the multidimensional integral. For this purpose, some studies rely upon the Bayesian Markov Chain Monte Carlo (MCMC) estimation (Chakir & Parent, 2009; Zhou et al., 2016), while others use the Maximum Approximate Composite Marginal Likelihood (MACML) (Sidharthan & Bhat, 2012; Bhat, 2015). Both estimations techniques offer significantly faster results than the conventional numerical integration, but their implementation is cumbersome (requiring a significant coding effort) and is outside the reach of the majority of the analysts.

c) GEV-GNL model to capture spatial correlation among possible outcomes

Considering spatial correlation among possible outcomes is straightforward when using the GNL model. Basically, the analysis simply has to specify the structure of $a_{i,ij}$, which are indicative for the interdependence between the outcomes i and j (see eq. 4). While the elements $a_{i,ij}$ and $a_{j,ij}$ do not directly correspond to the elements of the spatial weighting matrix is W or of the covariate matrix Ω, they are indeed related (Bhat & Guo, 2004 – to which the reader is referred for a detailed explanation – include a tabulation of the correlation for several combinations of μ, $a_{i,ij}$, and $a_{j,ij}$) and the considerations applying to the specification of W or Ω also apply to the specification of $a_{i,ij}$ and $a_{j,i}$. So, for instance, Bhat and Guo (2004) specified those elements exogenously, on the basis of adjacency, while Parady and Hato (2016) used an endogenous specification including two variables (Euclidean distance and altitude). Similarly, Sener et al. (2011) define a generalized parametric formulation, which they call Generalized Spatially Correlated Logit (GSCL).

Given that the models of the GEV family lead to closed-form kernels, the use of this approach to capture spatial correlation among possible outcomes is not doomed by the *curse of dimensionality*, or, more precisely, the GNL is affected to a much lesser extent than other models by the increase of dimensions. Consequentially, the computational costs are much less compared with any existent alternative and treating a large number of spatially correlated outcomes is not an impending limitation. For instance, Bhat and Guo (2004) considered more the 900 spatially correlated spatial units.

The main limitation of the model is, however, the lack of software availability, which implies that the analyst would necessarily be forced to code the model. Furthermore, while also an advantage regarding the numbers of parameters to be estimated, the GNL imposes a homoscedastic error structure (i.e., it imposes that the variability of the error is constant for all spatial unis), which can be disadvantageous when dealing with spatial correlation (especially when the correlation is induced by spatially correlated explanatory variables, which are not equally variated themselves).

d) GEV-GNL model to capture spatial correlation among observations

Given their structure, models of the GEV family can only be used to establish the probabilities of different possible outcomes within a given observation (the probability functions of the GEV family assume independent error terms across observations). Hence, the approach cannot be used to capture spatial correlation among observations.

e)   ML model to capture spatial correlation among possible outcomes

The approach closely resembles the treatment of spatial correlation given in the Probit model. Basically, the approach considers error terms of the form:

$$\tilde{\varepsilon} = (I - \rho \cdot W)^{-1} \cdot \upsilon + \varepsilon \qquad\qquad\qquad (8)$$

where ε is a vector of i.i.d errors following an EV1 distribution and υ is a vector of error terms following any desired distribution. It is customary to assume that υ follows a Normal distribution, which greatly increases the resemblance to the Probit treatment (even Bolduc et al., 1996, following the ML approach, claimed to rely on a Probit structure). As usual in ML models, the error term υ is considered via simulation leading to probabilities of the form depicted in eq. 6 and it is necessary to generate as many random disturbances as possible outcomes (or as latent variables to be considered). Consequentially, the approach is necessarily heteroscedastic and requires the computation of a *(J-1)*-dimensional integral.

While the computation of the integral depicted in eq. 6 is slightly easier than the computation of the integral associated with the Probit model, the approach still suffers from the *curse of dimensionality*, which greatly limits the number of possible outcomes than can be handled relying upon this structure (e.g., Bolduc, Fortin, & Fournier, 1996 considered 18 spatial units only). The treatment of the spatial weighting matrix W is analogous to the Probit model and it also can be defined exogenously or endogenously (using a parametric specification), although, to my best knowledge, for this approach no studies considering an endogenous representation of W have been reported in the literature.

