

# **Predicting the Distribution of Households and Employment: A Seemingly Unrelated Regression Model with Two Spatial Processes**

By

Bin (Brenda) Zhou  
Graduate Student Researcher  
The University of Texas at Austin  
6.508. Cockrell Jr. Hall  
Austin, TX 78712-1076  
brendazhou@mail.utexas.edu

and

Kara M. Kockelman  
(Corresponding author)  
Associate 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

The following paper is a pre-print and the final publication can be found in  
*Journal of Transport Geography* 17 (5):369-376, 2009.  
Presented at the 86th Annual Meeting of the Transportation Research Board,  
January 2007

## **ABSTRACT:**

Household and employment counts (by type) are key inputs to models of travel demand. For a variety of reasons, spatial dependence is very likely present in and across these counts. In order to identify the nature of these unobserved relationships, this study performs a series of Lagrange multiplier tests to confirm the co-existence of spatial lag and error processes within individual equations (6 household types and 3 employment categories). It then provides the first application of a feasible generalized spatial 3SLS estimation procedure for a seemingly unrelated regression (SUR) model of these equations.

In the resulting model of Austin, Texas data, local land use conditions offer substantial predictive power of households and jobs, and transportation access plays a role, as anticipated. The work demonstrates that SUR estimation of land use intensities from parcel-level data with two types of spatial dependence is feasible and meaningful. Coupled with an upstream model of land use type, this work offers the key inputs for travel demand analyses, with transportation system performance feedback.

*Key Words:* Spatial Econometrics, Seemingly Unrelated Regressions, Spatial Distribution of Households and Employment

## **INTRODUCTION**

As key inputs to travel demand models, zone-based household and employment counts (by type) are of interest to many. In-depth investigation of the factors that influence these variables and their spatial dependence is paramount for model predictions of job and population patterns.

Tobler's first law of geography states "everything is related to everything else, but near things are more related than distant things." (Tobler 1970 pp. 236) Essentially, the condition of one point or parcel or zone correlates with the conditions of its neighbors. Anselin (1988a) suggested that dependence is often present in cross-sectional data obtained using arbitrary delineation of spatial units (e.g., traffic analysis zone [TAZ] and census tract). Thus, zonal-based household and employment are likely to exhibit correlation, even after controlling for observable factors. Models without explicit treatment of these spatial dependencies may result in inappropriate inferences and conclusions.

Researchers in a variety of disciplines have investigated the interactions between jobs and population. Regional adjustment models (see, e.g. Steinnes and Fisher 1974; Mulligan et al. 1999) consist of two equations: one representing changes in population density (as a function of time-lagged population density and contemporaneous employment density, as well as other explanatory variables); the other tracking employment (as a function of time-lagged employment density, contemporaneous population density, and other explanatory variables). Carruthers and Vias (2005) examined whether population and employment growth are endogenously determined in the Rocky Mountain West. Their regional adjustment model suggests that there is a positive feedback at the county level, so that jobs follow people – in addition to people following jobs. However, other studies using different time lags and geographic scales (e.g. Clark and Murphy 1996, Henry et al. 1997, and Vias and Mulligan 1999), provide mixed evidence, and little work has considered the interactions between jobs and population at finer spatial scales, such as the TAZs used here. This work addresses that challenge, while allowing for spatial interactions and correlations among unobserved factors influencing household and employment intensities over space.

The following sections discuss highly related spatial econometric techniques, the data sets used here, the new spatial SUR model specification recommended for these data, model results and conclusions.

## **BACKGROUND**

Spatial dependence implies two-dimensional relationships across observations. Anselin and Bera (1998) summarized two strategies to address the restriction. One approach specifies the variogram as a continuous function of the distance between each pair observations. This is typically done in geostatistical studies (e.g. Cressie's [1993] Chapter 2). The other approach specifies a spatial stochastic process in which a random variable at a given location is related to its neighbors, using a weight matrix,  $\mathbf{W}$ . Each element of  $\mathbf{W}$ ,  $w_{ij}$ , is non-zero when observation  $i$  and  $j$  are "neighbors" (i.e., they impact one another's response, though they may be [and often are] non-contiguous). In this way, observed responses

and/or unobserved error components link to neighbors' contemporaneous responses.

The geostatistical approach is favored by mining engineers or geologists who seek to identify soil properties from a sample of known locations (Cressie 1993). It is seldom applied in studies when the data consist of counties or TAZs, because the continuous relation over space is hard to maintain. This study is in the line with the second strategy and considers spatial correlations among response variables and their error terms, while allowing cross-equation correlations in a system of linear regression equations.

## Two Spatial Processes

There are two ways to incorporate spatial dependence via a weight matrix. The first is known as a spatial lag model, and the latter is a spatial error model. The spatial lag model includes a lagged-response regressor, and is specified as  $\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\xi}$ , where  $\mathbf{y}$  is an  $n$  by 1 vector of the response variable,  $\rho$  is the spatial lag autoregressive coefficient,  $\mathbf{X}$  is an  $n$  by  $k$  matrix of explanatory variables,  $\boldsymbol{\beta}$  is a  $k$  by 1 vector of parameters to be estimated,  $\boldsymbol{\xi}$  is an  $n$  by 1 vector of independently and identically distributed error term. In contrast, the spatial error model specifies a spatial process for the error terms. This is often an autoregressive error process<sup>1</sup>, specified as follows:  $\boldsymbol{\varepsilon} = \lambda \mathbf{W} \boldsymbol{\varepsilon} + \boldsymbol{\xi}$ , where  $\lambda$  are spatial error autoregressive coefficient.

As Anselin (2001b, pp. 316) explains, a spatial lag is designed to reveal the “existence and strength of spatial interaction” while a spatial error seeks to correct for the “potentially biasing influences of the spatial autocorrelation”. A spatial lag model is appropriate when “spatial spillovers” exist across unit boundaries, or when the spatial scales for measurement and the true distance of interaction do not match. A spatial error model is best when the spatial dependence is associated with unobserved, spatially correlated factors. (Anselin and Bera, 1998)

Model specifications with spatial lag and spatial error allow for both processes, and therefore represent higher-order dependencies. An analyst's lack of strong prior beliefs regarding spatial relationships may require such flexibility (Anselin 2003a). Such a (single-equation) model may be defined as follows:

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} = \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \boldsymbol{\xi} \quad (1)$$

where  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are two  $n$  by  $n$  spatial weight matrices, and other terms are as previously defined.  $\mathbf{W}_1$  and  $\mathbf{W}_2$  were specified distinctly, in order to avoid identification problems. (See, e.g., Chapter 6 in Anselin [1980] as well as Anselin and Bera [1998].) Anselin (1988a) described single-equation models with both spatial processes in detail, and Case (1991, 1992) and Case et al. (1993) applied this formulation, for models of household rice demand, adoption of new technologies, as well as state government expenditures.

