Understanding the economic benefits of trails on residential property values in the presence of spatial dependence

June 13, 2011

Abstract

This paper examines the impacts of a multi-purpose trail on residential property values in a hedonic model. Using a large housing data set in combination with street network distances we show that proximity to trail entrances positively effects property values. Among other things, our study compares the hedonic model results from three different spatial specifications. We pay specific attention to the direct and indirect effects on residential property prices associated with potential changes in house characteristics. In addition, our study predicts property values around trail entrances using a 'modified spatial predictive process' approach that is well suited for capturing spatial dependence in large data sets.

KEYWORDS: hedonic models, spatial modeling, spatial predictive process.

JEL:C11,C21,R21,

1 Introduction

It is well documented in the relevant literature that open spaces, such as parks, urban forests, greenbelts, and multi-purpose trails, make communities more livable, provide opportunities to improve people's physical and mental health, and can boost local economies through tourism (Lindsey at al. 2004). For the "New Urbanist," multi-purpose trails provide the potential for bicycle commuting; help alleviate noise, pollution, and congestion, and, of course, expand the means for green transportation and a community's walkability. From a real estate perspective, trails, like many other amenities, can have significant spillover effects on residential property values when these properties are located within reasonable distances to the trails. But while an increasing number of studies in recent years explored the impacts of amenities onto residential property values, only very few studies are specifically devoted to trails.

Developed by Lancaster (1966) and Rosen (1974), the hedonic pricing model is the standard approach used to estimate the marginal implicit prices of individually selected housing characteristics. Housing price, based on people's willingness to pay for these intrinsic characteristics, is modeled as a function of a set of utility-bearing intrinsic properties that constitute it. However, the fact that home buying decisions are not solely based on structural features of the real estate (e.g., square footage, number of bedrooms), but are also influenced by community (e.g., school district), neighborhood (e.g., median household income), environmental (e.g.,traffic noise, air pollution), and locational (e.g., access to public transportation, distance to CBD, public parks) attributes, led to the development of a vast body of literature over time to account for this array of additional factors that influence house prices.

In addition to accounting for the contribution of non-structure related factors, this paper specifically focuses on the impact of the multi-purpose Little Miami Scenic Trail on neighboring house prices. Even though the literature is somewhat inconclusive about the impacts of trails on house prices (Mogush et al., 2005), we find that the Little Miami Scenic

Trail does have a positive impact on residential property values within close proximity, when using street network distances. We prefer street network distances between the residential properties and the closest trail entrance over the often used straight-line distances because potential bicyclists and pedestrians are most likely to travel along the street network to access the trail via the closest trail entrance. In addition to the standard parameter estimation process, we use our hedonic model results to predict the market values for all residential properties around these trail entrances.

While numerous studies do account for spatial dependence between locations when estimating the hedonic model parameters and when using these results to predicting real estate prices (Gelfand et al., 2004; Valente et al., 2005), many results presented in the relevant literature are somewhat restrictive as they have been derived using small data samples. Even though the use of a sparse spatial weight matrix tremendously decreases the computational complexity of spatial econometric models (see LeSage and Pace, 2009), many studies nevertheless give preference to explicitly modeling the decay in spatial correlations between locations through use of various functional forms (see Cressie, 1993). Until recently, these processes were rather restrictive in that they were applicable only to relatively small data sets. To overcome the data limiting factor, we apply the "modified" spatial predictive process as proposed by Banerjee et al. (2008) and as extended by Finley et al. (2009). Reducing the number of observations to a significantly smaller number of representative locations, we can overcome the computational challenge of our large data set during the estimation procedure. Through interpolation techniques, estimated parameters can be predicted from the representative smaller sample back to our large data set. After the parameter estimation, we apply the same predictive process to create a contour map consisting of predicted house prices around the trail entrances. To assess the quality of this estimation procedure, we will compare them with the results from more traditional specifications in the spatial econometric literature.

The outline of the paper is as follows. In Section 2, we discuss the relevant and still sparse literature on trails and greenbelts and their impacts on house prices within the hedonic framework. Section 3 presents the study area and the data set and Section 4 discusses in greater detail the Spatial Process as well as other widely used alternative methods for handling hedonic pricing models with georeferenced data sets. In Section 5, we discuss and compare all model results from our four empirical hedonic specifications. Finally, in Section 6 we present the contour map of predicted property values around the trail entrances.

2 Review of relevant literature on open spaces, trails, and greenbelts

There is a long list of factors that potential homebuyers take into consideration when looking for a desirable location to buy a house. The quality of the school district (Brasington, 1999; Clapp et al. 2008), the availability of public transport infrastructure (Hess and Almeida, 2007), and desirable neighborhood characteristics (Lynch and Rasmusson, 2001) have received much attention lately in the relevant literature. In addition, open spaces, e.g., parks, golf courses, water bodies,and trails among others, can, if regarded as an desirable amenity, add to residential property values. Assuming that potential homebuyers are willing to pay a premium for residential properties that are in close proximity to a park or a multi-purpose trail, their intrinsic values are included in property prices and can be estimated.

There is consensus in the relevant literature that parks and open spaces have significant effects on residential property prices as they provide improved access to recreational and fitness activities, protect ecosystems, wildlife, and watersheds, or are just of pure aesthetical value. In addition, parks and open spaces can provide the means to improve distressed housing markets, thereby generating additional property tax revenues, and therefore park and trail developments are viable investment strategies for improving the quality of life in cities in general. The fact that the effects of open spaces on residential property values vary widely by type, usage, size, and distance opens the gateway to a large body of literature. Of further importance is that the effects of open spaces co-vary with neighborhood characteristics such as population density, income, and crime (Anderson and West, 2006).

