THE DYNAMIC SPATIAL MULTINOMIAL PROBIT MODEL: ANALYSIS OF LAND USE CHANGE USING PARCEL-LEVEL DATA

Xiaokun (Cara) Wang, Ph.D., P.E.
Assistant Professor
Department of Civil and Environmental Engineering
Rensselaer Polytechnic Institute
4031 JEC Building,
110 8th Street, Troy, NY 12180-3590 wangx18@rpi.edu

Dr. Kara M. Kockelman, P.E., PhD
Professor & William J. Murray Jr. Fellow
Department of Civil, Architectural & Environmental Engineering
The University of Texas at Austin
1 University Station, #C1761 - (Office: 6.9 ECJ) - Austin, TX 78712-0278
kkockelm@mail.utexas.edu

Jason D. Lemp, PhD
Senior Professional
Cambridge Systematics
9015 Mountain Ridge Drive, Suite 210 - Austin, TX 78759
lemp5000@yahoo.com

ABSTRACT

Many transportation-related behaviors involve multinomial discrete response in a temporal and spatial context. These include quality of paved roadway sections over time, evolution of land use at the parcel level, vehicle purchases by socially networked households, and mode choices by individuals residing across adjacent homes or neighborhoods. Such responses depend on various influential factors, and can have temporal and spatial dependence or autocorrelation. In many cases, dynamic spatial-model specifications based on maximum fitness, profit or utility may be most appropriate.

This study develops a Dynamic Spatial Multinomial Probit (DSMNP) model by pivoting off the ordinary MNP model while incorporating spatial and temporal dependencies. The paper first explains how the model reflects the behaviors at play, and then describes estimation using Bayesian methods. Simulated data sets containing both generic and alternative-specific explanatory variables are used to validate the model's performance (and that of its associated code). Estimation efficiency issues and identification issues are discussed. The model is then applied to analyze parcel-level land use changes in Austin, Texas. It is found that better accessibility boosts the potential of residential development, with a statistically significant parameter. In the case of non-residential development, the effect is statistically insignificant. The effects of job and population density, neighborhood income and soil slope are also explored, and found to exert variable effects across space. It is also found that land development tends to cluster when existing development intensity in a neighborhood is low.

KEYWORDS: Multinomial Probit, Spatial Autocorrelation, Land Use Change, Discrete Panel Data

INTRODUCTION

Many transportation-related behaviors involve discrete response in a temporal and spatial context. For example, Potoglou and Kanaroglou (2008) found that car ownership was influenced by not only family structure, but a series of neighborhood characteristics derived from spatial data. Paez (2008) studied land use changes using a probit model with geographical weights. Sener et al. (2011) examined spatial autocorrelation in residential choice sets. In fact, other phenomena such as pavement conditions, trading partners, and mode choices all change over time, exhibiting some degree of temporal and spatial dependence or autocorrelation. The connections between observational units may be measured using network distances, peak-hour travel times, or degrees of social- (or business-) network separation. Of course, variables like slope and soil quality, access to trade routes, shared values, demographics, and various other attributes explain away much of the correlation witnessed over time and space. But significant uncertainty generally lingers in predictive models: unobservable yet influential factors remain. These may be the intensity of rainfall or marketing campaigns, or the pedestrian friendliness of different neighborhoods. The size of such contributions varies, often in a continuous fashion, over space. In contrast to time-series data, the dependencies are two dimensional. There is need to specify and estimate a behavioral-consistent model that can recognize temporal and spatial dependencies in discrete multiple-response data. The model specification and estimation techniques developed here seek to fill a key void in the spatial econometrics, where models of continuous response data and binary data are the norm.

The first spatial model for discrete response data was proposed by McMillen (1995) where he estimated a probit model with a spatial auto-regressive (SAR) process. The work in this field was further extended by Beron and Vijverberg (2004), LeSage (2000), and Kakamu and Wago (2007). In the review of spatial econometric methods for bio-economic and land use modeling, Holloway et al. (2007) mentioned several fresh studies involving spatial models for discrete response data. For example, Brasington and Sarama (2007) used a spatially correlated binary probit model to analyze deed restrictions in home sales, and Holloway and Lapar (2007) used similar techniques to estimate the magnitude of neighborhood effects.

In the latest review by Pinkse and Slade (2010), a wide range of related studies were discussed. The most intriguing ones among them are probably those by Kapoor et al. (2007) and Anselin et al. (2008), where the models are specified to handle spatial panel data. A limited number of studies have attempted to consider the spatial and temporal effects in multinomial discrete response data. For example, Wang and Kockelman's (2006) estimated urban land cover evolution using a mixed multinomial logit framework. However, the authors arbitrarily assumed a specific distance-decay function for inter-observational correlations, rather than using the more flexible SAR process. Wang and Kockelman (2009a, 2009b, 2009c) also developed a dynamic spatial ordered probit (DSOP) model and used it to analyze land development intensities (based on satellite image reflectances) and ozone concentration levels. Their model relies on a more flexible SAR process, but the response data are ordered, rather than unordered and categorical. The spatial multinomial probit (MNP) model proposed by LeSage and Pace (2009) is probably the most closely related to this paper's specification. However, their model is for cross-sectional data, without dynamic features. And it does not recognize correlation among alternatives[1], which can be critical for multinomial discrete response data.

Land use-transport modelers recognize that transport investments shape – to some extent – land use patterns (Knowles, 2006). Wang and Kockelman's DSOP study (2009b) analyzed the impact of distances to highways on grid cells' land development intensities. Travisi et al. (2009) developed a mobility impact index based on commute data and analyzed its impact on urban sprawl. Vandenbulcke et al. (2009)

---

[1] For example, factors that make one land use or vehicle type more attractive are likely to make one or more others more attractive, for the same observational unit (i.e., the parcel or household, in these two examples).

mapped travel access to towns and network nodes across Belgium, using spatially disaggregate data. However, no studies have managed to simultaneously analyze the spatial, temporal and other dependencies that are likely to exist across alternatives and their observational units in a statistically rigorous way. The transportation community and many others have great need for tools that tackle multinomial categorical data, where observational units have both dynamic features and spatial dependencies. This work is designed to address this need.

MODEL SPECIFICATION

This proposed model is based on the basic MNP model, but at the same time incorporates the effect of spatial and temporal dependencies. The observed dependent variable for observation i (1 to N) at time period t (1 to T) y_{it} takes possible values from {1,2, 3,…R+1}, or R+1 alternative responses. U_{ikt} indicates its utility for outcome k. The observed variable y_{it} will take value k if and only if U_{ikt} is the maximum utility among alternatives. Since an observation's preference is always a relative term, one response alternative needs to be chosen as the base. Here, the last (R +1) response alternative is used as the base, so the latent utility differences are as follows:

$$y_{ikt}^* = U_{ikt} - U_{i(R+1)t} \qquad\qquad k = 1,2,\ldots, R \qquad\qquad (1)$$

This latent utility difference is influenced by many factors, so that it takes the following form:

$$y_{ikt}^* = X_{ikt}'\beta + \varepsilon_{ikt} \qquad\qquad\qquad (2)$$

where X_{ikt}' is a $1 \times Q$ vector indicating the differences of explanatory variable values between alternative k and base alternative R+1. X_{ikt}' can be composed of the site-specific information, neighborhood information and observed dependent variable values in previous time periods. Some of these influential factors are invariant across alternatives, such as land use in the prior time period or local income levels, and must be interacted with alternative-specific indicators or other, generic attributes, such as costs and profits for developing certain parcel into different land use types.   All together, Q_1 denotes the total number of attributes or covariates describing each alternative. The location's land use type in the previous time period is not generic and can be accommodated by interaction with an indicator vector $\boldsymbol{\delta_{ikt-1}}$, which is a 1 x $(R+1)$ vector of zeros or ones so that $\delta_{ikt-1,m} = 1$ if the land use type at the previous period is of type m, or $y_{it-1} = m$. (where $m = 1,2, 3,\ldots, R+1$). The inclusion of $\boldsymbol{\delta_{ikt-1}}$ helps to identify how the observed outcome at previous time period affects alternatives' utilities in the subsequent time period. As one can expect, existing land use type will significantly affect the utilities in a subsequent period. If a parcel is currently used as residential, it will more likely favor the further development of commercial or residential, rather than industrial.