A reformulation of this model produces Eq. (2)'s reduced form, which emphasizes

---

<sup>1</sup> Two alternatives to the spatial autoregressive error process are a spatial moving average error process ( $\boldsymbol{\varepsilon} = \lambda \mathbf{W} \boldsymbol{\xi} + \boldsymbol{\xi}$ ) (Cliff and Ord 1981, Haining 1988) and a spatial error component process ( $\boldsymbol{\varepsilon} = \lambda \mathbf{W} \boldsymbol{\xi} + \boldsymbol{\psi}$ ) (Kelejian and Robinson, 1993, 1995).

the higher-order spatial process:

$$(\mathbf{I} - \lambda \mathbf{W}_2) (\mathbf{I} - \rho \mathbf{W}_1) \mathbf{y} = (\mathbf{I} - \lambda \mathbf{W}_2) \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\xi} \quad (2)$$

where  $\mathbf{I}$  is an  $n$  by  $n$  identity matrix.

The specification of the spatial weight matrix is rather arbitrary in practice. Many types of weight matrices have been suggested, and there is no agreement as to which form is most appropriate. (Anselin's [1988a] Chapter 3 provides a detailed discussion of this.) Anselin and Bera (1998) argue that a poorly specified weight matrix (i.e., one that differs from the true, data-generating process) results in the need for higher-order spatial process specifications, allowing the second weight matrix to represent some portion of remaining spatial dependence. Using Monte Carlo experiments, Florax and Rey (1995) demonstrated how weight-matrix misspecification can affect the power of tests for spatial dependence, along with parameter estimation, in the presence of small sample sizes. A more general model specification is preferable, to help avoid such issues.

In this study, four distinct weight matrices were specified in order to discern the most relevant ones. The first is a first-order contiguity matrix using queen criterion<sup>2</sup>. It was generated using GEODA<sup>3</sup> software (Anselin 2003b, 2005). The second weight matrix relies on the inverse of Euclidean distance (in miles) between TAZ centroids (rather than 0's and 1's, as in the first case), in order to reflect the decay of relationships with distance. The third and fourth matrices use the inverse of distance to the power of 1.5 and 2.0, respectively. These four weight matrices are defined as  $\mathbf{W}_{\text{con}}$ ,  $\mathbf{W}_{\text{dist1.0}}$ ,  $\mathbf{W}_{\text{dist1.5}}$ , and  $\mathbf{W}_{\text{dist2.0}}$ . All matrices were row-standardized (so that the row elements summed to one), in order to facilitate interpretation (Anselin 1988a). More specifically, a right-hand-side weight matrix represents a weighted average of the variable under consideration in neighboring units.

## Seemingly Unrelated Regression

In contrast to the widely used single-equation spatial specification, the model applied here recognizes dependencies emerging from three sources. First, response variables depend on neighboring unit responses, via a spatial lag component. Second, error terms are spatially correlated across observational units, via a weight matrix-based spatial error component. Third, these same error terms are correlated across equations, implying a seemingly unrelated regression (SUR) structure.

Anselin (1988a) derived likelihood functions for SURs that include either spatial error autocorrelation or spatially lagged response variables (but not both). Since then, a limited number of studies have applied spatial SUR in empirical situations. Rey and Montouri (1999) investigated U.S. regional income convergence using a spatial SUR with spatial error autocorrelation. Similarly, Gallo and Dall'erba (2003) addressed European regional GDP convergence with an emphasis on the existence of temporal and spatial heterogeneity. Lundberg (2005) investigated the spillovers of recreational and cultural services using a spatial SUR with spatially lagged response variables. However, to the

---

<sup>2</sup> Queen criterion defines neighbors as those that have either common boundaries or common corners.

<sup>3</sup> GEODA is free software developed by the University of Illinois at Urbana-Champaign's Spatial Analysis Laboratory (SAL).

authors' knowledge, no studies have applied a SUR model with both spatial error and lagged autocorrelation.

Kelejian and Prucha (2004) proposed an estimation approach for this triple-dependency in a simultaneous equations model (SEM), where response variables served as covariates in other equations. Their three-stage least-squares (3SLS) estimation approach derives from their earlier work, using a generalized method of moments (Kelejian and Prucha 1999) and spatial two-stage least-squares procedures for single-equation models (Kelejian and Prucha 1998; Das, Kelejian and Prucha 2003). Their SEM-based approach was adjusted to fit a SUR framework, as described below (in the section titled Model in SUR Form).

## **DATA DESCRIPTION**

In this study, TAZ-based densities<sup>4</sup> of household categories and employment types were investigated using a 9-equation spatial SUR. Households were classified into six categories (type I through type VI) according to the number of workers and the presence of children (under 18 years old) in each, and jobs were categorized as basic, retail and service types, according to the Standard Industrial Classification (SIC) codes. Table 1 provides the household and employment classifications used here.

The data for this endeavor come from several sources. Austin's Capital Area Metropolitan Planning Organization (CAMPO) provided employment estimates by type at the TAZ level for the year 2000. These derive from third-quarter ES202 reports maintained by the Texas Workforce Commission for the year 2000<sup>5</sup>. Fregonese Calthorpe Associates, an Oregon-based planning firm (contracted for the Envision Central Texas project [<http://envisioncentraltexas.org>]), provided household counts (by type) at the TAZ level.

Year 2000 land use parcel data was obtained from the City of Austin's Neighborhood Planning and Zoning Department (NPZD). The study area is an approximate circular area, covering 420 square miles and 473 TAZs. The City classifies parcels according to 15 distinct land use categories, as listed in Table 2. For ease of modeling and interpretation, these 15 were grouped into 11 types. All 611 parcels labeled as "unknown" were checked against 1995 and 2002 orthophotos<sup>6</sup>, and an appropriate land use code was determined. In addition, a GIS shape file for existing protected lands, future protected lands

---

<sup>4</sup> Since zone sizes can be quite arbitrary (and their variations can introduce heteroskedasticity in error term variation), density measures of households and jobs offer a more meaningful measure of land use intensity than counts. Since zone sizes are fixed, estimates of household and job counts emerge automatically from the model's specifications.

<sup>5</sup> According to CAMPO's Daniel Yang (May 30, 2006), American Business CD's were used to locate ES202 data employers that could not be address-matched (e.g., those using PO box numbers and parent-child companies). Addresses were confirmed via survey.

<sup>6</sup> The 1995 orthophotos provide images with a 1 m x 1 m resolution and were obtained from the Texas Natural Resource Information System (TNRIS) website. The 2000 orthophotos have a 2 ft x 2 ft resolution, and were downloaded from the Capital Area Council of Governments (CAPCOG) webpage.

and 100-year floodplain was assembled by the University of Texas at Austin's Dr. Barbara Parmenter. As with open space and water categories in the City's dataset, the protected lands and 100-year floodplains are excluded from future development.

In order to examine the impacts of a zone's general transportation access, Euclidean distances to the nearest major highway and average travel times to the region's CBD<sup>7</sup> during the AM and PM peak periods were calculated (from zone centroids). Figure 1 depicts the locations of these key elements.

The fractions of each land use type in each TAZ were assumed known, and used as explanatory variables, along with a measure of land use balance, or entropy. (These fractions are based on an upstream land use change model [see, e.g., Zhou and Kockelman 2006]. The land use balance variable is computed as suggested by Kockelman (1997):

$$Entropy = -\frac{1}{\ln(J)} \sum_j^J P_j \ln(P_j) \quad (3)$$

where  $J$  is the number of land use types under consideration and  $P_j$  is the fraction of land in use type  $j$ . Only 6 developed land uses (single-family, multi-family, commercial, office, industrial and civic) were included in this entropy equation, so there were no penalties for a large undeveloped area. This land use balance term ranges from 0 to 1, with 0 indicating that all developed land in the zone is of a single type and 1 indicating that all 6 are present equally.

