Transportation Research Part B 94 (2016) 440–455

Contents lists available at ScienceDirect

Transportation Research Part B

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

Discrete choice with spatial correlation: A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM)

Yiwei Zhou[a], Xiaokun Wang[b,*], José Holguín-Veras[c]

[a] Senior Traffic Engineer, Precision Systems, Inc., 80 M Street, SE, Washington, DC 20003, USA
[b] Department of Civil and Environmental Engineering, 4032 JEC Building, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590, USA
[c] Director of the Center for Infrastructure, Transportation, and the Environment, Department of Civil and Environmental Engineering, 4032 JEC Building, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590, USA

ARTICLE INFO

Article history:
Received 24 February 2016
Revised 17 October 2016
Accepted 18 October 2016
Available online 28 October 2016

Keywords:
Bayesian MCMC
Probit model
Endogenous weight matrix
Choice behavior

ABSTRACT

Discrete choice modeling is widely applied in transportation studies. However, the need to consider correlation between observations creates a challenge. In spatial econometrics, a spatial lag term with a pre-defined weight matrix is often used to capture such a correlation. In most previous studies, the weight matrix is assumed to be exogenous. However, this assumption is invalid in many cases, leading to biased and inconsistent parameter estimates. Although some attempts have been made to address the endogenous weight matrix issue, none has focused on discrete choice modeling. This paper fills an existing gap by developing a Spatial Autoregressive Binary Probit Model with Endogenous Weight Matrix (SARBP-EWM). The SARBP-EWM model explicitly considers the endogeneity by using two equations whose error terms are correlated. Markov Chain Monte Carlo (MCMC) method is used to estimate the model. Model validation with simulated data shows that the model parameters can converge to their true values and the endogenous weight matrix can be reliably recovered. The model is then applied to a simplified firm relocation choice problem, assuming that similar size firms influence one another. The model quantifies the peer effect, and takes into consideration other independent variables including industry type and population density. The estimation results suggest that peer influence among firms indeed affect their relocation choices. The application results offer important insights into business location choice and can inform future policy making. The sample size for applying the model is currently limited to hundreds of observations. This paper contributes to the existing literature on discrete choice modeling and spatial econometrics. It provides a new tool to discover spatial correlations that are hidden in a wide range of transportation issues, such as land development, location choice, and various travel behavior. Those hidden spatial correlations are otherwise difficult to identify and estimation results may be biased. Establishing a new model that explicitly considers endogenous weight matrix and applying the model to a real life transportation issue represent a significant contribution to the body of literature.

* Corresponding author.
E-mail addresses: zywadam@gmail.com (Y. Zhou), wangx18@rpi.edu (X. Wang), jhv@rpi.edu (J. Holguín-Veras).

http://dx.doi.org/10.1016/j.trb.2016.10.009
0191-2615/

Introduction

A classic assumption in travel behavior modeling is that observations are independent, which has proven to be unrealistic and restrictive in many transportation study contexts. Decision makers' behavior, ranging from long term, relatively stable activities (e.g., residential location choice Guo and Bhat, 2004), to daily activity patterns (e.g., making shopping trips with friends and family Zhou and Wang, 2014a, 2014b; Wang and Zhou, 2015), are correlated. Such correlations are attributed to explicit socioeconomic interaction with each other (Zhou and Wang, 2014a) or the shared unobserved effects in the data (Guo and Bhat, 2007). Unfortunately, previous studies mainly analyze travel behavior without considering the influence of such correlations in decision making. This prevented the research and practice communities from fully understanding people's activities, leading to biased estimation results, inaccurate interpretation and misleading policy measures (Cao, 2015).

