THE IMPACT OF WEIGHT MATRICES ON PARAMETER ESTIMATION AND INFERENCE: A CASE STUDY OF BINARY RESPONSE USING LAND USE DATA

Yiyi Wang
Assistant Professor
Civil Engineering Department
Montana State University
Bozeman, MT 59717-3900
yiyi.wang@ce.montana.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
6.9 E. Cockrell Jr. Hall
Austin, TX 78712-1076
kkockelm@mail.utexas.edu
Phone: 512-471-0210
FAX: 512-475-8744

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

ABSTRACT

This paper develops two new models and evaluates the impact of using different weight matrices on parameter estimates and inference in three distinct spatial specifications for discrete response. These specifications rely on a conventional, sparse, inverse-distance weight matrix for a spatial auto-regressive probit (SARP), a spatial autoregressive approach where the weight matrix

includes an endogenous distance-decay parameter (SARPα), and a matrix exponential spatial specification for probit (MESSP). These are applied in a binary choice setting using both simulated data and parcel-level land use data. Parameters of all models are estimated using Bayesian methods.

In simulated tests, adding a distance-decay parameter term to the spatial weight matrix improved the quality of estimation and inference, as reflected by a lower DIC value, but the added sampling loop required to estimate the distance-decay parameter substantially increased computing times. In contrast, the MESSP model's obvious advantage is its fast computing time, thanks to elimination of a log-determinant calculation for the weight matrix. In the model tests using actual land use data, the MESSP approach emerged as the clear winner, in terms of fit and computing times. Results from all three models offer consistent interpretation of parameter estimates, with locations farther away from the regional CBD and closer to roadways being more prone to (mostly residential) development (as expected). Again, the MESSP model offered the greatest computing-time savings benefits, but all three specifications yielded similar marginal effects estimates, showing how a focus on the spatial interactions and net (direct plus indirect) effects across observational units is more important than a focus on slope-parameter estimates, when properly analyzing spatial data.

Keywords: spatial autoregressive probit model, matrix exponential spatial specification, distance decay, Bayesian estimation, land use change

INTRODUCTION

Like many things in life, transportation involves spatial relationships. Whether one is investigating traffic counts, crash rates, vehicle ownership levels, or mode choices, a reflection

of spatial dependence is valuable, both for prediction and behavioral understanding. In the application of spatial statistics and spatial econometrics, weight matrices ($W = [w_{ij}]$) are crucial components; these represent the underlying spatial interdependence among proximate units , such as the simple inverse of network distances between traffic detectors, contiguity indicators of census tracts across a region, and who qualifies as a K-nearest neighbor within a social network.

The functional specification of appropriate weight matrices has long proven a controversial topic in spatial econometrics (as discussed in Anselin [1988] and Kostov [2010]). Nearly all weight matrices are specified a priori, simply as a function of distance or contiguity, raising the question of whether weight-matrix specification carries any important implications for interpretation of model results. In one of the literature's more unusual proposals, Cliff and Ord (1981) suggested combining an inverse-distance measure (or negative exponential) and the relative length of the common border between two spatial units: $w_{ij} = d_{ij}^{-a} \beta_{ij}^{b}$ (where d_{ij} is the distance between units i and j and β_{ij} as the proportion of the boundary of unit i shared by unit j). Another proposal, by Bodson and Peeters (1975), relies on a logistic function for relative levels of spatial interaction, with (a, b and c) parameters to be estimated via classical likelihood maximization: $w_{ij} = \sum_j \frac{a}{1 + b \cdot \exp(-c d_{ij})}$. These sorts of weight matrices were rarely used in practice due to estimation challenges and identification issues. In most applications, the weight matrix is more likely to be based on distance between units, or simply contiguity (Anselin 1988, Anselin 2002, LeSage and Pace 2009, Kostov 2010).

In practice, as noted earlier, weight matrices (and any associated parameters) are almost always assumed to be exogenous (see, e.g., Anselin [1988] and Anastasopoulos et al. [2010]), and

commonly rely either on distances between observational units in the data set, or their contiguity. By construction, all have zero diagonals (since the perfect correlation of a unit's error term with itself is implicit in the statistical model), and most are row-standardized (such that each row's values sum to one, in order to ensure that the largest eigenvalue of the weight matrix is 1 and the lowest eigenvalue is -1, facilitating maximum likelihood estimation or draws of the autocorrelation parameter ρ in a Bayesian scheme [LeSage and Pace 2009]). Row-standardization does not change the relative weight neighbors exert on other units, but it does alter the magnitude of the collective impact, which is then scaled appropriately by the multiplicative spatial autocorrelation parameter ρ (Parker 2011).

The focus of this paper is to explore how endogenous W's structure impacts discrete-response prediction. There have been several papers attempting to examine weight matrices' impact on model inference. Mizruchi and Neuman (2008) found that strongly connected (or highly dense/non-sparse) weight matrices tend to cause downward bias in the maximum-likelihood estimates of SAR's spatial autocorrelation parameter, ρ. Farber et al. (2008) found similar results when simulating how network topology influences spatial autocorrelation. However, their work compares specifications solely based on estimates of parameters, such as ρ, and these are subject to change under different assumptions of the underlying spatial data generating process. Thus, they should not be used as the yardstick for model comparisons. In addition, dense spatial matrices are rarely used in empirical studies for asymptotic theory to hold and out of consideration of behavioral realism (e.g., distant geographic units tend to exert little effect on one another), limiting the validity of these endeavors. In addition, both Mizruchi and Neuman

(2008) and Farber et al. (2008) treated the weight matrices (W) as fixed/exogenous, an assumption that is relaxed here.

LeSage and Pace's (2011) very recent study compares estimates and inferences of SAR models and spatial Durbin models (SDMs) for continuous response (with y equaling the share of adults voting across counties), with different spatial weight matrices achieved by varying the number of nearest neighbors (m) or the distance decay parameter (r). As LeSage and Pace (2011) observe, many spatial econometric papers focus on slope estimates, β, to represent the magnitude of covariate effects, thereby overlooking the important indirect effects that emerge through spatial associations. In other words, the true marginal effect of a covariate x_i – expressed as $\frac{\partial y}{\partial x_r}$ – requires far more than its associated β_r. Instead, it is the totaled set of direct and indirect effects that characterizes each attribute's (e.g., ground slope or distance to the nearest highway) effect on the response variable (as described in the Methodology section of this paper). As LeSage and Pace (2011) show, direct and indirect effects of SAR and SDM models are very stable/similar across different choices of m and r. Their results dispel the "myth" that it is useful or necessary to fine-tune one's spatial weight matrix (by altering m and r for example) because estimates and inferences are sensitive to moderate changes in these specifications. However, LeSage and Pace (2011) did note that significantly different matrix choices can indeed impact such inferences in meaningful ways.

Most of the relevant research relies on fixed weight structure, rather than allowing the data to explain their degree of connectedness, via the use of some parameters in the weight values' specification. This somewhat naïve and arbitrary approach to pre-assigning weights can call into

question the value and validity of spatial econometric specifications and results. In many cases, the dependence structure itself is a subject of interest. Kakamu (2005) suggested that fairly simple structures may mask spatial decay patterns. But his model focused on continuous response, rather than discrete response (e.g., land development, mode choice and other variables common to transportation studies). Thus*,* the long-pondered question remains: *what type or types of weight matrix should be used in spatial econometric applications* (Anselin 1988)?