In summary, the response variables are household and employment densities, and the explanatory variables include zonal land use conditions, travel time to the CBD, and distance to the nearest highway. Table 3 summarizes key statistics for all model variables.

## **MODEL SPECIFICATION**

Both the spatial lag and the spatial error processes are very likely present in and across the households and employment counts for several reasons. It is quite possible that the boundaries of TAZs delimit the extent of households and employment distributions in an arbitrary way, which necessitates the inclusion of a spatial lag in our model specification, due to spatial spillovers (including, for example, agglomeration economies, in business as well as household uses). Furthermore, not all factors influencing the land use intensities are measurable and readily available to the analyst (e.g., microclimatic conditions, local crime rates, continuous measures of grammar school quality), so spatial error may be indispensable to the model's specification. In addition, a pre-defined, arbitrary weight matrix also calls for use of higher-order spatial processes. Nevertheless, despite one's expectation of the co-existence of spatial lag and spatial error processes, diagnostic tests are needed to support the inclusion of both components. Therefore, a series of diagnostic tests were performed, before estimating a SUR model with both spatial processes.

---

<sup>7</sup> The CBD is defined as a 0.39-square mile rectangular area bounded by Guadalupe, Red River, Cesar Chavez and East 11th Streets.

## Tests for Spatial Dependence

A large number of diagnostic tests for spatial dependence have been proposed and their performance has been analyzed using Monte Carlo simulation in works by Anselin and Florax (1995), Anselin et al. (1996), and Anselin (2001b), among others. Recently, Florax and de Graaff (2004) provided a quantitative meta-analysis of Monte Carlo simulation results from a series of studies.

The most common test statistic for spatial dependence in a regression model is Moran's I (see, e.g. Cliff and Ord 1972). It is computationally simple and does not require that one specify a spatial process as the alternative hypothesis. Thus, it does not indicate the nature of the spatial process (e.g., whether it is due to an omitted spatially lagged response variable or a spatial error process). Anselin and Kelejian (1997) derived an asymptotic form of Moran's I for spatial error autocorrelation when endogenous variables and/or a spatially lagged dependent variable exists. Like Moran's I, the Kelejian-Robinson (K-R) test (1992) does not identify the cause of spatial dependence. However, it does not assume normality of the error term or linearity of models<sup>8</sup>, so it may be preferred in certain instances. Here, tests against specific spatial processes were used.

Tests against specific spatial processes generally are constructed as significance tests of the spatial autoregressive coefficient under consideration. These include the Wald, Likelihood Ratio (LR) and Lagrange Multiplier (LM) tests. The Wald and LR tests are relatively demanding because they require maximum likelihood estimation under the alternative hypothesis. Asymptotically equivalent to the Wald and LR tests, the LM test (e.g., see Burridge 1980; Anselin 1988a) does not require estimation of the more complicated model. The robust forms of LM tests can test for spatial lag dependence without assuming zero spatial error autocorrelation, and they can test for spatial error autocorrelation without assuming zero spatial lag autocorrelation. (Bera and Yoon 1993; Anselin et al. 1996)

Anselin (1988b) derived a LM test for spatial error autocorrelation in the presence of a spatially lagged response variable as a special case of a general LM test for spatial dependence and spatial heterogeneity. Specifically, for a candidate model specification like Equation (1), the null hypothesis is  $\lambda = 0$  in the presence of  $\rho$  (which is estimated by a MLE of the spatial lag model,  $\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\xi}$ ). The LM statistic is defined as follows:

$$LM_{\lambda|\rho} = \frac{\left( \hat{\mathbf{e}} \mathbf{W}_2 \hat{\mathbf{e}} / \hat{\sigma}^2 \right)^2}{tr(\mathbf{W}_2 \mathbf{W}_2 + \mathbf{W}_2' \mathbf{W}_2) - tr(\mathbf{W}_2 \mathbf{W}_1 \mathbf{A}^{-1} + \mathbf{W}_2' \mathbf{W}_1 \mathbf{A}^{-1})^2 \hat{V}(\hat{\rho})} \quad (4)$$

where  $\hat{\mathbf{e}}$  is a vector of residuals in the MLE results for the spatial lag model, and the "hats" (^) indicate values estimated using MLE. Under the null hypothesis, the LM statistics will

---

<sup>8</sup> Kelejian and Robinson (1998) extended the K-R procedure to a test for spatial autocorrelation and heteroskedasticity. Together with other tests for high-order spatial dependence (e.g. Anselin 2001b), this test is beyond the scope of this study.

converge to a chi-squared distribution,  $\chi^2$ , with one degree of freedom.

In addition, Anselin et al. (1996) and Anselin and Bera (1998) provided an LM test for spatial lag autocorrelation in the presence of a spatial error autocorrelation<sup>9</sup>. In considering candidate model specifications like Equation (1), the null hypothesis is  $\rho = 0$  in the presence of  $\lambda$  (which is estimated via MLE of the spatial error model,  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  with  $\boldsymbol{\varepsilon} = \lambda \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\xi}$ ). However, the equations given in these two papers neglect one term (the square of  $\hat{\sigma}^2$ )<sup>10</sup>. The correct LM statistic is defined as follows:

$$LM_{\rho|\lambda} = \frac{\left( \hat{\mathbf{u}}' \mathbf{B}' \mathbf{B} \mathbf{W}_1 \mathbf{y} / \hat{\sigma}^2 \right)^2}{\mathbf{H}_\rho - \mathbf{H}_{\theta\rho} \hat{\mathbf{V}} \left( \hat{\boldsymbol{\theta}} \right) \mathbf{H}'_{\theta\rho}} \quad (5)$$

where  $\hat{\mathbf{u}}$  is a vector of residuals in the MLE results for the spatial error model.

Furthermore,  $\mathbf{B} = \mathbf{I} - \hat{\lambda} \mathbf{W}_2$ ,  $\boldsymbol{\theta} = (\boldsymbol{\beta}', \lambda, \sigma^2)'$ , and  $\mathbf{H}_\rho$  and  $\mathbf{H}_{\theta\rho}$  are defined by Equations (6) and (7). Under the null hypothesis, the LM statistics also converge to a  $\chi^2$  distribution with one degree of freedom.

$$\mathbf{H}_\rho = tr(\mathbf{W}_1^2) + tr\left( (\mathbf{B} \mathbf{W}_1 \mathbf{B}^{-1})' (\mathbf{B} \mathbf{W}_1 \mathbf{B}^{-1}) \right) + \frac{\left( \mathbf{B} \mathbf{W}_1 \mathbf{X} \hat{\boldsymbol{\beta}} \right)' \left( \mathbf{B} \mathbf{W}_1 \mathbf{X} \hat{\boldsymbol{\beta}} \right)}{\hat{\sigma}^2} \quad (6)$$

$$\mathbf{H}_{\theta\rho}' = \begin{bmatrix} \frac{(\mathbf{B} \mathbf{X})' \left( \mathbf{B} \mathbf{W}_1 \mathbf{X} \hat{\boldsymbol{\beta}} \right)}{\hat{\sigma}^2} \\ tr\left( (\mathbf{W}_2 \mathbf{B}^{-1})' \mathbf{B} \mathbf{W}_1 \mathbf{B}^{-1} \right) + tr(\mathbf{W}_2 \mathbf{W}_1 \mathbf{B}^{-1}) \\ 0 \end{bmatrix} \quad (7)$$

