Measuring the Benefits of Neighborhood Park Amenities: Application and Comparison of Spatial Hedonic Approaches

Tadao Hoshino, Department of Geography and Environment, London School of Economics

Koichi Kuriyama, School of Political Science and Economics, Waseda University

Abstract

This paper uses the hedonic price method to estimate the influence of parks on the housing value of a residential neighborhood in Setagaya ward, Tokyo. When data are explicitly spatial, ordinary least squares (OLS) estimates are inefficient. To improve the accuracy of estimates, spatial autoregression models and kriging models were employed. The results show the clear validity of the spatial models compared to the OLS model. According to the results from the spatial models, the effect of parks on property values varies with the park's size: the number of medium-sized neighborhood parks is positively related to rental prices, while small- and large-neighborhood parks are not statistically positively influential by rental values. This casts doubt on the validity of past urban park planning policy, in which the primary concern was not for the size of individual open spaces but for the ratio of aggregated open spaces.

Keywords: Hedonic approach; Kriging; Spatial autocorrelation; Spatial regression models; Urban parks

1. Introduction

Urban parks have many different purposes, including improving urban environments, preventing disasters, and providing communication opportunities. Since Jacobs' monumental work in 1961, *The Death and Life of Great American Cities*, the provision of urban parks has been recognized as an important factor in urban planning. Now, the importance of urban park improvements is widely recognized by local municipalities as a primary attraction that consumers use to decide where to reside.

The hedonic price model has been used in many studies that evaluate urban park policies. Currently, the application of the hedonic approach to open spaces, including parks and forests, has been relatively minor compared to its application to air quality and hazardous waste sites because of the lack of detailed land-use data. Recently, because of improved computer technology and the rapid popularization of Geographic Information Systems (GIS), the problem of data availability for land use is no longer critical. Thus, the study of the hedonic approach to estimate the effect of open spaces has progressed remarkably by making use of geographical data (e.g., Cheshire and Shepard 1995, Geoghegan 2002, Irwin 2002, Irwin and Bockstael 2001, Lutzenhiser and Netusil 2001, Shultz and King 2001, and Smith et al. 2002.). A common finding of studies is that certain types of open space have positive effects on property prices.

According to the prevalence of statistical analyses using spatial data, the importance of spatial correlations among observations of the efficiency and consistency of the hedonic model estimates has recently received more attention (Kim et al. 2003). Spatial autocorrelation occurs when population members are related through their geographic locations (Dubin 1988). Spatial correlation is far from surprising in the hedonic model on housing because omitted variables will generally be spatially correlated. The specification of all of the many spatial characteristics affecting a property, however, would result in a function that is too complicated to compute. To solve this problem, studies have tried to visualize spatial aspects of the data in an empirically manageable form. Specifically, we focused on the *spatial autoregression models* that are commonly used in spatial econometrics (Anselin 1988), and the *kriging models* that are commonly used in spatial statistics (Cressie 1993). A hedonic approach making use of these explicit spatial regression techniques can be called *a spatial hedonic* approach. The number of empirical studies regarding the use of the spatial hedonic approach to estimate the effect of environmental quality is not large, but is steadily increasing (e.g., Acharya and Bennett 2001, Cho et al. 2006, Kim et al. 2003, Leggett and Bockstael 2000, and Paterson and Boyle 2002.)[1].

Urban parks can be categorized into various types based on their scale and use. The term "urban park" includes tiny parks that are exclusive to the people of a particular neighborhood, as well as large parks that offer sports facilities and museums to the people of an extensive region. The availability of small parks is quite limited, whereas that of large parks is quite widespread. Therefore, when studies deal with urban parks, they must pay great attention to the framework used to analyze these diverse parks. We particularly call parks in a residential neighborhood *neighborhood parks*. Neighborhood parks, as the name suggests, are available to all neighborhood residents who can access them. Additionally, because the size of neighborhood parks is usually small to medium, the parks are assumed not to be of value for people outside the immediate area because such parks have fewer facilities than large parks and thus attract fewer people from outside the immediate neighborhood. These features of most neighborhood parks are localized to a neighborhood and only provide locally beneficial external economies. In the case of local externalities, as Palmquist (1992) noted, the effects can be calculated by simply estimating the hedonic price function, without the need to conduct a complex two-step estimation procedure.

