A SPATIAL ANALYSIS OF A RURAL LAND MARKET USING ALTERNATIVE SPATIAL WEIGHT MATRICES

Patricia Soto, Louisiana State University Agricultural Center, Baton Rouge
Lonnie R. Vandeveer, Louisiana State University Agricultural Center, Baton Rouge
Steven A. Henning, Louisiana State University Agricultural Center, Baton Rouge
Huizhen Niu, Louisiana State University Agricultural Center, Baton Rouge

ABSTRACT

Recent literature in rural land value research suggests the use of spatial econometric procedures when spatial dependence exists within the data. These procedures require the use of a spatial weight matrix to adjust for dependence in the data. This research empirically tests two forms of spatial weight matrices. The first is the Delaunay method, which represents a rigid spatial weight matrix, while the second is the nearest neighbor method, which is dependent on a decay factor and a specified number of neighbors. Empirical results suggest the nearest neighbor method is the appropriate method for data used in this study.

INTRODUCTION

Several factors determine the value of rural real estate, including its inherent productive capacity, location, accessibility, and alternative uses. Additionally, continued economic and population growth increases the need for land, which puts upward pressure on rural land market values. With more rural land acres being converted at the urban fringe, buyers, sellers, planners, appraisers, tax assessors, and others are expected to have an increasing need for information related to the effect of location and economic development on rural land values. Generally, research aimed at identifying the effects of location and economic development on rural land market values is expected to provide improved information for both private and public decisions.

In previous studies, Vandeveer et al. (2002) showed the importance of applying spatial econometrics to a rural real estate market to estimate implicit prices. Geographical Information Systems (GIS) and spatial econometric procedures were used to model the rural land market in Louisiana. These procedures, along with spatial econometric procedures, are necessary for testing the data for spatial dependence and estimating models in the presence of spatial dependence. In addition, these procedures are important because modeling the real estate market in the presence of spatial autocorrelation using traditional OLS procedures may result in models with less than desirable statistical characteristics. Pace et al. indicated that real estate and spatial statistics complement each other, and employing spatial estimators provide benefits over ignoring dependencies in the data. The benefits include improved prediction, better statistical inference through unbiased standard errors, and better estimates because of the way that location is handled within the modeling procedure.

The general objective of this research is to empirically test two forms of spatial weight matrices in rural land market research. A spatial weight matrix in rural land markets is used in spatial econometric models to correct for spatial autocorrelation. In this analysis, spatial weight matrices are estimated using the Delaunay and the nearest neighbor methods. The Delaunay method is a rigid form of a spatial weight matrix, whereas the nearest neighbor is estimated to depend on a decay factor and a specified number of neighbors. With the nearest neighbor method it is expected that nearby rural land sales have a greater influence on a given sale than sales located farther away.

In the next section, hedonic procedures for modeling rural land market values are presented and discussed. The following section presents a discussion of spatial econometric procedures including a description of the two spatial weight matrices tested in this analysis. Data for this study comes from an area in Southeast Louisiana. Final sections provide a discussion of empirical results and conclusions.

HEDONIC MODEL

Previous studies have demonstrated the usefulness of using hedonic analysis in rural land market research (Palmquist). In an early study, Rosen defined hedonic prices as implicit prices of attributes and notes that they are revealed to economic agents from observed prices of differentiated products and the specific amounts of characteristics associated with them. Prices of these characteristics are implicit because there is no direct market for them. Palmquist provided a discussion of the theoretical basis for using hedonic analysis in rural land value studies and Danielson used the procedure to empirically analyze the rural land market in North Carolina.

This study follows the approach used by Danielson. Value in a rural land submarket is specified by the following transcendental function:

$$Price = \beta_o Z_1^{\beta_1} \exp\left[\sum_{i=1}^{m} \alpha_i X_i + \sum_{j=1}^{n} \gamma_j D_j + \varepsilon\right], \qquad (1)$$

where *Price* is the per acre price of land, Z_1 is the size of tract in acres, m is the number of additional continuous variables (X_i), n is the number of discrete (dummy) variables (D_j), and ε is a random disturbance term. Taking the natural logarithm of both sides of equation (1) gives:

$$\ln Price = \ln\beta_o + \beta_1 \ln Z + \sum_{i=1}^{m} \alpha_i X_i + \sum_{j=1}^{n} \gamma_j D_j + \varepsilon \qquad (2)$$

Because the price of land is hypothesized to decline as the size of tract (Z_1) increases, but at a decreasing rate, nonlinearities were incorporated for Z_1. Therefore, β_1 is hypothesized to be negative.