These two LM tests serve the investigative goal of this study and were applied for each of the 9 equations individually (since no spatial-SUR tests currently exist for this). Before calculating the LM statistics, spatial lag and spatial error models had to be estimated via MLE (using Anselin's [1980] likelihood functions). Each distinct pairing of the four weight matrices ( $\mathbf{W}_{\text{con}}$ ,  $\mathbf{W}_{\text{dist1.0}}$ ,  $\mathbf{W}_{\text{dist1.5}}$ , and  $\mathbf{W}_{\text{dist2.0}}$ ) was used in spatial lag and spatial error processes, representing a total of 12 weight-matrix model cases. Table 4 provides the LM test results. The high  $\chi^2$  values indicate that the unrestricted model specification is preferred to either the spatial lag or spatial error models for most equations in all 12 cases.

<sup>9</sup> Anselin and Bera (1998) also provided a joint test with a null hypothesis that both spatial autoregressive coefficients equal zero. This procedure is not relevant in this study, since it is not possible to identify whether the misspecification is due to lag or error dependence.

<sup>10</sup> The authors have informed Luc Anselin of this error.



More specifically, when spatial lag is the restricted model (the top division of the table), spatial error process exists in all of the 9 individual equations for 7 out of 12 cases. In the other 5 cases, only the basic employment equation has low  $\chi^2$  values. When spatial error is the restricted model (the bottom division of the table), spatial lag process exists in all of the 9 individual equations for 3 out of 12 cases. In the other 9 cases, at most 3 equations have low  $\chi^2$  values. In order to maintain the general SUR structure proposed here, all 9 equations were specified with spatial lags and spatial error terms, as described below.

### Model in a SUR Form

The SUR model specification, with both spatial lag and spatial error processes, is a multi-equation extension of Equation (1). The model is specified as follows

$$\begin{aligned} \mathbf{y}_m &= \rho_m \mathbf{W}_1 \mathbf{y}_m + \boldsymbol{\beta}_m \mathbf{X}_m + \boldsymbol{\varepsilon}_m = (\rho_m, \boldsymbol{\beta}_m)' (\mathbf{W}_1 \mathbf{y}_m, \mathbf{X}_m) + \boldsymbol{\varepsilon}_m = \boldsymbol{\delta}_m' \mathbf{Z}_m + \boldsymbol{\varepsilon}_m \\ \boldsymbol{\varepsilon}_m &= \lambda_m \mathbf{W}_2 \boldsymbol{\varepsilon}_m + \boldsymbol{\xi}_m \end{aligned} \quad (8)$$

where  $m = 1:M$ ,  $\boldsymbol{\xi} = (\xi_1', \xi_2', \dots, \xi_M')$ ,  $E[\boldsymbol{\xi} | X_1, X_2, \dots, X_M] = \mathbf{0}$ , and

$E[\boldsymbol{\xi} \boldsymbol{\xi}' | X_1, X_2, \dots, X_M] = \boldsymbol{\Omega}$ . In addition,  $\mathbf{y}_m$  is an  $n$  by 1 vector of response variables for equation  $m$ ,  $\mathbf{X}_m$  is an  $n$  by  $k_m$  matrix of explanatory variables for equation  $m$  ( $\mathbf{X}_m$  could have different numbers of columns for different equations),  $\boldsymbol{\beta}_m$  is a  $k$  by 1 vector of parameters to be estimated,  $\rho_m$  and  $\lambda_m$  are two scalars.

Kelejian and Prucha's (2004) model is a simultaneous equation model (SEM). It was altered here, to provide a SUR system, by disallowing all direct dependencies between response variables across equations and by assuming that each response variable depends only on its own spatial lag (rather than the spatial lag of other response variables in the equation system).

The proposed SUR model with two spatial processes was estimated in three steps. First, the model with spatial lag process was estimated via a two-stage least-squares (2SLS) process, using instrumental variables. Here, a matrix  $\mathbf{H}$  is composed of linearly independent columns  $\mathbf{X}$ ,  $\mathbf{W}_1 \mathbf{X}$ , and  $\mathbf{W}_2 \mathbf{X}$ , where  $\mathbf{X}$  is an  $n$  by  $k_{\max}$  matrix, and instrumental variables for equation  $m$  are  $\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}\mathbf{Z}_m$ .

Next, the residuals from this first step ( $\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_M$ ) were used to estimate the spatial error autoregressive coefficients ( $\lambda_m$ ) and the variances of error terms for each equation ( $\sigma_m^2$ ), using a generalized method of moments (GMM), as originally proposed by Kelejian and Prucha (1999). More specifically, three equations were constructed by squaring  $\boldsymbol{\varepsilon}_m - \lambda_m \mathbf{W}_2 \boldsymbol{\varepsilon}_m = \boldsymbol{\xi}_m$  and  $\mathbf{W}_2 \boldsymbol{\varepsilon}_m - \lambda_m \mathbf{W}_2 \mathbf{W}_2 \boldsymbol{\varepsilon}_m = \mathbf{W}_2 \boldsymbol{\xi}_m$  and then taking their product. Consistent with Kelejian and Prucha's (2004) notation, the two equations could be simplified as  $\boldsymbol{\varepsilon}_m - \lambda_m \bar{\boldsymbol{\varepsilon}}_m = \boldsymbol{\xi}_m$  and  $\bar{\boldsymbol{\varepsilon}}_m - \lambda_m \bar{\bar{\boldsymbol{\varepsilon}}}_m = \bar{\boldsymbol{\xi}}_m$ , where  $\bar{\boldsymbol{\varepsilon}}_m = \mathbf{W}_2 \boldsymbol{\varepsilon}_m$ ,  $\bar{\boldsymbol{\xi}}_m = \mathbf{W}_2 \boldsymbol{\xi}_m$  and  $\bar{\bar{\boldsymbol{\varepsilon}}}_m = \mathbf{W}_2 \bar{\boldsymbol{\varepsilon}}_m$ . Essentially,  $\lambda_m$  and  $\sigma_m^2$  are nonlinear least squares estimators, based on the

following equation:

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m \\ \tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m \\ \tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m \end{bmatrix} = \begin{bmatrix} 2\tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m & -\tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m & n \\ 2\tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m & -\tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m & \text{trace}(\mathbf{W}_2' \mathbf{W}_2) \\ \tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m + \tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m & -\tilde{\boldsymbol{\varepsilon}}_m' \tilde{\boldsymbol{\varepsilon}}_m & 0 \end{bmatrix} \begin{bmatrix} \lambda_m \\ \lambda_m^2 \\ \sigma_m^2 \end{bmatrix} + \boldsymbol{\zeta}_m \quad (9)$$

where  $\boldsymbol{\zeta}_m$  is a 3 by 1 vector of regression residuals.