However, very few studies attempt to assess the relationship between trails and residential property values within the hedonic price framework. In a recent study, Asabere and Huffman (2009) find that the impacts on home values resulting from trails, greenbelts, and trails with greenbelts are 2%, 4% and 5% respectively. Using a semi-logarithmic functional form in a non-spatial hedonic framework and qualitative predictors for the presence of trails and greenbelts, the authors show that house prices increase the most when greenbelts are used to buffer trails. Krizek (2006) argues that different types of bicycle facilities have different amenity values. Accordingly, the author explicitly distinguishes between three different types of bike trails: on-street bicycle lane, off-street bicycle trail (multi-purpose paths including rail trails), and roadside bicycle trail and between the city and its suburbs. Krizek's hedonic model does also not account for any spatial dependence, but does include distance measures for each home to the nearest trails. He concludes that suburban residents do not value bicycle facilities as a favorable amenity in Minneapolis-St. Paul, while off-street bicycle trails appreciates home values in the city. However, city homeowners regard roadside bicycle trails as a nuisance, thus having a negative impact on house prices. However, Mogush, Krizek, and Levinson (2005) explain this rather unexpected finding with the quantity and speed of the adjacent road traffic. Lindsey et al. (2004) in a semi-log, non-spatial hedonic model use a straight-line approach to identify properties that fall within a mile buffer zone around the trails included in their study. Supported by survey data, the authors argue that trail users beyond 1/2 mile Euclidean distance from the trail are more prone to drive to the trail. Of particular interest for our study are the results for the Moron Trail, which like the Little Miami Scenic Trail, is a heavily used converted rail-trail that runs from the center of the city north into the neighboring county. For properties located within 1/2 mile of the Moron Trail in Indianapolis, Lindsey et al. show that using mean values for all variables, a total of 14 percent ($13,056) of a predicted sales price of $93,283 is attributable directly to the Moon Trail. Altogether, this translates into a combined premium of $115.7 million in property values for the homes within one-half mile of the Moron Trail. A study by Nichols and Crompton (2005) is of relevance as it compares different proximity measures. Using a linear, non-spatial hedonic approach, physical proximity between properties and the greenbelts was established using street network distances and buffer zones based on street network distances. Though the results are not conclusive in that similar trends emerged for all three study areas, one can conclude that using network distances is superior to just using buffer zones. For instance, using network distances, Nichols and Crompton show that house prices fall by \$3.97 for every foot one moves away from the trail, the regression results become inconclusive when using buffer zones in that the results change widely with the established distance measures (i.e., $0 - 1/4$ mile, $1/4 - 1/2$ mile, $1/2 - 3/4$, and $3/4 - 1$ mile buffer zones).

Our study will significantly add to the literature on trails and greenbelts in that it is the first study that explicitly accounts for the phenomena of spatial dependence in house prices. Further, it discusses three different spatial modeling techniques and presents the corresponding results. Last, we contribute to the existing literature by predicting potential values of residential properties around the trail entrances and presenting them in a contour map.

3 Study Area, Data Sample, and Research Design

Our study area is the Little Miami Scenic Trail, a shared multi-purpose trail with equal rights for hikers, runners, skaters, bikers, and equestrians. Though the entire trail extends about 78 miles from the Little Miami Golf Center in Newton, Hamilton County, to Springfield, Clark County, our study focuses on the 12 miles most southern stretch, which lies in Hamilton County, the core county of the City of Cincinnati (Figure 1).

The section of the Little Miami Scenic Trail under study contains a total of 23 trailheads where recreational users can enter the trail. The Little Miami Scenic Trail is considered to be one of the main recreational facilities within the Cincinnati Metropolitan region, as a survey by the Friends of The Little Miami State Park, Inc., a non-profit organization with focus on updating and beautifying the trail, indicates. In just two days in July and August of 2010, a total of 4,979 users were counted on the trail at Loveland and 2,374 users at Milford.[1] The popularity of the trail suggests that in accordance with the hedonic price theory, its amenity value may be reflected in the form of a marginal price-the willingness of homeowners in close proximity of the trail to pay a premium for being close to the trail. The hedonic pricing technique (Lancaster (1966) and Rosen (1974)) is therefore the preferred method to estimate the marginal implicit (i.e., hedonic) prices of individually selected housing characteristics. Based on the notion that utility can be derived from commodities' intrinsic characteristics, the proposed framework allows us to explicitly estimate people's willingness to pay for these individual housing characteristics, including the proximity to the Little Miami Scenic Trail. Our data sample contains data for 1,762 single-family residential properties for the year 2005. The housing data were obtained from the Hamilton County Auditor's website and include (actual) sales prices as well as market values and structural characteristics of the properties, such as size (SQFT), age (AGE), and number of bedrooms/bathrooms. For this study, true market values are used instead of sales prices because not all sales prices were recorded for all properties included in the study. In fact, only 191 properties were sold in 2005. A large number of properties have not had recorded transactions in recent years and therefore, no up to date sales prices were available for these properties. For the purpose of this study, the use of market values is more appropriate because a larger sample size of $n = 1,762$ allows the construction of a weight matrix that includes a total of 10 neighbors for each property and, as such, goes beyond the more simplistic approach of only including the nearest neighbor in the weight matrix. Further, a larger sample size is essential for a precise prediction of all housing values in the study region (see section 6). In addition, Ventolo and Williams (1994) argue that the market value is the highest price that a property is to sell for in an open market, within a reasonable time frame. In other words, the true market value should come close to actual sales prices for all arm's length transactions and is generally representative of the sales price.[2] The housing data were

[1]'Friends' help out scenic Little Miami trail: Loveland - For supporters of the Little Miami Scenic Trail, their charge borders on a sacred trust, Cincinnati Enquirer August 21, 2010. The two days of count were Wednesday, July 28 and Sunday, August 8 2010.

[2]Not to be confused with the assessed value used for property tax calculations which in Ohio is simply

supplemented with data from the Ohio Department of Education, the Ohio Department of Transportation, and the Cincinnati Area Geographic Information System (CAGIS). Guided by the principle of parsimony, only statistically significant explanatory variables have been retained within the four developed hedonic pricing models. This is done to isolate the most important explanatory variables making it easier to describe the processes under study. The twelve most relevant variables explaining the variation in house prices that remained in our analysis are presented in Table 1 along with some descriptive statistics in Table 2.

Table 1: Data sources and definition

Variable	Description
PRICE	Market value of land and improvements in 2005 (Hamilton County Auditor)
TRAILD	Network distance between each property and the nearest trail entrance in feet (calculated using ArcView)
INC	Median household income by Census block group (Census Bureau, 2005)
SQFT	Finished square footage of the house (Hamilton County Auditor)
AGE	Age of the house in years (Hamilton County Auditor)
LOTSIZE	Lot size of the property in square feet (Hamilton County Auditor)
BASEMENT	Dummy variable denoting a full basement (Hamilton County Auditor)
BRICK	Dummy variable denoting exterior brick walls (Hamilton County Auditor)
FIRE	Dummy variable denoting at least one fireplace (Hamilton County Auditor)
MATH	State of Ohio 9^{th} grade math section proficiency test percentage passage rate for the 2005 school year (Ohio Department of Education)
TAXR	Gross tax rate by school district for the 2000 tax year (Ohio Department of Transportation)
NONRES	Percent of 2000 total property value by school district that is classified as: mineral, industrial, commercial, and railroad (Ohio department of Transportation)
CBDDIST	Shortest distance to Downtown Cincinnati (calculated using ArcView)