This paper takes off from LeSage and Pace (2011) by comparing the impacts of the weight-matrix specification in a standard SAR model (with a pre-determined/fixed weight matrix), a SAR model with an endogenously determined distance-decay parameter, and a matrix exponential spatial specification (MESS) – all within a binary-response setting. This paper aims to answer the question left in LeSage and Pace (2011) on the model inference impacts of spatial specifications with more distinct weight structures, and to provide evidence/guidance on weight matrix choice in modeling binary responses. The models were run using simulated data sets to match each model specification as well as year-2008 Austin land use data. Details about the sampling schemes are discussed below, followed by a description of the data sets used, modeling results, and paper conclusions.

METHODOLOGY

Use of Markov chain Monte Carlo (MCMC) sampling in a Bayesian estimation setting allows analysts to avoid the often impossibly complex computation of posterior distributions into simpler problems using parameters' conditional distributions, thus greatly facilitating model estimation and inferences (LeSage and Pace 2009, Gelman et al. 2004). Due to the discrete

nature of the responses being analyzed, the three models described and applied here rely on

Bayesian MCMC estimation techniques. The SAR binary Probit (SARP) specification follows

Chapter 10 of LeSage and Pace's (2009) book directly, and the SARPα model is an extension of

this. The MESS Probit (MESSP) model, described below, is an extension of the continuous

MESS model presented in their Chapter 9 (LeSage and Pace 2009).

The MESS was first introduced by Pace and LeSage in 2000. This specification enjoys an

important computational advantage over SAR approaches: it eliminates computation of the

likelihood function's log-determinant term, along with the spatial dependence parameter's

conditional posterior distribution (typically needed in the Bayesian estimation process). By

contrast, implementation of SAR models is impeded due to formidable computing efforts

involved in computing the log-determinant during Markov chain Monte Carlo (MCMC)

sampling (Authors et al. 2011). LeSage and Pace (2004) report MESS estimation to run

approximately 6 times faster than conventional SAR models in the MCMC paradigm for a

continuous response. LeSage and Pace (2004) adapted their MESS model to accommodate

binary and censored response data, and introduced hyperparameters to control the number of

neighbors selected and distance decay over space.

Bayesian methods are used to estimate parameters for the three binary-response model

specifications being compared here. This technique decomposes the complex estimation task into

much simpler conditional distributions of parameters and offers much specification flexibility

(Koop 2003). LeSage and Pace (2004) had also discussed Bayesian estimation for MESS Tobit

and probit models. As they point out, prior information regarding regression coefficients (β) is

unlikely to exert much influence on parameter estimates in large samples (typically available in transportation and land use contexts), but priors imposed on parameters in the weight structure could have a more noticeable influence since these dimensions of a spatial model retain much of their influences even in large samples.

Specification of the SARP Model (with Fixed Weight Matrix)

The SARP model takes the form: $y^* = \rho W y^* + X\beta + \varepsilon$ (LeSage and Pace, 2009), where y^* is the unobserved response variable with a positive value leading to a $y = 1$ outcome, and zero otherwise. ρ is the spatial autocorrelation coefficient characterizing the strength of spatial association in response values (after controlling for X factors), and W is an n by n row-standardized version of an initial weight matrix D, as described below. X is an n by K covariate matrix, with n denoting the number of observational units (e.g., parcels over space) under study and K the dimension of the parameter vector β. The error term, ε, is assumed to have an iid normal distribution: $\varepsilon \sim N(0, \sigma^2 I_n)$. To ensure identification (as present in any latent-response model), the homoscedastic error term is set to unity.

A qth nearest-neighbor setup is used here to define this exogenous matrix, where elements $D_{ij} = \frac{1}{d_{ij}}$, if $d_{ij} < d_{\max}$, and $D_{ij} = 0$, if $d_{ij} \geq d_{\max}$, with $d_{\max}$ being the distance of the qth nearest neighbor. This popular approach reflects the impacts of distance (rather than simple contiguity) while keeping the weight matrix reasonably sparse (with many zero-valued cells) and preserving data point boundaries. By construction, the diagonal elements of W are zeros. Row standardization (where each row's elements sum to 1) ensures that the largest and smallest eigenvalues are 1 and -1, respectively, facilitating ρ draws (Horn and Johnson 1993, LeSage and

Pace 2009).   MCMC sampling strategies for the posterior distributions of the conventional

SARP model's parameters are provided in LeSage and Pace (2009).

Specification of the SARPα Model

In empirical studies, the focus may not rest solely on the spatial interaction reflected by the

parameter ρ, but also on the spatial decay patterns evident in the matrix D. Kakamu (2005)

proposed and the estimated an exponent (α) on the distance values (d_{ij}) in a standard SAR model

(for continuous response). To the authors' knowledge, this work is the first to extend this idea

extension to the discrete-response setting.

The proposed (SARPα) model retains the basic structure of the SARP model: $y^* = \rho W_\alpha y^* +$

$X\beta + \varepsilon$, as described above. The only difference lies in the spatial weight matrix used. Here, W_α

is a function of the distance decay parameter α, such that $D_{ij} = d_{ij}^\alpha$, if $d_{ij} < d_{\max}$, and $D_{ij} = 0$

otherwise. W_α is the row-standardized version of D. The prior distributions for unknown

parameters $\beta, \alpha,$ and ρ are assumed independent, such that $\pi(\beta, \alpha, \rho) = \pi(\beta) \cdot \pi(\alpha) \cdot \pi(\rho)$.

Under assumptions of a normal prior for β and uniform priors for α and ρ, the conditional

posterior for the parameter β is a multivariate normal distribution:

$$p(\beta|\alpha, \rho, y^*) \sim MVN(c^*, T^*)$$

$$c^* = T^*[X^{'}(I - \rho W_\alpha)y^* + T^{-1}c]$$

$$T^* = (X^{'}X + T^{-1})^{-1}$$

where I is an n by n identity matrix, and c and T are the prior mean and variance for β. A rather diffuse prior can be used for β (as represented by a c of zero and a fairly large value for T), and is typically used for large spatial data sets (LeSage and Pace 2009). In contrast, the choice of priors for α and ρ tends to have a greater impact on the estimations of the whole model due to the positions they have in the weight structure $(I - \rho W_\alpha)$. The parameter ρ denotes the degree of spatial autocorrelation whereas α serves as the distance decay parameter. The posterior distribution of parameter ρ can be approximated by application of the Metropolis-Hastings algorithm or univariate integration (Gelman et al. 2004) and is expressed as follows:

$$p(\rho|\alpha,\beta,y^*) \propto |I_n - \rho W_\alpha| \cdot \exp\left[-\frac{1}{2}(Sy^* - X\beta)^{'}(Sy^* - X\beta)\right], \text{with } S = I_n - \rho W_\alpha.$$

Under the assumption of an uniform prior for the parameter α, its posterior takes the same form as that of ρ.

The last step is to update the latent response y^*, which follows a truncated multivariate normal distribution: $TMVN(S^{-1}X\beta, [S^{'}S]^{-1})$ with $= I_n - \rho W_\alpha$. Geweke (1991)'s m-step Gibbs sampler was used to accomplish this task.