Our primary objective was to measure the value of neighborhood parks using a spatial hedonic approach in conjunction with a basic ordinary least squares (OLS) hedonic approach and to compare the validity of these models. Very few hedonic studies have compared environmental benefits between the basic hedonic model and spatial hedonic models. In Section 2, we present a description of the econometric models used. Section 3 contains an overview of the study area and an explanation of the

1 All studies listed here used spatial econometric models, which use spatially weighted regression techniques in some form. Almost no hedonic studies have used the kriging (nonparametric) model to estimate the effect of environmental quality.

data used for the analysis. In Section 4, we discuss the estimation results. Finally, section 5 provides conclusions.

2. Econometric Models

Hedonic models were developed to deal with markets for differentiated products. The differentiated product will differ, depending on the specific characteristics or attributes that the model contains (Palmquist 1999). The theoretical framework of the market for heterogeneous goods was originally developed by Rosen (1974). In Rosen's (1974) framework, the price of any unit of a differentiated good is explained as a function of a bundle of characteristics. Housing prices are the most common example of this application. To specify the structure of housing characteristics, the characteristics are grouped into structural characteristics such as area, or number of rooms, as well as locational characteristics such as transportation accessibility or environmental amenities and disamenities.

The simplest and most general functional form used in hedonic studies is a simple linear regression model. If P is a vector of observed property values at N points on the plane, then the linear hedonic function is

$$P = X^s \beta^s + X^l \beta^l + \varepsilon , \qquad (1)$$

where P is the $N \times 1$ vector of property prices, X^s is the $N \times K^s$ matrix of structural variables with K^s characteristics, β^s is the $K^s \times 1$ vector of the coefficients for structural variables, X^l and β^l are locational variables and coefficients that are the same as X^s and β^s, and ε is the $N \times 1$ vector of errors. The coefficients are simply estimated by OLS. In particular, we are interested in the coefficients of variables regarding park amenities. As stated previously, the benefits of neighborhood parks seem to spill over exclusively into neighboring housing. Hence, an application of the hedonic approach in this case must be legitimate. However, with the presence of spatial autocorrelation the

OLS optimality unavoidably fails. In the following, the spatial econometric regression models and the kriging models are presented in order and are typical solutions to the issue of spatial autocorrelation.

2.1 Spatial autoregression models[2]

Two models are frequently used to represent the spatial autoregressive process: spatial error models and spatial lag models. In the former models, the spatial autocorrelation is considered to be caused by omitted variables, whereas in the latter models, it is caused by spatial interactions (endogeneity).

Spatial error models

Spatial autocorrelation caused by omitted spatially correlated variables does not bias the OLS estimates, but the estimates will be inefficient, and most troubling, the standard errors will be biased, leading to inaccurate hypothesis testing (Anselin 1988 and Legget and Bockstael 2000). This occurs because when there is spatial autocorrelation in the error term, $E(\varepsilon'\varepsilon)=\Omega$, an error variance–covariance matrix is formed with nonzero off-diagonal elements. In this situation, β should be estimated using a generalized least squares (GLS) estimator $\hat{\beta}_{GLS}=(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}P$ in terms of efficiency. The empirical issue is then to estimate the elements of Ω.

One way to obtain the structure of Ω directly is by modeling the covariance as a function of the Euclidean distance between two locations. One of the most commonly used spatial process specifications is the autoregression form used in disturbances. For a general case,

$$\begin{array}{l} P = X\beta+\varepsilon;\quad \varepsilon=\lambda W\varepsilon+u \\ u\sim N(0,\sigma^2 I) \end{array}, \text{ with } W=\begin{bmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{bmatrix}, \qquad (2)$$

2 Several discussions have been published on model specifications to incorporate spatial autocorrelation and spatial weights; see, e.g., Anselin 1988 and Cliff and Ord 1981.

