

A SPATIAL AUTOREGRESSIVE MULTINOMIAL PROBIT MODEL FOR ANTICIPATING LAND USE CHANGE IN AUSTIN, TEXAS

Yiyi Wang

The University of Texas at Austin
yiyiwang@mail.utexas.edu

Kara M. Kockelman

(Corresponding author)

Professor and William J. Murray Jr. Fellow
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
kkockelm@mail.utexas.edu
Phone: 512-471-0210

Paul Damien

B.M. (Mack) Rankin, Jr., Professor of Business Administration
Department of Information, Risk and Operations Management
The University of Texas at Austin

The following is a preprint, the final version can be found
in the *Annals of Regional Science*, Volume 52:251-278, 2014.

ABSTRACT

This paper develops an estimation strategy for and then applies a spatial autoregressive multinomial probit (SAR MNP) model to account for both spatial clustering and cross-alternative correlation. Estimation is achieved using Bayesian techniques with Gibbs and the generalized direct sampling (GDS). The model is applied to analyze land development decisions for undeveloped parcels over a 6-year period in Austin, Texas. Results suggest that GDS is a useful method for uncovering parameters whose draws may otherwise fail to converge using standard Metropolis-Hastings algorithms.

Estimation results suggest that residential and commercial/civic development tends to favor more regularly shaped and smaller parcels, which may be related to parcel conversion costs and aesthetics. Longer distances to Austin's central business district increase the likelihood of residential development, while reducing that of commercial/civic and office/industrial uses. Everything else constant, distances to a parcel's nearest minor and major arterial roads are estimated to increase development likelihood of commercial/civic and office/industry uses, perhaps because such development is more common in less densely developed locations (as proxied by fewer arterials). As expected, added soil slope is estimated to be negatively associated with residential development, but positively associated with commercial/civic and office/industry uses, though its effect on commercial/civic uses is not significant (perhaps due to some steeper terrains offering view benefits). Estimates of the cross-alternative correlations suggest that a parcel's residential use "utility" or attractiveness tends to be negatively correlated with that of commercial/civic but positively associated with that of office/industrial uses, while the latter two land uses exhibit some negative correlation. Using an inverse-distance weight matrix for each

parcel's closest 50 neighbors, the spatial autocorrelation coefficient is estimated to be 0.706, indicating a marked spatial clustering pattern for land development in the selected region.

Key words: spatial autoregressive models, multinomial probit, Bayesian estimation, generalized direct sampling, land use change.

INTRODUCTION

The development of land impacts travel choices and traffic patterns, and transportation system investments and travel decisions affect land use change. For example, more compact and mixed land development may propel people to choose non-motorized modes. Lower density development decisions contribute to longer travel distances and more vehicle-miles traveled per capita (Litman 2012, Cervero and Kockelman 1997).

More accurate forecasts of land development improve long-run travel forecasts. One difficulty associated with such forecasts relates to the categorical nature of land use change, for example, industrial, office, residential, and other uses. Random utility theory supports models for these discrete unordered responses (McFadden 1986). The theory assumes that decision makers are rational and select alternatives that yield the maximum (latent) benefits for them. A multinomial logit (MNL) specification with independent (Gumbel-type) error terms is imperfect in many settings. For example, if construction costs for commercial development are high in one location, the costs for residential development are probably high too.

The multinomial probit model (MNP) allows for cross-alternative correlations. These correlations can be ascribed to missing variables characterizing choice alternatives. However, the independence across observation units remains problematic for many contexts. It is quite likely that one unit (e.g., a parcel) is influenced by its neighbors due to missing variables and/or spatial and other interactions (e.g., lighting conditions that affect decisions to use transit).

The spatial MNP model developed here accounts for both cross-individual interactions (emerging from physical proximity) and cross-alternative correlations. This paper is organized as follows: the SAR MNP's mathematical formulation is presented first, followed by a section on Bayesian MCMC estimation. Austin, Texas's land use data are used, with parameter estimates and inference summarized in the results section followed by some final remarks.

LITERATURE REVIEW

Discrete choice models are common in land use modeling. Examples include series of binomial logit models (Verburg et al. 2004) for residential, industrial/commercial, and recreational land uses on a 500m by 500m grid-cell map, Zhou and Kockelman's (2008) logit models for parcel subdivision, and UrbanSim's simulation code (Waddell et al. 2003). Even after controlling for a host of local, neighborhood attributes around grid cells and parcels, much spatial autocorrelation can remain in unobserved factors (Miaou et al. 2003). Few existing studies attempt to account for such effects, since these imply two-dimensional dependence across, potentially, thousands of observations, requiring manipulation of large matrices and high dimensional multivariate distributions.

As with various other socio-economic factors (including home prices, poverty levels, travel distances, and election outcomes), land use patterns tend to be correlated across space. Work on

discrete states of land use change with such specifications can be found in Munroe et al.'s (2002) series of binary probit and random-effect probit models using panel techniques, and Wang and Kockelman's (2009a, 2009b) spatially ordered probit model with a temporal component.

Spatial extensions of the multinomial probit model can be found in Chakir and Parent's (2009) analysis of France's land use change (including a fairly balanced choice set with urban, agricultural, forest, and vacant types). Their model assumed that spatial autocorrelation occurred across the error terms in a standard multinomial probit model for a total of 3,116 parcels across 164 counties. Only parcels within the same county were treated as correlated. Using maximum approximate composite marginal likelihood (MACML) techniques, Sidharthan and Bhat (2012) estimated a temporal spatial MNP model with spatial autocorrelation occurring across latent utilities, with application to land use changes for a 395 by 395 gridded neighborhood in the suburbs of Austin.

Anselin et al. (2006) noted how a spatial lag model (i.e., spatially autocorrelated response) is characteristic of a spatial or social interaction process, in which the value of the response variable at one location is jointly determined by its neighboring agents. In the empirical literature, it is often used to analyze interaction among local governments (e.g., taxation and nearby jurisdictions' expenditures). By contrast, a spatial error specification (such as the one used in Chakir and Parent [2009]) does not assume an underlying spatial or social interaction process, but rather a sort of nonspherical error covariance matrix due to omitted variables (Anselin et al. 2006). In empirics, the spatial error model is suitable for cases where spatial autocorrelation occurs in a subtle manner, such as when missing variables (like soil quality and rainfall) exhibit spatial clustering.

This paper builds upon the spatial MNP specification proposed by LeSage and Pace (2009) with spatially autocorrelated response and successfully estimates such models by incorporating advanced Bayesian techniques. An example is provided using Austin's parcel-level land use change data, which offers more behavioral realism than gridded data.

METHODOLOGY

This section details the mathematics of the spatial autoregressive multinomial probit (SAR MNP) model. Throughout the paper, i is used to indicate observations (or land parcels), j denotes alternatives, and k indicates the k^{th} covariate for observation i . The SAR MNP model assumes that the latent utilities in location i associated with land use types j can be expressed as a weighted sum of i 's neighbors' latent utilities associated with the same land use type¹. In other words, the $NJ \times 1$ vector of utilities $\tilde{\mathbf{y}}^* = (\mathbf{y}_1^{*'}, \mathbf{y}_2^{*'}, \dots, \mathbf{y}_N^{*'})$ and each of the $J \times 1$ vectors $\mathbf{y}_i^* = (y_{i1}^*, \dots, y_{ij}^*)'$ can be framed as a continuous SAR specification (LeSage and Pace 2009), as follows:

$$\tilde{\mathbf{y}}^* = \rho \tilde{W} \tilde{\mathbf{y}}^* + \tilde{X} \boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

¹ A total of $(J+1)$ alternatives is considered with indices $\{j = 0, 1, \dots, J\}$, where $\{j=0\}$ is the base alternative for identification purposes. Thus, y_{ij}^* is the difference between the j^{th} alternative's utility and that of the base alternative.

where ρ is the spatial autocorrelation coefficient. For dimension conformity, the $NJ \times KJ$

covariate matrix $\tilde{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}$, where $X_i = \begin{pmatrix} \mathbf{X}_{i,1} & \cdots & \cdots & \mathbf{0}_K \\ \vdots & \mathbf{X}_{i,2} & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{0}_K & \cdots & \mathbf{X}_{i,J} & \vdots \end{pmatrix}$, and $\tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix}$, with

$\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ})'$. Note that in the absence of generic variables (which vary by alternative – unlike, say, parcel size and distance to the nearest highway), the $K \times 1$ vector $\mathbf{X}_{i,j}$ will be identical across alternatives, for each observation i . One may consider using different numbers of covariates for each alternative by supplying an $\mathbf{X}_{i,j}$ vector of variable length, leading to a nearly identical model, still estimable using the Bayesian procedure laid out in this section (but with non-conjugate posterior, making parameter draws more challenging). For identification purposes, the $KJ \times 1$ vector $\boldsymbol{\beta}$ contains the stacked, alternative-specific parameter vectors: $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_J)'$. The $NJ \times NJ$ weight matrix is denoted as $\tilde{W} = W \otimes I_J$, where W is a typical $N \times N$ row-standardized weight matrix (with zero-valued diagonal elements, by construction²), I_J is a $J \times J$ identity matrix, and the symbol “ \otimes ” indicates a Kronecker product (where each element in the first matrix is multiplied by the entire second matrix, one at a time). The covariance matrix

for $\tilde{\boldsymbol{\varepsilon}}$ is $I_N \otimes \Sigma = \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma \end{bmatrix}$, where the $J \times J$ matrix Σ indicates the cross-alternative covariance matrix for error terms across alternatives. Since spatial autocorrelation already exists across the y_{ij}^* terms, this SAR MNP specification assumes independent and identical error terms ($\boldsymbol{\varepsilon}_i$) over space, but not alternatives. The closest applications in published work are Wang et al.’s (2011) dynamic spatial MNP and Chakir and Parent’s (2009) paper, which also assume that the cross-alternative covariance structure is identical over space/across observational units. This paper differs from the other spatial MNP models by allowing for spatially autocorrelated latent responses. This paper’s Appendix provides an alternative representation of this model, which may yield computational advantages by employing the matrix-variate distribution (Kadiyala and Karlsson 1997).