Matrix Exponential Spatial Specification Binary Probit (MESSP)

For model comparability, a similar q^{th}-nearest-neighbor row-standardized weight matrix W is used in the MESSP model. As noted in the Introduction, a key benefit of the MESS approach is that the analysis avoids computation of the logarithmic Jacobian term (i.e., $\ln[|I_n - \rho W_\alpha|]$). A MESS model assumes a matrix exponential decay pattern for neighbors, with diagonal weights

exceeding 1. The MESS model for a *continuous* response vector (y) takes the following form (LeSage and Pace, 2007):

$$Sy = X\beta + \varepsilon$$

$$y = S^{-1}X\beta + S^{-1}\varepsilon$$

where X, β and ε are as defined earlier. The matrix S relies on the matrix exponential operation, whose inverse can be expressed as follows:

$$S^{-1}(\alpha) = e^{-\alpha W} = I - \alpha W + \frac{\alpha^2}{2}W^2 + \cdots + \frac{(-\alpha)^t}{t}W^t + \cdots$$

where W is as defined earlier and α is a distance decay parameter, with $\alpha < 0$ when positive spatial autocorrelation exists (the standard condition). By construction, $W_{ii} = 0$, $W_{ij} > 0$ if i and j are neighbors, and $W_{ij} = 0$ otherwise. W^k is a weight matrix based on k^{th}-order neighbors. An α value close to zero indicates the absence of spatial autocorrelation in the underlying data generating process (since the matrix exponential of a zero matrix will result in an identity matrix for $S(\alpha)$).

As desired, the matrix exponential form ensures less influence for higher-order/more-distant neighbors (LeSage and Pace 2004). A positive spatial autocorrelation, reflected by a positive ρ (in SAR models) or a negative α term (in the MESS model), is more common than negative spatial autocorrelation and is expected in transport and land use applications (since similar behaviors and preferences can "spill over" to nearby sites and decision makers, and unobserved variables causing one section of roadway, person, or parcel to behave in a certain way are likely to influence nearby units similarly).

For the binary setup, the conditional posteriors take the form[1]:

$$p(\beta|\alpha, y^*, \sigma = 1), p(\alpha|\beta, y^*, \sigma = 1), \text{ and } p(y^*|\alpha, \beta, \sigma = 1, y).$$

Here, the latent response y^* is treated as unknown parameter to be estimated, a procedure known as data augmentation (Robert and Casella, 2004). It follows a truncated normal distribution: $y^* \sim TMVN\{S^{-1}X\beta, (S'S)^{-1}\}$. Geweke's m-step procedure is used to draw the n-variate latent response y^*. A multivariate normal prior is assigned for β with fairly large variance and zero means, and arbitrary uniform prior for α. The error term is forced to unity ($\sigma = 1$) for parameter identification. The kernel posterior is then expressed as: $p(\beta, \alpha|y^*) \propto \exp[-\frac{1}{2}(Sy^* - X\beta)'(Sy^* - X\beta)] \cdot \pi(\alpha) \cdot \pi(\beta)$. When computing the posterior of the parameter β conditional on α and y^*, the prior $\pi(\alpha)$ can be regarded as a constant thereby omitted from the kernel posterior. Using a procedure termed "completing the square" (LeSage and Pace 2009), the conditional posterior can be written as:

$$p(\beta|\alpha, y^*) \sim MVN(c^*, T^*)$$

$$c^* = T^*[X' \cdot S \cdot y^* + cT^{-1}]$$

$$T^* = (X'X + T^{-1})^{-1}$$

The conditional posterior for the parameter α is expressed as:

$$p(\alpha|\beta, y^*) \propto [y^{*'}S'MSy^*]^{-(n-k)/2} \cdot \pi(\alpha)$$

where k is the number of parameter to be estimated and M is an idempotent matrix expressed as $(I_n - \frac{X'X}{n})$. The Metropolis-Hastings algorithm was used to draw this parameter.

---

[1] Notes that the binary MESS model takes the form: $Sy^* = X\beta + \varepsilon$, where $\varepsilon \sim N(0,1)$ and $y = 1$, if $y^* > 0$ ($y = 0$ otherwise).

Computing Direct and Indirect Effects of Covariates across Space

In a spatial context, the expectation of the marginal effects (on all y values) of perturbing the kth

covariate's value (at any of the n locations) should be represented as an n by n matrix:

$$\frac{\partial E(y|X)}{\partial x_r} = (I_n - \rho W)^{-1} I_n \beta_r$$

as shown in LeSage and Pace (2011). The cell in the ith row and j^{th} column denotes the change in

the dependent variable (y) at location i in response to a one-unit change in the kth covariate at

location j. In the binary-response version of this setup for the SARP and SARPα models, the

marginal effects also are represented by an n by n matrix (LeSage and Pace 2009):

$$\frac{\partial E(y|X)}{\partial x_r} = \phi[(I_n - \rho W)^{-1} I_n \bar{x}_r \beta_r] \odot (I_n - \rho W)^{-1} I_n \beta_r$$

where ϕ is the standard normal distribution's density function, and the i,j cell value indicates the

change in the *probability* that $y = 1$ is observed at the ith geographic unit corresponding to a unit

change in the covariate x_r at the jth location. The operator $\odot$ denotes element-by-element

multiplication between two matrices.

Direct effects lie on the diagonal of these matrices, and indirect effects lie off the diagonals. The

average total effect of a covariate x_k is the mean of all the rows' sums. And the average *direct*

effect is the mean of just the *diagonal* elements. The difference between the averaged total effect

and direct effect yields what is called the averaged indirect effect (LeSage and Pace 2011).

Model Fit Statistics

To aid in evaluation of the distinct models, deviance information criteria (DIC) values – which are highly regarded Bayesian goodness-of-fit statistics (Carlin and Louis 2000) – were computed, along with other measures of model performance, such as log-likelihood values evaluated at the mean values of all parameter estimates. The DIC statistic is computed as: $DIC = p_D + \overline{D}$, where p_D is a penalizing term and $\overline{D}$ is the average of all posterior likelihoods across sampling iterations. p_D, also known as the number of effective parameters, is defined as $\overline{D} - D(\overline{\theta})$, with $D(\overline{\theta})$ computed as the posterior likelihood at the mean posterior of parameters. The deviance term D is defined as: $D(\theta) = -2\log[L(y|\theta)] + C$, where y is observed data, θ is the set of unknown parameters (to be estimated), and $L(y|\theta)$ denotes the likelihood function. C is a constant that cancels out when comparing DIC values among different models. Models with smaller DIC value should be preferred over models with larger DIC values (Carlin and Louis 2000).

In the case of a continuous SAR model, the likelihood can be written as: $L(y|\beta,\sigma,\rho) = (2\pi\sigma^2)^{-n/2} \cdot |A| \cdot \exp[-\frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta)]$, as described in LeSage and Pace (2009). For the binary SARP, SARPα and MESSP models, the sampled latent variable y^* is used in the likelihood computation.