Since there are $R+1$ alternatives (hence R latent utility differences relative to the base alternative), if there are also Q_2 alternative-specific variables, the total number o explanatory variables is $Q = \big(Q_1 + (R + 1)\big) \times R + Q_2$. If the explanatory variables are felt to be spatially related, the data can be pre-processed by multiplying by W_X , which is an analyst-defined matrix that spatially relates observed attributes of nearby observations to other observations.

β is a $Q \times 1$ vector containing the corresponding parameters. The error term ε_{ikt} captures all effects not explained by the control variables. In order to accommodate the dynamic and spatial features in the uncontrolled terms, the error is specified as follows:

$$\varepsilon_{ikt} = \lambda_k \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\varepsilon_{jkt} + \theta_{ikt} \qquad\qquad\qquad (3)$$

$$\theta_{ikt} = \rho\theta_{ikt-1} + \eta_{ikt} \tag{4}$$

In other words, the error term is composed of two parts: $\sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ijk}\varepsilon_{jkt}$ captures the spatial dependence of the error term and the remaining uncertainty is reflected in θ_{ikt}. Here, w_{ijk} is the exogenous value indexing the relative locations between observational unit i and observational unit j. It is also set to be alternative-specific, since different alternatives may exhibit different patterns of spatial dependence. λ_k is the correlation parameter for alternative k; and θ_{ikt} is an error term specific to observational unit i at time t for alternative k, and not influenced by its neighbors. Furthermore, θ_{ikt} is temporally dependent on observational unit i's site-specific error term for the same alternative at a previous time period, θ_{ikt-1}. ρ is the temporal autocorrelation coefficient. Finally, η_{ikt} is the remaining uncertainty not explained by spatial or temporal dependencies. It captures dependencies between alternatives, and so is assumed to be iid normal (over space and time periods, but not alternatives), as follows:

$$E(\eta_{ikt}) = 0 \tag{5}$$

$$cov(\eta_{ikt}, \eta_{ilt}) = \sigma_{kl} \tag{6}$$

First, all observations at the same time period and for the same alternative are stacked:

$$\boldsymbol{\varepsilon_{kt}} = \lambda_k W_k \boldsymbol{\varepsilon_{kt}} + \boldsymbol{\theta_{kt}} \tag{7}$$

$$\text{or} \quad \boldsymbol{\varepsilon_{kt}} = H_k{}^{-1}\boldsymbol{\theta_{kt}} \tag{8}$$

where $\boldsymbol{\varepsilon_{kt}} = \begin{bmatrix} \varepsilon_{1kt} \\ \vdots \\ \varepsilon_{Nkt} \end{bmatrix}$ and $\boldsymbol{\theta_{kt}} = \begin{bmatrix} \theta_{1kt} \\ \vdots \\ \theta_{Nkt} \end{bmatrix}$

W_k is the $N \times N$ weight matrix for alternative k, with entry value w_{ijk} on row i and column j, and 0 on its diagonal, and $H_k = I_N - \lambda_k W_k$.

Next, one stacks observations at all time periods for the same alternative, and then stacks all alternatives, leading to a vector of all $N \times T \times R$ errors terms:

$$\boldsymbol{\varepsilon} = H^{-1}\boldsymbol{\theta} \tag{9}$$

where $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_R \end{bmatrix}, \varepsilon_k = \begin{bmatrix} \varepsilon_{k1} \\ \vdots \\ \varepsilon_{kT} \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_R \end{bmatrix}, \theta_k = \begin{bmatrix} \theta_{k1} \\ \vdots \\ \theta_{kT} \end{bmatrix}$

and

$$H = \begin{bmatrix} I_T \otimes H_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & I_T \otimes H_R \end{bmatrix} \tag{10}$$

The above form is quite similar to the spatial SUR model developed by Wang and Kockelman (2007). Essentially, the error term is influenced by spatial dependency (represented by H), temporal autocorrelation (captured by $\boldsymbol{\theta}$), and the cross-alternative correlations (captured by $\boldsymbol{\eta}$).

If the error terms η_{ikt} between all alternatives follow the following multivariate normal distribution:

$$\boldsymbol{\eta_{it}} \sim N(0, B) \tag{11}$$

where B is an $R \times R$ matrix containing elements σ_{kl}, then the covariance matrix of the AR(1) process θ_{ikt} for certain observational unit i is

$$\boldsymbol{\theta_i} \sim N(0, \Sigma) \tag{12}$$

where $\Sigma = \mathrm{B} \otimes \mathrm{P}, \mathrm{P} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^T \\ \rho & 1 & \rho & \cdots & \rho^{T-1} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^T & \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix} \tag{13}$

for all observational units, and

$$\boldsymbol{var(\theta)} = (\Sigma \otimes \mathrm{I}_N). \tag{14}$$

The covariance matrix of $\boldsymbol{\varepsilon}$, Ω, is then given by the following

$$\Omega = \mathrm{H}^{-1}(\Sigma \otimes \mathrm{I}_N)\mathrm{H}^{-1'} = \left[\mathrm{H}'(\Sigma^{-1} \otimes \mathrm{I}_N)\mathrm{H}\right]^{-1} \tag{15}$$

Bayesian Estimation

The joint posterior distribution for all parameters can be written as follows:

$$p(y^*, \beta, \lambda_k, \rho, \mathrm{B}|Y, X) \propto p(Y|y^*)\pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\beta)\pi(\lambda_k)\pi(\rho)\pi(\mathrm{B}) \tag{16}$$

As is standard in Bayesian estimation, the conditional posterior distributions of all parameters can be derived by extracting only items that contain them, as follows:

$$p(\beta| \dots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B}) \, \pi(\beta) \tag{17}$$

$$p(y^*| \dots) \propto p(Y|y^*)\pi(y^*|\beta, \lambda_k, \rho, \mathrm{B}) \tag{18}$$

$$p(\rho| \dots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\rho) \tag{19}$$

$$p(\mathrm{B}| \dots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\mathrm{B}) \tag{20}$$

$$p(\lambda_k| \dots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\lambda_k) \tag{21}$$

In an MCMC framework, the values of parameters can be drawn sequentially from these distributions. The following describes the set of conditional posterior distributions.

Conditional Posterior of β

If it is assumed that β has normal priors, i.e.,

$$\pi(\beta) \sim N(\beta_0, \Theta_\beta) \tag{22}$$

then, the conditional posterior distribution can be written as follows :

$$p(\beta|Y, X, U, \lambda_k, \rho, \mathrm{B}) \propto \exp\left(-\frac{1}{2}\left([U - X\beta]'\Omega^{-1}[U - X\beta] + [\beta - \beta_0]'\Theta_\beta^{-1}[\beta - \beta_0]\right)\right) \tag{23}$$

In other words, the conditional posterior distribution is multivariate normal with mean μ_β and covariance matrix Λ_β, as follows:

$$\mu_\beta = \Lambda_\beta\left(X'\Omega^{-1}y^* + \Theta_\beta{}^{-1}\beta_0\right) \tag{24}$$

$$\Lambda_\beta = \left(X'\Omega^{-1}X + \Theta_\beta{}^{-1}\right)^{-1} \tag{25}$$

Conditional Posterior of y^*

From equation 18, the following can be easily derived:

$$p(y^*|\dots) \propto p(Y|y^*)\pi(y^*|\beta,\lambda_k,\rho,\mathrm{B}) \sim TMVN\{X\beta,\Omega\} \tag{26}$$

The sampling of y^* is truncated so that y^*_{ikt} is the largest component of $\boldsymbol{y}^*_{it}$ if $y_{it} = k$, or all y^*_{ikt} are negative if $y_{it} = R+1$. These conditions can be expressed as drawing from the non-truncated multivariate normal distribution subject to the following constraints:

$$Dy^* > 0 \tag{27}$$

where D is an $NTR \times NTR$ matrix with diagonal elements equal to -1 and off-diagonal elements equal to 0, with the following exception:

$$D_{[N\times T\times(l-1)+(t-1)\times N+i],[N\times T\times(k-1)+(t-1)\times N+i]} = 1, \text{ for all } l = 1,2,\dots R \text{ if } y_{it} = k \tag{28}$$

for any $i = 1,2,\dots N$; $t = 1,2,\dots T$; $k = 1,2,\dots R$.