For a case with 4 alternatives (with one alternative serving as the base, so $J = 3$), Σ is a 3 by 3 covariance matrix: $\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$. The observed response values (y_i) are as follows:

$$y_i = j, \text{ if } y_{ij}^* = \max[y_{i,1}^*, y_{i,2}^*, \dots, y_{i,J}^*] > 0, \text{ and } y_i = 0 \text{ if } y_{ij}^* < 0 \text{ for all } j = 1, \dots, J.$$

The latent utility $\tilde{\mathbf{y}}^*$ follows a truncated NJ -dimension multivariate normal distribution:

$\tilde{\mathbf{y}}^* \sim MVN \left[(I_{NJ} - \rho \tilde{W})^{-1} \tilde{X} \boldsymbol{\beta}, (I_{NJ} - \rho \tilde{W})^{-1} (I_N \otimes \Sigma) (I_{NJ} - \rho \tilde{W})^{-T} \right]$. Its probability density

² $D_{ij} = \frac{1}{a_{ij}}$ if $d_{ij} < d_{\max}$, and $D_{ij} = 0$ if $d_{ij} \geq d_{\max}$, where d_{\max} is the threshold distance associated with the q^{th} -nearest neighbor to geographic unit i). For the $J=2$ case, the i^{th} and $(i+1)^{\text{th}}$ row of \tilde{W} is expressed as: $\tilde{W}_i = \begin{pmatrix} W_{i1} & 0 & W_{i2} & 0 & \cdots & W_{in} & 0 \\ 0 & W_{i1} & 0 & W_{i2} & \cdots & 0 & W_{in} \end{pmatrix}$.

function $p(\tilde{\mathbf{y}}^* | \rho, \boldsymbol{\beta}, \Sigma)$ is proportional to $|A^{-1}(I_N \otimes \Sigma)A^{-T}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(\tilde{\mathbf{y}}^* - A^{-1}\tilde{X}\boldsymbol{\beta})' \cdot [A^{-1}(I_N \otimes \Sigma)A^{-T}]^{-1}(\tilde{\mathbf{y}}^* - A^{-1}\tilde{X}\boldsymbol{\beta})\right\}$, where $A = (I_{NJ} - \rho\tilde{W}) = (I_N - \rho W) \otimes I_J$, and $-T$ denotes the transpose of the inverse (or the inverse of the transpose, thanks to their operational interchangeability). This density expression can be simplified, thanks to associated with matrix determinants and Kronecker products:

$$p(\tilde{\mathbf{y}}^* | \rho, \boldsymbol{\beta}, \Sigma) \propto |I_N - \rho W|^J \cdot |\Sigma|^{-N/2} \cdot \exp\left\{-\frac{1}{2}(A\tilde{\mathbf{y}}^* - \tilde{X}\boldsymbol{\beta})' \cdot (I_N \otimes \Sigma^{-1}) \cdot (A\tilde{\mathbf{y}}^* - \tilde{X}\boldsymbol{\beta})\right\}.$$

Truncation occurs because outcomes (y_i) dictate the ranges of values that \mathbf{y}_{ij}^* 's can take. For example, $y_i = 0$ leads to the restriction $\mathbf{y}_{ij}^* < 0$ for $j = 1, \dots, J$, which can be expressed as a series

of linear constraints: $\begin{bmatrix} -\infty \\ -\infty \\ -\infty \end{bmatrix} < D_i \cdot \mathbf{y}_i^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \mathbf{y}_i^* < \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. Likewise, the linear constraints

when $y_i = 1, 2$, or 3 are expressed as:

$$\begin{aligned} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} < D_i \cdot \mathbf{y}_i^* &= \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix} \cdot \mathbf{y}_i^* < \begin{bmatrix} +\infty \\ +\infty \\ +\infty \end{bmatrix} \text{ if } y_i = 1 \\ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} < D_i \cdot \mathbf{y}_i^* &= \begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \end{bmatrix} \cdot \mathbf{y}_i^* < \begin{bmatrix} +\infty \\ +\infty \\ +\infty \end{bmatrix} \text{ if } y_i = 2 \\ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} < D_i \cdot \mathbf{y}_i^* &= \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \mathbf{y}_i^* < \begin{bmatrix} +\infty \\ +\infty \\ +\infty \end{bmatrix} \text{ if } y_i = 3 \end{aligned} \quad (1)$$

These 3 by 3 matrices, D_i , will be used to draw truncated multivariate normal values in the Markov Chain Monte Carlo (MCMC) series. Such ideas are also discussed in LeSage and Pace's (2009) book (which proposed the SAR MNP specification used here), but with modifications.

Since it is unobservable, the latent response $\tilde{\mathbf{y}}^*$ is treated as an unknown parameter using data augmentation techniques, during the estimation process (as commonly done in Bayesian model settings: Carlin and Louis [2009] and McCulloch et al. [2000]).

BAYESIAN MCMC ESTIMATION

Using Bayes' classic theorem, the posterior distribution of unknown parameters is proportional to the response values' likelihood (or the data function) times any prior assumptions or understanding an analyst may have for all unknowns. Notationally, $p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\mathbf{y} | \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})}{p(\mathbf{y})} \propto p(\mathbf{y} | \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})$, with $\boldsymbol{\theta}$ encompassing all unknown model parameters and $\pi(\boldsymbol{\theta})$ indicating prior distributions or prior knowledge for the unknown parameters (Carlin and Louis 2009). In spatial settings for discrete response, $\boldsymbol{\theta}$ tends to be of high dimension with a posterior distribution whose form is unknown.

MCMC sampling is commonly used to estimate such models, by recasting the high-order integration into a series of simpler – and sometimes known – *conditional* distributions. This section discusses the techniques used to draw the unknown parameters $\boldsymbol{\beta}$, ρ , Σ , and $\tilde{\mathbf{y}}^*$.

If the latent responses were observable, the model's likelihood function could be written as $p(\tilde{\mathbf{y}}^* | \rho, \boldsymbol{\beta}, \Sigma) \propto |I_N - \rho W|^J \cdot |\Sigma|^{-N/2} \cdot \exp\left\{-\frac{1}{2}(\mathbf{A}\tilde{\mathbf{y}}^* - \tilde{\mathbf{X}}\boldsymbol{\beta})' \cdot (I_N \otimes \Sigma^{-1}) \cdot (\mathbf{A}\tilde{\mathbf{y}}^* - \tilde{\mathbf{X}}\boldsymbol{\beta})\right\}$. And the posterior distribution would then be: $p(\rho, \boldsymbol{\beta}, \Sigma | \tilde{\mathbf{y}}^*) \propto p(\tilde{\mathbf{y}}^* | \rho, \boldsymbol{\beta}, \Sigma) \cdot \pi(\rho) \cdot \pi(\boldsymbol{\beta}) \cdot \pi(\Sigma)$. Estimation can be conducted using Gibbs sampling, which utilizes conditional posterior distributions (e.g., $p(\rho | \tilde{\mathbf{y}}^*, \boldsymbol{\beta}, \Sigma)$ and $p(\boldsymbol{\beta} | \tilde{\mathbf{y}}^*, \rho, \Sigma)$), rather than trying to compute the posteriors (e.g., $p(\rho | \tilde{\mathbf{y}}^*)$, $p(\boldsymbol{\beta} | \tilde{\mathbf{y}}^*)$, $p(\Sigma | \tilde{\mathbf{y}}^*)$), which involve non-closed-form integrals. The following sub-sections present all conditional distributions used here.

Conditional Distribution of $\boldsymbol{\beta}$

Assuming a normal prior for $\boldsymbol{\beta}$ (i.e., $\boldsymbol{\beta} \sim MVN(\boldsymbol{\beta}_0, T_0)$),

$$\begin{aligned} p(\boldsymbol{\beta} | \rho, \tilde{\mathbf{y}}^*, \Sigma) &\propto p(\tilde{\mathbf{y}}^* | \boldsymbol{\beta}, \rho, \Sigma) \cdot \pi(\boldsymbol{\beta}) \\ &\propto MVN[\mathbf{A}^{-1}\tilde{\mathbf{X}}\boldsymbol{\beta}, \mathbf{A}^{-1}(I_n \otimes \Sigma)\mathbf{A}^{-T}] \cdot MVN(\boldsymbol{\beta}_0, T_0) \\ &\propto MVN[\mathbf{C}^*, \mathbf{T}^*] \end{aligned} \quad (2)$$

where $\mathbf{C}^* = \mathbf{T}^*[T_0^{-1}\boldsymbol{\beta}_0 + \tilde{\mathbf{X}}'(I_n \otimes \Sigma^{-1})\mathbf{A}'\tilde{\mathbf{y}}^*]$ and $\mathbf{T}^* = [\tilde{\mathbf{X}}'(I_n \otimes \Sigma^{-1})\tilde{\mathbf{X}} + T_0^{-1}]^{-1}$.

Using a mathematical trick called ‘‘completing the square’’, the conditional distribution is shown to be conjugate (i.e., the conditional distribution and its prior come from the same distribution type, with updated parameters). Here diffuse or uninformative priors are used for all parameters to avoid biasing estimation and inference by placing too much prior ‘‘knowledge’’ on parameters: $\boldsymbol{\beta}_0 = \mathbf{0}$ and $T_0 = a \cdot I_K$ with a represented by a very large real number (such as 10,000) and I_K serving as a K by K identity matrix.

Conditional Distribution of ρ

$$\begin{aligned} p(\rho | \boldsymbol{\beta}, \tilde{\mathbf{y}}^*, \Sigma) &\propto p(\tilde{\mathbf{y}}^* | \boldsymbol{\beta}, \rho, \Sigma) \cdot \pi(\rho) \\ &\propto |A| \cdot \exp\left\{-\frac{1}{2}(\tilde{\mathbf{y}}^* - \mathbf{A}^{-1}\tilde{\mathbf{X}}\boldsymbol{\beta})' \cdot \mathbf{A}'(I_N \otimes \Sigma^{-1})\mathbf{A} \cdot (\tilde{\mathbf{y}}^* - \mathbf{A}^{-1}\tilde{\mathbf{X}}\boldsymbol{\beta})\right\} \cdot U(0, 1) \\ &\propto |A| \cdot \exp\left\{-\frac{1}{2}(\tilde{\mathbf{y}}^* - \mathbf{A}^{-1}\tilde{\mathbf{X}}\boldsymbol{\beta})' \cdot \mathbf{A}'(I_N \otimes \Sigma^{-1})\mathbf{A} \cdot (\tilde{\mathbf{y}}^* - \mathbf{A}^{-1}\tilde{\mathbf{X}}\boldsymbol{\beta})\right\} \end{aligned} \quad (3)$$

Stationarity of the response-generating process (over space) requires that $\frac{1}{\omega_{\min}} < \rho < \frac{1}{\omega_{\max}}$ (Elhorst 2003), where ω_{\min} and ω_{\max} denote W 's minimum and maximum eigenvalues. While the maximum eigenvalue of a row-standardized W is guaranteed to be one (Elhorst 2003), empirical work also tends to restrict ρ 's lower bound, via the range $[-1, 1]$. In reality, no general rule yet defines a lower bound, but negative spatial autocorrelation tends to be rare, and this is especially true in terms of land development (where like land uses tend to cluster across parcels). Thus, a standard uniform prior is used here for ρ , with range $[0, 1]$.