Prior to developing a rural land market hedonic model, it is necessary to test for spatial dependence in the data. Diagnostic tests are used to test for spatial dependence and to provide guidance in identifying the correct model (i.e., spatial lag or spatial error) for estimation (Anselin). For this analysis, Lagrange Multiplier tests and Likelihood Ratio tests are used to test for spatial dependence (Anselin).

SPATIAL PROCEDURES

With spatial autocorrelation in the data, hedonic model estimation using OLS procedures generate estimates that are inefficient. Inefficient estimates could result on creating misleading model inferences. Following Anselin, spatial autocorrelation occurs when the dependent variable or error term at each

location is correlated with observations for the dependent variable or error term at other locations. This means that for neighboring locations *i* and *j*:

$$E(y_i y_j) \neq 0 \quad (3)$$

or

$$E(\varepsilon_i \varepsilon_j) \neq 0 \quad (4)$$

where (3) is defined as a spatial lag situation (Anselin). When the dependent variable exhibits spatial autocorrelation, the simultaneous spatial autoregression estimator corrects the usual prediction of the dependent variable, $y = X\beta + \varepsilon$, by using a weighted average of the values on nearby observations, Dy. The spatial lag situation is specified by the following model:

$$y = \alpha Dy + X\beta + \varepsilon \quad (5)$$

where:

y	=	vector dependent observations,
α	=	spatial autoregressive coefficient,
Dy	=	spatially lagged dependent variable,
D	=	*n* by *n* weight matrix,
X	=	matrix of explanatory variables,
β	=	vector of regression coefficients, and
ε	=	vector of error terms.

The spatial lag situation assumes that the residuals, ε, are independently and normally distributed. These assumptions are the following:

(*i*) $D_{ii}=0$ for all *i*

(*ii*) $\sum_{j=1}^{n} D_{ij} = 1$ for all *i*

(*iii*) $0 \leq \alpha \leq 1$

(*iv*) $\varepsilon \sim N(0, \sigma^2 I)$

When spatial dependence occurs in the error, as defined in (4), a regression specification with a spatial autoregressive error term is used to develop model estimates. The spatial error model is:

$$y = X\beta + \varepsilon \quad (6)$$

$$\varepsilon = \alpha D\varepsilon + \xi \quad (7)$$

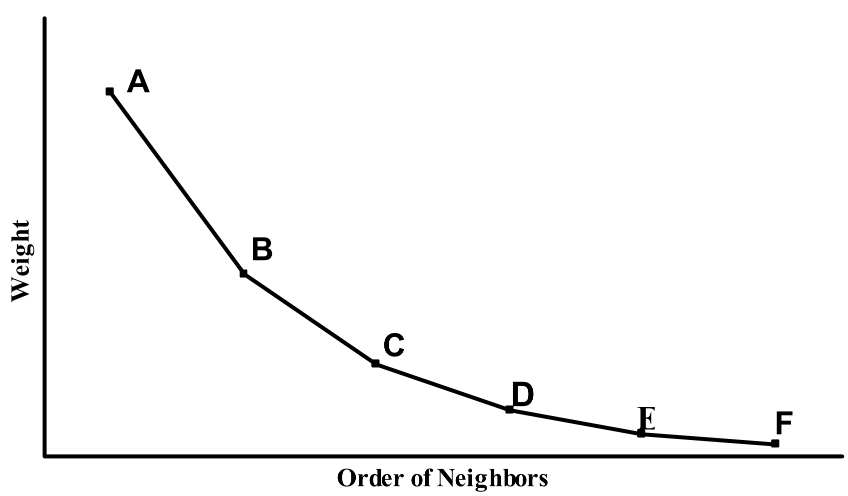
where:

y	=	vector dependent observations,
X	=	matrix of explanatory variables,
β	=	vector of regression coefficients,
ε	=	vector of error terms,
$D\varepsilon$	=	spatial lag for error terms,
α	=	Autoregressive coefficient, and
ξ	=	error term.