where λ is the spatial autoregression parameter, W is the $N \times N$ spatial weight matrix, and u is the $N \times 1$ vector of the i.i.d. normal error term. The structure of the variance–covariance matrix becomes a nonzero off-diagonal matrix. In many previous empirical studies, each element of W was defined simply as an inverse of the Euclidean distance, consistent with Tobler's (1970) first law of geography, which states, "Everything is related to everything else, but near things are more related than distant things." Alternatively, we can assume that a boundary exists, the greatest distance over which the value at one point is related to the value at another point. In this framework, a spatial interdependence relationship between pairs of points that are farther apart than this distance equals zero. This concept is the same as the *range* in the kriging model explained below. In either definition, the important point to note is that both definitions only consider a distance between two points and do not consider the direction.[3] Accordingly, the covariance of any two points that have the same distance between them will be the same, depending solely on the distance. These assumptions are called *isotropy* and *stationarity*, respectively.[4] Depending on these assumptions, we define two practical and empirically manageable weight matrixes to be examined: one has a simpler and traditional form, and the other is more flexible:

Simple *Flexible*

$$w_{ij} = \begin{cases} \dfrac{1}{d_{ij}^2} & i \neq j \\ 0 & i = j \end{cases} \qquad w_{ij} = \begin{cases} 1 - \left(\dfrac{d_{ij}}{h}\right)^2 & if\ d_{ij} < h,\ i \neq j, \\ 0 & otherwise \end{cases} \qquad (3)$$

where d_{ij} is the distance between points i and j, and h is the boundary distance to be estimated.

3 How to define and measure a distance is also something to be considered. Not only geographical distance but economic and social distance can be applicable. The difference in significance between distance measures should be a future research topic.

4 However, because it is not surprising that roads and other components of transport infrastructure are not radially symmetric(Cheshire and Sheppard 1995), using these assumptions might be too simplistic.

The weight falls to zero when the distance between i and j equals h or longer. All unknown coefficients and parameters can be estimated using the maximum-likelihood method. The log-likelihood function is thus:

$$L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(P - X\beta)'[(I - \lambda W)'(I - \lambda W)] \times (P - X\beta) + \ln|I - \lambda W|. \quad (4)$$

Eq. (4) will be maximized with respect to β, σ^2, λ, and h simultaneously (or just λ and h by concentrating on β and σ^2) to obtain the maximum-likelihood estimates.

Spatial lag models

Unlike the case of spatial error autocorrelation, spatial autocorrelation caused by spatial interactions (endogeneity) biases the OLS estimates. To represent this type of spatial process, the spatial lag model is given as:

$$\begin{aligned} &P = \rho WP + X\beta + u \\ &u \sim N(0, \sigma^2 I) \end{aligned}, \text{ with } W = \begin{bmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{bmatrix}, \quad (5)$$

assuming homoskedasticity in disturbances, where ρ is the spatial lag parameter. Similarly, all unknown coefficients and parameters can be estimated using the maximum-likelihood method. The log-likelihood function can be written as:

$$L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}[P'(I - \rho W)'M(I - \rho W)P] + \ln|I - \rho W|, \quad (6)$$

where $M = I - X(X'X)^{-1}X'$. With regard to the spatial weight matrixes for the spatial lag model, we only present the case of the simple form in Eq. (3), because of the low significance of the spatial lag process that was suggested by the specification tests, which are described next.

Specification tests

To detect misspecification due to spatial autocorrelation processes and to distinguish the nature of the spatial autocorrelation in the error term (spatial error) and in the dependent variable (spatial lag), the modified Lagrange multiplier (LM) tests developed by Anselin et al. (1996), which are computationally simple and robust, are conducted. These tests are simple in the sense that they are based on OLS residuals. The test for both spatial error and spatial lag processes, H_0: $\lambda = 0$, $\rho = 0$, takes the form with an assumption that both weight matrixes are identical:

$$LM_{\lambda.\rho} = \frac{(\hat{u}'W\hat{u}/\hat{\sigma}^2)^2}{T} + \frac{(\hat{u}'WP/\hat{\sigma}^2 - \hat{u}'W\hat{u}/\hat{\sigma}^2)^2}{NJ_{\rho.\beta} - T}, \tag{7}$$