DATA SETS FOR MODEL EVALUATIONS

Two styles of data were used to compare model results here, a set of three simulated data sets (one for each model type), and a land use data set from Austin, Texas. The first set (of simulated data) were generated with known/assumed parameter values and sample size $n = 1,000$, in order

to evaluate each model's proper recovery of parameter values[2]. Rather than relying on purely random spatial coordinates to define observational units in this simulated data set, these data points' coordinates were randomly drawn from the centroids of year-2003 undeveloped parcels, thus offering a somewhat realistic geographic setting. Three sets of data values (each model specific), including 4 covariates (X's) and error terms (ε's) (in each model), were generated from iid standard normal distributions. In addition to a simulated set of data (for model-estimation testing purposes), the set of Texas parcel data applied in the work by Authors et al. (2011) was used here. Only 42,589 of the undeveloped parcels across the County of Travis that were privately held in 2003 and had the potential to develop but showed no signs of subdivision (or merge) through 2008 were used here. Their year-2008 land use categories were collapsed into "developed" (residential, commercial, and industrial land use types) and "undeveloped" categories. And their land use-change status (remaining as undeveloped, $y = 0$, or changing to developed status, $y = 1$) were modeled as functions of eight regressors: network distance (in miles) to Austin's central business district (*DistCBD*), Euclidean distances (in miles) to the nearest minor arterial, major arterial and freeway (*DistMnrArt*, *DistMajArt*, and *DistFwy*), soil slope (*Slope*) in percent rise, a logsum measure of accessibility (*Access*), parcel size in acres (*Area*), and ratio between parcel perimeter and size (*PeriArea*) in 1/ft. These are described in more detail below.

Parcels with the potential to develop exclude parks, preserved land, greenbelts and water. Among the sample's 42,589 parcels, 64.8 percent remained undeveloped during the 5-year period. Among those that developed by 2008, the vast majority (98.7%) developed into residential (either single-family or multi-family) uses.

---

[2] A corresponding model for the continuous response y^* was also specified and estimated first, to provide an initial check on the estimation code.

Covariates for Binary Models of Land Use Change

The two land use types described above serve as response categories for land use change from an undeveloped state in 2003. A variety of attributes or "covariates" are expected to influence the various likelihoods of development, including soil slope and parcel geometry, distance to the region's CBD, distances to various roadway types and regional accessibility.

Here, slopes first took the form of a raster layer (at 10 m resolution) in percent. The *Slope* attribute was averaged over each parcel's spatial extent, to use in the regression models. Distance to the region's CBD regularly is a powerful covariate in models of land value and land use (Zhou and Author 2008, Haider and Miller 2000, Srour et al. 2002). Here, this attribute was computed as the shortest-path network distance from each parcel's centroid to the Texas State Capitol, based on Travis County's 2005 coded network. Distances to the nearest freeway, major arterial and minor arterial can also play important roles in a site's viability for development (Srour et al. 2002, Iacono et al. 2008), with access of interest to most developers. The expected-maximum-utility or logsum measure obtained from discrete choice models of destination choice can account for the behavioral nature of such choices (Niemeier 1997), so such a measure was calibrated by running a conditional logit model of destination TAZ choice for all 13,942 trips in the 2006 Austin Travel Survey. Detailed descriptions of these covariates can be found in Authors et al.'s (2011) work on geographically weighted multinomial logit regression (GWR MNL).

DATA ANALYSIS AND MODEL COMPARISONS

The three model specifications were compared – first using simulated data and then using the Travis County land use data, as described earlier.

Table 1. SARP Model Estimates Using Simulated Data.

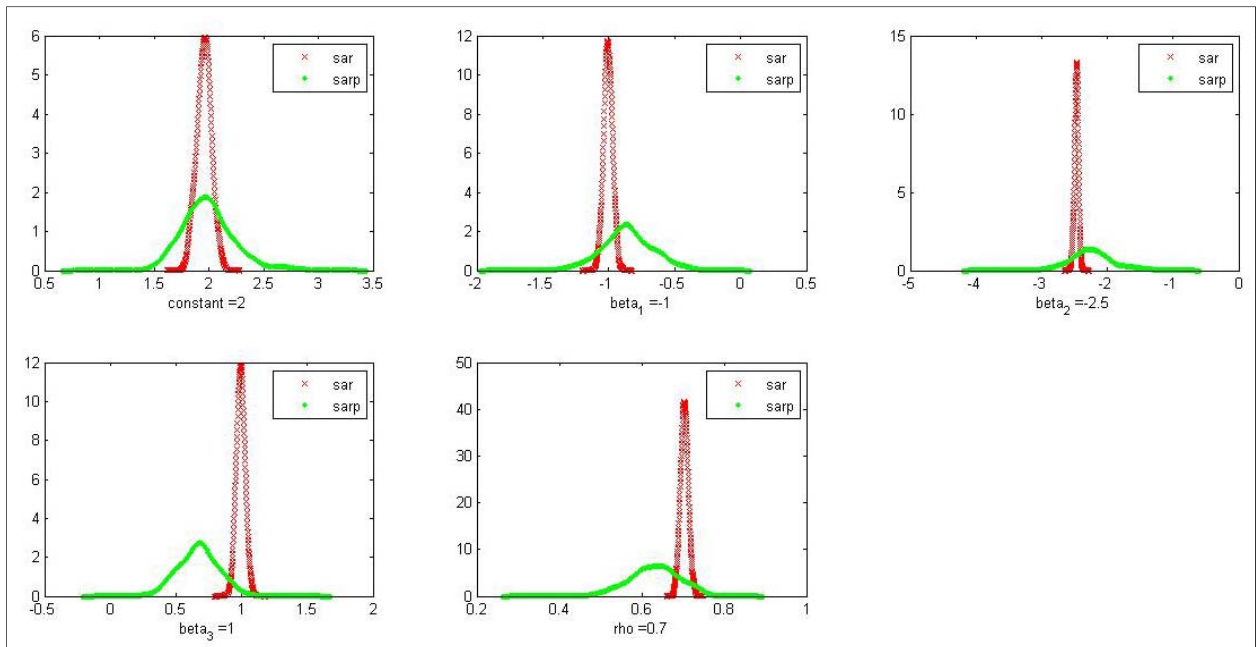| Parameter	SAR					SARP							
	Mean	Monte Carlo Error	Pseudo T-stat.	Avg. Dir. Effect	Avg. Indir. Effect	Mean	Monte Carlo Error	Pseudo T-stat.	MSE	Avg. Dir. Effect	Avg. Dir. Effect	Avg. Indir. Effect	Avg. Dir. Effect
$\beta_0=2$	1.955	0.067	29.412	-	-	1.986	0.238	8.337	0.057	-	-	-	-
$\beta_1=-1$	-1.005	0.034	-29.894	-1.141	-2.253	-0.863	0.191	-4.506	0.056	-0.029	-0.029	-0.045	-0.451
$\beta_2=-2.5$	-2.471	0.031	-80.547	-2.804	-5.537	-2.259	0.338	-6.675	0.173	-0.075	-0.075	-0.118	-1.127
$\beta_3=1$	0.993	0.033	29.904	1.126	2.224	0.669	0.156	4.292	0.134	0.022	0.022	0.035	0.452
$\rho=0.7$	0.704	0.009	76.016	-	-	0.629	0.060	10.507	0.009	-	-	-	-
$\sigma=1$	1.002	0.045	22.334	-	-	-	-	-	-	-	-	-	-
# of Obs.	1000					1000							
# of Draws	2000					2000							
Total Time of Sampling*	6 sec.					1,340 sec.							
R^2	0.901					-							
Adjusted R^2	0.900					-							
Log-Lik at Means	-6287					-5362							
DIC	12,708					11,198							
p_D	67.3					237.0							
$\bar{D}$	12,640					10,960							
$D(\bar{\theta})$	12,573					10,723							