The average house in our study area is about 40 years old, has on average 2,203 square feet of living space, and is built on a lot of 4,185 square feet (0.096 acres). About 15% of the properties are built of bricks, 43% have a full basement, and 72% have a fireplace. With respect to the 2000 gross tax rate by school district, the mean value was 71 mills, with a minimum of 47 mills and a maximum of 85 mills. Local school achievements have

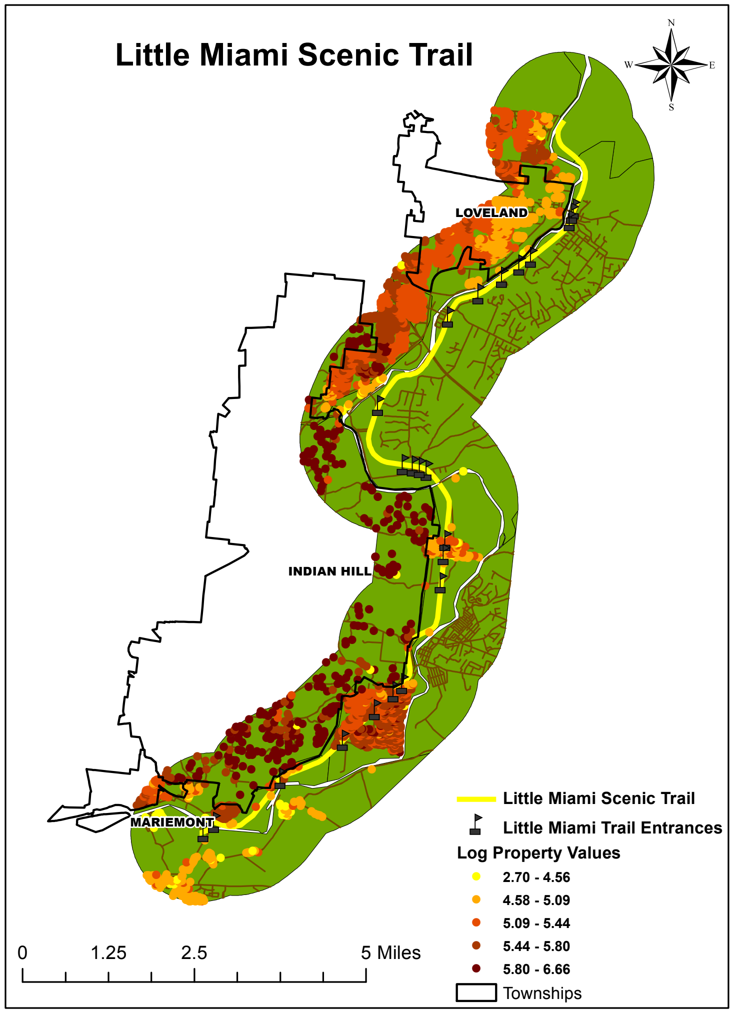
defined as 35 percent of the true market value.

Table 2: Data - Summary statistics

Variable	Mean	S.D.	Min	Max
PRICE	263517.82	280043.12	25500.00	3448600.00
TRAILD	5772.39	2560.28	137.62	9882.53
INC	83602.96	34525.03	35417.00	191974.00
SQFT	2202.91	1177.90	525.00	13235.00
AGE	40.43	26.02	2.00	174.00
LOTSIZE	4185.01	13776.71	687.00	440067.00
BASEMENT	0.432	0.495	0.00	1.00
BRICK	0.15	0.36	0.00	1.00
FIRE	0.72	0.45	0.00	1.00
MATH	90.91	4.17	87.10	98.30
TAXR	70.58	12.36	47.10	85.15
NONRES	14.98	10.30	6.62	39.55
CBDDIST	22.34	4.82	12.81	29.02

been accounted for by the 9th grade math section proficiency passage rate for the 2005 school year, which has a mean of 91 percent, compared to 84.5 percent statewide. Another important component of our regression models is the percentage of property value in 2000 that is classified as mineral, industrial, commercial, and railroad real estate. These non-residential properties make up on average 15 percent of all property values by school district in our sample data. Further, the average household income by Census block group is $86,603. Using the CAGIS data in conjunction with the housing data from the Hamilton County Auditor, two distance variables were generated within ArcInfo 9.3 (ESRI). First, the distances between all single-family residential properties and their nearest trailhead was calculated. A reasonable cut-off point of 10,000 feet was used (i.e., 1.89 miles), which gave us a total of 1,762 residential properties. The choice of network distances over buffer zones or straight-line distances was made to account for the fact that trail users are most likely to follow the street network to the nearest trail entrance. Network distances were calculated within ArcInfo's Network Analyst. Using the "closest facility" command, the trailheads were uploaded as facilities and the residential properties as incidents. Executing the "solve" command in a second step then identified the shortest routes and calculated the

Figure 1: Observed housing prices



distances from each property to the closest trailhead. Hierarchies among the street network dataset were used to disallow the use of interstates and highways by bicyclists, though we allow bicyclists to ride one way streets in the wrong direction. Following the notion of the location rent model (Chesire and Sheppard 1995) in that property values decline with increasing distance from the Central Business District, straight-line distances for each of the 1,762 properties to the CBD, i.e., downtown Cincinnati, was calculated and included as an explanatory variable.

4 Model Specifications and Comparison

The main objective of the paper is to explain the variation in housing prices around the Little Miami Scenic Trail. The empirical hedonic model specification used in presented research is:

$$\ln(PRICE) = \beta_0 + \beta_1(SQFT) + \beta_2(AGE) + \beta_3(LOTSIZE) + \beta_4(BASEMENT) + \beta_5(BRICK) + \beta_6(FIRE) + \beta_7(MATH) + \beta_8(TAXR) + \beta_9(NONRES) + \beta_{10}(INCOME) + \beta_{11}(CBDDIST) + \beta_{12}(TRAILD) + \epsilon \quad (1)$$

where PRICE refers to the market values of the included single-family residential properties and the explanatory variables are defined as in Table 1 above. We adopted the semi-logarithm (log-linear) functional form; partly due to its dominance in the relevant literature and partly to control for the large variation in house prices. To avoid multicollinearity problems, some of the highly correlated explanatory variables were excluded from the models. Altogether, twelve explanatory variables remained in our empirical model specification containing six structural housing characteristics (i.e., square footage, age, lot size, basement, brick construction, and fireplace), three school district variables (i.e., math proficiency test results, school district tax rate, percent of non-residential property values), two neighborhood variables (i.e., median household income, distance to CBD), and the network distance between residential properties and the Little Miami Scenic Trail. Altogether, we used four different Bayesian estimation techniques, namely Ordinary Least Square (OLS), Spatial Autogressive Regression (SAR), Spatial Error Model (SEM), and Spatial Process (SP). While we will present all results from all four model specifications, we will pay specific attention to the impact the Little Miami Scenic Trail has on property values. The OLS model follows closely the standard hedonic pricing approach and as such does not account for spatial dependence of any kind.