Since $p(\rho | \boldsymbol{\beta}, \tilde{\mathbf{y}}^*, \Sigma)$ does not follow a known distribution for ready draws, Gibbs sampling cannot be used. Instead, the Metropolis-Hastings (M-H) algorithm was first used here, to draw ρ , as suggested by LeSage and Pace (2009). Since the M-H algorithm with a uniform or beta prior for ρ behaved poorly, with no sign of convergence, a relatively new technique, called generalized

direct sampling (GDS), was used (Braun and Damien 2011). The GDS method generates independent samples from a target posterior distribution in parallel, whereas Gibbs and M-H strategies rely on sequential sampling. The GDS procedure for drawing ρ values largely follows steps suggested by Braun and Damien (2011), with some minor adaptations. It can be described as follows:

1. Find the mode, ρ^* , of $\log[p(\rho|\boldsymbol{\beta}, \tilde{\mathbf{y}}^*, \Sigma)]$ and compute the (unnormalized) logarithm of its posterior density value, $c = \ln[p(\rho^*|\boldsymbol{\beta}, \tilde{\mathbf{y}}^*, \Sigma)]$. Optimization of this constrained density (to find the mode) can be conducted using MATLAB's nonlinear-optimization-with-bounds routine (*fmincon*). Supplying the first derivative (gradient) and second derivative (Hessian) of the Lagrangian (to reflect the bounds on ρ) facilitates MATLAB's search. For problems involving only bounds or linear constraints, the Lagrangian's and original objective function's Hessian are the same:

$$\text{Gradient} = J \cdot \text{Trace}[(I_N - \rho W)^{-1} W] - (\tilde{W} \tilde{\mathbf{y}}^*)' (I_N \otimes \Sigma^{-1}) [(I_{N_J} - \rho \tilde{W}) \tilde{\mathbf{y}}^* - \tilde{X} \boldsymbol{\beta}]$$

$$\text{Hessian} = J \cdot \text{Trace}[(I_N - \rho W)^{-1} W (I_N - \rho W)^{-1} W] - (\tilde{W} \tilde{\mathbf{y}}^*)' (I_N \otimes \Sigma^{-1}) (\tilde{W} \tilde{\mathbf{y}}^*)$$

2. Define a truncated normal distribution, $g(\rho)$ with the same mode, ρ^* , and let $k = g(\rho^*) = \text{Normal density } N(\rho^* | \text{mean} = \rho^*, \text{var} = s \cdot \text{Hessian}^{-1})$. Here, "s" is a scaling factor to fine-tune the proposal distribution to ensure regular acceptance of the proposed ρ values, in step 6 of this GDS process.
3. Sample $\rho_1, \rho_2, \dots, \rho_M$ independently from $g(\rho)$. Compute $\log[\Phi(\rho_m)] = \log[p(\rho_m|\boldsymbol{\beta}, \tilde{\mathbf{y}}^*, \Sigma)] - \log(c) + \log(k) - \log[g(\rho_m)]$ for each of the proposed draws. If $\log[\Phi(\rho_m)] > 0$ for any of these draws, repeat step 2 and choose another proposal distribution for which $\log[\Phi(\rho_m)] \leq 0$ holds.
4. Let Z be the minimum $\log[\Phi(\rho_m)]$ from the proposal draws. Evaluate $q_M(\frac{i}{N}) = \frac{1}{M} \sum_m 1 \left[\frac{\log(\Phi(\rho_m))}{|Z|} + 1 > \frac{i}{N} \right], i=1, \dots, N$, where q_M is a monotonic function of i (by construction) with values ranging between 0 (when $i = 1$) and 1 (when $i = N$).
5. Sample v from beta($i+1, N-i+1$), choosing i with probability proportional to $q_M(i/N)$. Compute $\log(u) = |Z|(v - 1)$.
6. Draw ρ from $g(\rho)$ until $\log[\Phi(\rho)] > \log(u)$. Accept the proposed ρ as a single draw from the target posterior distribution $p(\rho|\boldsymbol{\beta}, \tilde{\mathbf{y}}^*, \Sigma)$.

Conditional Distribution of Σ

Sampling the cross-alternative covariance matrix, Σ , is by no means a trivial task. From a strictly econometric viewpoint, this quantity is not identified. Thus, one must impose some restrictions for identification to hold. Usually, the first element on the diagonal line of Σ is set to one so that the other parameters can be identified (McCulloch et al. 2000, Koop 2003). Using procedures proposed by McCulloch et al. (2000) and summarized in Koop (2003), here the covariance matrix is first partitioned, as follows:

$$\Sigma = \begin{bmatrix} 1 & \eta' \\ \eta & \Phi + \eta\eta' \end{bmatrix}$$

For identification purposes, the first-row and first-column element is set to one, $\varepsilon_{i1} \sim N(0,1)$, and $(v_i = \varepsilon_{i2}, \dots, \varepsilon_{ij}) | \varepsilon_{i1} \sim N(\eta \cdot \varepsilon_{i1}, \Phi)$. Usually a multivariate normal prior is assigned for η and a Wishart prior³ for Φ^{-1} , which are mathematically expressed as: $\pi(\eta) \sim N(\eta_0, V_0)$ and $\pi(\Phi^{-1}) \sim W(p_0, \Phi_0^{-1})$. McCulloch et al. (2000) showed how the full conditionals for η and Φ^{-1} are conjugate:

$$p(\eta | \tilde{y}^*, \Phi, \beta) \sim MVN(\mu_\eta, \Omega_\eta)$$

where $\Omega_\eta = (V_0^{-1} + \Phi^{-1} \sum_{i=1}^N \varepsilon_{1i}^2)^{-1}$ and $\mu_\eta = \Omega_\eta (V_0^{-1} \eta_0 + \Phi^{-1} \sum_{i=1}^N v_i \varepsilon_{1i})$.

And $p(\Phi^{-1} | \tilde{y}^*, \eta, \beta) \sim W(p_1, \Phi_1^{-1})$

where $p_1 = N + p_0$ and $\Phi_1^{-1} = [\Phi_0 + \sum_{i=1}^N (v_i - \varepsilon_{1i} \eta) (v_i - \varepsilon_{1i} \eta)']^{-1}$.

Koop (2003) reported some empirical studies where the Gibbs sampler worked quite slowly for MNP models. Imai and Dyk (2005) proposed a sampler using marginal data augmentation, providing better convergence and efficiency. Dyk and Meng (2001) also provide some more efficient algorithms. This portion of the SAR MNP code is provided by LeSage and Pace (2009).

Conditional Distribution of \tilde{y}^*

The latent variable \tilde{y}^* follows a truncated multivariate normal distribution because \tilde{y}^* reflects the actual outcomes (y_i), as shown in Equation 1. In this SAR MNP case, one can draw samples from $\tilde{y}^* \sim MVN[\mu, \Omega]$, with $\mu = (I_{NJ} - \rho \tilde{W})^{-1} \tilde{X} \beta$ and $\Omega = (I_{NJ} - \rho \tilde{W})^{-1} (I_N \otimes \Sigma) (I_{NJ} - \rho \tilde{W})^{-T}$, subject to $a < D \tilde{y}^* < b$, where $D = \text{diag}(D_i)$ is an $NJ \times NJ$ block diagonal matrix to ensure that the linearly transformed \tilde{y}^* meets required constraints and D_i is a $J \times J$ matrix for each observation unit/site (as shown in Equation 1).

Geweke (1991) proposed an m -step approach to draw sequentially from the transformed MVN distribution $Z \sim N(0, T)$ subject to the constraint $\alpha \leq Z \leq \beta$, where $T = D \Omega D'$, $\alpha = a - D \mu$, $\beta = b - D \mu$, and $\tilde{y}^* = \mu + D^{-1} Z$. This approach is based on the fact that each element of Z is univariate truncated normal conditional on Z 's other elements. In other words, each element z_i can be expressed as a weighted average of the other elements z_{-i} plus a noise term, as follows: $z_i = \sum_{q \neq i}^{NJ} c_{iq} z_q + h_i \varepsilon_i$, where $c_{iq} = -(T_{ii})^{-1} T_{iq}$, $h_i^2 = (T_{ii})^{-1}$, $\varepsilon_i \sim N(0, 1)$ subject to the constraint $\frac{(\alpha_i - \sum_{q \neq i}^{NJ} c_{iq} z_q)}{h_i} < \varepsilon_i < \frac{(\beta_i - \sum_{q \neq i}^{NJ} c_{iq} z_q)}{h_i}$, with T_{ii} and T_{iq} indicating the ii^{th} and iq^{th} elements of T (where $q \neq i$).

Thus, each pass (out of a total of m passes) draws one element of Z each time, conditional on the other z_{-i} 's, until the series has sampled all NJ z_i 's. The algorithm terminates after the m^{th} step where m is predetermined (by the analyst) and the final pass of Z values are used to impute \tilde{y}^*

³ The Wishart distribution is any of a family of probability distributions defined over symmetric and nonnegative-definite matrix-valued random variables (i.e., "random matrices"). It is particularly valued in Bayesian inferences, since it leads to a conjugate prior for the inverse of the covariance matrix of a multivariate normal distribution (Robert and Casella 2004).

based on the equality $\tilde{y}^* = \mu + D^{-1}Z$. Geweke (1991) found that relatively small m values can result in fairly good estimates, so $m = 10$ was used here.

DATA SETS

The land use maps of Austin and its extraterritorial jurisdiction (ETJ) in 2000 and 2006 are used here for a demonstration of the SAR MNP model. There are 10,169 undeveloped parcels, out of a total of 242,631 parcels, in 2000. Among these undeveloped parcels, 7,219 experienced no physical changes over the 6-year period so can be traced to year 2006's parcels by a common ID field (as used by the City of Austin). Land use types were determined based on the City's land use map, for the entire city. To ease computational burdens, a 1,500-parcel sample was randomly selected for analysis, across the city (as illustrated in Figure 1).

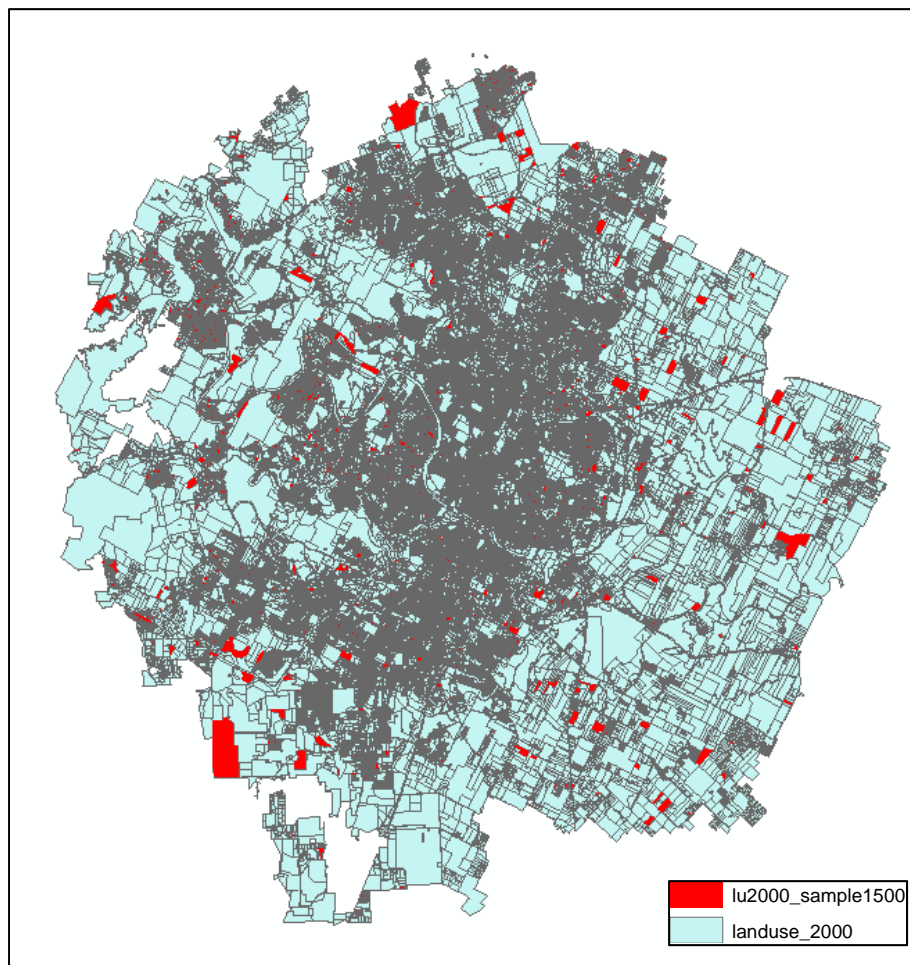


Figure 1. Austin and its ETJ Parcels in 2000 and Selected Sample

Water covered parcels, parks and green belts are excluded from analysis while golf courses and camp grounds are tagged as commercial uses. Ninety agricultural land parcels sprung up during the six-year period, with the majority located in the northeast side of Austin. Since the purpose of this paper is to study urban land development and the sample size for agricultural land is severely limited, this land category was also removed from the sample.