Figure 1. Posterior Density of Parameters for SARP and SAR (latent response y^*) Models Using Simulated Data
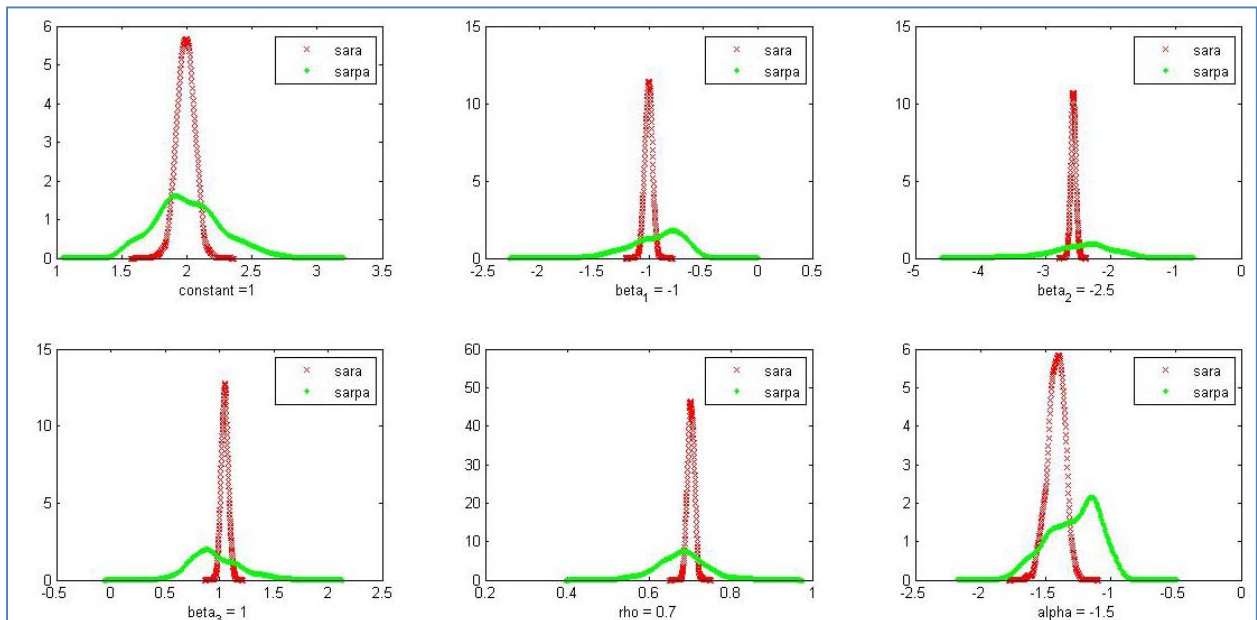

Figure 2. Posterior Density of Parameters for SARPα and SARα (latent response y^*) Models Using Simulated Data.
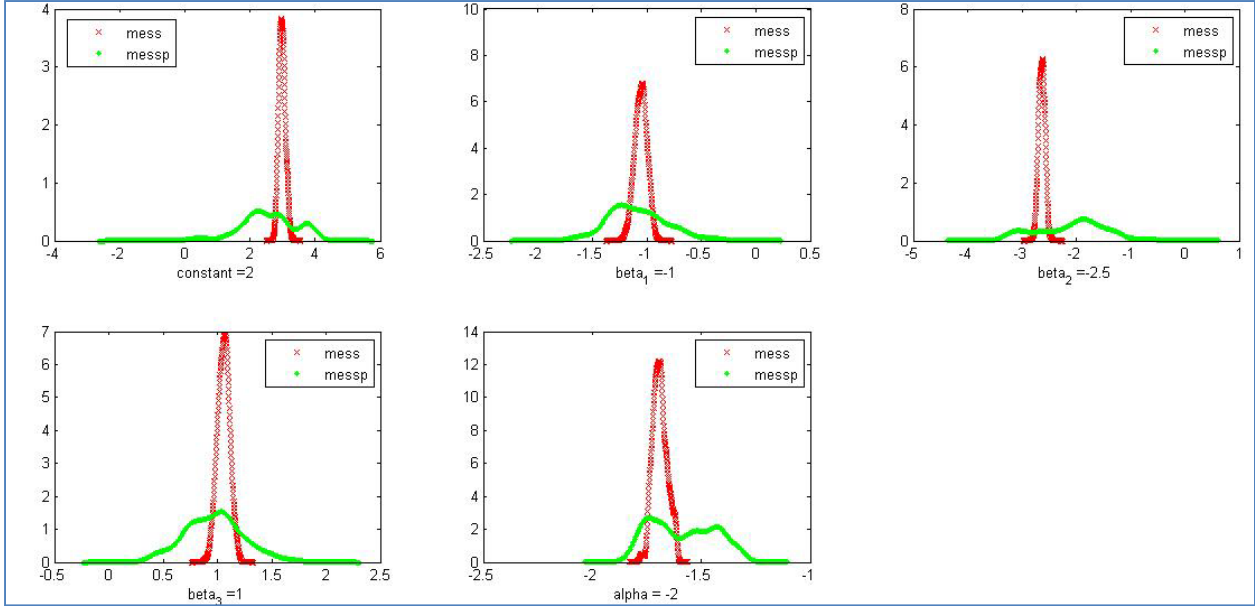
Figure 3. Posterior Density of Parameters for MESSP and MESS (latent response y^*) Models Using Simulated Data.

Numeric Results for Simulated Data Sets

Table 1 summarizes the results for the SARP model and the associated continuous SAR model (for y^*) using the simulated data set. The SAR model generated parameter estimates much closer to their true values than the SARP model, as reflected by the average parameter draws (after the 500-draw burn-in period [which allowed draws for the posterior distribution of parameter estimates to stabilize]), and shown in Figure 1. Results show how access to the latent response, y^*, reduces uncertainty in those mean estimates (as shown in Figures 1, 2 and 3), as expected, as along with variance in the averaged direct and indirect effects. Sampling time was remarkably (but not prohibitively) higher for Bayesian estimation of the SARP model, relative to its continuous counterpart (as expected), for the same 2000 sampling draws. Such computing time differences are a result of the burden of drawing the latent response, requiring the n x n inverse term $(I_n - \rho W)^{-1}$.

Moran's I statistic (as discussed in Anselin [1988] and Banerjee et al. [2004] for example) can be used to characterize any spatial autocorrelation remaining in the models' residual terms. In a binary response setting, the residual is defined as the difference between the observed response and predicted response, which will result in 1's, 0's, and -1's. Here, Moran's I was computed for the real land use data set, to get a sense of how much spatial autocorrelation may remain even after applying a spatial model specification. To test the statistical significance of Moran's I, its value was divided by its variance, resulting in relatively low Z-scores (or t-statistics) in Tables 4 through 6, suggesting no statistically significant spatial autocorrelation.

Table 2. SARPα Model Estimates Using Simulated Data.