In the third step, a Cochrane-Orcutt transformed regression model

$\mathbf{y}_m^* = \boldsymbol{\delta}_m' \mathbf{Z}_m^* + \boldsymbol{\xi}_m$  (with  $\mathbf{y}_m^* = \mathbf{y}_m - \lambda_m \mathbf{W}_2 \mathbf{y}_m$  and  $\mathbf{Z}_m^* = \mathbf{Z}_m - \lambda_m \mathbf{W}_2 \mathbf{Z}_m$ ) was estimated

using feasible generalized least squares techniques. (Readers may consult Greene's (2000) Chapter 14 for more details on this estimation procedure.)

The above estimation procedure was applied using each of the 12 combinations of weight matrices described earlier. Equation (10)'s goodness-of-fit measure for the SUR model (McElroy 1977) were calculated for each case:

$$R^2 = 1 - \frac{\hat{\boldsymbol{\xi}}' \hat{\boldsymbol{\Omega}}^{-1} \hat{\boldsymbol{\xi}}}{\sum_{i=1}^M \sum_{j=1}^M \hat{\sigma}^{ij} \left( \sum_{t=1}^n (y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) \right)} \quad (10)$$

where  $\hat{\sigma}^{ij}$  is the  $ij$ th element in matrix  $\hat{\boldsymbol{\Omega}}^{-1}$ ,  $\bar{y}_i$  is the mean of response values for

equation  $i$ , and other terms are as previously defined. The case with  $\mathbf{W}_{\text{con}}$  and  $\mathbf{W}_{\text{dist}1.0}$  for spatial lag and error yielded the highest goodness-of-fit value (at 0.674, as compared to 0.422 through 0.635 for the other 11 cases). Therefore, this pair of weight matrices was used in the final specification, and the model was calibrated by first incorporating all explanatory variables, and then refining the specification using a process of stepwise deletion (removing explanatory variables with t-statistics less than 1.645 [p-values over 0.10]).

## MODEL RESULTS

Table 5 provides the SUR model's final estimation results. Single-family, mining and utility land-use proportions were statistically insignificant in all 9 equations, and thus were removed from the SUR system.

As expected, all of the spatial lag autocorrelation coefficients ( $\rho$ ) were estimated to be positive and statistically significant, confirming the existence of strong spatial dependence. 7 out of 9 spatial error autocorrelation coefficients ( $\lambda$ ) were estimated to be negative, with one corresponding to the LM statistics that did not support inclusion of spatial error components (the basic employment density). The statistical significance of the error autocorrelation coefficients ( $\lambda$ ) was unknown from the three-step estimation procedure, which means the inference on co-existence of two spatial processes could not be derived directly. However, the series of LM tests, before application of the spatial 3SLS SUR model, confirmed the co-existence of the two processes in the final model specification (the case with  $\mathbf{W}_{\text{con}}$  and  $\mathbf{W}_{\text{dist}1.0}$ ), except for the spatial lag model of basic

employment.

The estimates of parameter coefficients are tangible and reasonable. Most have expected signs. For example, the multi-family land-use proportion is estimated to exert a positive impact on household densities of all types, though with different magnitudes. In contrast, commercial, office, industrial and civic use proportions are estimated to generally decrease household densities (as they compete for space). Peak-period travel time to the regional CBD is estimated to have a positive impact on household type III densities (i.e., those with two-plus workers and children). This type of household tends to live at the city periphery, where larger home sizes prevail.

Somewhat interesting is the fact that household densities do not increase with the percentage of excluded land (open space and protected areas). While such spaces may be attractive to certain household types, land development intensities nearby are generally lower.

In terms of employment densities, all three respond positively to increases in the proportion of office land use. Retail density also was found to rise with commercial land use, in a form of complementarity. In addition, retail densities were estimated to rise with reductions in the land use entropy term, indicating that zones with a single, relatively homogeneous land use types (perhaps mostly commercial in nature) tend to have higher numbers of retail jobs. Service-job densities were found to fall in response to multi-family use proportions. This may be due to the fact that Austin's service facilities are focused on the region's core (where the State capitol and other government facilities exist, amid other business uses).

Retail densities were estimated to decline with travel times to the CBD, suggesting that shopping is rather centralized, everything else constant. It is surprising that all three job densities are estimated to increase with distance to the region's major highways (after controlling for all other explanatory variables, two spatial processes and error-term interactions across equations). Office proportions, which positively affect these three job densities, may be heavily attracted to the major highways, resulting in a marginally positive effect.

## **CONCLUSIONS**

This paper investigates the spatial distribution of jobs and households at the level of TAZs in the Austin region. These variables are key inputs to travel demand and air quality models as well as regional policymaking. Thus, their accurate estimation is of great interest.

There is little doubt that spatial dependence exists across spatially proximate zones. However, it is unclear what kinds of spatial processes are most appropriate in capturing the underlying relationships. Prior studies (e.g., Clark and Murphy 1996; Deller et al. 2001; Carruthers and Vias 2005) have demonstrated interactions between jobs and population at relatively large geographic scales, like counties. Few have considered such interactions at geographic scales comparable to TAZs, while recognizing spatial correlations. This study applied LM tests for the co-existence of spatial lag and spatial error processes. It then

estimated 9 equations simultaneously, using feasible generalized spatial 3SLS, permitting, for the first time, the presence of both such processes in a SUR model. This procedure was derived from Kelejian and Prucha's (2004) SEM results, by disallowing direct dependencies between response variables across equations and by assuming that each response variable depends only on its own spatial lag.

The model includes a series of explanatory variables, including local land use composition (which can be generated, for example, from Zhou and Kockelman's [2005] upstream models of land use change), land use balance, travel time to the region's CBD, and Euclidean distances to the nearest major highway. As expected, knowledge of local land use conditions offers substantial predictive power for equations of land use intensity. Estimation results are reasonable with relatively high  $R^2$  statistics when two spatial processes and equation-system effects are recognized. As expected, all spatial lag autocorrelation coefficients are statistically significant (and all are positive), confirming the existence of spatial lag effects. Spatial error effects also exist, but estimates of t-statistics for their associated coefficients are unidentifiable.<sup>11</sup>

This SUR model was designed to work together with an upstream land use change model (Zhou and Kockelman 2006) in predicting the spatial distributions of households and employment. The land use change model provides the dynamic nature of the modeling system, while the SUR model utilizes only cross-sectional data to allocate jobs and populations in a projection year. Therefore, this study analyzes related, but distinctive, data situations; direct dependencies between response variables across equations were eliminated and correlations in the error terms were designed to take care of the interactions between households and employment counts. It may not completely address the endogeneity issue (for example, service jobs may follow residential development in population counts, rather than developed area). However, land use conditions that were determined before our household and employment SUR model were used as explanatory variables, so the endogeneity is not nearly so much an issue as it would be in a simple model of household and employment totals/densities, without an upstream model of land use change.

Key extensions of this work are felt to lie in the use of panel data, to permit an additional layer of correlation across observed values, and the use of full-information maximum likelihood estimate (MLE) approaches that are mathematically tractable for such complex specifications. Additionally, application of the model, in concert with the upstream model of land use change (Zhou and Kockelman 2006), and validation relative to current and coming land use patterns should prove insightful, particularly in comparison with predictions from simpler model specifications.