Seven land use types can be generalized from the City’s land use map codes: undeveloped, residential, commercial, civic, office, industrial, and transportation (including railroad facilities, utilities, and any parking lots labeled as the parcels’ predominant use). Given the scarcity of some use types, further grouping was conducted, resulting in four general land use types: undeveloped, residential, commercial/civic, office/industrial/transportation, with Table 1 summarizing their 2006 status.

Table 1 Development Status of 2000’s Vacant Parcels in 2006.

<i>Land Use (LU)</i>	<i>LU Code</i>	<i>Original Sample</i>		<i>Selected Region Sample</i>	
		<i>Counts</i>	<i>Shares</i>	<i>Counts</i>	<i>Shares</i>
Undeveloped	0	3,909	0.541	805	0.537
Residential	1	2,961	0.410	614	0.409
Commercial/Civic	2	169	0.023	41	0.027
Office/Industry	3	180	0.025	40	0.027
<i>Totals</i>		<i>7,219</i>	<i>1</i>	<i>1,500</i>	<i>1</i>

Explanatory variables (covariates) include parcel area (measured in square feet), parcel perimeter-to-area ratio (in ft⁻¹), network distances to Austin’s Central Business District (in miles), Euclidean distances to the nearest minor arterial, major arterial, and freeway, along with soil slope (represented by percent rise, or vertical rise per one unit increase in horizontal length), and Census block-group-level population density in year 2000 (measured in persons per acre). Table 2 provides summary statistics for these covariates. Network distances were computed using TransCAD’s shortest path function for each parcel’s centroid. Austin’s 2005 network file (from the Capital Area Metropolitan Planning Organization) was used. The rest of the variables were computed using ArcGIS and MATLAB, including Euclidean distances between all parcel centroids; these inter-parcel distances allowed for the construction of an inverse-distance weight matrix (*W*), reflecting the 50 nearest neighbors of each parcel (to provide some sparsity in the 1500x1500 weight matrix).

Table 2. Summary Statistics of Covariates of Parcels in Selected Neighborhood.

	<i>Area</i>	<i>PeriArea</i>	<i>DistCBD</i>	<i>DistMin</i>	<i>DistMaj</i>	<i>DistFwy</i>	<i>Slope</i>	<i>PopDens 2000</i>
Min	935.8	0.001	7.000	0.016	0.012	0.030	0.000	0.107
Max	9.817E+06	0.245	14.060	4.697	0.906	1.824	16.30	49.98
Mean	9.729E+04	0.033	10.800	2.294	0.227	0.603	2.652	5.645
Std Dev	4.754E+05	0.020	1.779	1.003	0.166	0.434	2.010	5.597
Units	Sq. Feet	1/Feet	Miles	Miles	Miles	Miles	Percent	Pop/Acre

A distance-based weight matrix was used to describe proximity between parcels. For each parcel, the first 50 nearest neighbors were identified and their inverse Euclidean distances were computed as the cell values in the matrix. Parcels that lie outside the 50-nearest-neighbor range were assigned zero weights (but they still can have effects on those parcels, thanks to shared neighbors, or neighbors of neighbors, etc. The diagonal elements take zero values by construction, with the final weight matrix row-standardized to facilitate estimation. The 50-

nearest-neighbor distances in this 1,500-parcel sample range from 1.51 miles to 6.43 miles, with an average “neighborhood” size of about 3.27 miles (and a standard deviation of 1.12 miles).

RESULTS AND ANALYSIS

Parameter estimates were obtained using the MCMC Bayesian estimation scheme, with the series of 3,000 draws shown in Figure 2 and posterior densities shown in Figure 3. The first 1,000 draws were treated as burn-in, a period when Bayesian learning is underway and MCMC chains evolve to a convergent series of distributions. Parameter point estimates are the average of the last 2,000 draws, and the draws provide the standard errors, with results shown in Table 3. Parcel size is estimated to be negatively correlated with the utilities of residential and commercial/civic development, while exerting a positive impact on a change to office/industry uses (from an undeveloped state) in a statistically significant way. The Perimeter-to-area ratio (*PeriArea*) characterizes parcel shape irregularity; the larger this value, the less regular (i.e., less square or circular) a parcel’s shape looks.⁴ The *PeriArea* variable exhibits a strong negative association with residential and commercial/civic development, and a positive association with office/industry uses. Thus, it appears that residential and commercial/civic development tends to favor more regularly shaped (and smaller, as noted above) parcels, which may be related to parcel conversion costs and aesthetics. In contrast, office/industry users may be more adaptable and place less emphasis on parcel configuration.

A parcel’s (network) distance to Austin’s CBD is estimated to affect land development outcomes in a statistically significant way. Estimates of marginal effects suggest that longer distances increase the likelihood of residential development, while reducing those of commercial/civic and office/industry uses, as expected. Having controlled for *DistCBD* (and *DistFwy*), distances to a parcel’s minor and major arterial roads are estimated to increase commercial/civic and office/industry development likelihood, perhaps because such developments are more common in less densely developed locations (as proxied by fewer arterials). Distance to the nearest freeway facility (*DistFwy*) is estimated to have a statistically significant – and positive – effect on the likelihood of residential development, suggesting that residential developers avoid proximity to freeways, *ceteris paribus*. Freeway distance exerts a strong negative effect on non-residential uses, after controlling for other distance effects (as described above), along with parcel size, shape, and slope.

Soil slope may be regarded as a proxy measure for construction costs (e.g., the steeper the slope, the more costly is land development). Here, it is estimated to be negatively associated with residential use but positively associated with commercial/civic and office/industry uses, though its effect on commercial/civic use is not significant (perhaps due to some steeper terrains offering view benefits). Interestingly, population density is estimated to have no statistically significant effects on commercial/civic and office/industry development likelihoods, though its negative impact on residential development is quite noticeable. Perhaps a block-group-level density variable is too coarse (and spatially arbitrary) for a parcel-level model; as an improvement, buffer-area densities (with annuli of ¼-mile and ½-mile, for example) could be built around each parcel, for future models of this type.

⁴ For example, longer parcels (with a high aspect ratio) and those with an “L” shape will have larger *PeriArea* values.

The SAR MNP model also generates estimates of cross-alternative correlations. For example, a parcel's residential use potential or (latent) "utility" is estimated to exhibit positive correlation with that of commercial/civic, but negative correlation with that of office/industry uses, while the latter two land uses' utilities (or latent attractiveness or profitability) exhibit some negative correlation. However, none of the cross-alternative covariance estimates is statistically significant, which may be due to a small and rather unbalanced sample, imperfect grouping of alternatives, and/or the latent nature of the model (and its underlying utility terms).

The spatial autocorrelation coefficient is estimated to be +0.706, with strong statistical significance, demonstrating the likely clustering of urban land development in the selected region. One expects this, since isolated developments are not as meaningful to their buyers/users as are coordinated developments, with greater neighborhood interaction, due to lowered travel costs and higher agglomeration benefits

Marginal effects are computed as the averaged changes in probabilities (by land use types) per one standard deviation (SD) increase in one of the covariates, holding everything else constant. The SAR MNP model allows one to analyze direct and indirect spatial effects separately, as opposed to the models developed by Chakir and Parent (2009) and Wang et al. (2011), which yield marginal effects indistinguishable from those of the standard MNP model (since they assume that spatial autocorrelation occurs only across the error terms, rather than across latent response terms). Unlike LeSage and Pace's (2009) bivariate spatial example, the SAR MNP model's marginal effects (for a particular land use type k and covariate p) are represented as a series of $K \cdot P \cdot n \times n$ matrices (where n is the total number of observations/locations, or 1,500 in this paper), with the ij^{th} cell value indicating changes in land use k 's probability at location i following a one-SD increase in the p^{th} covariate's value at location j (K is the total number of land use alternatives and P the total number of covariates minus the constant term). Unlike the bivariate (probit) spatial model, no closed-form expression exist for the first derivative of $p(y_i = k)$ with respect to x_{jp} , so the difference in probabilities (by land use types) is calculated as the departure of the new probabilities (per one SD increase in the covariates) from the baseline probabilities computed at the original covariates values. The averaged direct effects, shown in Table 4, are calculated as the average of the diagonal elements of this square matrix, while the averaged indirect effects result from averaging off-diagonal elements.

Here, a one-SD increase in the *Area* (parcel size) covariate of all parcel-level values is estimated to decrease the residential outcome probability (averaged across all sample records' predictions) by 0.014 and the average commercial/civic use outcome probability by 0.001, while raising the undeveloped and office/industry probabilities. Unlike most other covariates, parcel *Area* is estimated to exert more influence over a parcel's land use outcomes than this covariate's spillover (indirect) effects on its neighbors. In other words, home developers favor smaller parcels, and appear to avoid proximity to larger parcels, since larger parcels are more likely to become industrial and office uses, from which many home owners presumably desire spatial buffers. Both the *Area* and the *PeriArea* (ratio of parcel perimeter to area) variable show similar patterns when comparing their direct and indirect effects, suggesting that larger and more irregular-looking parcels are less attractive for residential and commercial/civic developments. Table 4 values show how a one-SD increase in the *DistCBD* variable lowers the model-predicted

probability of commercial/civic and office/industry uses for a parcel's own development. Nevertheless, this variable's spillover effect on office/industry use is estimated to be positive, suggesting that a one SD increase in *DistCBD* may be associated with office/industry development among neighbors (though dampening a parcel's own probability of such development). Such competing effects may emerge from the fact that large industrial land parcels are unlikely to locate near one another (a repelling pattern), in contrast to a cooperative (and clustering) pattern more evident among residential and commercial/civic parcel developments. While direct and indirect effects present a local picture of certain land development mechanisms (e.g., cooperative versus competitive behaviors within a neighborhood), one should use the *total* average (computed as the sum of the two separate effects) to compare the overall marginal effect of covariates (across the n=1,500 sample). For example, a one-SD increase in the *DistMin* variable across all parcels in the sampled data set is estimated to raise the sample-average predicted office/industry and residential outcome probabilities by 0.01 and 0.22, respectively, while decreasing commercial/civic likelihood by 0.0408. A one SD increase in *DistMaj* is estimated to lower the sample average residential and commercial/civic outcome probabilities by 0.10 and 0.003, while increasing office/industry outcome by 0.03. In contrast, a one-SD increase in the *DistFwy* variable predicts 0.2 and 0.09 increases in the samples' (overall/average) residential and commercial/civic probabilities, and a -0.12 decrease in office/industry likelihood.