Parameter	SAR with distance decay parameter					SARPα							
	Mean	Monte Carlo Error	Pseudo T-stat.	Avg. Dir. Effect	Avg. Indir. Effect	Mean	Monte Carlo Error	Pseudo T-stat.	MSE	Avg. Dir. Effect	Avg. Indir. Effect	True Avg. Dir. Effect	True Avg. Indir. Effect
$\beta_0=2$	1.990	0.068	29.376	-	-	2.011	0.253	7.947	0.064	-	-	-	-
$\beta_1 = -1$	-1.003	0.034	-29.286	-1.277	-2.101	-0.935	0.243	-3.846	0.063	-0.459	-0.707	-0.509	-0.820
$\beta_2 = -2.5$	-2.574	0.037	-69.653	-3.277	-5.392	-2.461	0.471	-5.227	0.223	-1.207	-1.862	-1.271	-2.051
$\beta_3 = 1$	1.049	0.032	33.041	1.336	2.198	0.961	0.227	4.225	0.053	0.472	0.727	0.509	0.821
$\rho =0.7$	0.703	0.008	85.720	-	-	0.680	0.058	11.729	0.004	-	-	-	-
$\alpha = -1.5$	-1.421	0.065	-21.723	-	-	-1.373	0.194	-7.076	0.054	-	-	-	-
$\sigma=1$	2.968	0.050	59.007	-	-	1.000	-	-	-	-	-	-	-
# of Obs.	1000					1000							
# of Draws	2000					2000							
Total Time of Sampling*	8,398 sec.					9,855 sec.							
R^2	0.887					-							
Adjusted R^2	0.887					-							
Log-Lik at Means	-3099.7					-1036.3							
DIC	3124.0					3924.8							
p_D	12.1					926.1							
$\bar{D}$	3111.9					2998.7							
$D(\bar{\theta})$	3099.7					2072.6							

As shown in Table 2, the SARPα model's inclusion of a distance decay parameter (α) does improve the accuracy of parameter estimates to their true values to some extent. Not surprisingly, with the new weight-matrix structure used in Table 4's model specification, all three covariates' marginal effects seem to increase, as compared to those estimated in the SARP model. In addition, it took around 2.74 hours (on an Intel Core i5 2.27GHz laptop with 4.00 GB RAM) to complete the 2000 draws for each of the six parameters in the SARPα model for simulated data (with a sample size of 1000), 7 times the computing time needed for the SARP specifications, which shows how adding a distance decay parameter to the weight structure can require substantially more computing time and efforts. One clear advantage of the MESSP model is its fast estimation: here it required 42% and 92% less computing time than SARP and SARPα models, respectively.

Table 3. MESSP Model Estimates Using Simulated Data.

Parameter	MESS					MESSP						True Avg. Dir. Effect	True Avg. Indir. Effect
	Mean	Monte Carlo Error	Pseudo T-stat.	Avg. Dir. Effect	Avg. Indir. Effect	Mean	Monte Carlo Error	Pseudo T-stat.	MSE	Avg. Dir. Effect	Avg. Indir. Effect		
$\beta_0 = 2$	2.984	0.099	30.075	-	-	2.549	0.843	3.025	1.011	-	-	-	-
$\beta_1 = -1$	-1.056	0.057	-18.628	-1.300	-4.423	-1.090	0.268	-4.069	0.080	-0.307	-0.080	-0.541	-2.406
$\beta_2 = -2.5$	-2.635	0.058	-45.200	-3.244	-11.041	-2.093	0.633	-3.309	0.565	-0.773	-0.202	-1.342	-6.003
$\beta_3 = 1$	1.056	0.055	19.172	1.300	4.426	0.968	0.268	3.608	0.073	0.258	0.067	0.541	2.405
$\alpha = -2$	-1.690	0.033	-51.789	-	-	-1.188	0.146	-8.155	0.681	-	-	-	-
$\sigma = 1$	3.258	0.146	22.395	-	-	1.000	-	-	-	-	-	-	-
# of Obs.	1000					1000							
# of Draws	2000					2000							
Total Time of Sampling*	4 sec.					832 sec.							
R^2	0.9287					-							
Adjusted R^2	0.9285					-							
Log-Lik at Means	-2159.4					-1792.4							
DIC	4319.2					5978.3							
p_D	0.2					1196.8							
$\bar{D}$	4319.04					4781.5							
$D(\bar{\theta})$	4318.9					3584.7							

As shown in Table 3, the continuous-response (y^*) and discrete-response (y) parameter estimates for the MESS-based specifications differ noticeably from their true values. A plausible cause is the assumption of a matrix-exponential decay pattern, which is conceptually and mathematically distinct from the standard decay patterns assumed in the SARP[3] and SARPα models.

In sum, all three specifications managed to recover their underlying (simulated-data) parameter values, using Bayesian estimation methods. As expected, the continuous model structures, which lacked any latent structure, offered far tighter estimation of slope and autocorrelation or distance-decay parameters, as expected. Interestingly, adding the distance decay parameter α did not much alter estimation and inference of the spatial autocorrelation term (denoted as ρ), with the SARP and SARPα models producing similar mean values and variance of sampling draws and 95% intervals for ρ. The DIC values and marginal effects are not directly comparable across these distinct data sets (since y^* and y values generated differ, due to the different spatial structures), though it is interesting to see how much they can vary among seemingly similar model specifications. Run times are meaningful for comparison, with the α term's sampling loop resulting in a substantial increase, and the MESSP specification dominating, thanks to its avoidance of any log-determinant calculations. Model comparisons using a single, actual data set offer another opportunity for model comparisons, as discussed in the next section.

Estimation Results using Austin's Land Use Data

These three model specifications were evaluated using Austin's land use data set. To moderate estimation times, a sample of 1000 parcels was randomly drawn from the 42,589 undeveloped

---

[3] The SARP specification also relies on such a distance-decay weight structure, with α = -1.

parcels described earlier. Among these, 452 had been developed over the 5-year period (from

2003 to 2008), whereas the remaining 548 lay undeveloped through 2008.

Table 4. SARP Model Estimates for Austin Land Data.

Covariates	SARP				
	Mean	Monte Carlo Error	Pseudo T-stat	Avg. Dir. Effect	Avg. Indir. Effect
Constant	0.735	2.101	0.350	-	-
DistCBD	0.052	0.038	1.354	0.019	0.011
DistMin	-0.154	0.062	-2.473	-0.058	-0.033
DistMaj	0.454	0.171	2.660	0.171	0.097
DistFwy	0.193	0.074	2.606	0.073	0.041
Slope	-0.034	0.014	-2.473	-0.013	-0.007
Access	-0.446	0.704	-0.633	-0.168	-0.093
PeriArea	2.760	1.632	1.691	1.041	0.587
Area	-6.13E-07	2.11E-07	-2.901	-2.31E-07	-1.27E-07
ρ	0.377	0.060	6.299	-	-
# of Obs.	1,000				
# of Draws	2,500				
Total Time of Sampling	2,859 seconds				
Log Likelihood at Means	-624.8				
Likelihood Ratio	$487.5 > \chi^2_{0.05}(9)=16.9$				
DIC	4609.3				
p_D	1679.8				
$\bar{D}$	2929.5				
$D(\bar{\theta})$	1249.6				
Moran's *I*	-0.012				
Var(*I*)	0.023				
Z-score of I	0.52				

Notes: Moran's *I* is computed as $I = \frac{n \sum_i \sum_j w_{ij}(\varepsilon_i-\bar{\varepsilon})(\varepsilon_j-\bar{\varepsilon})}{(\sum_{i \neq j} w_{ij}) \sum_i (\varepsilon_i-\bar{\varepsilon})^2}$, where w_{ij} denotes the ij^{th} element of the weight

matrix $\mathbf{W}$, ε_i is the observed residual for site *i*, and $\bar{\varepsilon}$ indicates the average of the *n* ε_i's. This statistic's

variance is formulated as $\text{var}(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2}$, where $S_0 = \sum_{i \neq j} w_{ij}$, $S_1 = \frac{1}{2} \sum_{i \neq j} (w_{ij} +$

$w_{ji})^2$, and $S_2 = \sum_k (\sum_j w_{kj} + \sum_i w_{ik})^2$ (Banerjee et al. [2004]).