---

<sup>11</sup> T-statistics for the spatial error autocorrelation coefficients are not directly available from the estimation procedure. Thus, proper tests, as used here, are needed to confirm the co-existence of spatial lag and spatial error before applying a SUR with the two spatial processes.

## ***ACKNOWLEDGEMENTS***

We wish to thank Dr. Parmenter at the University of Texas at Austin and Daniel Yang at CAMPO for the valuable data sets, Darla Munroe at the Ohio State University and Bill Pan at Johns Hopkins University for insightful advice on model specification, and Annette Perrone for editing assistance. We also want to thank the Environmental Protection Agency for financially supporting this study under its STAR (Science to Achieve Results) Grant program. The multi-team project is titled “Predicting the Relative Impacts of Urban Development Policies and On-Road Vehicle Technologies on Air Quality in the United States: Modeling and Analysis of a Case Study in Austin, Texas.”

## REFERENCES

- Anselin, L. (1980). *Estimation methods for spatial autoregressive structures*. Regional Science Dissertation and Monograph Series 8. Field of Regional Science, Cornell University, Ithaca, N.Y.
- Anselin, L. (1988a). *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.
- Anselin, L. (1988b). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity, *Geographical Analysis* 20, 1-17.
- Anselin, L. (2001a) Spatial econometrics. In *A companion to theoretical econometrics*, edited by B. Baltagi, 310-330. Oxford, England: Blackwell.
- Anselin, L. (2001b). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference* 97, 113–139.
- Anselin, L. (2003a). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, 153–166.
- Anselin, L. (2003b). GeoDa 0.9 User's Guide. Retrieved June, 2006 from <https://geoda.uiuc.edu/pdf/geoda093.pdf>.
- Anselin, L. (2005). Exploring Spatial Data with GeoDa: a Workbook. Retrieved June, 2006 from <https://geoda.uiuc.edu/pdf/geodaworkbook.pdf>.
- Anselin, L. and Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah and D. E. A. Giles, Eds., *Handbook of Applied Economic Statistics*, p 237–289. New York: Marcel Dekker.
- Anselin, L. and Florax, R.J.G.M (1995). Small sample properties of tests for spatial dependence in regression models: some future results. In L. Anselin and R. Florax (Eds.), *New Directions in Spatial Econometrics*, p 21–74. Berlin: Springer-Verlag.
- Anselin, L. and Kelejian H.H. (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review* 20, 153–182.
- Anselin, L., Bera, A.K., Florax, R.J.G.M., and Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26, 77–104.
- Bera, A. and Yoon, M. (1993). Specification testing with locally misspecified alternatives, *Econometric Theory* 9, 649–658.
- Burridge, P. (1980). On the Cliff–Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society B* 42, 107–108.
- Carruthers, J.I. and Vias, A.C. (2005) Urban, suburban, and exurban sprawl in the Rocky Mountain West: evidence from regional adjustment models. *Journal of Regional Science* 45, 21–48.

- Capital Area Council of Governments (CAPCOG). Retrieved June, 2005 from [http://www.capcog.org/Information\\_Clearinghouse/Orthoimagery\\_main.asp](http://www.capcog.org/Information_Clearinghouse/Orthoimagery_main.asp).
- Case, A. (1991). Spatial patterns in household demand. *Econometrica* 59, 953–965.
- Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics* 22, 491–508.
- Case, A., Rosen, H.S., and Hines, J.R. (1993). Budget spillovers and fiscal policy interdependence: evidence from the States. *Journal of Public Economics* 52, 285–307.
- Clark, D.E. and Murphy, C.P. (1996) Countywide Employment and Population Growth: An Analysis of the 1980s. *Journal of Regional Science* 36, 235–256.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.
- City of Austin, Land Use Survey Project Description. Retrieved June, 2005 from <http://www.ci.austin.tx.us/landuse/lupd.htm>.
- Clark, D.E. and Murphy, C.A. (1996). Countywide employment and population growth: an analysis of the 1980s. *Journal of Regional Science* 36, 235–256.
- Cliff, A. and Ord, J.K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4, 267–284.
- Cliff, A. and Ord, J.K. (1981). *Spatial Processes: Models and Applications*. London: Pion.
- Das, D., Kelejian, H., and Prucha, I. (2003). Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82, 1–26.
- Deller, S.C., Tsai, T.H., Marcouiller, D.W., and English, D.B.K. (2001). The role of amenities and quality of life in rural economic growth. *American Journal of Agriculture Economics* 83, 352–365.
- Florax, R.J.G.M., and Rey, S. (1995). The impacts of misspecified spatial interaction in linear regression models. In L. Anselin and R. Florax (Eds.), *New Directions in Spatial Econometrics*, p 111–135. Berlin: Springer-Verlag.
- Florax, R.J.G.M. and de Graaff, T. (2004). The performance of diagnostic tests for spatial dependence in linear regression models: a meta-analysis of simulation studies. In L. Anselin, R. Florax and S.J. Rey (Eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, p 29–66. Berlin: Springer-Verlag.
- Gallo, J.L. and Dall’erba, S. (2003). Evaluating the temporal and the spatial heterogeneity of the European convergence process, 1980-1999. Presented at presented at the 50th North American Meetings of the Regional Science Association International (RSAI), Philadelphia, November 20-22, 2003.
- Greene, W. (2000). *Econometric Analysis*. Upper Saddle River: Prentice-Hall.
- Haining, R. (1988). Estimating spatial means with an application to remotely sensed data. *Communications in Statistics: Theory and Methods* 17, 573–597.

- Henry, M.S., Barkley, D.L., and Bao, S. (1997). The Hinterlands' Stake in Metropolitan Growth: Evidence from Selected Southern Regions. *Journal of Regional Science* 37, 479–501.
- Kelejian, H.H. and Prucha, I. (1998). A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17, 99–121.
- Kelejian, H.H. and Prucha, I.R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40, 509–533.
- Kelejian, H.H. and Prucha, I.R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Economics* 118, 27–50.
- Kelejian, H.H. and Robinson, D.P. (1992). Spatial autocorrelation: A new computationally simple test with an application to per capita country police expenditures. *Regional Science and Urban Economics* 22, 317–333.
- Kelejian, H.H. and Robinson, D.P. (1993). A suggested method of estimation for spatial independent models with autocorrelated errors, and an application to county expenditure model. *Papers in Regional Science* 72, 297–312.
- Kelejian, H.H. and Robinson, D.P. (1995). Spatial autocorrelation: a suggested alternative to the autoregressive model. In L. Anselin and R. Florax (Eds.), *New Directions in Spatial Econometrics*, p 75–93. Berlin: Springer-Verlag.
- Kelejian, H.H. and Robinson, D.P. (1998). A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results. *Regional Science and Urban Economics* 28, 389–417.
- Kockelman, K.M. (1997). Travel behavior as a function of accessibility, land use mixing, and land use balance: evidence from the San Francisco Bay Area. *Transportation Research Record* 1607, 117–125.
- Lundberg, J. (2005). A Spatial Interaction Model of Spillovers from Locally Provided Public Services. Forthcoming in *Regional Studies*.
- McElroy, M. (1977). Goodness of fit for seemingly unrelated regressions: Glahn's  $R^2_{y,x}$  and Hooper's  $\bar{r}^2$ . *Journal of Econometrics* 6, 381–387.
- Mulligan, G. F., Vias, A.C. and Glavac, S.M. (1999). Initial Diagnostics of a Regional Adjustment Model. *Environment and Planning A*, 31, 855–876.
- Rey, S. and Montouri, B (1999). U.S. regional income convergence: a spatial econometric perspective. *Regional Studies* 33, 143–156.
- Steinnes, D.N. and Fisher, W.D. (1974). An Econometric Model of Interurban Location. *Journal of Regional Science* 14, 65–80.
- Texas Natural Resource Information System (TNRIS). Retrieved June, 2005 from <http://www.tnr.is.state.tx.us/DigitalData/doqs.htm>.