Two additional observations can be made of the marginal effects table values. The first is the oscillating pattern found for the direct and indirect effects of some covariates for a particular use type (e.g., a negative direct effect accompanied by a positive indirect effect of *DistCBD* on office/industry development, or a positive direct effect accompanied by a negative indirect effect of *DistMaj* on residential development). Such patterns suggest locally "competitive" or "restrictive" behaviors among some land use types. In contrast, a "cooperative" pattern exists in neighborhoods where the direct and indirect effects are estimated to have the same sign. These differ from the more commonly used terms of "dispersion" and "clustering" (see, e.g., Anselin 2006, LeSage and Pace 2009), which reflect the sign of the model's spatial autocorrelation coefficient, which provides a measure of overall spatial dependence for the region under study. The second observation is that indirect effects dominate direct effects (in terms of absolute values) for half of the covariates considered here. Covariates such as population density, area size, and slope tend to yield indirect effects that are less than the corresponding direct effects.

Model Comparisons

This section compares parameter inference and goodness-of-fit statistics across the SAR MNP model (estimated above), an independent SAR MNP model (in which the covariance matrix is represented as an identity matrix), and their corresponding aspatial models (an independent MNP model that assumes the latent errors associated with each choice are i.i.d standard normal, and a standard MNP model – both estimated using STATA software, and requiring 28 and 32 minutes for estimation, respectively). Using the same n=1,500 land development data set, Tables 5 and 7 summarize results of these four models. Deviance information criteria (DIC) values are used to compare the two spatial models, as estimated using Bayesian methods, while the likelihood ratio index (LRI) statistic can be used to compare all four model fits.

The SAR MNP and independent SAR MNP models provide higher levels of significance for the covariates while yielding higher (as expected) mean and median log-likelihood values (as compared with the log-likelihood values at convergence for the two aspatial models). Likelihood ratio tests suggest that the SAR MNP model yields a (statistically) significantly better fit than the independent SAR MNP model ($LRI = -2 \times (-2461.62 + 2451.88) = 19.486 > \chi^2(6) = 12.59$), which is consistent with DIC value comparisons (since lower DIC values indicated better fit for Bayesian models). Prediction results for a hold-out sample of 3,000 parcels are summarized in Table 8, along with pseudo- R^2 values. The SAR MNP model yields the highest correct prediction rates, followed by the independent SAR MNP, MNP, and independent MNP models. However, improved parameter inference and fit for the two spatial models come at the expense of markedly longer run times, as shown in Tables 3 and 5.

CONCLUSION

This paper builds upon the spatial autoregressive MNP specification proposed by LeSage and Pace (2009) and specifies a four-alternative (four-category) SAR MNP model, to account for both spatial clustering and cross-alternative correlations. Two issues have plagued successful estimation and application of such models. One relates to the difficulty of sampling the spatial autocorrelation coefficient, ρ . Conventional M-H algorithms were initially tested here, using a random-walk proposal, followed by more sophisticated alternatives (such as the beta proposal, whose distribution parameters are determined using the mode and Hessian of the maximum log posterior of $p(\rho | \beta, \tilde{y}^*, \Sigma)$). Test runs showed how these M-H algorithms failed to converge, as reflected in ρ 's trace plot oscillations. Generalized direct sampling or GDS (Braun and Damien 2011) was then used to sample ρ , relying on independent draws from a judiciously selected proposal distribution that shares the same mode with the target posterior. GDS offers some computational advantages -- thanks to parallel programming, and it avoids convergence concerns that many associate with Bayesian MCMC methods (thanks to independent samples from the proposal distribution). Codes of the GDS sampling method are available upon request.

A more fundamental question relates to whether the resulting posterior is proper. As Nobile (2000) noted, improper posteriors (or posteriors that are not bounded in the real domain) may result even when the conditional posterior distributions are well defined, a problem exacerbated by the fact that the trace plots of the Gibbs sampling output may not be able to hint at posterior impropriety (see Hobert and Casella [1996] for more details). Even though the estimation results seem successful, a theoretical proof of the propriety of the joint posterior would be helpful, to further validate the SAR MNP specification.

This work's SAR MNP model was applied to a sample of 1,500 parcels randomly selected across Austin, Texas. Bayesian MCMC sampling produced a series of 3,000 draws of slope parameters, covariance terms, and the spatial autocorrelation coefficient. The perimeter-over-area (*PeriArea*) variable was found to exhibit a strong negative association with Residential and Commercial/Civic development (suggesting that oddly shaped parcels are less likely for such developments), and a statistically significant positive association with Office/Industry and Undeveloped uses. Longer distances to CBD were estimated to increase the likelihood of all three types of land development. And, ceteris paribus, Euclidean distances to a parcel's closest minor and major arterial roads were shown to increase commercial and industrial development probabilities, perhaps because such developments are more common in less densely developed

locations (as reflected by fewer arterials). The nearest-freeway's distance was estimated to have a strong positive effect on the likelihood of residential development (suggesting that, *ceteris paribus*, residential developers avoid proximity to freeways), while exerting a strong negative effect on non-residential uses, after controlling for other distance-based covariates, along with parcel size, shape, and slope. Steeper soil slopes are estimated to reduce the attractiveness of Residential uses while increasing the likelihood of Commercial/Civic and Office/Industry uses, and block-group-level population density has no statistically significant effects with the exception of a fairly strong negative effect on Residential development, *ceteris paribus*. A comparison of direct and indirect effects of covariate changes suggests that indirect effects generally outweigh direct effects among the covariates, with the exception of a parcel's *Area* variables, which exert more influence over a parcel's own land use outcomes than their spillover (indirect) effects on its neighbors.

Another interesting observation relates to the oscillating pattern found for the direct and indirect effects of some covariates for a particular use type. This pattern indicates "competitive" or "restrictive" behaviors of some land use type, as opposed to a "cooperative" pattern or behavior for neighborhoods, where the direct and indirect effects are estimated to have the same sign. This differs from "dispersion" and "clustering" behaviors already prevalent in the spatial analysis literature (to describe an overall/regional spatial relationship).

An inverse-distance weight matrix (W) was used here, for the 50 nearest neighbors of each parcel. As expected, clustering effects were estimated to be significant (with an average spatial autocorrelation coefficient of +0.706), and aspatial error correlations were positive across the commercial use types (and negative across residential and non-residential development types). The time-consuming nature of draws for larger data sets is an issue, along with imbalance in outcomes, common among land use data sets. A more judicious way to select data points, to achieve tractability and better balance, would be useful, along with higher-speed sampling procedures. Nevertheless, this work was able to successfully navigate the sampling necessary for Bayesian estimation of an MNP model, while permitting spatial autocorrelation through spatially lagged latent utility terms. This style of a SAR MNP model significantly improves model fit, as compared to the aspatial and independent SAR MNP models (based on likelihood ratio tests and hold-out sample's prediction results). Such specifications and estimation techniques hold promise for a variety of travel behavior and other choice-related, discrete-response data sets.

ACKNOWLEDGEMENTS

The authors thank Drs. Michael Braun, James LeSage, and Olivier Parent for their insightful comments and suggestions, and Ms. Annette Perrone for all her administrative and editorial support. They also thank the National Science Foundation for Award SES-0818066, which financially supported this work.

REFERENCES

Anselin, L. (2006) Spatial Econometrics. In: Mills T., Patterson K. (eds) *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Palgrave Macmillan, New York: 901-969.

- Braun, M. and Damien, P. (2011) Generalized direct sampling for hierarchical Bayesian models. McCombs Research Paper Series No. IROM-02-11. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.1907835>
- Carlin, B. P. and Louis, T. A. (2009) *Bayesian Methods for Data Analysis 3rd Edition*, CRC Press, Boca Raton, FL.
- Cervero, R. and Kockelman, K. (1997) Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment* 2(3): 199-219.
- Chakir, R., and Parent, O. (2009) Determinants of Land Use Changes: a Spatial Multinomial Probit Approach. *Papers in Regional Science* 88 (2):327-344.
- Elhorst, P. J. (2003) Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review* 26 (3): 244-268.
- Geweke, J. (1991) Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities. Proceedings of 23rd Symposium on the Interface Between Computing Science and Statistics. Seattle, Washington.
- Hobert, J., and Casella, G. (1998) Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors, *Journal of Computational & Graphical Statistics* 7(1): 42-60.
- Imai, K., and van Dyk, D. (2005) MNP: R Package for Fitting the Multinomial Probit Model. *Journal of Statistical Software* 14(3): 1-32.
- Kadiyala, K. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR models, *Journal of Applied Econometrics*, 12: 99-132.
- Koop, G. (2003) *Bayesian Econometrics*, John Wiley & Sons Ltd, West Sussex, England.
- LeSage, J., and Pace, K. (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL.
- Litman, T. (2012) Land use impacts on transport. Victoria Transport Policy Institute. URL: <http://www.vtpi.org/landtravel.pdf>
- 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(1): 173-93.
- McFadden, D. (1986) The choice theory approach to market research. *Market Science* 5 (4): 275-297.
- Munroe, D. K., J. Southworth, C. M. Tucker. (2002) The Dynamics of Land-Cover Change in Western Honduras: Exploring Spatial and Temporal Complexity. *Agricultural Economics* 27 (3):355-369.
- Nobile, A. (2000) Comment: Bayesian multinomial probit models with normalization constraint. *Journal of Econometrics* 99: 335-345.

Robert, C., and Casella, G. (2004) *Monte Carlo Statistical Methods*, 2nd Edition, Springer, New York.

Sidharthan, R. and Bhat, C.R. (2012) Incorporating Spatial Dynamics and Temporal Dependency in Land Use Change Models, forthcoming, *Geographical Analysis*.

van Dyk, D., and Meng, X. (2001). The Art of Data Augmentation. *Journal of Computational and Graphical Statistics* 10: 1-111.

Verburg, P.H., J. R. Van Eck, T. C. Nijs, and M. J. Dijst. (2004) Determinants of Land-Use Change Patterns in the Netherlands. *Environmental and Planning B* 31:125-150.

Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and F. Ulfarsson. (2003) Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics* 3 (1):43-67.

Wang, X., and K. M. Kockelman. (2009a) Application of the Dynamic Spatial Ordered Probit Model: Patterns of Land Development Change in Austin, Texas. *Papers in Regional Science* 88 (2): 345-366.

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

Wang, X., Kockelman, K.M., and Lemp, J. (2011) The Dynamic Spatial Multinomial Probit Model: Analysis of Land Use Change Using Parcel-Level Data. *Journal of Transport Geography* 24: 77-88.

Zhou, B., and K. M. Kockelman. (2008) Neighborhood Impacts on Land Use Change: A Multinomial Logit Model of Spatial Relationships. *Annals of Regional Science* 42 (2):321-340.

APPENDIX

An alternative way to represent the SAR MNP model uses matrix-variate priors. Kadiyala and Karlsson (1997) provided the technicalities of the matrix-variate t and normal distributions, and these are used to guide the derivation of ρ 's posterior for the SAR MNP model.