Weiss and Habib (2017) present a specification based on a Cholesky decomposition of W, which they call Spatially Weighted Error Correlation (SWEC) error model, but it basically follows the same principles proposed by Bolduc et al. (1996; both considered exogenously defined W matrix). Weiss, Hasnine, and Habib (2019) showed that, for their dataset, the ML treatment outperforms the GNL treatment, but that is probably due to the heteroscedastic structure of the model (opposite to the homoscedastic GNL), which adds a substantial number of degrees of freedom (parameters to be estimated).

While in terms of computational costs the ML treatment does not majorly differ from the Probit treatment, it certainly offers (in comparison with the Probit) a potential advantage that (to my best knowledge) has not yet been considered in the literature. If the analyst defines a structure of block-correlation (e.g., by grouping together spatial units expected to have a high degree of correlation, given their generalized distance or other considerations, such as political divisions) with all possible outcomes belonging to a given block being affected by the same disturbance (note that it does not imply perfect correlation within the block, given the existence of the error ε), the analyst would be able to reduce the number of elements to be considered (and the dimensions of the integral) to the number of blocks (as it would be necessary to estimate only one parameter per block and not one parameter per each pair of outcomes). The block-correlation structure can also be overlapping. This treatment, while an exogenous approximation of the approach proposed by Bolduc et al. (1996), would offer significant gains in terms of computational costs and allow considering a much larger number of spatially correlated outcomes as it has been possible in the past.

f)   ML model to capture spatial correlation among observations

Given that the ML treatment requires to generate a random disturbance for any element to be correlated in the covariance matrix, it implies that using the ML to capture spatial correlation among observations requires generating a random disturbance for all observations, which would completely exhaust the degrees of freedom of the model. Consequentially, the (fully specified) ML treatment cannot be used to capture spatial correlation among observations and it has not been used in the past.

Nevertheless, if, akin to the block-correlated representation introduced in the previous point, the modeler would be able to exogenously define a structure of block correlated observations affected by the same disturbance, they would be able to reduce of number of disturbances to be considered via simulation, enabling the treatment of spatially correlated observations on the basis of the ML. This approach has not been previously reported in the literature.

g) Capturing spatial correlation induced by spatially correlated independent variables

To my best knowledge, no existent study has directly and explicitly dealt with correlation induced by spatially correlated independent variables.[3] However, depending on the kind of correlation being induced, the treatment should be akin to the treatment of spatial correlation among possible outcomes or observations.

Basically, problems associated with spatial correlation among independent variables will only lead to biased estimators when they also include spatially correlated measurement/imputation errors. Let's consider a model including a vector of explanatory variables X_{in} with spatially correlated measurement/imputation errors and a vector Y_{in} of independent variables no with spatially correlated measurement/imputation errors, so that:

$$\tilde{X}_{in} = X_{in} + \upsilon_{in} \tag{9}$$

where $\tilde{X}_{in}$ stands for the values available to the modeler and X_{in} for the actual values without errors, while υ_{in} is a vector of spatially correlated measurement/imputation errors. In this case, the latent variable U_{in} associated with outcome i and observation n can be expressed as:

$$U_{in} = f\left(X_{in}, Y_{in}\right) + \varepsilon_{in} \quad => \quad U_{in} = f\left(\tilde{X}_{in} + \upsilon_{in}, Y_{in}\right) + \varepsilon_{in} \tag{10}$$