This constraint ensures that

$$
\begin{aligned}
y^*_{ikt} - y^*_{ilt} &> 0 \qquad &\forall l \quad if \quad y_{it} = k, k = 1,2,\dots R \\
or \; y^*_{ilt} &< 0 \qquad &\forall l \quad if \quad y_{it} = R+1
\end{aligned}
\tag{29}
$$

With this transformation, the y^* can be derived by first drawing $z \sim N(0, D\Omega D')$ subject to linear restrictions $z > -DX\beta$ and $y^* = X\beta + D^{-1}z$. Such a truncated multivariate normal distribution can be sampled based on rejection sampling, as discussed in Geweke et al. (1994) and Laud et al. (2010). The approach used here was proposed by Stefan and Manjunath (2008), based on the method suggested by Horrace (2005) and Jayesh and Petar (1999). Unfortunately, the method does not seem numerically efficient for problems of high dimensionality (as in the case of land use over space and time). As possible, more approaches should be investigated.

Conditional posterior of $\mathbf{B}$

From equation (20), it is clear that $\mathbf{B}$'s conditional of can be written as follows:

$$p(\mathrm{B}|Y,X,y^*,\beta,\lambda_k,\rho) \propto \pi(y^*|\beta,\lambda_k,\rho,\mathrm{B})\pi(\mathrm{B}) \tag{30}$$

It is easier to specify the conjugate prior if one samples B^{-1} rather than $\mathbf{B}$ itself. Here, it is assumed that the prior on B^{-1} (which has dimension R x R) follows a Wishart distribution with prior parameters B_0^{-1} and b_0, where B_0 is the scale matrix (a R x R positive definite matrix and) and $b_0 \geq R - 1$ indicating the degree of freedom. Thus, B^{-1}'s conditional posterior of is given by:

$$p(\mathrm{B}|\dots) \propto |\Omega|^{-1/2}\exp\left(-\frac{1}{2}[y^* - X\beta]'\Omega^{-1}[y^* - X\beta]\right)|\mathrm{B}^{-1}|^{(b_0-R-1)/2}\exp\left(-\frac{1}{2}\mathrm{trace}[\mathrm{B}_0\mathrm{B}^{-1}]\right)$$

$$\propto |\mathrm{B}^{-1}|^{(NT+b_0-R-1)/2}\exp\left(-\frac{1}{2}\mathrm{trace}[(C + \mathrm{B}_0)\mathrm{B}^{-1}]\right) \tag{31}$$

6

As a result, under the conditional posterior distribution, B^{-1} is Wishart distributed with scale matrix $(C + B_0)^{-1}$ and degrees of freedom $NT + b_0$. Here, C is a R x R matrix given by the following:

$$C = \begin{bmatrix} Z_1'(P^{-1} \otimes I_N)Z_1 & \cdots & Z_1'(P^{-1} \otimes I_N)Z_R \\ \vdots & \ddots & \vdots \\ Z_R'(P^{-1} \otimes I_N)Z_1 & \cdots & Z_R'(P^{-1} \otimes I_N)Z_R \end{bmatrix} \tag{32}$$

where $\quad Z = H[y^* - X\beta], Z_k = \begin{bmatrix} Z_{k1} \\ \vdots \\ Z_{kT} \end{bmatrix}, Z_{kt} = \begin{bmatrix} Z_{kt1} \\ \vdots \\ Z_{ktN} \end{bmatrix}$ $\tag{33}$

A key issue that needs to be discussed here is the identification problem. In an MNP (or multinomial logit) model, a scale shift (factoring both parameters and errors) will not change observed responses. In the case of dynamic spatial multivariate probit model, this feature implies that

(1) One of the diagonal element of the covariance matrix B (here σ_{RR}) needs to be set to unity; and
(2) Though theoretically identifiable, in practice the model needs to be identified with a long and large panel (large T and large N).

Many existing studies have discussed this identification issue as well as approaches that can be used to solve this issue (see, e.g., McCulloch et al. [2000], Koop [2003], Imai and Dyk [2005], Nobile [2000], and LeSage and Pace [2009].) Each approach has its strengths and limitations. In this work, Nobile's (2000) proposed method is used.

Conditional Posterior of ρ

From equation (19), one can write ρ's conditional posterior as follows:

$$p(\rho|Y, X, y^*, \beta, \lambda_k, B) \propto \pi(y^*|\beta, \lambda_k, \rho, B)\pi(\rho) \tag{34}$$

Here, a uniform prior is assumed for ρ. To ensure stationarity, ρ is truncated, so that $-1 \leq \rho \leq 1$. In other words, its prior only contributes to the posterior distribution via truncation. The conditional posterior can be written as follows:

$$p(\rho|Y, X, y^*, \beta, \lambda_k, B) \propto |\Omega|^{-1/2} \exp\left(-\frac{1}{2}[y^* - X\beta]'\Omega^{-1}[y^* - X\beta]\right) \tag{35}$$

$$p(\rho|Y, X, y^*, \beta, \lambda_k, B) \propto |P^{-1}|^{NR/2} \exp\left(-\frac{1}{2}Z'([B^{-1} \otimes P^{-1}] \otimes I_N)Z\right) \tag{36}$$

The conditional posterior distribution reduces to the following:

$$p(\rho|\dots) \propto (1 - \rho^2)^{NR/2} \exp\left(-\frac{1}{2}Z'([B^{-1} \otimes (\rho^2 P_1^{-1} - \rho P_2^{-1})] \otimes I_N)Z\right)$$

$$\propto (1 - \rho^2)^{NR/2} \exp\left(-\frac{1}{2}\frac{(\rho - \bar{\rho})^2}{\sigma_\rho^2}\right) \tag{37}$$

where $\bar{\rho} = \frac{1}{2}\sigma_\rho^2 Z'([B^{-1} \otimes P_2^{-1}] \otimes I_N)Z$ $\tag{38}$

$$\sigma_\rho^2 = \left(Z'([B^{-1} \otimes P_1^{-1}] \otimes I_N)Z\right)^{-1} \tag{39}$$

$$P_1^{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{40}$$

$$P_2^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The additional multiplicative component in front of the exponential makes this distribution non-standard. An MH step or numerical integration can be used, and MH is used here.

Conditional Posterior of λ_k

From equation (21), the conditional posterior of ρ can be written as follows:

$$p(\lambda_1, \dots, \lambda_R | Y, X, y^*, \beta, B, \rho) \propto \pi(y^* | \beta, \lambda_k, \rho, B)\pi(\lambda_1, \dots, \lambda_R)$$

A uniform prior is assumed on λ_k, with bounds $\varphi_{1k}, \varphi_{2k} \ \forall \ k$. Like ρ, the prior is flat, and only contributes to the posterior distribution via truncation. The conditional posterior can be written as follows, with $y^* - X\beta = V$:

$$p(\lambda_1, \dots, \lambda_R | \dots) \propto \left(\prod_k |H_k|^T\right)\exp\left(-\frac{1}{2}V'H'\left(\Sigma^{-1} \otimes I_N\right)HV\right) \tag{41}$$

For such a multivariate non-standard distribution, it is expected that the most efficient to use MH sampling, thus generating all λ_k's at once. Sequential sampling based on numerical integration may be another option. Either way, this sampling is expected to be a very computationally intensive process. Potentially better approaches for this should be explored.

MODEL VALIDATION

Data were simulated to examine the performance of the above model and associated code. The simulated data contains 100 observational units, observed over 4 time periods. Each observational unit at each time period can make 4 possible responses ($N=100$, $T=4$, $R=3$). There are 2 generic explanatory variables and 2 alternative-specific explanatory variables. Together with the observed dependent variable values from previous period, the vector of explanatory variable for each observation has totally 20 elements.

The two invariant variables are generated from a gamma distribution with shape and scale parameter 5 and 2. The two alternative-specific variables come from a gamma distribution with parameters 4 and 3. The observed Y values at time zero are generated from a uniform distribution within $[0, R+1]$ and all are rounded up, so that the frequency of initial Y values is evenly distributed among alternatives. (It should be noted that the initial condition problem needs further discussion; here, it is simply assumed that Y_0 is strictly exogenous.)