If one writes the stacked latent response as $\tilde{\mathbf{y}}^* = (\mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_J^*)'$, with $\mathbf{y}_j^* = (y_{1j}^*, \dots, y_{Nj}^*)'$, then $\tilde{\mathbf{X}}$ should take the form: $\tilde{\mathbf{X}} = \begin{bmatrix} X & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X \end{bmatrix} = I_J \otimes X$, with the $N \times K$ covariate matrix X repeating itself across different alternatives (due to a lack of generic and/or alternative-specific variables).

Similarly, $\tilde{\mathbf{W}} = I_J \otimes W$ and $\tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_J \end{pmatrix}$, with $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \dots, \varepsilon_{Nj})'$. Here, $\tilde{\boldsymbol{\varepsilon}}$ is assumed to follow a

$N(\mathbf{0}, \Sigma \otimes I_N)$ distribution, since the correlated error terms of each observation are separate. $\boldsymbol{\beta}$ assumes the same form as shown in the body of this paper. This alternative representation offers

the possibility of utilizing the matrix-variate distribution, which may facilitate estimation of ρ 's posterior distribution, by integrating out the other unknown parameters.

Given the SAR assumption, where $\tilde{\mathbf{y}}^* = \rho\tilde{W}\tilde{\mathbf{y}}^* + \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$, the likelihood function follows a matrix-variate distribution, with density function

$$p(\tilde{\mathbf{y}}^*|\rho, \boldsymbol{\beta}, \Sigma) = (2\pi)^{-nJ/2}|\Sigma|^{-n/2}\exp\left\{-\frac{1}{2}\text{tr}[\Sigma^{-1}(A\tilde{\mathbf{y}}^* - X\boldsymbol{\beta})'(A\tilde{\mathbf{y}}^* - X\boldsymbol{\beta})]\right\}, \text{ where } A = I_n - \rho W.$$

Assuming a diffuse prior for Σ and $\boldsymbol{\beta}$, such that $\pi(\boldsymbol{\beta}, \Sigma) \propto |\Sigma|^{-\frac{J+1}{2}}$, the joint posterior can be written as follows:

$$p(\rho, \boldsymbol{\beta}, \Sigma|\tilde{\mathbf{y}}^*) \propto |\Sigma|^{-\frac{J+1}{2}}|\Sigma|^{-\frac{n-k}{2}}\exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(A\tilde{\mathbf{y}}^* - X\hat{\boldsymbol{\beta}})'(A\tilde{\mathbf{y}}^* - X\hat{\boldsymbol{\beta}})\right]\right\}|\Sigma|^{-\frac{k}{2}}\exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\}$$

Thus, the conditional posterior for $\boldsymbol{\beta}$ follows the matrix-variate normal distribution:

$$p(\boldsymbol{\beta}|\rho, \Sigma, \tilde{\mathbf{y}}^*) \sim MN_{kJ}(\hat{\boldsymbol{\beta}}, \Sigma, (X'X)^{-1}).$$

One advantage of assuming matrix-variate priors lies in the ease in integrating out $\boldsymbol{\beta}$ and Σ , and then obtaining ρ 's marginal posterior:

$$p(\rho, \boldsymbol{\beta}|\tilde{\mathbf{y}}^*) \propto \int p(\rho, \boldsymbol{\beta}, \Sigma|\tilde{\mathbf{y}}^*)d\Sigma = \left|S + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right|^{-\frac{n}{2}}, \text{ with } S = (A\tilde{\mathbf{y}}^* - X\hat{\boldsymbol{\beta}})'(A\tilde{\mathbf{y}}^* - X\hat{\boldsymbol{\beta}}).$$

$$p(\rho|\tilde{\mathbf{y}}^*) = \int p(\rho, \boldsymbol{\beta}|\tilde{\mathbf{y}}^*)d\boldsymbol{\beta} \propto |S|^{-k/2} \text{ when assuming } \boldsymbol{\beta} \text{ follows matrix-variate } t \text{ distribution.}$$

One may also consider using the griddy Gibbs sampling technique (LeSage and Pace 2009) to draw directly from the marginal posterior of ρ , as shown above.

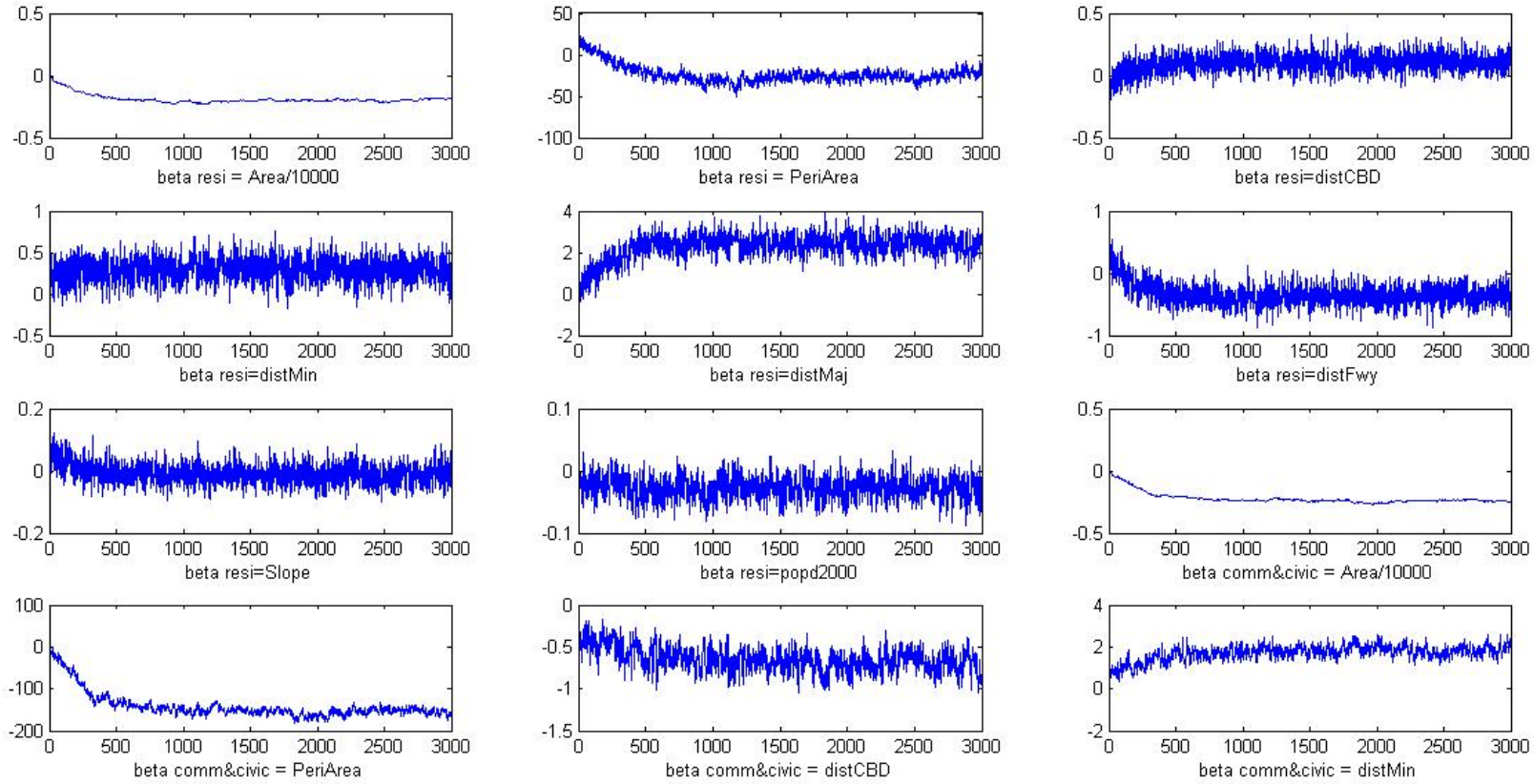


Figure 2a. Trace Plots of Parameter Estimates for SAR MNP.

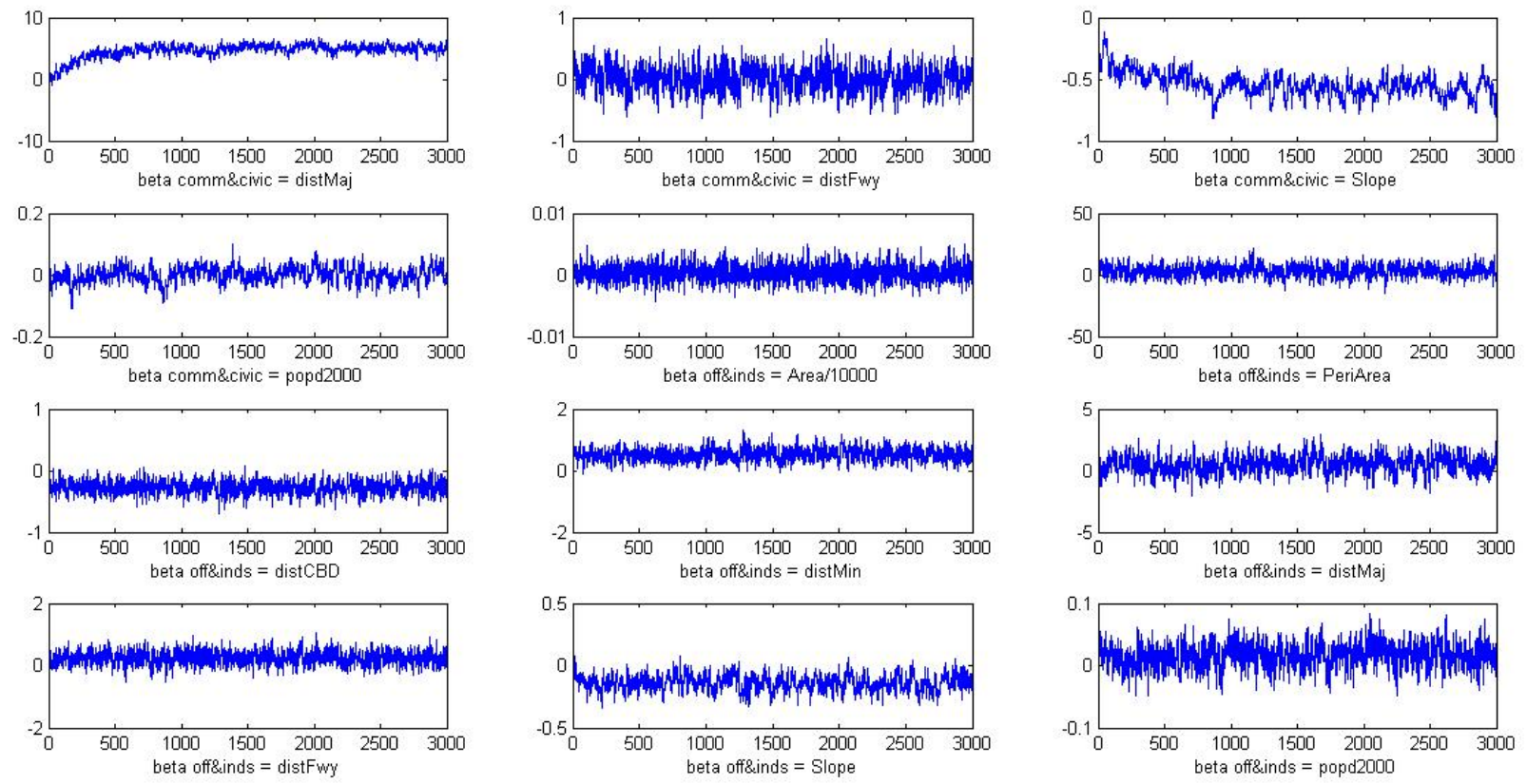


Figure 2b. Trace Plots of Parameter Estimates for SAR MNP.