Table 4 summarizes parameter and marginal effects estimates, as well as goodness-of-fit statistics, for the SARP model. ρ is estimated to be a statistically and practically significant, with a mean value of 0.38. As parameter signs suggest, undeveloped parcels farther from Austin's CBD appear to have experienced more land development, ceteris paribus. But the influence is not statistically significant or practically significant, as reflected by the 95% interval's inclusion of zero. Distance-to-roadway terms have more statistically significant effects, as reflected by Table 4's fairly large pseudo-t-statistics. Proximity to minor arterials tends to increase the likelihood of land development, while proximities to major arterials and freeways have the opposite effect. A one-mile increase in distance to the nearest minor arterials is estimated, on average, to reduce the probability a parcel being developed by 0.06 whereas a one mile increase in proximity to major arterial and freeway tends to boost the average parcel's development probability by 0.17 and 0.07, respectively. The estimated direct and indirect effects for transportation access measures appear to be larger than those for soil slope and parcel geometries, suggesting that transport access tends to play a more important role in Austin's land development. Overall, however, the estimated direct and indirect effects are quite small, suggesting that far more than transport access and location centrality may be at play in parcel choices for new development in the Austin region. As expected, milder slopes tend to enhance the likelihood of a property's development, in statistically significant ways (but with only moderately practical significance: Average Direct Effect = -0.0126 probability per percent slope). Interestingly, a larger perimeter-to-area ratio is positively associated with development, with strong practical significance but only weak statistical significance. Parcel area also was estimated to have a positive effect, though its practical significance (as reflected by estimates of marginal effects) is at most moderate. Somewhat unexpectedly, the accessibility index was negatively related with

land use development in a practically significant way. Perhaps land values and development

constraints in less accessible locations are such that they attract more development. A more-

detailed, market-based model for development choices would be valuable here.

Table 5. SARPα Model Estimates for Austin Land Use Data.

Covariates	SARPα				
	Mean	Monte Carlo Error	Pseudo T-stat	Avg. Dir. Effect	Avg. Indir. Effect
Constant	-3.535	2.097	-1.686	-	-
DistCBD	0.151	0.037	4.052	0.033	0.005
DistMin	-0.218	0.060	-3.615	-0.084	-0.007
DistMaj	0.585	0.166	3.515	0.232	0.018
DistFwy	0.165	0.072	2.296	0.065	0.005
Slope	-0.040	0.015	-2.655	-0.016	-0.001
Access	0.924	0.704	1.312	0.023	0.027
PeriArea	5.006	1.956	2.560	2.037	0.160
Area	-7.85E-07	2.39E-07	-3.290	-3.10E-07	-2.39E-08
ρ	0.062	0.064	0.973	-	-
α	-4.216	2.969	-1.420	-	-
# of Obs.	1000				
# of Draws	2500				
Total Sampling Time	11,965 seconds				
Log Likelihood at Means	-1342.3				
Likelihood Ratio	$133.3 > \chi^2_{0.05}(10)=18.3$				
DIC	3517.5				
p_D	417.0				
$\bar{D}$	3101.5				
$D(\bar{\theta})$	2685.6				
Moran's I	0.115				
Var(I)	0.151				
Z-score of I	0.76				

Table 5's SARPα model results deliver a very minor spatial autocorrelation estimate (averaging

just 0.062), in large part because the specification allows for a sizable distance decay parameter

(mean α of -4.2). However, neither of these mean spatial-parameter estimates is far from zero, in

a statistical sense. The lost significance seems to have been picked up by the covariates, with

pseudo-*t* statistics for all covariates slightly larger than those delivered by the SARP model. Signs on all parameter estimates are consistent across Tables 4 and 5, except that the sign on the *Access* variable changed, though it is not quite statistically significant. Distances to roadways are shown to have statistically significant impact on land development decisions, with somewhat larger direct and indirect effects than those with slope and parcel size measures. Proximity to minor arterials tends to encourage development, while proximity to higher-class (and higher-speed) roadways is likely to dampen a parcel's developmental chances. The SARPα model yielded a lower (better) DIC value than the SARP specification, but required approximately 4.2 times the computing time. Moran's I is not statistically significant, as reflected by its small t-statistic (0.115/0.151=0.76).

Table 6. MESSP model estimates for Austin land use data.

Covariates	MESSP				
	Mean	Monte Carlo Error	Pseudo T-stat.	Avg. Dir. Effect	Avg. Indir. Effect
Constant	-2.356	3.190	-0.739	-	-
DistCBD	0.161	0.047	3.451	0.034	0.009
DistMin	-0.261	0.105	-2.482	-0.082	-0.013
DistMaj	0.661	0.216	3.059	0.226	0.035
DistFwy	0.128	0.100	1.284	0.064	0.010
Slope	-0.037	0.018	-2.041	-0.017	-0.003
Access	0.518	1.090	0.475	0.019	0.056
PeriArea	4.324	2.368	1.826	1.848	0.283
Area	-9.13E-07	2.71E-07	-3.36E+00	-3.40E-07	-5.12E-08
α	-2.480	0.608	-4.079	-	-
# of Obs.	1000				
# of Draws	2500				
Total Sampling Time	1,962 seconds				
Log-Lik at Means	-1429				
Likelihood Ratio	$662.4 > \chi^2_{0.05}(9)=16.9$				
DIC	3039				
p_D	91				
$\bar{D}$	2948				
$D(\bar{\theta})$	2857				

Moran's I	0.238
Var(I)	0.183
Z-score of I	1.30

Table 6's MESSP parameter estimates are very consistent with those of the SARPα model (including marginal effects). Interestingly, the MESSP model enjoys the lowest DIC among all models, as well as the greatest-time savings (running in just over 30 minutes on a standard desktop, suggesting it is a winner. More importantly, parameter estimation and inference of covariates do not vary much under different weight matrix choices, at least for the two more flexible specifications as reflected by the closely resembled mean of draws after burn-in and average direct effects in Tables 4 and 5.

CONCLUSIONS

This paper develops and then compares the specifications, parameter estimates, and implications of SARP, SARPα and MESSP models, using simulated data sets and actual land use data (at the parcel level, from Austin, Texas). Marginal effects of various covariates were computed reflecting the spatial relationships that exist in these three distinct model types, all to anticipate binary response over space. DIC values and other goodness-of-fit statistics were used to compare model applications, along with run times, recovery of true parameters (in the case of the simulated data sets), and estimates of marginal effects.