Weight matrices for all alternatives are set to be the same in this simulation dataset. And the values of $w_{ij} \ \forall i \neq j$ are set to be as follows:

$$\begin{cases} w_{ij} = \frac{1}{d_{ij}} & \text{if } d_{ij} \leq 1 \text{ mile} \\ w_{ij} = 0 & \text{if } d_{ij} > 1 \text{ mile} \end{cases} \tag{42}$$

The distance variable w_{ij} was drawn randomly from a gamma distribution with shape and scale parameters 2 and 0.4, respectively. The temporal autocorrelation coefficient ρ was set to equal 0.5. The spatial autocorrelation coefficient was set to equal 0.25 for the first alternative (λ_1), 0.2 for the second (λ_2) and 0.3 for the third (λ_3), meaning that they all tend to exhibit spatial clustering, but type 3 preference is the most spatially autocorrelated. Parameter values (β) for the two invariant variables are 0.5 and 1.1 for the first alternative; 0.2 and -0.1 for the second alternative; and 1.2 and -0.8 for the third alternative. Parameter values for the alternative-specific variables are -0.6 and 1.4. For lagged Y values, the parameters' values were set to equal 1.1,-0.8,-0.6, and -1.1 for the first alternative; 1.5,-0.5,-0.6, and -1.0 for the second alternative; and 0.8, 1.2, 0.7,and -0.8 for the third alternative.

In a land use example, if alternative 1 is residential, type 2 is industrial, type 3 is commercial and type 4 is institutional, these parameters for lagged Y values can be interpreted as follows: If the previous land use type is industrial in nature (type 2), the parcel tends to be developed into commercial (type 3, with a coefficient 1.2), followed by institutional (type 4, coefficient 0 for base alternative), then industrial (type 2, with a coefficient of -0.5), and finally residential as least favored (type 1, with a coefficient of -0.8). Of course, these coefficients are only part of the contributive factors to the mean (or expected) value of utilities. The total utility values are also greatly influenced by those spatially and temporally autocorrelated error terms.

The three alternatives also are specified to be positively correlated, so that

$$B = \begin{bmatrix} 2.0 & 0.2 & 0.3 \\ 0.2 & 1.8 & 0.6 \\ 0.3 & 0.6 & 1.0 \end{bmatrix}$$

The simulated dataset was estimated using the above model, with 2000 draws. Unfortunately, the current estimation efficiency does not prove satisfactory with such a high-dimension data set: With a 2.66GHz PC, estimation time is approximately 5 seconds per draw. Previous studies in spatial econometrics (e.g., LeSage and Pace, 2009) indicate that run times will increase exponentially with N (the so-called n-cubed problems), yet land use analysis often involves millions of observational units (e.g., the number of parcels in a metropolitan area). Such long run times may render the model infeasible for reasonable sample sizes and computing capabilities. This suggests that N must be limited. On the other hand, N cannot be so small that parameter identification issues emerge (e.g., there is not enough variation in responses and attributes, and/or enough autocorrelation present, for associated parameters to be estimated). As one example of this identification issue, the estimation results of a small-sample experiment (with T=10, R=3 and N=30) with the same coefficients as specified above are shown in Figure 1.

Figure 1.a indicates that the β estimates are not converging, especially coefficients on y_{t-1}. Since $\mathbf{Y}_{t-1}$ is only a vector of indicator variables (with values 1 and 0), inadequate variation may have aggravated the identification issues caused by small sample size. As shown in Figure 1.a, the parameters for $\mathbf{Y}_{t-1}$ do not converge: they are large yet seem to cancel one another. The large βs also exaggerated the magnitude of error terms, clouding ρ (as shown in Figure 1.b) and rendering λ_k estimates negative (Figure 1.c). As can be expected, because the coefficient estimates do not converge, predicted responses also vary significantly from actual y values (Figures 1.e and 1.f).

Given such identification problems, the sample size will need to be kept reasonably large, while bearing the estimator efficiency problem in mind. One potential solution can be used: For the large-sample problem, instead of considering the spatial dependence across all observations, relatively localized dependence can be considered, which leads to a method resembling a combination of Geographically Weighted Regression (GWR) and SAR: One can select $M < 100$ roughly equally spaced points and a small sample of N_m (approximately 100) neighbors around these points to construct M different models, with each model returning a set of estimates. The magnitude and spatial variation of such estimates

should provide interesting insights about the relationship over the entire sample space. While this model's estimation computationally intensive, its specification is quite flexible in that it explicitly incorporates both spatial and temporal autocorrelation. It is also straightforward to implement this combination approach of GWR with SAR, as shown in the following section.

LAND USE DATA DESCRIPTION

This study's analysis area covers all privately held parcels in Travis County, Texas. Eight years of data (2000 to 2007) are available, each containing 320,550 year-2007 parcels. (Over the eight-year period, many parcels subdivided, rendering almost all 2007 parcels the smallest observational units in the panel.)

The Travis County Appraisal District shows 96 different improvement types, and these were merged into 3 categories of land use (*LU*), with 1 indicating residential, 2 for all other non-residential development types, and 3 for undeveloped (and used as the base category). While just three categories may not offer many interesting insights into the land use evolution process, the combined categories help to ensure enough observations in each group for practical model identification. (In the study area, the great majority of parcels are residential [65 percent], with "other" developed land use types accounting for less than 5% of the parcels).

Influential factors at both local and regional levels are considered here, as covariates. Local factors include the average slope of each parcel, as well as local job and population densities, and per-capita incomes. Slope values were derived from USGS elevation data at 30 meter resolution. A parcel's slope was calculated as the average value of the 30 meter resolution data points lying within that parcel[2]. Both densities are "net" densities: The employment density (*ED*) is the ratio of total TAZ job count and area of land occupied by commercial, industrial and other employment types in the TAZ[3]. The population density (*PD*) was first computed at the Census tract level, by dividing population count[4] by land area (as opposed to water area). Each TAZ's population density was then calculated as the weighted average of tract-level population densities, with overlapping land areas as the weights. These TAZ-level population densities were finally assigned to all parcels within each TAZ's boundaries. The per capita income (*INC*) for each parcel is derived in the similar way (first at the tract level, and then at the TAZ level). In addition to these local factors, this study also controls for an "Accessibility Index" (*AI*) as an indicator of the parcel's regional access to jobs, with distances being the shortest-path network distances. More details on variables can be found in Wang et al. (2010).

As discussed above, it is computationally infeasible to consider spatial dependence across all parcels. Instead, an approach integrating GWR and SAR notions is used. Centroids of Austin's 48 zip code areas are used here to center the *M* neighborhoods. These often represent neighborhood planning areas (NPAs), which are the smallest administrative unit for land use development in Austin and thus make good sense from a land development-study standpoint. The number of neighboring parcels around these centroids was set to 100. Of course, by selecting the 100 nearest neighbors, sampling rates across zip codes differ: parcel numbers range from 2,828 to 16,164, with parcels in the more densely developed areas of Austin being underrepresented. However, recognizing that land use change often occurs at a higher rate along an urban area's periphery, such a sampling scheme may be more reasonable than uniform-count sampling.

---

[2] If no data point falls in the boundary of a parcel, the parcel will obtain slope value from the closest point.
[3] One limitation here is that the original data set did not offer information on land areas of civic uses (since these are non-taxable). Yet such employment is counted. Thus, the area of land for employment is underestimated in some zones, leading to somewhat higher-than-actual net employment density values.
[4] Population counts are only available for years 2000, 2006 and 2007. The in-between years' values were interpolated based on an exponential growth assumption.

Table 1 provides summary statistics for all 320,550 parcels and the 4,800 sampled parcels. In general, the 4,800 parcels well represent the larger data set. The only exception is the maximum value of employment density, where parcels with extremely high values are not found in the 4,800 sample; however, the sample's average employment densities still remain close to the complete data set's value. The population densities in the selected subsamples also seem to be higher, though the parcels with highest values are not included. Figure 2 also shows the land use pattern for the entire area as well as the selected subsamples in years 2000 and 2007. As shown, over the data set's eight year period, intensive land development has occurred in both Austin's downtown area and the county's more peripheral areas.