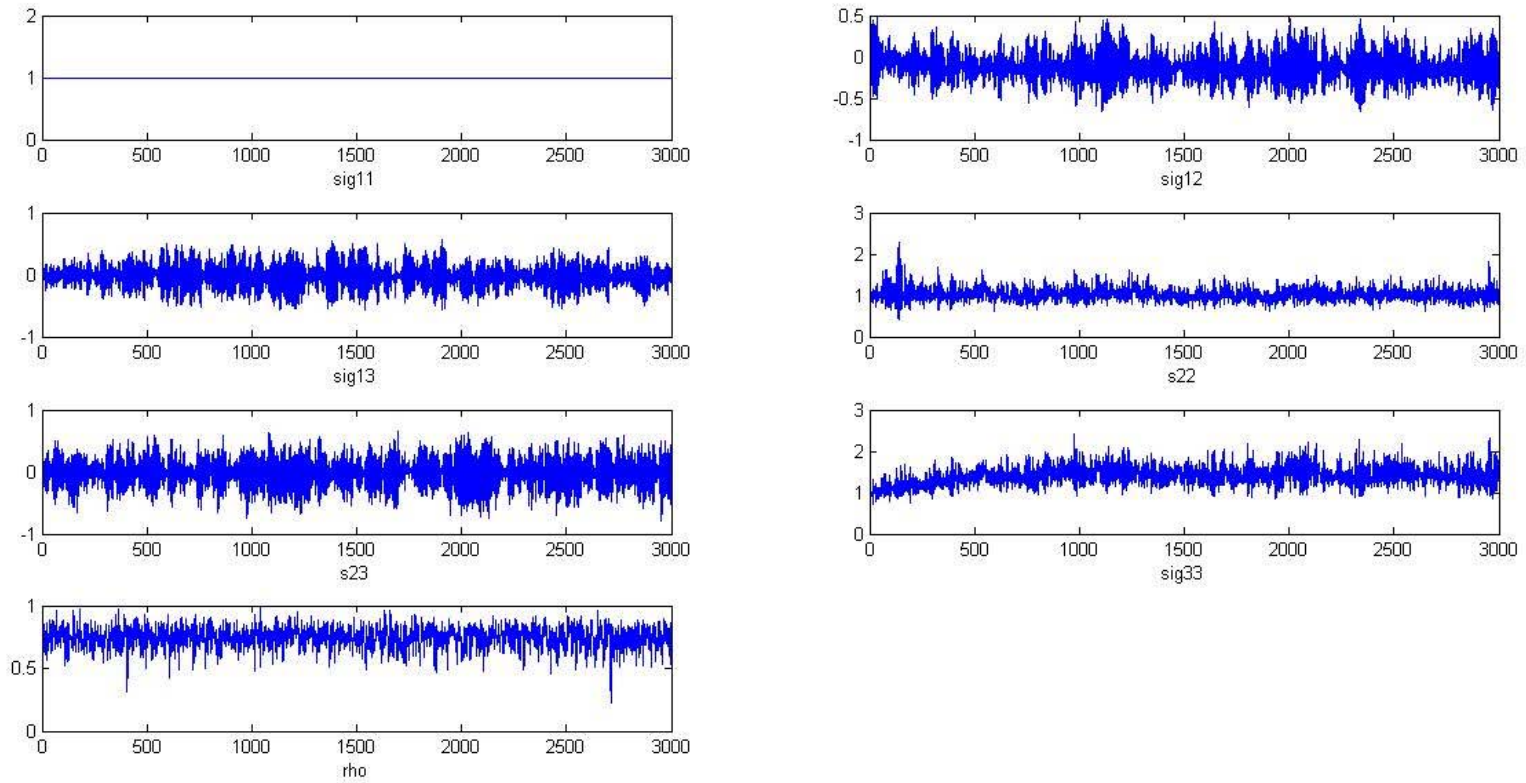


Figure 2c. Trace Plots of Parameter Estimates for SAR MNP.