Then, considering the following Taylor expansion:

$$U_{in} \approx f\left(\tilde{X}_{in}, Y_{in}\right) + \frac{\partial f}{\partial X_{in}}\left(\tilde{X}_{in}, Y_{in}\right) \cdot \upsilon_{in} + \varepsilon_{in} \tag{11}$$

the underlying latent variable can be expressed as:

$$U_{in} \approx f\left(\tilde{X}_{in}, Y_{in}\right) + \tilde{\varepsilon}_{in} \quad with \quad \tilde{\varepsilon}_{in} = \frac{\partial f}{\partial X_{in}}\left(\tilde{X}_{in}, Y_{in}\right) \cdot \upsilon_{in} + \varepsilon_{in} \tag{12}$$

where $\tilde{\varepsilon}_{in}$ is a spatially correlated error term, despite the fact that ε_{in} may have been not spatially correlated (note that the correlation is only due to the spatially correlated measurement/imputation errors υ_{in}). Hence, the problem is akin to the previous cases, depending on whether the correlation has been induced at the level of the possible outcomes or of the observations. Nevertheless, the nature of the

---

[3] Although Czajkowski, Budzinski, Campbell, Giergiczny, and Hanley (2017), and successive studies by the same group, introduced a method to capture spatial correlation among parameters to be estimated, which lead to the same mathematical representation, but the interpretation of the model is different (see Ortúzar & Willumsen, 2011 for a good discussion on random error components and random coefficients).

correlation must be taken into account in the definition of the spatial weighting matrix W.

h) Spatial correlation among possible outcomes and observations

Again, to my best knowledge, no study reported in the discrete model literature has considered spatial correlation among possible outcomes and observations simultaneously. While, as previously outlined, such situations indeed occur in reality, their treatment is extraordinarily complex and computationally prohibitive.

From the previously discussed approaches, only two alternatives are suitable to treat this kind of correlation: the Probit specification considering the MCMC or the MACML estimation techniques and the ML specification considering spatially correlated blocks for outcomes and observations. The computational costs of both approaches would be extremely large, although the method of correlation blocks can reduce the number of blocks to any number desired by the modeler at expenses of the precision.

5   Conclusions

The treatment of spatial correlation in multinomial discrete models is not a "sexy" problem: its treatment is complicated, is associated with large computational costs, requires the use of advanced econometric techniques (many of them no easily available from an operational viewpoint), and the results are not particularly appealing given that the focus of the analysis is mostly not set upon demonstrating the existence or quantifying the spatial correlation, but to establish causal relationships in data that happens to be spatially correlated. Therefore, spatial correlation has become a stumbling block for many modelers, which, for the sake of simplicity, simply prefer to ignore the issue. The goal of this paper is certainly not to overcome these issues, which are real and substantial, but to offer a discussion on the advantages and limitations of different approaches that have been proposed to deal with spatial correlation in mutlinomial discrete models.

When the possible outcomes of a discrete model exhibit spatial correlation, the most appealing option, from a computational viewpoint, seems to be relying on the GNL structure, as it is associated with a closed-form probability kernel, which offers substantial computational advantages. Moreover, the Generalized Spatially Correlated Logit (GSCL), which allows for a parametric specification of the correlation structure is highly appealing, given the endogenous estimation of the spatial dependence. The main inconvenient of the GNL is its lack of computational availability and, therefore, the necessity of coding the algorithms. Alternative approaches (Probit, ML), while adding flexibility (the GNL is characterized by a homoscedastic error structure) are computationally costly and can only handle a limited number of possible outcomes. As a compromise solution between complexity, computational costs, and realism, this paper introduces the use of an exogenous block correlated ML, which allows reducing the number of dimensions over which the ML probabilities must be integrated.

In the case of spatial correlation at the level of the observations, no closed-form expression can be derived, and the computation of multidimensional integrals associated with the Probit and ML models are usually costlier than in the previous case, given the large dimensions of the covariance matrix. The only plausible ways to treat spatial correlation among alternatives seems to be to rely on advanced numerical integration techniques (requiring a substantial coding effort) such as MCMC or MACML, or to (over)simplify the correlation structure using a block correlated ML structure.

Correlation at the level of explanatory variables does not have modeling implications, unless they are subject to measurement/imputation errors. In such case, it would result in either correlation at the level of the outcomes, at the level of the observations or both. Finally, if spatial correlation affects both outcomes and observations, the options are the same but the computational costs are even larger.

Further research is required to assess the different approaches discussed in this work with regard to the efficiency of the estimators as well as their resilience to misspecification. Along these lines, it is necessary to define measures of spatial correlation specific to the particularities of multinomial discrete models.