Spatial dependence is based on the fact that economic actors (buyers, sellers, and realtors) take the values of neighboring residential properties into consideration when pricing a property. Though each property differs with respect to structural characteristics, each house shares with its neighbors those influences that are generated from almost identical "location" factors. Accordingly, nearby properties tend to be more similar than those that are located further away. In practical terms, this logic of a spatial autoregressive structure is implemented using a spatial weight matrix W that identifies neighboring observations. For presented research, we chose a row normalized weight matrix W based on the 10 nearest neighbors. Further, we define a location index s for each property which varies continuously over D, the set of all possible locations in our study region. We define housing prices as $y(s)$ for all properties in our finite set of locations $s_1, s_2, \ldots, s_n$. One of our main focii of this research is to define different measurements for the covariance $Cov(y(s_i), y(s_j)) = C(h)$, where h is the distance between site s_i and s_j. The interpretation of the spatial dependence as a consequence of omitted variables-a structural process-is the foundation of the Spatial Autoregressive (SAR) model (LeSage and Pace, 2009). More specifically, latent influences not included in the study (i.e., omitted variables), such as lack of privacy along the trail, noise and crime issues, and/or the perceived distance to the trail, could influence residential property values. Given that it is very unlikely to account for all possible influences on property values, the SAR approach does account to some extent for these unobservable factors. The Spatial Autoregressive is expressed in its matrix form as:

$$Y = \rho WY + X\beta + \epsilon, \tag{2}$$

where $Y = (y_{s_1}, \ldots, y_{s_n})$ is an $n \times 1$ vector containing the housing prices, W is the $n \times n$ spatial weight matrix, β is the $k \times 1$ vector of parameters to be estimated and $X = (X_{s_1}, \ldots, X_{s_n})'$ is the $n \times k$ matrix of explanatory variables, including an intercept term. Each error term ϵ is normally and identically distributed with a zero mean and a variance σ^2. The scalar ρ measures the strength of the spatial dependence.

Through spatially structured random effects in the disturbance process, the Spatial Error Model (SEM) assumes spatial autocorrelation only in the unobserved random part of the specification. The SEM in matrix notation is defined as:

$$
\begin{aligned}
y &= X\beta + u, \\
u &= \lambda W u + \epsilon.
\end{aligned} \tag{3}
$$

where λ measures the strength of the spatial dependence in the spatial lag of the error terms and each error term ϵ is normally and identically distributed with a zero mean and a variance σ^2.

A more complex spatial dependence structure is introduced in the Spatial Process (SP). As Valente et al. (2005) emphasized, standard spatial econometric techniques (i.e., SAR and SEM) can have practical limitations as the correlation between locations is solely dependent on the initial definition of W and the inversion of W during the estimation procedure. The Spatial Process, on the other hand, is more flexible in that it allows modeling of complex spatial dependence structures beyond the standard approach. In our research, we followed geostatistical modeling techniques. Our spatial process model is of the following functional form:

$$
y \sim N(X\beta, \tau^2 I_N + \sigma^2 H(\phi)), \tag{4}
$$

where τ^2 is called the nugget, σ^2 the partial-sill, and ϕ the decay parameter. The spatial connectivity structure is embedded in the covariance function $C(h) = \sigma^2 H(\phi)$. More specifically, the $n \times n$ spatial correlation matrix H is defined by an isotropic function which depends only on the distance between locations, but is independent of the direction. In fact, the covariance function $C(h)$ between any two locations s_i and s_j depends only on the separation vector h. We model the covariance function using the Matérn form which is defined as:

$$C(h) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(2\sqrt{\nu}h\phi)^{\nu}K_{\nu}(2\sqrt{\nu}h\phi) & \text{if } h > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases} \tag{5}$$

where the parameter ν controls the smoothness of realizations and ϕ is the spatial range parameter. The functions $\Gamma(.)$ and K_ν are the gamma function and the modified Bessel function of order ν, respectively. Depending on the parameter ν, the Matérn form can encompass different classes of covariance functions, including the exponential covariance function ($\nu = 1/2$) and the Gaussian covariance function ($\nu \to \infty$). The Matérn covariance function is crucial for the described Spatial Process (SP) in that it allows the estimation of the smoothness parameter ν. The differentiability of this function influences widely the outcomes of predicted. Thus, choosing a suitable information prior for ν for the Bayesian estimation procedure is of major importance. Despite its more complex functional specification, the SP's significant advantage over the SAR and SEM models is that it allows the user to estimate the distance over which the spatial correlation is defined. The use of a separate covariance function, however, increases the computational complexity within the SP specification, as the spatial dependence between locations is not defined anymore through a sparse spatial weight matrix (see LeSage and Pace, 2004). More specifically, the number of necessary computations increases to n^3 for n observations. To overcome indicated computational hurdle, we follow closely Banerjee's (2008) spatial predictive process approach for large data sets. The basic idea of the predictive process is to reduce the dimension of the covariance matrix by selecting a smaller data set with m representative observations from the originally observed n data points. The key is to select a small enough data set to simplify the computational process, while containing enough information to estimate the underlying spatial process for the full data set of n observations. We will refer again to this selection process of representative locations in the section on predictions.