LAND USE CHANGE ANALYSIS

Each of the subsamples was analyzed using the DSMNP model specification presented earlier. The weight matrix of each subsample was generated based on Euclidean distances between parcels, and spatial autoccorrelation with each parcel's nearest 50 neighbors were considered. Among these 48 zipcode-based samples, 6 did not converge due to a lack of variation in the data (similar to the singular Hessian problem in maximum likelihood estimation). Two lack enough land-use-type variation types (even with the grouped categories), with category 2 ("other" development types) containing less than 1% of observations. The other four subsamples do not have adequate variation in their explanatory variables: most of the parcels fall within the same TAZ and hence share the same values. Table 2 provides averages of parameter estimates from the remaining 42 subsamples.

On average, higher slope is favored by both residential and non-residential developments, but these effects are statistically insignificant at the 90% confidence level (especially for non-residential development). High employment densities are estimated to encourage further development, but again the effects appear to largely be practically insignificant. Population density is estimated to have a negative effect on land development, with a much higher magnitude on residential types. Better accessibility boosts the potential of residential development, with a statistically significant parameter. In the case of non-residential development, better accessibility is estimated to have a negative effect on average, but the effect is statistically insignificant. The effects of existing land use types (at the parcel itself) are statistically insignificant too on average, as is the temporal autocorrelation coefficient ρ, indicating weak temporal dependence in these subsamples. The spatial dependences as indicated by λs are positive for both residential and non-residential development, and are statistically significant. The average spatial autocorrelation for residential development is higher than that for non-residential (0.192 versus 0.039), suggesting a stronger tendency of clustering development for residential units. Apparently, 100 points seems not to be enough to get statistical significance in the 42-case GWR contexts.

Estimator variation across the GWR runs can also be illustrated, and some interesting trends observed. The coefficient estimates vary significantly across these subsamples. Depending on locations, the same factor's effect can be negative or positive, highly statistically significant or insignificant. For example, Figure 3 shows accessibility's effect on residential development across the county. Better accessibility most favors residential development in the remote areas and the heavily populated areas surrounding downtown region (but apparently not the downtown region itself). These two types of locations value access more than an access-rich location, such as downtown Austin, where lots of roads and business already exist. However, the accessibility index shows either statistically insignificant or negative effects around the Lake Travis, West Lake Hills, Shady Hollow and McKinney Falls State Park regions, where most of the residential developments are high-end and residents probably more privacy and tend to devalue such access (due to noise, view impacts, and/or other concerns). For non-residential development, accessibility has positive and statistically significant effects in most regions, with exception of the County's east edge, where land is largely zoned for agricultural and is considered "undeveloped" in this dataset. The locations with the most positive effects again appear to be the more remote areas and the areas *surrounding* Austin's downtown. As discussed earlier, owners of properties in these neighborhoods tend to value relative changes in their access the most.

Slope's coefficient estimates also exhibit some interesting pattern. For residential development, it is mostly negative or statistically insignificant within the City of Austin's boundary. Higher slopes tend to require extra building costs and thereby deter development. However, in more remote areas, slope appears favored for residential use, perhaps because it provides more interesting topographies and views. Non-residential development in general does not appear to value added slope; industrial and commercial users may well prefer flat terrain for easy access (especially for heavy freight vehicles), easy site design and less-expensive drainage structures. This can be confirmed by the largely negative values of slope's effects within the City boundary. Outside of the City, where non-residential development tends to be more recreational use in nature (such as the area north of Lake Travis), the effect of slope is estimated to be highly positive.

Though the average effects of employment density are positive for both types of development, their estimated effects vary significantly over the region. Sites that benefit residential development most are those where new business is being developed and residential development tends to fall behind, such as the Round Rock region and south Austin along Interstate 35. For non-residential development, increases in employment density seem to positively affect the downtown most, followed by other places City neighborhoods with mixed land use patterns (such as the Barton Hills neighborhood).

Two types of locations seem to benefit most from high population density: regions currently with low population densities and no special attractions (such as the Bee Cave area), and places that already have very high population densities (north of the Colorado River and along the Loop 1/MoPac freeway corridor). The former sites are in the starting stage of residential development, and thus presumably more sensitive to population density. The latter is designed to be heavily residential and favors higher population density. In other location, high population density is estimated to suppress residential development. The effect of population density on non-residential development also seems intriguing: the effects are negative in the city center, then positive in the outside ring, and becoming negative or insignificant again along the County's borders. It seems that when the existing population density is very high or very low, increasing population density tends to make non-residential development less attractive to owners/developers. In the case where population density is already quite high, higher population density probably implies even higher land acquisition (and presumably building) costs and possibly greater resistance of mixed development from local residents, making non-residential development more difficult. In the peripheral areas, where existing population density is extremely low, residents may well prefer the low-density nature of their environment, and will resist industrial or commercial development. In such cases, the increase of population density can thus imply a lower chance of non-residential development.

The impacts of per-capita incomes on both types of land development are similar. In most parcels, high income indicates higher chance of development over the 8-year period. The exceptions are places that are zoned for state park and agricultural land (and coded as undeveloped), where income has either insignificant or negative effects.

The spatial autocorrelation parameter is positive for residential land development in most subsamples, but areas that are already heavily residential tend to have a negative coefficient for this important parameter, indicating resistance potentially caused by too much competition and a stark lack of land use mixing (an attribute presumably required, at some minimum level, reasonably functioning neighborhoods). A similar trend can be observed for non-residential development: Most areas present clustering effects, while a few areas, including the downtown, appear to oppose such development. These patterns may portray an interesting phenomenon in regional development, whereby similar land use types tend to cluster until near-saturation, at which point developers (and/or their neighbors) start to seek dissimilar (or mixed)

12

development. The temporal autocorrelation parameter is consistently insignificant across all subsamples, suggesting the negligible temporal dependence.

Of course, as discussed above, the lack of variation in both dependent variables (y) and explanatory variables may cause identification challenges. Some of these subsamples may still suffer from such identification issues, and results may not be derived from converged estimates. Though with 2000 draws they exhibit converging trends, it is possible that with more draws the estimates will eventually diverge. Recognizing that some of the above findings seem counter-intuitive, model performance with small sample sizes may be questionable.

CONCLUSIONS

This study proposes a dynamic spatial multinomial probit (DSMNP) model that simultaneously captures observations' dependences across space, over time and between alternatives. The model was estimated within a Bayesian framework, and validated using a simulated dataset with 100 data points, 3 levels, and 4 time periods. To ensure computational feasibility, it appears that small samples are needed for such a spatially complex model of discrete response over time. On the other hand, the sample size needs to large enough to ensure identifiability. One potential solution is a combination of GWR and SAR, which is capable of accommodating both spatial variation and spatial autocorrelation. This form of model was applied to explore land use change patterns in Austin, Texas. 48 subsamples were generated based on Travis County's zip code areas, each containing 100 parcels, and estimation results of 42 are reported here. In addition to coefficient estimates for control variables, the model also provides estimates of the temporal and spatial autocorrelation coefficients for each zip code area. The distributions of these estimates further illustrate the spatial variation of land use change processes and suggest a need for behavioral flexibility in model specifications.

While the parameters in the land use change example appeared to converge over the Bayesian MCMC sampling process, convergence is not guaranteed. They, too, may suffer from convergence issues, and results could be misleading. In addition to the dilemma between efficiency and identifiability, observations in these subsamples also share similar features (since several attributes are derived at the TAZ level, for example), thus reducing variation in input information. The dependent variable, land use type, is also highly unbalanced: many (about 85% of) parcels remain "residential" (or undeveloped) over the 8 year panel period, which does not offer much useful information to the model estimation process. Strategically determining the alternative set (e.g., changing the total number of alternatives) across space may be useful in future investigations. A regional-level spatial autocorrelation parameter (across subsamples) should also be considered in future work (perhaps by specifying a layered spatial autocorrelation structure). In addition, the current model setup assumes that the data generation process for the first time period (determining Year 2000 land use types) is independent of later periods – and thus exogenous, which generally is unrealistic and can cause an initial-conditions problem. Ideally, this problem should also be remedied, by treating the first time period as endogenous (as proposed by LeSage and Parent [2010]).