References

Bhat, C. (2015). A new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation. *Transportation, 42*(5), 879–914.

Bhat, C. R., & 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.

Bierlaire, M. (2001). A general formulation of the cross-nested logit model. 1st Swiss Transport Research Conference, Ascona, Switzerland, March 1-3.

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

Cardell, N. S., & Dunbar, F. C. (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research Part A: General, 14*(5–6), 423–434.

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

Chu, C. A. (1989). Paired combinatorial logit model for travel demand analysis. *Proceedings of the Fifth World Conference on Transportation Research, 4,* 295–309.

Cressie, N. (1993). Statistics for spatial data (Fourth ed.). Hoboken, NJ: John Wiley and Sons.

Czajkowski, M., Budziński, W., Campbell, D., Giergiczny, M., & Hanley, N. (2017). Spatial heterogeneity of willingness to pay for forest management. *Environmental and Resource Economics, 68*(3), 705–727.

Daly, A., & Bierlaire, M. (2003). *A general and operational representation of GEV models* (Technical report RO-030502). Lausanne, Switzerland: Institute of Mathematics, Operations Research Group ROSO, EPFL.

Daly, A. J., & Zachhary S. (1978). Improved multiple choice models. In D. A. Hensher, & M. Q Dalvi (Eds.), *Determinants of travel choice* (pp. 335–357). Westmead, UK: Saxon House.

Domencich, T., & McFadden, D. (1975). *Urban travel demand–A behavioral analysis.* Amsterdam: North-Holland Publishing.

Garrido, R. A., & Mahmassani, H. S. (2000). Forecasting freight transportation demand with the space–time multinomial probit model. *Transportation Research Part B: Methodological, 34*(5), 403–418.

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

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlquist, L. Lundquist, F. Snickars, & J. Weibull. (Eds.) *Spatial interaction theory and residential location* (pp. 75–96). Amsterdam: North-Holland Publishing.

Ortúzar, J. de D., & Willumsen, L. G (2011). *Modelling transport* (Fourth ed.). Chichester, UK: John Wiley and Sons.

Parady, G. T., & Hato, E. (2016). Accounting for spatial correlation in tsunami evacuation destination choice: A case study of the Great East Japan Earthquake. *Natural Hazards, 84*(2), 797–807.

Schnier, K. E., & Felthoven, R. G. (2011). Accounting for spatial heterogeneity and autocorrelation in spatial discrete choice models: Implications for behavioral predictions. *Land Economics, 87*(3), 382–402.

Sener, I. N., Pendyala, R. M., & Bhat, C. R. (2011). Accommodating spatial correlation across choice alternatives in discrete choice models: An application to modeling residential location choice behavior. *Journal of Transport Geography, 19*(2), 294–303.

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

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(1), 234–240.

Train, K. E. (2009). *Discrete choice methods with simulation* (Second ed.) Cambridge, UK: University Press.

Ward, M. D., & Gleditsch, K. S. (2002). Location, location, location: An MCMC approach to modeling the spatial context of war and peace. *Political Analysis, 10*(3), 244–260.

Wen, C. H., & Koppelman, F. S. (2001). The generalized nested logit model. *Transportation Research Part B: Methodological, 35*(7), 627–641.

Weiss, A., & Habib, K. N. (2017). Examining the difference between park and ride and kiss and ride station choices using a spatially weighted error correlation (SWEC) discrete choice model. *Journal of Transport Geography, 59,* 111–119.

Weiss, A., Hasnine, S., & Habib, K. N. (2019). *A comparative study of alternative methods for capturing spatial correlations in discrete choice models through an empirical application on school choice location modelling.* Paper presented at the 98th Annual Meeting of the Transportation Research Board, Washington, DC, January 13–17.

Williams, H. C. W. L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A, 9*, 285–344.

Zhou, Y., Wang, X., & Holguín-Veras, J. (2016). Discrete choice with spatial correlation: A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM). *Transportation Research Part B: Methodological, 94*, 440–455.