Again, D is an n by n weighting matrix with 0's on the diagonal. In this spatial autoregressive model α is restricted to lie within the interval [0,1], and the errors ξ are independently and normally distributed.

According to Anselin and Hudak, the first stage to implement a spatial econometric strategy is the construction and estimation of the weight matrix, given the spatial arrangement of the observations. While a spatial weight matrix may take on several different forms, two are of concern for this study. The first is the Delaunay triangular fixed weight matrix of a symmetric form that leads to a variance covariance matrix that is dependent upon the autoregressive parameter α alone. The other is a flexible spatial weight matrix. The nearest neighbor method is a flexible weight matrix that assumes that spatial dependence depends on a decay relationship and the number of neighbors. Specifically, the nearest neighbor weight covariance matrix is an asymmetric matrix that depends upon three parameters; α, the autoregressive parameter; m, the number of neighbors; and ρ, the rate weight decline, also referred as the decay parameter.

Figure 1.
Hypothetical decay relationship of decay weights according to the order of neighbors.



The conceptual relationship embedded in the nearest neighbor method is illustrated in Figure 1. Weight in the vertical axis represents the weight given by ρ to the power of the order of neighbors. For this example, let's assume that ρ=0.5 and *m*=6 neighbors (points A through F). Therefore, a ρ of 0.5 indicates that the first neighbor will give half the weight of the first neighbor (point A), the second neighbor a quarter of the weight of the first neighbor (point B), and so on. Nearest neighbor point F in Figure 1 does not significantly influence a given observation.

DATA

Data for this study are based on rural land market sales for the southeast area of Louisiana that were collected using mail survey techniques. The study area was estimated from previous research using multivariate procedures (Kennedy, et al.). These data represent a subset of a larger database collected for the state of Louisiana for the period January 1993 through June 1998 (Vandeveer, et al.). The rural land market survey was mailed to state certified appraisers, officers in commercial banks, Farmers Service Agency personnel, Federal Land Bank personnel, Production Credit personnel, members of the Louisiana Chapter of the American Society of Farm Managers and Rural Appraisers, and members of the Louisiana Realtors Land Institute.

The study area consisted of eight parishes in southeast Louisiana. A total of 257 rural land sales were included in this study. Each rural land sale was ten acres or more in size, included attachments to the surface such as buildings and other improvements, and located outside major metropolitan areas in Louisiana. Rural land sales ranged from 10 acres to 1,250 acres in size and per acre values ranged from \$475 to \$12,908 per acre.

Table 1
Variables used in hedonic model estimation, Southeast area, Louisiana rural land market survey, 1993- 1998.

Variable	Description	Expected Sign
Continuous Variables		
PRICE	Per acre price of land (\$)	
SIZE	Size of the tract (acres)	(-)
VALUE[a]	Value of improvements (\$)	(+)
RDNT	Road distance to nearest town (miles)	(-)
TTNC	Travel time to nearest city (minutes)	(-)
TIME	Month of sale	(+)
Discrete Variables (1,0)		
RT	Paved road access	(+)
REC	Influence for purchase: recreational	(-)
NOMSA	New Orleans MSA	(+)

[a] Value of improvements is a transformed variable (the number in dollars to the 0.01 power).

Variables hypothesized to influence per acre rural land values are defined in Table 1. PRICE in Table 1 is the dependent variable used in the hedonic model and represents the per acre selling price for each tract of rural land and improvements. Continuous variables expected to have an inverse relationship with per acre selling price include size of tract (SIZE), road distance to the nearest town (RDNT), and travel time to the nearest city (TTNC). There is generally a negative relationship between size of tract and per acre selling price because fewer buyers compete in markets for larger tracts; whereas, many buyers compete in markets for smaller tracts. For locational variables including travel time, location theory generally suggests an inverse relationship between distance to markets and per acre selling prices. TTNC and RDNT were computed using the Street Atlas USA computer software.

Continuous variables expected to positively influence rural land values include value of improvements (VALUE), and the time of sale (TIME). These variables represent positive attributes of rural land and hence, are hypothesized to have a positive influence on per acre rural land values.