In response to the research needs, some studies sought to address the correlations in the decision making process. Such form of dyadic dependency between agents in close social or spatial proximity is also referred to as the endogenous interaction effect (Elhorst, 2010 and Manski, 1993). There are many valid approaches to capture such endogenous interactions. Several studies used linear-in-means model and assumed peer social interaction effects within exclusive groups (Blume et al., 2011). Lee et al. (2010) used a spatial network autoregressive weight structure for correlations within network, as well as group-specific unobserved effects. At the same time, some researchers start considering spatial correlations in discrete choice models (Brock and Durlauf, 2001, 2006, 2007; Soetevent and Kooreman, 2007; Krauth, 2006; Zhang and Wang, 2016a, 2016b; Zou et al., 2015). Recently, Bhat (2015) accommodated spatial correlation effects, while allowing a global spatial structure on the individual-specific unobserved response sensitivity to exogenous variables. The latter two effects are referred to as spatial drift effects. In most of these studies, the weight matrix represent the spatial correlations. The weight matrix indicates relative weight of social or spatial interactions between agents. Because of the high dimension of correlation caused by spatial interactions, most of these spatial models have difficulty handling large sample sizes.

In general, there has been limited empirical work on the observation correlation, particularly in discrete form (Bhat et al., 2015). Two major challenges exist in addressing observation correlations: First, in order to consider extensive observation correlations, the model structure requires a new theoretical framework, and the model estimation and interpretation could become extremely difficult (Bhat et al., 2014a). Second, it is difficult to measure the level of connection between two observations. Spatial econometrics often relies on Tobler's first law of geography and assumes that close objects are more related than distant objects. Most spatial econometric models then assume an exogenous weight matrix. However, in many transportation problems, the weight matrix entry should not be exogenous, especially when the dependency is caused by social interaction or if it is subject to the influence of many unobserved effects. The misspecification of weight matrix may lead to erroneous estimation results, and possibly misleading policy assessments. Many studies have acknowledged these problems and highlighted the importance of tackling them (Anselin, 2010; Corrado and Fingleton, 2011; Pinkse and Slade, 2010), but due to the complication resulting from the unit dependency, model nonlinearity, and weight matrix endogeneity, few studies have successfully addressed these challenges.

In this paper, we propose to capture the observation interaction effect through the definition of an endogenous weight matrix in a spatial lag structure. The spatial lag structure has been used in many previous studies to analyze observation interdependency (Chakir and Parent, 2009; Bhat, 2011; Sidharthan and Bhat, 2012; Zhang and Wang, 2016a, 2016b; Ni et al., 2016). A typical example is the spatial autoregressive model (SAR), where the dependent variable is function of a linear combination of neighboring observations, as well as exogenous variables and an error term. The weight matrix entries are used to measure the relative level of social and spatial interactions between agents. There are many approaches to define the weight matrix (LeSage and Pace, 2009). It can be defined based on geographic approximation or social connection depending on need (Dugundji, 2013; Leenders, 2002), which makes it applicable to many social and economic issues. For example, in social networking, the weight matrix can be defined as binary peer matrix, which measures peer effect (LeSage and Pace, 2009). In this paper, we go further to allow the weight matrix entries to be endogenous. The weight elements are specified as functions of exogenous variables and a stochastic term. By allowing correlation between the weight elements' stochastic term and the error terms in the decision making models, it relaxes the constraint of an exogenous weight matrix.

Consequently, this paper develops a Spatial Autoregressive Binary Probit Model with Endogenous Weight Matrix (SARBP-EWM). Bayesian Markov Chain Monte Carlo (MCMC) method is used for estimation. The model is developed based on SAR model while allowing the weight matrix to be endogenous. It can be applied to many transportation phenomena with individuals choosing among alternatives in discrete form, including impact of transportation infrastructure on urbanization, social network, travel behavior, residential location choice, etc. This paper also applies the model to a firm relocation issue to investigate the observation interaction and other influential factors such as industry type and population density in firm relocation decision. The application of the SARBP-EWM model can quantify intensity of interaction among agents, and identify influential factors in the decision-making process. It will add to the existing literature by addressing observation correlation in discrete choice models.

Literature review

In a spatial econometric model, weight matrix represents the relative strength of connection between each pair of spatial units. Traditionally, the weight matrix in spatial econometric models is treated as exogenous. This assumption is true in certain circumstances. For example, when spatial weight is defined as the geographic distance between different spatial units,