We have given preference to the Bayesian estimation procedure over more traditional estimation procedures, for instance Maximum Likelihood, as these procedures cannot estimate the smoothness parameter ν in the spatial process (see LeSage and Pace, 2009, for further information about the estimation procedure). The remaining three models were then estimated within the Bayesian framework to maintain comparability among all individual model specifications. For the Bayesian estimation process, the following hierarchical specifications were applied: the parameters β are all normally distributed, but non informative. The measures for the strength of the spatial dependence, λ and ρ, as well as the range parameter ϕ, follow uniform distribution. We assigned inverse gamma priors for σ^2 and τ^2. Of course, all parameters are assumed to be independent. For assigning a prior for the smoothness parameter ν in the Matérn correlation function, we were guided by the fact that data seldom suggest a prior for ν of orders greater than 2 and accordingly assigned a uniform prior distribution of (0, 2) for ν. We further follow closely Banerjee et al. (2008) and define a lower cut-off value of 0.05 in the variance-covariance matrix for the effective range of the spatial dependence. In other words, correlations of less than 0.05 do not suggest significant spatial dependences and are therefore replaced by zeros in the matrix. Lastly, we implement a vague prior on ϕ with a uniform distribution is defined on the interval (0.5,30) and which corresponds to an effective spatial range between 100 and 6,000 feet for $\nu = 0.5$. We use the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002) as our Bayesian model selection criterion to compare our four presented models with each other. The DIC is well suited whenever Markov Chain Monte Carlo simulations were used to obtain posterior distributions of the models. As such, the DIC is easily calculated from posterior samples and should be used only with Gaussian likelihoods such as described here. The DIC is defined as the sum of the deviance (a measure of model fit) and the effective number of parameters (a penalty term for model complexity). Lower DICs indicate better model performance and are preferred.

5 Analysis and Discussion of Empirical Results

A cross-comparison of the DIC across all four empirical models highlights the superior overall model fit of the spatial process which has by far the lowest DIC with -2,700.4

Table 3: Estimation Results for OLS, SAR, SEM and SP - 20,000 iterations

	SP		SEM		SAR		OLS	
Parameter	Mean	p-value	Mean	p-value	Mean	p-value	Mean	p-value
Constant	7.744	0.000	6.241	0.000	4.307	0.000	7.046	0.000
	(7.096, 8.395)		(5.559, 6.922)		(3.805, 4.848)		(6.538, 7.530)	
TRAILD	-3.40E-05	0.000	-2.20E-05	0.000	-8.80E-06	0.005	-1.000E-05	0.000
	(-4.22E-05, -2.57E-05)		(-3.10E-05, -1.20E-05)		(-1.40E-05, -3.10E-06)		(-1.55E-05, -4.22E-06)	
INC	2.65E-06	0.000	2.00E-06	0.061	-6.60E-08	0.400	2.000E-06	0.064
	(2.06E-06, 3.24E-06)		(1.20E-06, 2.60E-06)		(-5.00E-07, 3.80E-07)		(1.45E-06, 2.36E-06)	
SQFT	3.23E-04	0.000	3.30E-04	0.000	3.00E-04	0.000	3.880E-04	0.000
	(3.06E-04, 3.40E-04)		(3.10E-04 , 3.40E-04)		(2.90E-04 , 3.10E-04)		(3.72E-04, 4.04E-04)	
AGE	-2.83E-03	0.000	-2.30E-03	0.000	-1.80E-03	0.000	-0.002	0.000
	(-0.003, -0.002)		(-2.90E-03, -1.80E-03)		(-2.20E-03, -1.40E-03)		(-0.001,-0.002)	
LOTSIZE	4.86E-06	0.000	4.00E-06	0.000	4.00E-06	0.000	1.100E-05	0.000
	(3.77E-06, 5.94E-06)		(3.10E-06 , 6.00E-06)		(3.30E-06 , 4.70E-06)		(9.47E-06,1.33E-05)	
BASEMENT	0.038	0.005	0.052	0.000	0.021	0.040	0.046	0.000
	(0.009, 0.067)		(0.031, 0.072)		(0.002, 0.041)		(0.025,0.066)	
BRICK	0.033	0.155	0.03	0.030	0.051	0.000	0.011	0.212
	(-0.007, 0.073)		(0.004, 0.056)		(0.026, 0.075)		(-0.0134,0.037)	
FIRE	0.280	0.000	0.201	0.000	0.213	0.000	0.223	0.000
	(0.241, 0.320)		(0.169, 0.234)		(0.186, 0.241)		(0.193,0.252)	
MATH	0.031	0.000	0.048	0.000	0.033	0.000	0.034	0.000
	(0.025, 0.037)		(0.04 , 0.054)		(0.028, 0.037)		(0.029,0.0391)	
TAXR	0.010	0.000	0.007	0.000	0.005	0.000	0.011	0.000
	(0.009, 0.011)		(0.005, 0.009)		(0.004, 0.006)		(0.010,0.012)	
NONRES	0.001	0.200	0.005	0.000	0.002	0.023	0.003	0.000
	(-0.001, 0.002)		(0.003, 0.006)		(0.001, 0.004)		(0.002,0.004)	
CBDDIST	0.002	0.174	0.01	0.000	0.006	0.001	0.004	0.004
	(-0.002, 0.006)		(0.004, 0.015)		(0.003, 0.009)		(0.001,0.008)	
σ^2	0.040	0.022	0.037	0.000	0.021	0.000	0.027	0.000
	(0.031, 0.048)		(0.034, 0.041)		(0.02 , 0.023)		(0.025,0.029)	
τ^2	0.039	0.000						
	(0.031, 0.049)							
ϕ	12.920	0.039						
	(2.028, 28.250)							
ν	1.168	0.017						
	(0.555, 1.893)							
λ			0.538	0.000				
			(0.479, 0.583)					
ρ					0.298	0.000		
					(0.260, 0.331)			
DIC	-2700.4		-2690.7		-2691.3		-2659.9	

95% HPDI in parentheses

(Table 3) and as such is our preferred model.[3] The SEM and the SAR model perform very

[3]Markov Chain Monte Carlo estimation results are based on a simulated chain where the first 5,000 samples are discarded as a 'burn-in' period, followed by 15,000 iterations that were collected to produce posterior summaries for the parameters of interest.

similar as indicated by their DIC of -2,690.7 and -2,691.3 respectively. All three spatial model variants outperform the standard OLS model, which has the highest DIC with -2,659.9.

Turning to the parameters for the structural, neighborhood, and school district variables in Table 3, we conclude that all estimated parameters have the expected signs. The median household income by Census block group variable, however, is statistically insignificant in the SAR model. In addition, three of the explanatory variables are statistically insignificant in the Spatial Process (SP) model: the exterior wall structure (BRICK), the percent of property value that is class 2 (NONRES), and the distance to the Central Business District (CBDDIST). However, the SAR results are not directly comparable to the SEM and SP results and precaution is thus necessary with respect to their interpretation. In fact the SEM and SP specifications do not account for spillover effects onto neighboring properties. As such, a change in an explanatory variable in the SEM and SP models leads to a change only in the price for one particular property. Because of the presence of the inverse matrix $(I_n - \rho W)^{-1}$ in the reduced form of the SAR specification, spillover effects onto neighboring properties are implicitly included when changing an explanatory variable for a given location as specified in the spatial weight matrix W. To account for these spillover effects, LeSage and Pace (2009) propose a scalar summary measure that is composed of a direct plus an indirect effect, which allow further insight on the magnitude of the feedback effects through the spatial connectivity structure. The direct, indirect, and total effects for the SAR model are shown in Table 4 below. For instance, adding 100 square feet to the footprint of the structure would directly increase its value by 3.1%, but also significantly impact the neighboring house prices by the magnitude of 1.2%.