As expected, direct estimation of simulated latent response values (y^*'s) produced far more precise and accurate parameter estimates (when using simulated data, where parameter values were given) than their corresponding binary-response models (which operate under a blinder of discretized information). And, with the single land use data, the more flexible SARPα and

MESSP models out-performed the SARP model, though results (measured using marginal effects, since slope parameters are deceiving in a context with spatial autocorrelation) are largely consistent (in sign and magnitude) across all three specifications.

The MESSP out-performed the others, particularly in its computational complexity and associated run times, suggesting it may be of great value for further spatial econometric specifications, with large data sets and both continuous and discrete response types. The magnitudes of direct and indirect marginal effects appear quite small in the model for land use change, as compared to the SARP and SARPα model results, suggesting that the MESSP model may not be appropriate for this particular data set. Results of covariate effects appear consistent with Author et al.'s (2011) findings, which calibrated a GWR MNL model for the Austin data set (across six land use categories, with residential dominating here).

In conclusion, adding a distance decay parameter to a spatial autoregressive model is likely to boost model performance, as reflected by a lower DIC value, but can potentially present computing-time problems in large data sets. An MESSP structure appears easiest to work with and perform better in prediction, but all models, with their different spatial weighting structures, deliver – as LeSage and Pace (2009) predicted – the same behavioral implications for marginal effects. This is reassuring, though relatively few spatial modelers compute marginal effects (and may be distracted by slope coefficients).

This work operationalizes SARPα for the first timwith a demonstration to land use change data, which are important for long-term transportation planning applications, among other motivations. Clear extensions of this work include applications for multinomial-response settings (where

statistical identification of some parameters can pose problems, as seen in the dynamic

multinomial probit spatial models developed in Author et al. [2011]) and more rigorous

Bayesian analysis, with modified (reduced) autocorrelation between MCMC draws.

Additionally, the endogeneity issue should be considered because, in theory, an SAR

specification leads to a non-diagonal variance-covariance matrix for response values and

the error terms, and because an instrumental variable (IV) matrix should be used in order to

achieve consistent estimation. Endogeneity and identification issues remain central to debates

and sophisticated contributions on-going in the spatial econometrics field. A great deal of work

remains for spatial modelers, and transportation is a terrific context for such investigations.

ACKNOWLEDGEMENTS

REFERENCES

Anastasopoulos, R. Ch., Florax, R., Labi, S., and Karlaftis, M. G. (2010) Contracting in highway maintenance and rehabilitation: are spatial effects important? *Transportation Research Part A* 44: 136-146.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Anselin, L. (2002) Under the hood: issues in the specification and interpretation of spatial regression models. *Agricultural Economics*. Vol. 27(3): 247-267.

Bodson, P. and Peters, D. (1975) Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation: an application to a Belgian labor-demand function. *Environment and Planning A*. Vol.7: 455-472.

Bunch, D. S. (1991) Estimability in the multinomial probit model. *Transportation Research Part B*. Vol. 25 (1): 1-12.

Carlin, B. and Louis, T. (2000) Bayes and Empirical Bayes Methods for Data Analysis, 2nd Edition. Boca Raton: Chapman & Hall.

Cliff, A. and Ord. J. (1981) *Spatial Processes, Models and Applications*. London: Pion.

Dacey, M. (1968) *A Review of Measures of Contiguity for Two and K-Color Maps*. In Spatial Analysis: A Reader in Statistical Geography. Englewood Cliffs, N.J.: Prentice-Hall.

Franzese, R. J. and Hays, J. C. (2008) Interdependence in comparative politics: substance, theory, empirics, substance. *Comparative Political Studies*. Vol. 41(4/5): 742-780.

Farber, S., Páez, A., and Volz, E. (2008) Topology and dependency tests in spatial and network autoregressive models, *Geographical Analysis*, Vol. 41(2): 158-131.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2004) *Bayesian Data Analysis, 2nd Edition*. Boca Raton: Chapman & Hall/CRC.

Goetzke, F. and Andrade, P. M. (2010) Walkability as a summary measure in a spatially autoregressive mode choice model: an instrumental variable approach. In Paez, A., Gallo, J. L., Buliung, R. and Dall'Erba, S. *Progress in Spatial Analysis*. New York: Springer.

Higham, N. J. (2005) The Scaling and Squaring Method for the Matrix Exponential Revisited, *SIAM J. Matrix Anal. Appl.*, Vol. 26 (4): 1179–1193.

Haider, M., and Miller, E.J. Effects of Transportation Infrastructure and Location on Residential Real Estate Values. *Transportation Research Record* No. 1722: 1-8.

Horn, R. and Johnson, C. (1993) *Matrix Analysis*. Cambridge University Press, New York.

Iacono, M., Levinson, D., and El-Geneidy, A. (2008) Models of Transportation and Land Use Change: A Guide to the Territory. *Journal of Planning Literature*. Vol. 22: 323-340.

Kostov, P. (2010) Model boosting for spatial weighting matrix selection in spatial lag models, *Environment and Planning B: Planning and Design*. Vol.37(3): 533-549.

Koop, G. (2003) *Bayesian Econometrics*. West Sussex: Wiley.

Leenders, R. T. (2001) Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, Vol. 24 (1): 21-47.

LeSage, J.P and Pace, R.K. (2004) Using matrix exponential to estimate spatial probit/tobit models, in Getis, A., Lacambra, J. M., and Zoller, H. G. (2004) *Spatial Econometrics and Spatial Statistics*. New York: Palgrave Macmillan.

LeSage, J. P. and Pace, K.R. (2007) A matrix exponential spatial specification, *Journal of Econometrics*, Vol. 140: 190-214.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*, Boca Raton: CRC Press/Taylor & Francis.

LeSage, J.P and Pace, K.R. (2011) The biggest myth in spatial econometrics. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1725503. Accessed June 1st 2011.

Lin, H., Kang, S. L., Espey, M. and Allen, J. (2005) Modeling Urban Sprawl and Land Use Change in a Coastal Area- A Neural Network Approach. Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, Jul. 24-27, 2005.

Mizruchi, M.S., and Neuman, E.J. (2008) The effect of density on the levels of bias in the network autocorrelation model, *Social Networks*, Vol. 31(3): 190-200.

Niemeier, D. A. Accessibility: An Evaluation Using Consumer Welfare. *Transportation*, Vol. 24, 1997: 377-396.

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

Parker, D. A Brief Introduction to Spatial Econometrics. Lecture notes. Accessed July 7, 2011. URL: classweb.gmu.edu/dparker3/lumta_present/spat_econ_intro.pdf

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

Srour, I., Author, and Dunn, T. (2002) Accessibility Indices: A Connection to Residential Land Prices and Location Choice. *Transportation Research Record* No. 1805: 25-34.

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

Authors and Lemp, J. D. (2011) The Dynamic Spatial Multinomial Probit Model: Analysis of Land Use Change Using Parcel-Level Data. Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, D.C. URL: http://www.ce.utexas.edu/prof/kockelman/public_html/TRB08_DSOPmethods.pdf Accessed July 3 2011.

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

Authors (2011) Anticipating Land Use Change Using Geographically Weighted Regression Models for Discrete Response. Forthcoming in *Transportation Research Record*.

Whittle, P. (1954) On stationary process in the plane. *Biometrika*. Vol. 41: 434-449.

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