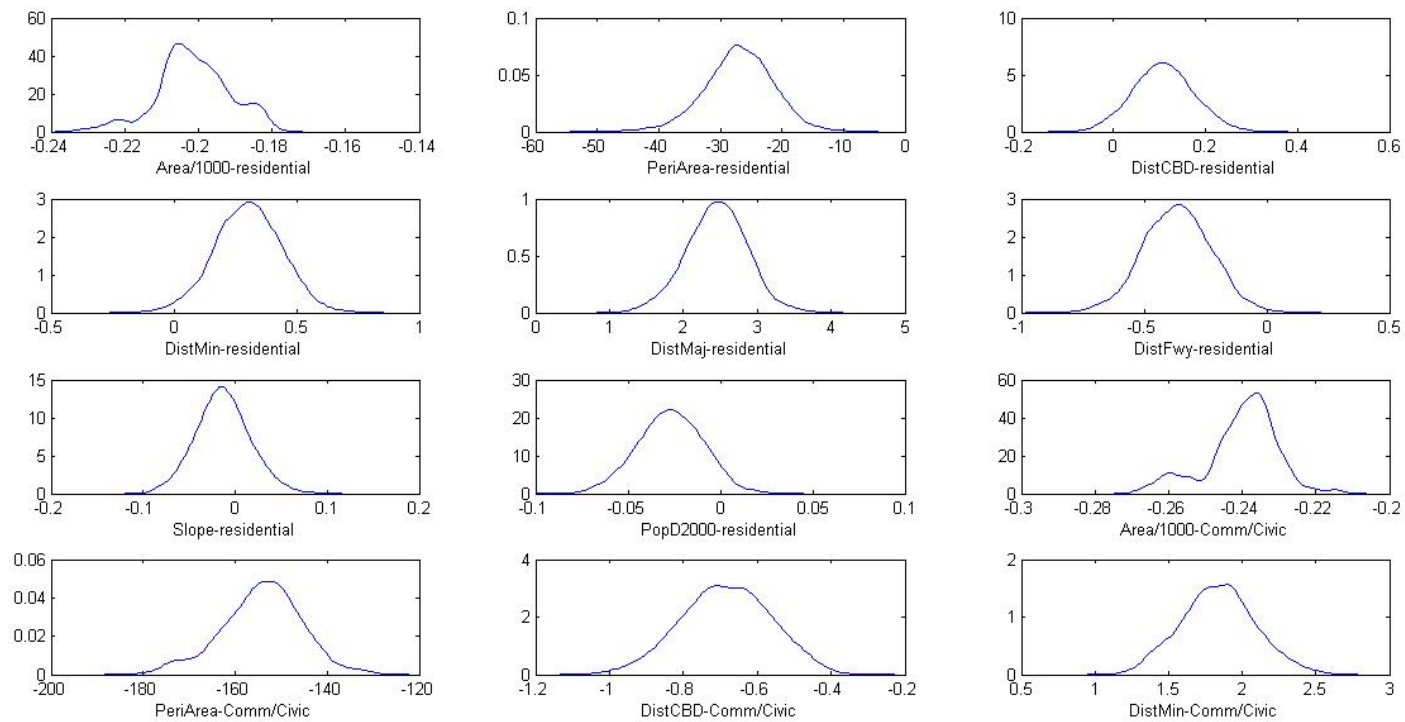


Figure 3a. Posterior Densities of Parameter Draws

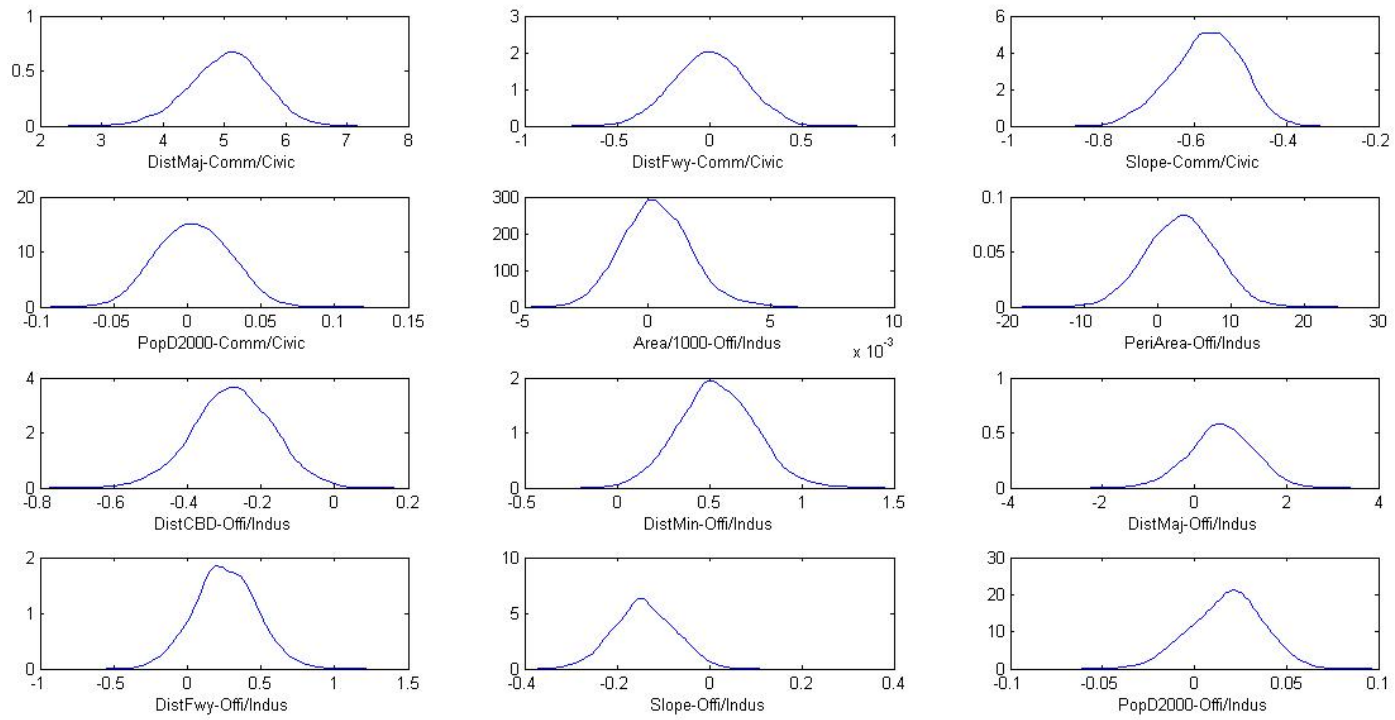


Figure 3b. Posterior Densities of Parameter Draws

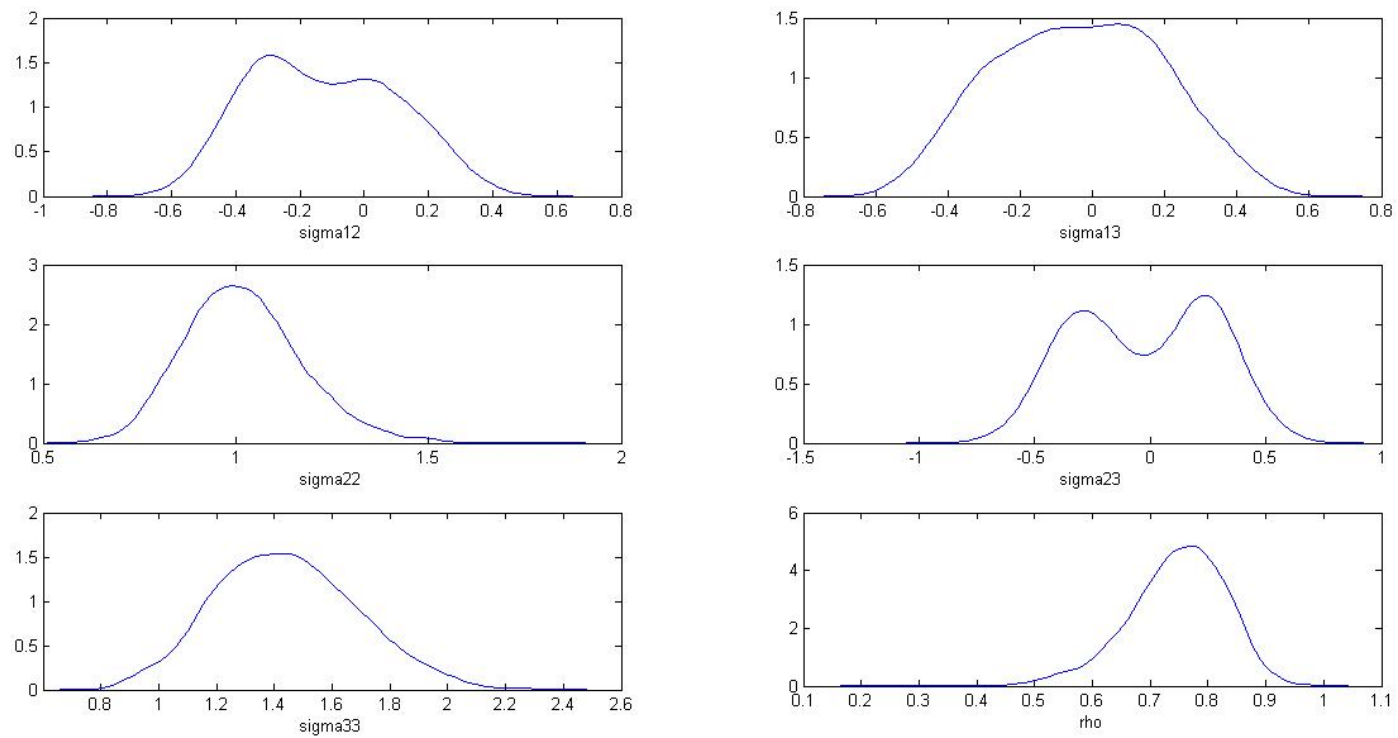


Figure 3c. Posterior Densities of Parameter Draws

Table 3. Parameter Estimates of SAR MNP Model (sample size=1,500).

	Residential	t-Stat	Commercial/civic	t-Stat	Office/Industry	t-Stat
<i>Constant</i>	-0.328	-2.11	-0.472	-2.66	-0.934	-5.35
<i>Area/1000</i>	-0.005	-6.50	-0.001	-1.99	0.001	3.17
<i>PeriArea</i>	3.932	2.02	6.336	2.78	11.17	4.69
<i>DistCBD</i>	0.043	3.39	-0.075	-3.32	-0.085	-2.93
<i>DistMIN</i>	-0.118	-2.70	0.297	3.87	0.163	1.61
<i>DistMAJ</i>	0.584	4.44	0.348	1.52	0.090	0.39
<i>DistFWY</i>	0.100	2.08	-0.204	-2.29	-0.114	-1.24
<i>Slope</i>	-0.022	-1.86	0.001	0.04	0.074	4.03
<i>Popd2000</i>	-0.012	-2.25	-0.005	-0.81	0.000	0.08
	Estimate	t-Stat				
σ_{11}	1.000	-				
σ_{12}	0.004	0.02				
σ_{13}	-0.015	-0.14				
σ_{22}	1.012	5.63				
σ_{23}	-0.018	-0.11				
σ_{33}	1.015	7.03				
ρ	0.706	8.03				
<i>DIC</i>	5204.46					
<i>Median LogLik</i>	-2451.88					
<i>Mean LogLik</i>	-2571.23					
<i>Run times</i>	47.3 hrs					
# of draws=3000, burnin=1000						

Notes: DIC is a goodness-of-fit measure for Bayesian models and is computed as the sum of the effective number of parameters (p_D) and the expectation of deviance (\bar{D}), with $p_D = \bar{D} - D(\hat{\theta})$ and $D(\theta) = -2 \log(p(y|\theta)) + C$, where θ denotes the unknown parameters, y represents data, and C is a constant which cancels when comparing different Bayesian models.

Table 4. Marginal Effects of Parameter Estimates.

	Avg. Effect	Undeveloped	Residential	Comm/Civic	Office/Industry
<i>Area/1000</i>	Direct	-0.495	-0.014	-0.001	0.510
	Indirect	-0.343	-0.007	-8.11E-07	0.350
<i>PeriArea</i>	Direct	-0.100	-0.015	-0.005	0.120
	Indirect	-0.142	-0.083	-0.007	0.232
<i>DistCBD</i>	Direct	0.018	0.034	-0.046	-0.006
	Indirect	0.012	0.025	-0.039	0.002
<i>DistMIN</i>	Direct	-0.025	-0.001	0.020	6.01E-03
	Indirect	-0.205	0.222	-0.028	0.011
<i>DistMAJ</i>	Direct	-0.054	0.040	0.009	0.005
	Indirect	0.127	-0.136	-0.012	0.021
<i>DistFWY</i>	Direct	0.236	0.048	-0.257	-0.027
	Indirect	-0.408	0.156	0.347	-0.095
<i>Slope</i>	Direct	0.211	-0.267	2.20E-03	0.054
	Indirect	0.014	-0.097	2.96E-03	0.080
<i>Popd2000</i>	Direct	0.168	-0.149	-0.058	0.039
	Indirect	-0.036	-0.014	-0.001	0.051

Table 5. Parameter Estimates of SAR MNP Model with Homoskedastic, Uncorrelated Error Terms.

	Residential	t-Stat	Comm/Civic	t-Stat	Office/Industry	t-Stat
<i>Constant</i>	-0.490	-3.30	-0.506	-2.71	-0.770	-4.44
<i>Area/1000</i>	-0.004	-5.62	-0.001	-1.51	0.001	3.15
<i>PeriArea</i>	5.657	3.12	7.470	3.27	10.705	4.69
<i>DistCBD</i>	0.046	3.67	-0.079	-3.53	-0.110	-3.85
<i>DistMIN</i>	-0.149	-3.66	0.311	4.07	0.252	2.49
<i>DistMAJ</i>	0.546	4.20	0.502	2.33	-0.262	-1.10
<i>DistFWY</i>	0.132	2.78	-0.260	-2.90	-0.115	-1.20
<i>Slope</i>	-0.018	-1.59	-0.005	-0.23	0.091	4.70
<i>Popd2000</i>	-0.011	-2.06	-0.003	-0.47	0.000	0.07
ρ	0.751	6.21				
<i>DIC</i>	5241.76					
<i>Median LogLik</i>	-2461.62					
<i>Mean LogLik</i>	-2581.88					
<i>Run times</i>	35.2 hrs					
# of draws=3000, burnin=1000						

Table 6. Parameter Estimates of Standard MNP Model.

Multinomial probit regression						Number of obs = 1500
						Wald chi2(24) = 229.32
Log likelihood = -3074.31						Prob > chi2 = 0.0000
	Residential	t-Stat	Comm/Civic	t-Stat	Office/Industry	t-Stat
<i>Constant</i>	-1.751	-6.21	-1.504	-1.57	-0.793	-2.60
<i>Area/1000</i>	-0.007	-3.04	-0.004	-0.65	0.001	1.61
<i>PeriArea</i>	3.547	1.07	4.329	1.18	-7.920	-0.53
<i>DistCBD</i>	0.214	8.16	0.158	1.02	-0.091	-0.55
<i>DistMIN</i>	-0.471	-5.22	-0.361	-1.17	-0.620	-1.61
<i>DistMAJ</i>	0.920	3.37	0.198	2.12	-2.710	-1.62
<i>DistFWY</i>	0.287	2.30	-0.047	-1.07	-0.021	-2.10
<i>Slope</i>	-0.031	-1.51	-0.014	-2.4	-0.081	-1.76
<i>Popd2000</i>	-0.015	-0.71	0.021	1.13	0.171	1.21
	Estimate	t-Stat				
σ_{11}	1	-				
σ_{12}	1.219	1.52				
σ_{13}	-0.637	-0.56				
σ_{22}	1.237	0.50				
σ_{23}	-0.822	-0.43				
σ_{33}	1.293	0.08				

Table 7. Parameter Estimates of an Independent MNP Model.

Independent Multinomial probit regression						Number of obs = 1500
						Wald chi2(24) = 194.59
Log likelihood = -3079.41						Prob > chi2 = 0.0000
	Residential	t-Stat	Comm/Civic	t-Stat	Office/Industry	t-Stat
<i>Constant</i>	-0.933	-4.95	-0.453	-1.31	-0.953	-2.91
<i>Area/1000</i>	-0.006	-5.56	0.000	0.03	0.001	2
<i>PeriArea</i>	1.449	0.65	-7.184	-1.69	-4.174	-0.95
<i>DistCBD</i>	0.130	7.23	0.010	0.26	0.003	0.08
<i>DistMIN</i>	-0.294	-4.95	0.031	0.25	-0.238	-1.57
<i>DistMAJ</i>	0.508	3.11	-0.885	-1.83	-0.835	-1.76
<i>DistFWY</i>	0.209	3.15	-0.104	-0.72	0.078	0.58
<i>Slope</i>	-0.016	-1.13	-0.268	-3.6	-0.050	-1.48
<i>Popd2000</i>	-0.008	-1.31	0.006	0.87	0.008	1.47

Table 8. Correct Predication Rates among Candidate Models.

Models	Pseudo- R ²	Land Type	Undev.	Residen.	Comm/ Civic	Office/ Indus.	Observed Totals	% Correctly Predicted
SAR MNP	0.354	Undev.	739				1620	43.5
		Residen.		517			1230	
		Comm/Civic			24		69	
		Office/Indus.				26	81	
Indept. SAR MNP	0.322	Undev.	652				1620	39.1
		Residen.		469			1230	
		Comm/Civic			28		69	
		Office/Indus.				23	81	
MNP	0.28	Undev.	324				1620	18.8
		Residen.		209			1230	
		Comm/Civic			14		69	
		Office/Indus.				17	81	
Indept. MNP	0.23	Undev.	281				1620	15.5
		Residen.		156			1230	
		Comm/Civic			13		69	
		Office/Indus.				14	81	