Regarding the Little Miami Scenic Trail, we can conclude that access to the trail does have a significant effect on single-family residential property values as long as these lie within 10,000 feet network distance to one of the twenty-three trail entrances. For the Spatial Process model, an increase of one foot from a trail entrance means a reduction in house price of 0.0034 percent. For a property priced at the mean value of $263,518, this translates into a reduction of $8.96 for each foot increase in distance, or $8,960 for each 1,000 foot increase. For the same mean property value, the SEM and the SAR model show percent reductions in house prices of $5.80 and $3.43, respectively, for a one foot change in network distance corresponding to reductions of $5,797 and $3,426 in mean property values for 1,000 foot distance increase. Thus, for a house at a distance of 1,000 feet to a trail entrance, we estimated a reduction in value of 1.3-3.4 percent. With respect to the structural characteristics, we find that adding 100 square feet to the footprint of the mean-priced house adds between $7,906 and $8,696 (3.0-3.3 percent), depending on which of the three model results we use. A one year increase in age of house, on the other hand, reduces the house price by $474 to $746 (0.180-0.283 percent) and as such indicates a minimal influence of age on housing values. Also, the lot size is of lesser influence and adds a marginal $105 to $128 (0.04-0.05 percent) to the mean house price for adding 100 square feet to it. Further, a full basement adds as much as 2.1 to 5.2 percent to house prices, while an exterior brick wall adds another 3.0 to 5.1 percent to it. However, we see the fireplace result (i.e., 20.1 to 28.0 percent) with much reservation. Though results in this magnitude for qualitative indicator variable estimates are not unknown, their interpretation differs as they only refer to an upward or downward shift of the intercept. Confirming to prior expectations, the neighborhood and the school district are significant determinants of house prices. According to the spatial process, an increase in median household income by $10,000 adds as much as $6,983 to the mean house (2.65 percent). Using the State of Ohio 9th grade math test rate and the gross tax rate by school district as indicators for the quality of the school districts, our results show clearly a positive relationship with house prices. More specifically, a 1 point increase in the 9th grade test rate and a 1 mill increase in the gross school tax rate are equivalent to a 3.1 to 4.8 percent and 0.5 to 1.0 percent increase in house prices.

Our empirical results do further support the hypothesis of spatial dependence in house prices as shown in the results of the three spatial model specifications. Comparing the spatial strength parameters of the two spatial econometric models reveals that the strength of the spatial dependence λ in the SEM and ρ in the SAR model are significant and positive with estimated values of 0.538 and 0.298 respectively. With respect to the error structure of the Spatial Process, Table 3 shows that all four of the relevant parameters are statistically significant as well. The estimated partial sill (σ^2)-part of the spatial error component-is 0.040 at its mean and is statistically significant at the 5 percent level (p-value = 0.022). The nugget (τ)-the non-spatial error component-is 0.039 and statistically significant at the 1 percent level (p-value = 0.000). Comparing the partial sill to the nugget for $h = 0$, which refers to the main diagonal of $C(h)$ and denotes the variance in Y, implies that the spatial and non-spatial error components are approximately of equal importance. The parameter ν estimate, a reflector of the smoothness of the spatial process, is 1.168. Comparing this estimate of 1.168 to ν values of 0.5 for the traditional exponential distribution and of 2.0 for the Gaussian distribution, we conclude that our estimated covariance function is smoother than an exponential specification, but not as smooth as the Gaussian specification. The parameter ϕ, which controls for the decay in the spatial correlation, has an estimated value of 12.920. This estimate reveals a rather small range over which the spatial correlation is defined. For instance, using the two closed forms of the Matérn correlation function for $\nu = 0.5$ and $\nu = 1.5$, we obtain a range of 232 feet and 367 feet, respectively. Given the high level of heterogeneity in house prices that exists between the neighborhoods along the trail, this strong decay of spatial correlation was somewhat expected.

Table

6 House Price Prediction

The prediction of the housing price for any location in the study area using estimated predictors is quite intuitive from a Bayesian point of view. This spatial process is a widely used prediction technique as its estimation is based on a sampling scheme. In other words, the estimation of the full posterior predictive distribution for any desired location conditioned on the observed distribution can be accomplished through the use of a set of representative

Table 4: Direct and indirect effects for SAR model - 20,000 iterations

	Direct			Indirect			Total		
Parameter	2.5%	Mean	97.5%	2.5%	Mean	97.5%	2.5%	Mean	97.5%
TRAILD	-1.6E-05	-9.0E-06	-2.0E-06	-6.0E-06	-4.0E-06	-1.0E-06	-2.2E-05	-1.3E-05	-3.0E-06
INC	-4.8E-07	-6.3E-08	3.6E-07	-2.4E-07	-3.1E-08	1.8E-07	-7.1E-07	-9.4E-08	5.4E-07
SQFT	3.0E-04	3.1E-04	3.2E-04	1.0E-04	1.2E-04	1.4E-04	4.0E-04	4.3E-04	4.6E-04
AGE	-0.002	-0.002	-0.001	-0.001	-0.001	-0.001	-0.003	-0.003	-0.002
LOTSIZE	3.0E-06	4.0E-06	5.0E-06	1.0E-06	2.0E-06	2.0E-06	5.0E-06	6.0E-06	7.0E-06
BASEMENT	-0.002	0.021	0.045	-0.001	0.009	0.018	-0.003	0.030	0.063
BRICK	0.021	0.051	0.081	0.009	0.021	0.034	0.030	0.072	0.115
FIRE	0.184	0.217	0.251	0.068	0.087	0.108	0.254	0.304	0.353
MATH	0.028	0.033	0.038	0.011	0.013	0.016	0.039	0.047	0.054
TAXR	0.004	0.006	0.007	0.002	0.002	0.003	0.006	0.008	0.010
NONRES	0.001	0.002	0.004	0.000	0.001	0.002	0.001	0.003	0.005
CBDDIST	0.002	0.006	0.009	0.001	0.002	0.004	0.003	0.008	0.013