The discrete recreational influence variable (REC) is hypothesized to have a negative relationship with per acre land values because much of the data in this analysis represents marginal marshland and upland well suited for hunting, trapping, and other outdoor uses, but with limited development potential. Paved road access (PR) is expected to have a positive effect on per acre land values. Finally, tracts located close to New Orleans metropolitan statistical area (NOMSA) are expected to have a higher value.

EMPIRICAL RESULTS

Prior to developing hedonic models of the rural land market area, data were tested for spatial autocorrelation. In this study, spatial autocorrelation occurs if the price variable is correlated with itself over space. Knowledge of spatial autocorrelation is of concern because its presence means there is interdependence in the data, whereas most statistical methods assume independence in the data. Ignoring spatial autocorrelation in a hedonic analysis of real estate values may result in inefficient and biased econometric results.

Diagnostic statistical tests were used to test for spatial dependence and to identify the correct model (i.e., spatial lag or spatial error) for estimation (Anselin). For this analysis, Lagrange multiplier (LM) tests were used to test for spatial dependence. Both Lagrange multiplier tests had high values and were statistically significant meaning that there was spatial autocorrelation in the data. The spatial lag model had the highest Lagrange multiplier value, indicating to be the appropriate model chosen for this study.

Once the spatial lag model was selected, the next step was to choose the value of ρ and m for the spatial weight matrix that best fit the data. A maximum likelihood spatial lag model and the nearest neighbor weight matrix was conducted using an iterative process with ρ ranging from 0.4 to 1 and m ranging from 1 to 30 and log likelihood values, for the models used, to evaluate the results. Results from the log likelihood values indicated that decay parameter of 0.5, along with 12 neighbors, best fit the data.

The results of the three hedonic models estimated are presented in Table 2. The first model represents estimates based on traditional ordinary least squares (OLS) procedures. The second model is based on maximum likelihood procedures using a fixed Delaunay weight matrix (from equation 4). The third model, also estimated using equation 4 is based on maximum likelihood procedures using the flexible asymmetric nearest neighbor matrix.

Table 2
Estimated coefficients[a] for Hedonic OLS and ML models with Delaunay matrix and nearest neighbor matrix, southeast Louisiana rural land market area, 1993-1998.

Item	OLS model	Spatial weight matrix for ML model	
		Delaunay	Nearest Neighbor (NN)
Variable			
α **(lag dependent variable coefficient)**		0.5400	0.4700
Ln SIZE	-0.2234 (62.5121)***	-0.2372 (79.0148)***	-0.2539 (85.9905)***
TIME	0.0081 (22.1438)***	0.0093 (34.5698)***	0.0100 (37.7033)***
NOMSA	0.6563 (53.1229)***	0.7310 (24.0079)***	0.7255 (26.5820)***
RT	0.1776 (10.7683)***	0.1609 (10.8837)***	0.1458 (9.1921)***
REC	-0.1687 (3.5690)**	-0.0923 (1.3821)	-0.0959 (1.5296)
RDNT	-0.0100 (10.9627)***	-0.0145 (11.1011)***	-0.0130 (8.8455)***
TTNC	-0.4336 (15.9144)***	-0.3631 (5.1631)**	-0.3602 (4.7276)**
VALUE	0.1833 (12.9911)***	0.2128 (20.9486)***	0.2362 (25.4714)***
INTERCEPT	8.6784 (642.1644)***	8.6913 (75.5003)***	8.7093 (150.2954)***
m			12
ρ			0.5
Log likelihood	-482.7292	-463.1123	-461.6046

[a] In the variables section, likelihood ratio values are in parentheses, ***denotes statistical significance at the 1% level, ** denotes statistical significance at the 5% level, and * denotes statistical significance at 10%.

Results presented in Table 2 indicate that all variables were found to be statistically significant in explaining per acre rural land values. When using the OLS model procedures, all the variables are statistically significant at the five percent level, and all variables had the expected sign. Results presented in Table 2 indicate that hedonic models estimated by maximum likelihood methods differ from the model estimated by OLS. Specifically, the variable recreational influence (REC) was found to be statistically significant in the OLS model, while this was not the case regarding the ML spatial models. These results suggest that using OLS model results could lead one to make incorrect conclusions regarding the effect recreational influences on rural per acre land values.

Log likelihood values presented in Table 2 were estimated to provide measures of goodness of fit for the estimated hedonic models. Log likelihood values are -463.11 and -461.60 for the Delaunay and the nearest neighbor models, respectively; which are larger than the same value for the OLS model (-482.73) indicating a better statistical fit of the data by the maximum likelihood models.