- Tobler W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234–240.
- Vias, A.C. and Mulligan, G.F. (1999). Integrating Economic Base Theory with Regional Adjustment Models: The Nonmetropolitan Rocky Mountain West. *Growth and Change* 30, 507–525.
- Zhou, B. and Kockelman, K.M. (2006). Neighborhood Impacts on Land Use Change: A Multinomial Logit Model of Spatial Relationships. Presented at the 52<sup>nd</sup> Annual North American Meeting of the Regional Science Association International, Las Vegas.

## **TABLES AND FIGURES**

Table 1. Employment and Household Classification

<b>Category</b>	<b>Definition</b>
Type I household	0-worker household, with at least one child under 18 years of age
Type II household	1-worker household, with at least one child under 18 years of age
Type III household	2 or more-worker household, with at least one child under 18 years of age
Type IV household	0-worker household, with no children
Type V household	1-worker household, with no children
Type VI household	2 or more-worker household, with no children
Basic Employment	Division A (agriculture, forestry, and fishing) Division B (mining) Division C (construction) Division D (manufacturing) Division E (transportation, communications, electric, gas, and sanitary services) Division F (wholesale trade)
Retail Employment	Division G (retail trade)
Service Employment	Division H (finance, insurance and real estate) Division I (services) Division J (public administration)

Table 2. Land Use Categories

<b>Original Land Use Classification</b>	<b>Description</b>	<b>Final Classification</b>
Large Lot Single-family	Single-family detached, two-family attached with lot size bigger than 10 acres	Single-family
Single-family	Single-family detached, two-family attached	
Mobile Homes	Mobile homes	Multiple-family
Multi-family	Three/fourplex, apartment/condo, group quarters, retirement	
Commercial	Retail and general merchandise, apparel and accessories, furniture and home furnishings, grocery and food sales, eating and drinking, auto related, entertainment, personal services, lodgings, building services	Commercial
Office	Administrative offices, financial services (banks), medical offices, research and development	Office
Industrial	Manufacturing, warehousing, equipment sales and service, recycling and scrap, animal handling	Industrial
Civic	Semi-institutional housing, hospital, government services, educational meeting and assembly, cemetery	Civic
Mining	Resource extraction	Mining
Utilities	Utility services	Utility
Undeveloped/Rural	Rural uses, vacant land, land under construction	Undeveloped
Open Space	Parks/greenbelts, golf courses, camp grounds and open spaces set aside for preservation or protection.	
Water	Water	
Transportation	Railroad facilities, transportation terminal, aviation facilities, marina parking facilities	Transportation

Source: Land Use Survey Project Description, City of Austin

Table 3. Summary Statistics for All Variables

	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Std. Deviation</b>
<b>Response Variables</b>				
Type I household density (HHs/mile <sup>2</sup> )	0	80.3	5.92	8.23
Type II household density (HHs/mile <sup>2</sup> )	0	870.4	96.4	115.7
Type III household density (HHs/mile <sup>2</sup> )	0	1,044	171.8	179.6
Type IV household density (HHs/mile <sup>2</sup> )	0	1,341	128.9	190.3
Type V household density (HHs/mile <sup>2</sup> )	0	7,167	539.2	815.2
Type VI household density (HHs/mile <sup>2</sup> )	0	2,584	336.4	368.1
Basic employment density (jobs/mile <sup>2</sup> )	0	56,498	918.9	3,609
Retail employment density (jobs/mile <sup>2</sup> )	0	20,041	696.6	1,866
Service employment density (jobs/mile <sup>2</sup> )	0	116,916	2,723	10,966
<b>Explanatory Variables</b>				
Single-family use fraction	0	0.808	0.271	0.202
Multi-family use fraction	0	0.559	0.049	0.076
Commercial use fraction	0	0.850	0.050	0.085
Office use fraction	0	0.626	0.033	0.068
Industrial use fraction	0	0.596	0.035	0.080
Civic use fraction	0	0.929	0.058	0.134
Undeveloped land fraction	0	0.868	0.153	0.184
Mining use fraction	0	0.504	0.0057	0.042
Utility use fraction	0	0.412	0.0056	0.033
Excluded land fraction	0	0.979	0.171	0.199
Land use entropy	0.1	0.830	0.391	0.170
Distance to nearest highway (miles)	0	6.57	1.20	1.27
Travel time to the CBD (minutes)	0	37.2	12.4	8.30
Number of observations	473			

Table 4. Lagrange Multiplier Test Results

Equations	LM Statistics (spatial error autocorrelation in the presence of spatial lag)											
	W <sub>con</sub>	W <sub>con</sub>	W <sub>con</sub>	W <sub>dist1.0</sub>	W <sub>dist1.0</sub>	W <sub>dist1.0</sub>	W <sub>dist1.5</sub>	W <sub>dist1.5</sub>	W <sub>dist1.5</sub>	W <sub>dist2.0</sub>	W <sub>dist2.0</sub>	W <sub>dist2.0</sub>
Spatial lag matrix	W <sub>dist1.0</sub>	W <sub>dist1.5</sub>	W <sub>dist2.0</sub>	W <sub>con</sub>	W <sub>dist1.5</sub>	W <sub>dist2.0</sub>	W <sub>con</sub>	W <sub>dist1.0</sub>	W <sub>dist2.0</sub>	W <sub>con</sub>	W <sub>dist1.0</sub>	W <sub>dist1.5</sub>
Type I household density (HHs/mile <sup>2</sup> )	262.8	620.0	633.2	123.0	246.9	204.3	239.1	2,646	598.9	433.4	12,398	3,633
Type II household density (HHs/mile <sup>2</sup> )	268.6	642.4	715.4	155.4	307.3	263.0	180.6	327.0	308.6	449.9	8,193	2,935
Type III household density (HHs/mile <sup>2</sup> )	370.2	755.1	786.7	266.2	1,511	794.8	535.2	12,975	1,849.1	296.3	1,746	1,231
Type IV household density (HHs/mile <sup>2</sup> )	59.0	128.1	127.9	39.2	102.5	66.8	51.8	498.3	109.8	73.6	241.2	670.1
Type V household density (HHs/mile <sup>2</sup> )	581.2	1,031	919.5	529.7	4,096	1,579	12,706	917,056	67,807	2,402	138,185	34,276
Type VI household density (HHs/mile <sup>2</sup> )	132.1	262.0	265.1	122.7	485.6	273.1	788.5	37,717	3,371	621.2	33,443	8,253
Basic employment density (jobs/mile <sup>2</sup> )	<b>0.790</b>	9.67	18.6	<b>1.281</b>	9.74	11.8	<b>1.34</b>	8.04	17.2	<b>1.59</b>	<b>1.81</b>	18.7
Retail employment density (jobs/mile <sup>2</sup> )	53.6	173.9	205.0	61.8	477.9	343.8	401.8	25,712	2,265	71.2	484.1	605.1
Service employment density (jobs/mile <sup>2</sup> )	112.6	399.8	492.4	204.9	1,340	709.4	3,223	214,229	17,485	510.2	19,978	6,025