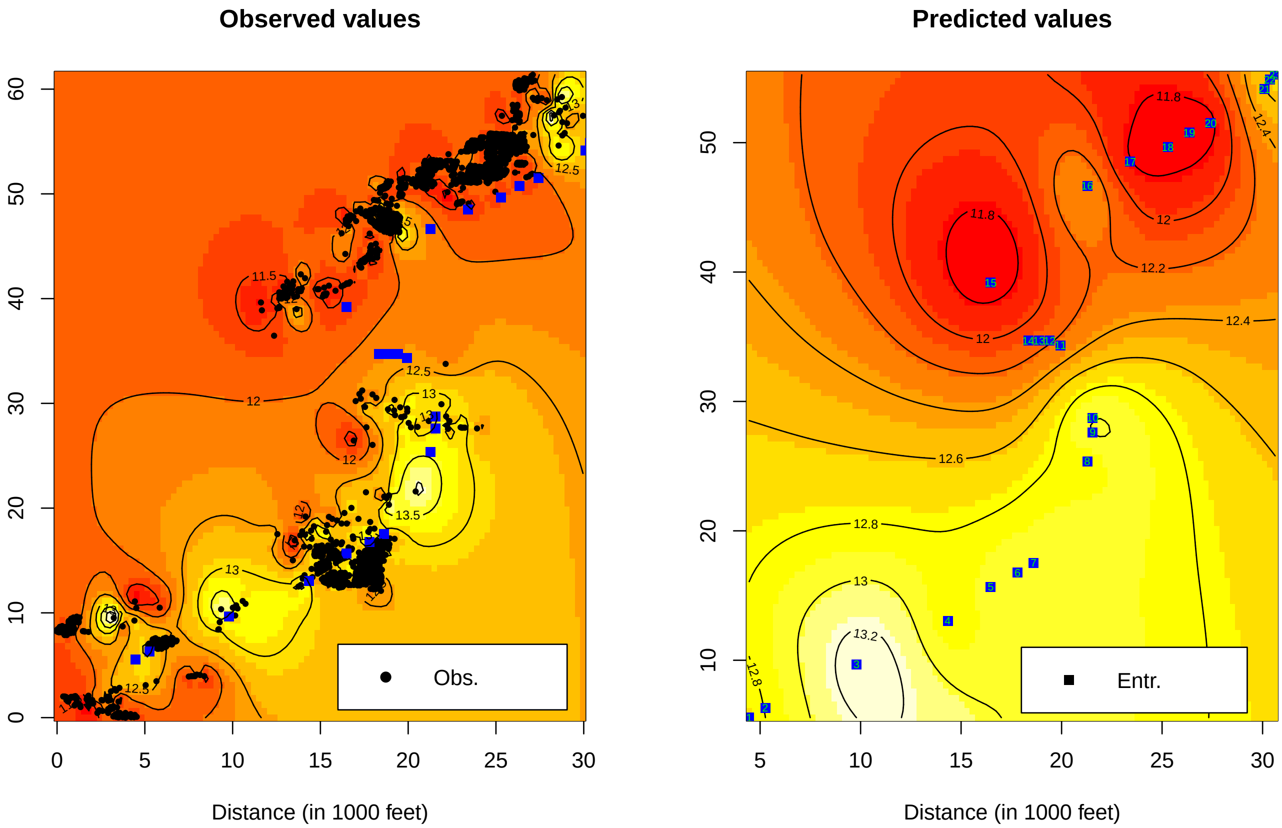
sample locations: the knots. For this study, we implement a prediction technique first introduced by Banerjee et al. (2008) and subsequently improved by Finley et al. (2009) to forecast the house prices near the trail entrances. A modified predictive process is applied to large data sets in order to reduce the bias for the non-spatial variance term τ^2. This is based on the reduction of the original data set to a smaller representative set of knots through a process having spatially adaptive variances, in order to guarantee adequate properties for the covariance of the unobserved spatial effect for these sample locations. More specifically, we modified the predictive process by randomly selecting 558 knots from existing locations and then performed the spatial predictions with respect to the nearest trail entrances.[4] Figure 2 shows the distribution of the observed house prices and the distribution of the estimated house prices (in log) for the parcels around the entrances.

The fact that house prices can be predicted for the sample locations, results in an estimated property value surface map, which, in return, allows us then to obtain the property prices for any property that lies on the map. The contour lines are computed using bivariate linear interpolation. For comparison, the predictions for each trail entrance are presented

[4]Since the spacing of the locations is relatively irregular, we could use a space-covering design (see Royle and Nychka, 1998). To overcome this issue we implement a larger numbers of knots making sure that results are robust to the selection of knots.

in Table 5 below.

Figure 2: Observed and Predicted housing prices (in log) around the trail entrances



Housing prices can be predicted at arbitrary locations and hence new properties prices can be obtained through an estimated property value surface. The contour lines are computed using bivariate linear interpolation. Predictions for each entrance are presented in Table 5

According to Figure 2, we identify the highest predicted property values with a mean value of \$580, 706 ($exp(13.272)$) around the third trail entrance in the Village of Mariemont, Ohio. With a mean value of \$444, 631 ($exp(13.005)$), the second highest predicted property prices are in close proximity to the ninth trail entrance in the Village of Indian Hill. The lowest predicted valued properties with a mean value of \$123, 254 ($exp(11.722)$) lie around

Table 5: Prediction for property values around trail entrances

Entrances	2.5%	Mean	97.5%	Entrances	2.5%	Mean	97.5%
Entr.1	11.638	12.749	13.921	Entr.13	10.800	12.140	13.453
Entr.2	11.783	12.806	13.853	Entr.14	10.823	12.099	13.434
Entr.3	12.162	13.272	14.305	Entr.15	10.526	11.722	12.930
Entr.4	11.875	12.835	13.744	Entr.16	11.131	12.325	13.503
Entr.5	12.006	12.920	13.878	Entr.17	10.766	11.810	12.860
Entr.6	11.916	12.967	14.019	Entr.18	10.511	11.763	12.943
Entr.7	11.879	12.924	13.978	Entr.19	10.692	11.770	12.845
Entr.8	11.543	12.818	14.083	Entr.20	10.604	11.763	12.854
Entr.9	11.893	13.005	14.141	Entr.21	11.316	12.592	13.921
Entr.10	11.778	12.988	14.066	Entr.22	11.380	12.598	13.780
Entr.11	10.777	12.259	13.563	Entr.23	11.515	12.618	13.912
Entr.12	10.908	12.181	13.632				

the fifteenth trail entrance in the City of Loveland. The comparison of observed and predicted house prices in Figure 2 confirms our findings. Comparing predicted house prices in Table 5 with the summary statistics of observed prices in Table 2, we observe a smoothing out effect as predicted values do not show extreme outlying house prices as indicated in the summary statistics. Though we present a contour map of predicted house prices we want to emphasize that this prediction process allows us to obtain a joint predictive distribution for any location around the trail based on the means of the associated predictive distributions.