Although not shown in Table 2, the likelihood ratio test was used to evaluate the two spatial models. The results indicated a likelihood ratio[1] of 3.01 for the difference between the Delaunay and the nearest neighbor method. This result is statistically significant at the 10 percent probability level, suggesting the nearest neighbor better fits the data and is the better model for this analysis.

SUMMARY AND CONCLUSIONS

The general objective of this discussion was to develop a hedonic model of the land market in Southeast Louisiana and to compare and evaluate estimates based on a Delaunay versus a nearest neighbor spatial weight matrix. Spatial lag hedonic models were estimated using both a fixed Delaunay matrix and a nearest neighbor flexible weight matrix.

Tests from the OLS model indicated spatial dependence in the data. The benefits of using spatial models include improved prediction, better statistical inference through unbiased standard errors, and better estimates because of the way location is handled within the modeling procedure.

All hypothesized statistics tested in the hedonic models had the excepted signs and were statistically significant at the five-percent level, except for the recreational influence. OLS results indicate that the variable recreational influence (REC) was statistically significant in explaining per acre rural land values in the study area. However, when adjustments were made for spatial autocorrelation in the maximum likelihood spatial models, recreational influence was not significant. This suggests spatial autocorrelation in the data could have caused one to make erroneous conclusions concerning the effect of location on per acre land values in the study area.

Statistical tests indicated that spatial econometric models better fit the data than OLS. Additionally, with this particular set of data, log likelihood values suggest that the nearest neighbor spatial model is the appropriate model. Further research should continue to test for other forms of spatial weight matrices.

ENDNOTES

[1] The likelihood ratio test is calculated as two times the difference between the log likelihood value of the unrestricted and the restricted model that follows a chi square distribution with one degree of freedom.

REFERENCES

Anselin, Luc. "SpaceStat Version 1.80 User's Guide." Regional Research Institute, West Virginia State University, 1995. Danielson, Leon E. "Using Hedonic Pricing to Explain Farmland Prices." The Farm Real Estate Market, Proceedings of a Regional Workshop, Southern Natural Resource Economic Committee, p. 57-74, 1984.

Danielson, Leon E. "Using Hedonic Pricing to Explain Farmland Prices." *The Farm Real Estate Market, Proceedings of a Regional Workshop.* Southern Natural Resource Economic Committee. (1984):57-75.

Kennedy, G., S. Henning, L. Vandeveer, and M. Dai. "Multivariate Procedures for Identifying Rural Land Submarkets," *Journal of Agricultural and Applied Economics*, 29(1997):373-383.

Lesage, J. and R. K. Pace. "Using Matrix Exponentials to Explore Spatial Structure in Regression Relationships." Manuscript available at: http://www.spatial-statistics.com. 2002.

Pace, R. K. Spatial Statistics Toolbox 2.0. Available at: http://www.spatial-statistics.com. 2002.

Pace, R. K., R. Barry, and C. F. Sirmans. "Spatial Statistics and Real Estate," *Journal of Real Estate Finance and Economics,* 17(1998):5-13.

Pace, R. K. and R. Barry. Spatial Statistics Toolbox 1.1. Manuscript available at: http://www.spatial-statistics.com. 2002.

Pace, R. K., R. Barry, O.W. Gilley, and C.F. Sirmans. "A Method for Spatial-Temporal Forecasting with an application to Real Estate Prices, *"International Journal of Forecasting,"* 16(2000):229-246.

Palmquist, R. B. "Heterogeneous Commodities, Hedonic Regressions, and the Demand for Characteristics." The Farm Real Estate Market. Proceedings of a Regional Workshop, Southern Natural Resource Economics Committee, 1984:45-56.

Rosen, Sherwin M. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, 82(1974): 34-55.

Vandeveer, L., S. Henning, and G. Kennedy. "An Economic Analysis of the Rural Real Estate Market in Louisiana," Hatch Research Project LAB03421, 1999-2004.

Vandeveer, L., P. Soto, and H. Niu. "A Statewide Spatial Analysis of the Effects of Location and Economic Development on Rural Land Values." *Southwestern Economic Review*, 29(Spring 2002):1-20.