Table 4. Lagrange Multiplier Test Results (continued)

Equations	LM Statistics (spatial lag autocorrelation in the presence of spatial error)											
	W <sub>con</sub>	W <sub>con</sub>	W <sub>con</sub>	W <sub>dist1.0</sub>	W <sub>dist1.0</sub>	W <sub>dist1.0</sub>	W <sub>dist1.5</sub>	W <sub>dist1.5</sub>	W <sub>dist1.5</sub>	W <sub>dist2.0</sub>	W <sub>dist2.0</sub>	W <sub>dist2.0</sub>
Spatial lag matrix	W <sub>con</sub>	W <sub>con</sub>	W <sub>con</sub>	W <sub>dist1.0</sub>	W <sub>dist1.0</sub>	W <sub>dist1.0</sub>	W <sub>dist1.5</sub>	W <sub>dist1.5</sub>	W <sub>dist1.5</sub>	W <sub>dist2.0</sub>	W <sub>dist2.0</sub>	W <sub>dist2.0</sub>
Spatial error matrix	W <sub>dist1.0</sub>	W <sub>dist1.5</sub>	W <sub>dist2.0</sub>	W <sub>con</sub>	W <sub>dist1.5</sub>	W <sub>dist2.0</sub>	W <sub>con</sub>	W <sub>dist1.0</sub>	W <sub>dist2.0</sub>	W <sub>con</sub>	W <sub>dist1.0</sub>	W <sub>dist1.5</sub>
Type I household density (HHs/mile <sup>2</sup> )	82.4	44.8	14.4	<b>2.59</b>	42.5	7.86	8.14	124.6	12.4	8.68	114.8	54.2
Type II household density (HHs/mile <sup>2</sup> )	85.2	48.1	13.5	<b>1.91</b>	48.7	6.49	7.73	131.7	10.6	8.72	136.1	68.3
Type III household density (HHs/mile <sup>2</sup> )	72.8	38.9	8.98	7.95	57.3	8.41	24.8	148.1	13.8	30.5	153.2	80.0
Type IV household density (HHs/mile <sup>2</sup> )	15.7	5.53	<b>1.77</b>	<b>0.407</b>	<b>1.64</b>	<b>0.435</b>	<b>0.468</b>	16.6	<b>0.680</b>	<b>0.0491</b>	13.1	<b>1.39</b>
Type V household density (HHs/mile <sup>2</sup> )	133.8	94.0	51.4	3.86	82.8	27.9	7.92	154.6	30.4	5.09	131.5	75.5
Type VI household density (HHs/mile <sup>2</sup> )	107.4	74.4	37.4	13.6	92.2	38.5	28.3	155.0	40.4	25.4	136.4	84.7
Basic employment density (jobs/mile <sup>2</sup> )	4.86	<b>0.455</b>	<b>0.0320</b>	6.56	3.25	<b>0.0362</b>	20.1	26.0	3.36	25.0	24.0	8.16
Retail employment density (jobs/mile <sup>2</sup> )	24.9	5.59	<b>0.129</b>	16.8	21.3	<b>0.0241</b>	40.5	94.2	<b>1.10</b>	34.0	77.9	24.5
Service employment density (jobs/mile <sup>2</sup> )	117.1	69.5	27.4	8.01	78.8	15.3	30.7	213.1	34.3	36.2	204.8	115.4

Notes: 1. Numbers in bold represent statistically insignificant LM statistics at 90% confidence level.

2. The critical value for a 90% confidence level in a  $\chi^2$  distribution with one degree of freedom is 2.706.

Table 5. Result of Seemingly Unrelated Regression with Spatial Lag and Spatial Error Processes

Variables	Type I Household	Type II Household	Type III Household	Type IV Household	Type V Household	Type VI Household	Basic Employment	Retail Employment	Service Employment
Constant	2.28 (3.12)	52.73 (4.49)	117.7 (6.40)	37.60 (1.77)	187.7 (2.96)	207.5 (6.86)	-521.9 (-4.59)	183.7 (1.09)	-1,171.3 (-3.47)
Multi-family use fraction	13.32 (4.15)	446.3 (9.27)	410.4 (5.12)	472.9 (5.34)	4,468.2 (14.89)	1,469.4 (10.75)			-7,318.6 (-2.25)
Commercial use fraction	-10.39 (-3.39)	-139.3 (-3.24)	-384.1 (-5.62)	-147.3 (-1.77)	-628.7 (-2.51)	-456.0 (-4.05)		9,072.9 (13.89)	
Office use fraction	-12.99 (-3.81)	-296.8 (-5.80)	-502.5 (-6.20)		-649.7 (-2.20)	-864.0 (-6.59)	14,323.3 (6.58)	2,798.2 (2.85)	40,466.6 (8.42)
Industrial use fraction	-6.00 (-2.06)	-118.5 (-2.81)	-336.9 (-4.99)	-222.3 (-2.80)	-935.2 (-3.67)	-579.7 (-5.23)			
Civic use fraction	-6.06 (-3.20)	-86.54 (-3.12)	-253.2 (-5.60)	-192.9 (-3.65)	-1,088.2 (-6.72)	-547.0 (-7.65)			
Undeveloped land fraction	-3.90 (-2.65)	-76.02 (-3.54)	-238.6 (-6.49)	-80.18 (-1.84)	-381.9 (-2.87)	-342.0 (-5.77)			
Excluded land fraction	-4.36 (-3.19)	-97.94 (-5.08)	-284.2 (-9.05)	-77.71 (-2.04)	-493.3 (-4.24)	-358.3 (-6.83)			
Land use entropy								-1,594.4 (-4.78)	
Dist to nearest highway (miles)							153.15 (2.22)	98.19 (2.13)	341.4 (2.05)
Travel time to the CBD (minutes)			0.95 (2.22)					-10.07 (-1.86)	
$\rho$	1.00 (24.02)	0.813 (13.91)	0.995 (12.56)	0.952 (13.64)	0.798 (15.80)	0.823 (17.53)	0.851 (14.60)	0.864 (16.07)	0.923 (31.89)
$\lambda$	-4.39	-0.170	8.66	-2.90	0.092	-0.805	-5.08	-2.32	-4.57
$R^2$	0.678 <sup>12</sup>								
No. of Observation	473								

<sup>12</sup> The  $R^2$  is slightly higher than that with all the explanatory variables. Greene (2000) argued there is no assurance that dropping one variable will result in a decrease in the  $R^2$ , because the estimation procedure is based on minimizing the generalized sum of squares in stead of sum of squares. (page 209)

Figure 1. Austin's Major Highways and Central Business District