In a last step, we measure how well centered our predicted results for the OLS and the SP specifications are. To do so, we compare the OLS with the SP predictions using the value of mean square predictive error (MSPE) $\sum_{i=1}^{n}[(\widehat{Y}_i - Y_i)^2]/n$. For Y_i, we randomly selected 80% of the original observations, while the remaining 20% of the sample data (i.e., the hold out data) were used for our predictions and refer to $\widehat{Y}_i$. Our results show a MSPE for the OLS model as 0.0483, compared to a MSPE of 0.0404 for the spatial process. We conclude that using the SP specification is preferred over the OLS specification as indicated by the significant reduction in the predictive SSE. In other words, the predicted house prices in the SP specification are closer to the observed values than for the OLS specification indicating the superior performance of the spatial process.

7 Conclusion

It is well documented in the relevant literature that location matters for home buyers and as such is a major component in determining property values. In the presented paper, we showed that multi-purpose trails have a significant influence on the price of houses when they lie within close proximity to the trail; where the distance to the trail is calculated based on street network distances. More specifically, we estimated the influence of the Little Miami Scenic Trail in Hamilton County, Ohio, to devalue the average priced house in our sample by $8,960 when moving away from the trail by 1,000 feet.

In this paper we compared the estimation results of four different procedures within the Bayesian framework. Overall, we conclude that all spatial model variants outperform the non-spatial OLS specification. In addition to the more widely used Spatial Autoregressive (SAR) model and the Spatial Error Model (SEM), we implemented the Spatial Process (SP). This more recent geostatistical specification has been developed specifically to be used with larger data sets, while at the same time implicitly modeling the underlying spatial structure in the data set. In this sense, the SP allows the implementation of a functional form which helps to understand the spatial relationships between properties based on distances from one another. Our results indicate that the Spatial Process improves the estimation results when compared to the more traditional spatial econometric models. However, the SP, like the SEM falls short of capturing spillover effects between neighboring properties, whereas the SAR model does capture these direct and indirect effects.

In the last section of the research we presented a house price contour map for our study region. In other words, we predicted, based on a selection of representative knots, the conceptual prices for all properties included in our sample around all 23 trail entrances. Using the MSPE as a performance criteria we show that the SP predictions of house prices are closer to the observed values than using the OLS predictions.

We conclude our research with the finding that trails do have a significant impact on the prices of surrounding properties. While we were able to account for spatial dependence in our sample data, we suggest that future research employs a more dynamic setting with time dependent variables and house prices being influenced by their neighboring properties.

References

[1] Anderson, S. T. and S. E. West. (2006). Open space, residential property values, and spatial context. *Regional Science and Urban Economics*, 36 (6):773-789.

[2] Asabere, P. and F. Huffman. 2009. The relative impacts of trails and greenbelts on home prices. *The Journal of Real Estate Finance and Economics*. 38 (4): 408-419.

[3] Banerjee, S., Gelfand, A.E., Finley, A.O. and Sang, H. (2008). Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society, Series B*, **70**, 825848.

[4] Brasington, David M. (1999). "Which Measures of School Quality Does the Housing Market Value?," *Journal of Real Estate Research*, **18**, 395-414.

[5] Cheshire, P. and S. Sheppard (1995) On the prices of land and the value of amenities. *Economica* 62 (246): 247-267.

[6] Clapp J. M., A. Nanda, and S. L. Ross. (2008) "Which School Attributes Matter? The Influence of School District Performance and Demographic Composition on Property Values". *Journal of Urban Economics*, **63**(2), 451-66.

[7] Finley, A.O., Sang, H., Banerjee, S. and Gelfand, A.E. (2009). "Improving the performance of predictive process modeling for large datasets", *Computational Statistics and Data Analysis*, **53**, 2873-2884.

[8] Gelfand, A E, Ecker M D, Knight J R, Sirmans C F. (2004). "The Dynamics of Location in Home Price," *Journal of Real Estate Finance and Economics*, **29**, 149-166

[9] Hess, D. B. and T. M. Almeida. (2007) "Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York". *Urban Studies*, **44**(5), 1041-1068.

[10] Krizek, K. (2006) Two approaches to valuing some of bicycle facilities' presumed benefits: Propose a session for the 2007 National Planning Conference in the City of Brotherly Love. Journal of the Amaerican Planning Association, 72(3): 309-320.

[11] Lancaster, K. J. (1966) A new approach to consumer theory, *The Journal of Political Economy*, 74(2), 132-157.

[12] LeSage, J. P. and R. K. Pace (2004). "Models for Spatially Dependent Missing Data, *Journal of Real Estate Finance and Economics*, **29**(2), 233-254.

[13] LeSage, J. and R. K. Pace (2009). *An Introduction to Spatial Econometrics*, Chapman Hall/CRC Press.

[14] Lindsey, G., J. Man, S. Payton, and K. Dickson. 2004. Property Values, Recreation Values, and Urban Greenways. *Journal of Park and Recreation Administration*, 22(3): 69-90.

[15] Lynch, A. K. and D. W. Rasmussen. (2001) Measuring the impact of crime on house prices. *Applied Economics*, 33(15): 1981-1989

[16] Mogush P., K. J. Krizek, and D. Levinson (2005) The Value of Trail Access on Home Purchases. Presented at 84th Annual Meeting of Transportation Research Board in Washington, DC, January 9-13th 2005.

[17] Nicholls, S. and Crompton, J. L. (2005) The impact of greenways on property values: Evidence from Austin, Texas. *Journal of Leisure Research*, 37(3), 321-341.

[18] Rosen, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *The Journal of Political Economy*, 82(1): 34-55.

[19] Royle, J.A. and Nychka, D. (1999). "An algorithm for the construction of spatial coverage designs with implementation in SPLUS", *Computers and Geosciences*, 24, 479488.

[20] Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin and Angelika van der Linde (2002). "Bayesian measures of model complexity and fit", *Journal of the Royal Statistical Society, Serie B*, **64**, 1-34.

[21] Valente, J. Wu, S.S. Gelfand, A. and Sirmans, C. F.(2005). "Apartment rent prediction using spatial modeling", *Journal of Real Estate Research*, **27**, 105-136.

[22] Ventolo, W. L. and M. R. Williams. (1994). Fundamentals of Real Estate Appraisal. Chicago, Illinois: Real Estate Education Company.