Other, more specific improvements that may enhance the routine's overall estimation efficiency can also be investigated. These include: (1) alternative code for sampling the latent $y*$ values should be tested (e.g., matrix D can be broken down for every observation unit at a certain time period, reducing the dimension to $R \times R$), (2) an adaptive MH process to draw the parameter ρ (its logconcave conditional distribution ensures the feasibility of an adaptive rejection approach), and (3) a parallel multivariate slice sampler should be investigated for draws of λ_k.

In summary, the model specification proposed in this study offers a new means for more appropriately reflecting a variety of discrete response behaviors with spatial and temporal autocorrelation. However, estimation-process efficiency and identification issues create a central dilemma for such a complex

model. To accommodate larger sample sizes, significant work and innovative problem-solving are needed. Continuing work in this area should be proved challenging yet exciting.
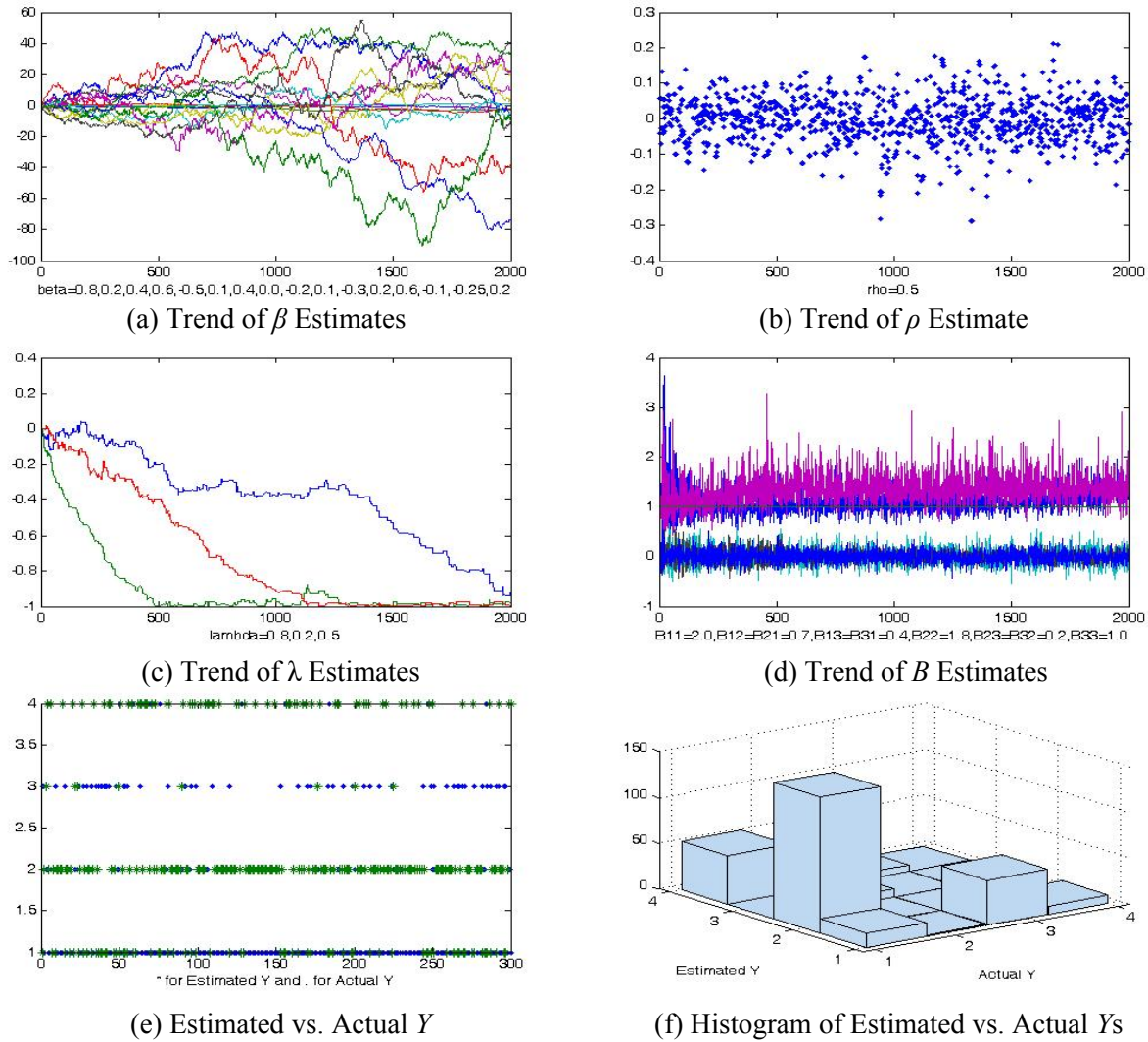
ACKNOWLEDGEMENTS

REFERENCES

Anselin, L. (2001) "Issues in spatial probit models." Workshop on Qualitative Dependent Variable Estimation and Spatial Effects, College of ACES, University of Illinois, April 20, 2001.

Anselin, L., Gallo, J.L. and Jayet, H. (2008) *Spatial Panel Econometrics*. New York: Springer.

Beron, K. and Vijverberg, W. (2004) "Probit in a spatial context: a Monte Carlo analysis". In Anselin, L. Florax, R. and Rey, S. (Eds.), *Advances in Spatial Econometrics*. Heidelberg: Sringer-Verlag.

Brasington, D. and Sarama, R. F. (2007) *Deed Types, House Prices, and Mortgage Interest Rates*. Working Paper. Department of Economics, Louisiana State University, 2007.

Geweke, J., Keane, M. and Runkle, D. (1994) "Alternative computational approaches to inference in the multinomial probit model." The Review of Economics and Statistics 76: 609-632.

Holloway, G. J. and Lapar, L. (2007)"How big is your neighbourhood? Spatial implications of market participation among Phillipino smallholders." *Journal of Agricultural Economics* 58 (1): 37-60.

Holloway,G. Lacombe,D. and LeSage, J. P.(2007) "Spatial Econometric Issues for Bio-Economic and Land-Use Modeling." *Journal of Agricultural Economics* 58 (3): 549-588.

Horrace, W. (2005). "Some results on the multivariate truncated normal distribution." *Journal of Multivariate Analysis* 94: 209–221.

Imai, K. and van Dyk, D. A. (2005) "A Bayesian analysis of the multinomial probit model using marginal augmentation." *Journal of Econometrics* 124: 311-334.

Jayesh H. Kotecha and Petar M. Djuric (1999). "Gibbs sampling approach for generation of truncated multivariate Gaussian random variables." *IEEE Computer Society*: 1757–1760 .

Kakamu, K. and Wago, H. (2007) "Bayesian spatial panel probit model with an application to business cycle in Japan." Working paper. Accessed May 10, 2007: http://www.mssanz.org.au/modsim05/proceedings/papers/kakamu_2.pdf

Kapoor, M., H. Kelejian, and Prucha, I. (2007) "Panel Data Models with Spatially Correlated Error Components." *Journal of Econometrics* 140(1): 97–130.

Knowles, R.D. (2006) "Transport shaping space: differential collapse in time-space." *Journal of Transport Geography* 14(6): 407-425.

Koop, G. (2003). *Bayesian Econometrics*, West Sussex, England: John Wiley & Sons.

Laud, P.W., Damien, P. and Shively, T.S. (2010) *Sampling truncated distributions via rejection algorithm*. Working paper.

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

LeSage, J. P., Fischer, M. and Scherngell, T. (2007) "Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects." *Papers in Regional Science* 86 (3): 393-421.

LeSage, J. and Pace, R. K. (2009). *An Introduction to Spatial Econometrics*, Boca Raton, FL: CRC Press.

LeSage, J. and Parent, O. (2010). "A space-time filter for panel data models containing random effects." Working paper.

McCulloch, R.E., Polson, N.G. and Rossi, P.E. (2000). "A Bayesian analysis of the multinomial probit model with fully identified parameters," *Journal of Econometrics* 99: 173-193.

McMillen, D. P. (1995) "Spatial effects in probit models: A Monte Carlo investigation." In Anselin, L. and Florax, R. (Eds.), *New Directions in Spatial Econometrics*. Heidelberg: Springer-Verlag.

Nelson, G. C., Harris, V. and Stone, S. W. (1999) "Spatial Econometric Analysis and Project Evaluation: Modeling Land Use Change in the Darién." Accessed May 26, 2010: https://netfiles.uiuc.edu/gnelson/www/papers/paperindex.htm.

Nobile, A. (2000). "Comment: Bayesian multinomial probit models with a normalization constraint." *Journal of Econometrics* 99: 335-345.

Paez, A. (2006) "Exploring contextual variations in land use and transport analysis using a Probit model with geographical weights." *Journal of Transport Geography* 14(3): 167-176.

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

Potoglou, D. and Kanaroglou, P.S. (2008) "Modelling car ownership in urban areas: a case study of Hamilton, Canada." *Journal of Transport Geography* 16(1): 42-54.

Sener, I., Pendyala, R., and C. Bhat (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): 293-303.

Stefan, W. and Manjunath, B.G. (2010) *Package 'tmvtnorm'* . User Handbook. Accessed May 30, 2010: http://cran.r-project.org/web/packages/tmvtnorm/tmvtnorm.pdf

Sutradhar, B.C. and Kumar, P. (2003) "The Inversion of Correlation Matrix for MA(1) Process." *Applied Mathematics Letters* 16: 317-321.

Vandenbulcke, G., Steenberghen, T., and Thomas, I. (2008) "Mapping accessibility in Belgium: a tool for land-use and transport planning?" *Journal of Transport Geography* 17(1): 39-53.

Wang, X. and Kockelman, K. (2006) "Tracking land cover change in a mixed logit model: recognizing temporal and spatial effects." *Transportation Research Record* 1977: 112-120.

Wang, X. and Kockelman, K. (2007) "Specification and estimation of a spatially and temporally autocorrelated seemingly unrelated regression model: application to crash rates in China." *Transportation* 34(3):281-300.

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

Wang, X. and Kockelman, K. (2009c) "Application of the Dynamic Spatial Ordered Probit Model: Patterns of Ozone Concentration in Austin, Texas." *Transportation Research Record* No. 2132: 13-24.

Wang, Y., Kockelman, K. and Wang, X. (2010) "Anticipating land use change using Geographically Weighted Regression models for discrete response." Under review for publication in *Transportation Research Record.*

(a) Trend of β Estimates

(b) Trend of ρ Estimate

(c) Trend of λ Estimates

(d) Trend of B Estimates

(e) Estimated vs. Actual Y

(f) Histogram of Estimated vs. Actual Ys

Figure 1. Convergence Issues with Small-Sample Estimation (T=10, R=3 and N=30)

So sorry to see such poor prediction in (f)! we should see tall bars down the middle diagonal; no? of course, all the trend lines of estimators are wandering around, and demonstrating an obvious lack of convergence, so the poor fit example is very consistent. How can we read anything into the estimates if we don't believe we're anywhere near a converged/stable state? KARA: THIS IS AN EXTREME CASE OF UNCONVERGED SAMPLE. IT IS NOT THIS BAD FOR MOST LU APPLICATION SUBSAMPLES, AT LEAST IN THE SUBSAMPLES THAT I RANDOMLY CHECKED, THE TREND LINES SEEM TO BE CONVERGING. BUT AGAIN, IT COULD DIVERGE LATER…

Table 1. Summary Statistics for Entire Dataset and Subsamples

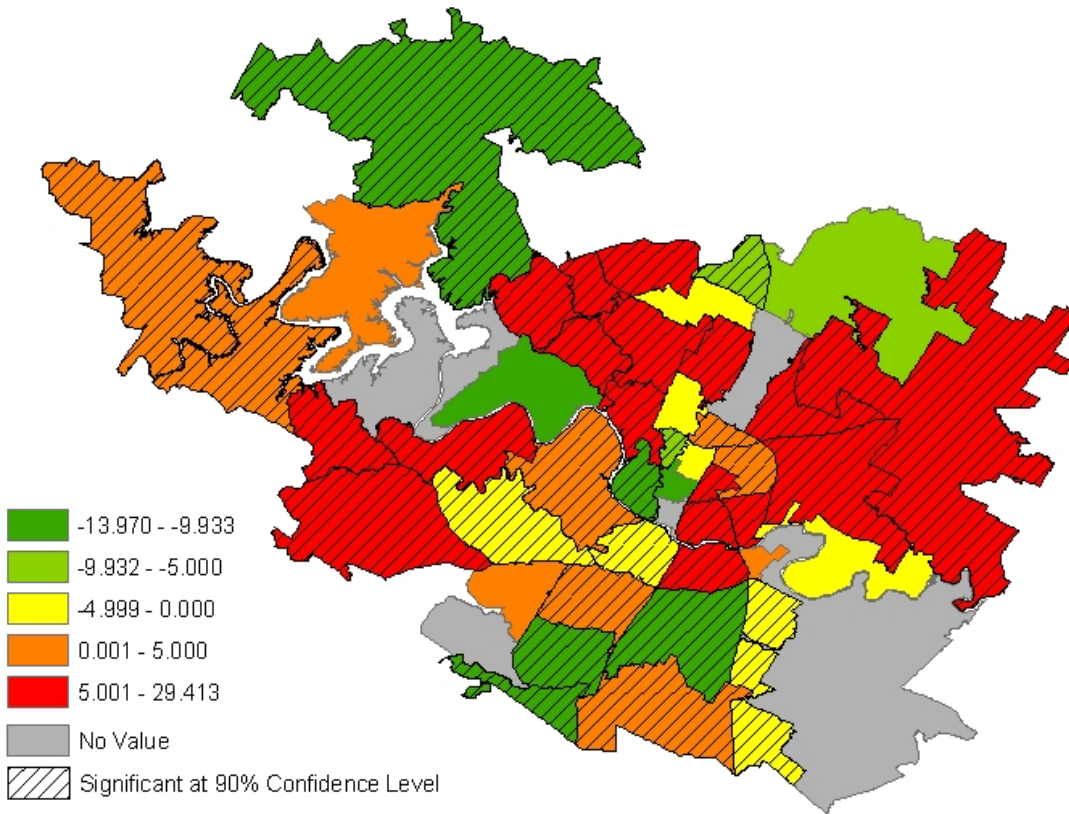|	Year	Entire Dataset				Subsamples			
N_{obs}		320550				4800			
	Year	Min	Max	Mean	SD	Min	Max	Mean	SD
LU (1= residential, 2= non-residential, 3=as undeveloped)	2000	1	3	1.790	0.939	1	3	1.775	0.935
	2001	1	3	1.720	0.933	1	3	1.691	0.926
	2002	1	3	1.689	0.925	1	3	1.680	0.924
	2003	1	3	1.655	0.914	1	3	1.623	0.904
	2004	1	3	1.634	0.903	1	3	1.628	0.898
	2005	1	3	1.609	0.887	1	3	1.609	0.884
	2006	1	3	1.581	0.866	1	3	1.589	0.865
	2007	1	3	1.543	0.839	1	3	1.552	0.845
Slope (%)	---	0	40.6	2.97	3.01	0	24.3	2.82	2.91
ED (jobs per acre in TAZ)	2001	0	9550	30.50	99.90	0	392.7	32.08	53.09
	2002	0	8146	29.20	87.60	0	386.0	30.77	49.79
	2003	0	6743	27.90	75.90	0	393.8	29.46	47.07
	2004	0	5340	26.60	65.10	0	401.6	28.15	45.06
	2005	0	3936	31.60	63.20	0	409.0	30.34	49.02
	2006	0	7220	31.50	83.40	0	459.4	33.81	57.10
	2007	0	11999	44.00	134.30	0	651.0	45.90	83.68
PD (persons per acre in TAZ)	2001	0	150.1	4.200	8.800	0	116.0	5.057	8.694
	2002	0	150.0	4.300	8.900	0	115.9	5.139	8.811
	2003	0	149.9	4.300	8.900	0	115.8	5.223	8.932
	2004	0	149.8	4.400	9.000	0	115.7	5.310	9.058
	2005	0	149.7	4.400	9.100	0	115.7	5.400	9.189
	2006	0	149.6	4.500	9.200	0	115.6	5.493	9.324
	2007	0	150.9	4.600	9.400	0	116.6	5.652	9.588
INC (average annual income per household per year in TAZ)	2001	0	91925	28711	14878	0	91925	26795	13532
	2002	0	93239	29355	15203	0	93239	27382	13829
	2003	0	94572	30015	15538	0	94572	27983	14135
	2004	0	95924	30691	15885	0	95924	28599	14453
	2005	0	97296	31384	16243	0	97296	29230	14781
	2006	0	98687	32095	16613	0	98687	29877	15120
	2007	0	105955	34193	17682	0	105955	31828	16144
AI (Accessibility Index)	2001	0	6.629	5.278	1.367	0	6.627	5.225	1.424
	2002	0	6.647	5.301	1.367	0	6.645	5.248	1.423
	2003	0	6.670	5.327	1.367	0	6.668	5.274	1.424
	2004	0	6.698	5.357	1.370	0	6.697	5.304	1.427
	2005	0	6.729	5.395	1.368	0	6.728	5.341	1.425
	2006	0	6.755	5.434	1.357	0	6.754	5.380	1.417
	2007	0	6.773	5.456	1.355	0	6.772	5.402	1.416

(a) Land Use in 2000   (b) Land Use in 2007

(c) Land Use of Subsamples in 2000   (d) Land Use of Subsamples in 2007

Legend

● Undeveloped

● Residential

● Other Developed
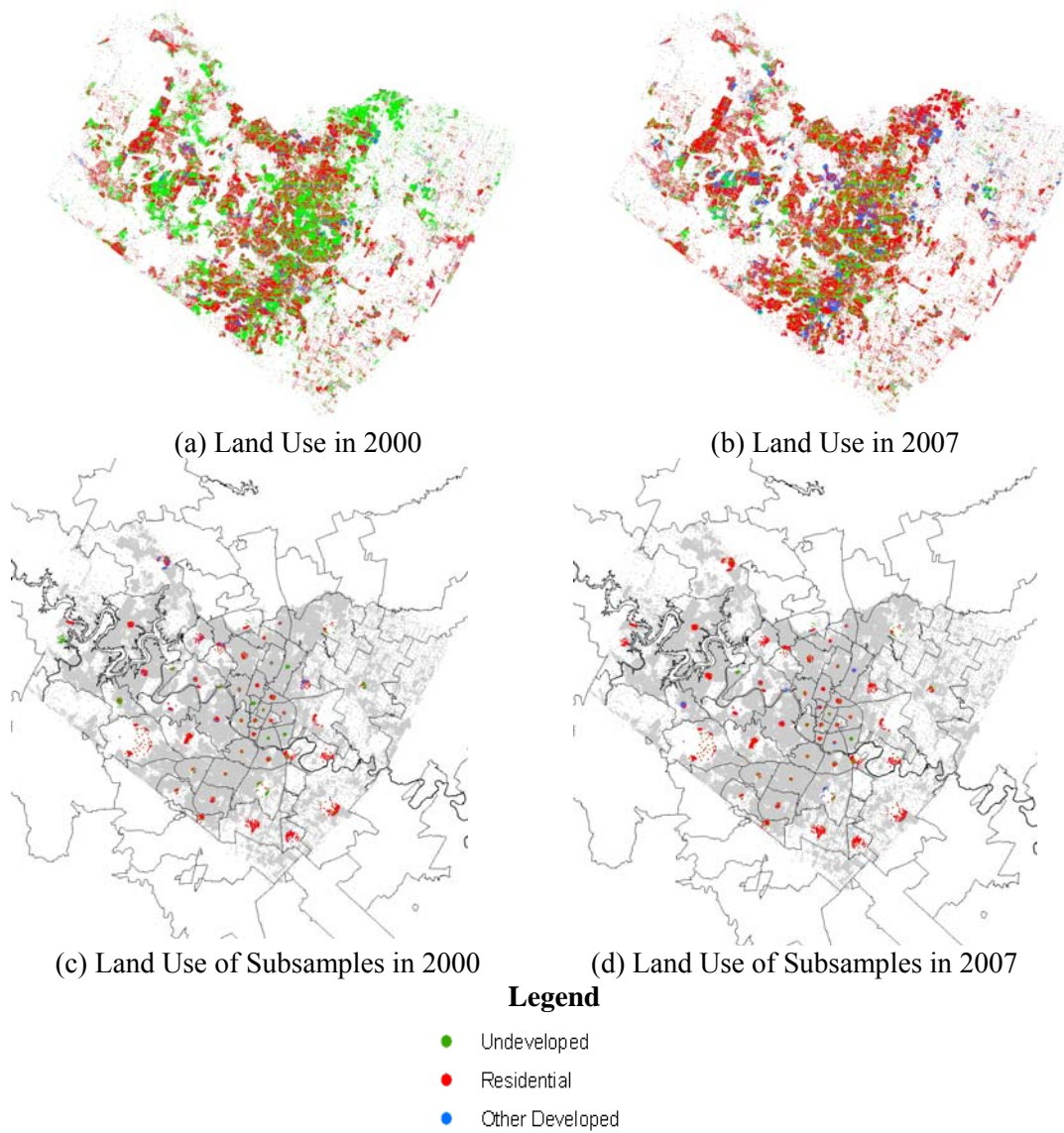
Figure 2. Land Use Patterns for Entire Area and Selected Subsamples

Table 2. Summary Statistics of Subsamples' Estimation Results

	Coefficient (β)			Significance Level (p-value)		
	Mean	Min	Max	Mean	Min	Max
Coefficients for Residential Development						
Constant	0.3249	-4.852	5.942	0.6445	3.593E-07	0.9975
Slope	0.6621	-26.89	27.93	0.1083	1.606E-12	0.6044
ED	3.300E-02	-4.714	8.344	9.626E-02	0.0000	0.8552
PD	-0.7319	-18.71	22.92	0.2440	0.0000	0.9983
INC	-9.806E-04	-8.040E-03	6.565E-03	0.1042	0.0000	0.8853
AI	3.319	-13.97	29.41	9.713E-02	0.0000	0.9901
$LU_{t-1}=1$ (Previously Residential)	-1.176	-17.57	15.83	0.1820	1.332E-15	0.9402
$LU_{t-1}=2$ (Previously Non-residential)	0.2648	-11.52	11.98	0.3563	2.908E-12	0.9494
$LU_{t-1}=3$ (Previously Undeveloped)	1.213	-16.08	14.54	0.3039	2.658E-13	0.9947
Coefficients for Other Types of Development						
Constant	0.5096	-2.324	18.51	0.7486	0.0000	0.9996
Slope	1.510	-17.67	24.14	0.2340	4.441E-16	0.9688
ED	4.304E-02	-5.084	2.965	8.121E-02	7.310E-10	0.8294
PD	-3.643E-02	-21.14	26.72	0.1985	0.0000	0.9985
INC	3.042E-04	-6.387E-03	8.807E-03	0.1135	0.0000	0.9349
AI	-0.723	-52.84	23.40	0.1547	0.0000	0.8860
$LU_{t-1}=1$ (Previously Residential)	0.293	-16.55	14.13	0.3217	0.0000	0.9856
$LU_{t-1}=2$ (Previously Non-residential)	0.637	-17.80	22.85	0.3518	0.0000	0.9772
$LU_{t-1}=3$ (Previously Undeveloped)	-0.395	-17.67	14.06	0.4416	4.441E-16	0.9936
Parameters of Error Terms						
λ for Residential (Spatial Autocorrelation)	0.1918	-1.000	0.999	1.226E-03	0.0000	0.0515
λ for Non-residential (Spatial Autocorrelation)	3.909E-02	-1.000	0.999	0.0000	0.0000	0.0000
ρ (Temporal Autocorrelation)	-4.654E-04	-3.713E-02	2.330E-02	0.8353	2.695E-01	0.9982
B11 (Variance for Residential)	1.051	0.951	1.113	0.0000	0.0000	0.0000
B12 (Covariance of Alternatives)	1.605E-03	-3.129E-02	2.852E-02	0.8892	4.310E-01	0.9992
B22 (Variance for Non-residential)	1.000	1.000	1.000	---	---	---

Figure 3. Spatial Distribution of β_{AI} for Residential Land Development

Legend:
- $-13.970 - -9.933$
- $-9.932 - -5.000$
- $-4.999 - 0.000$
- $0.001 - 5.000$
- $5.001 - 29.413$
- No Value
- Significant at 90% Confidence Level