where $\hat{u}$ is the $N \times 1$ vector of OLS residuals, $\hat{\sigma}^2$ is the maximum likelihood variance $\hat{u}'\hat{u}/N$, $J_{\rho.\beta} = [(WX\hat{\beta})M(WX\hat{\beta}) + T\hat{\sigma}^2]/N\hat{\sigma}^2$, $\hat{\beta}$ is the vector of OLS parameters, and $T = trace[W'W + W^2]$. This statistic is distributed as $\chi^2_{(2)}$. The statistic can test for both processes as a two-directional test statistic; however, it cannot distinguish between the two and results in a loss of power when only one of the two misspecifications is present. Therefore, one-directional tests should be conducted. The test for a spatial error process, H_0: $\lambda = 0$, in the presence of misspecification involving a spatial lag process is given by:

$$LM_{\lambda} = \frac{1}{T - T^2(NJ_{\rho.\beta})^{-1}} \left(\frac{\hat{u}W\hat{u}}{\hat{\sigma}^2} - T(NJ_{\rho.\beta})^{-1} \frac{\hat{u}'WP}{\hat{\sigma}^2} \right)^2. \tag{8}$$

The test for a spatial lag process, $H_0: \rho = 0$, in the presence of misspecification involving a spatial error process is given by:

$$LM_{\rho} = \frac{1}{NJ_{\rho.\beta} - T}\left(\frac{\hat{u}WP}{\hat{\sigma}^2} - \frac{\hat{u}'Wu}{\hat{\sigma}^2}\right)^2. \qquad (9)$$

Both statistics are distributed as $\chi^2_{(1)}$.

If both one-directional tests are significant, one can follow a classical approach, i.e., select the model corresponding to the test with the highest value, or use an approach inspired by Hendry's work, i.e., start with a very general model and reduce the model on the basis of significance tests (Florax and Nijkamp, 2003).

2.2 Kriging models[5]

Kriging is a minimum mean square error statistical procedure for spatial prediction that assigns a differential weight to observations that are closer to the dependent variable's location (Dubin et al. 1999). The prediction procedure corresponds to a spatial version of Goldberger's best linear unbiased prediction (BLUP) method. Because kriging is not so much an estimation method as a prediction method, one cannot observe the error term that corresponds to u of the spatial error model at a disaggregated level while ε is also observable.

The spatial weights are computed from the estimated *semivariogram* or *covariogram*, and the covariogram corresponds to the variance–covariance function used to estimate Ω. To estimate the covariogram, several steps are required.

We now consider ε to be a spatially stochastic variable specifically tied to a particular location.

5 See, for example, Basu and Thibodeau 1998, and Cressie 1993, for a more detailed discussion on this issue.

Because it is almost impossible to model the joint distribution function of ε for every location point by point, some assumptions are needed to simplify it. Then, the stationary assumption can again be applied. Let x_i represent the location of property i. Thus, $S(x_i)$ denotes the stochastic error of the hedonic price equation for a property located at x_i. If a stochastic process is *second-order stationary*, then

$$\begin{aligned} &E[S(x_i)] = \mu \\ &Cov[S(x_i),\ S(x_j)] = C(x_i - x_j) \end{aligned}, \tag{10}$$

where C is called a covariogram. Eq. (10) means that the first and second moments of distribution can be the same at all locations, and the spatial interdependence is described solely as a function of the Euclidean distance between two points. This is a case of isotropy. Of course, $C(0)$, the variance, is constant at any location. This is a very useful assumption, especially when the distribution is normal. Dependence could vary with the direction, as well as the distance, and this is the case in anisotropy. Finding a practical way to model anisotropic dependence for real-estate data is both a challenge and an opportunity for improving hedonic estimation (Dubin et al. 1999). Alternatively, we can also assume that the difference between two values of $S(x_i)$ is a stationary distribution, as follows:

$$\begin{aligned} &E[S(x_i) - S(x_j)] = 0 \\ &Var[S(x_i) - S(x_j)] = 2\gamma(x_i - x_j) \end{aligned}, \tag{11}$$

where γ is called a semivariogram. This type of stationary phase is called the *intrinsic stationary*. Apparently, if the process S is second-order stationary, it implies that it is also intrinsic stationary. Under the second-order stationary process, the relationship between the semivariogram and the covariogram can be written as:

$$
\begin{aligned}
\gamma(d) &= \frac{1}{2} Var[S(x+d) - S(x)] \\
&= \frac{1}{2} Var[S(x+d)] + \frac{1}{2} Var[S(x)] - Cov[S(x+d),\, S(x)], \\
&= C(0) - C(d)
\end{aligned}
\tag{12}
$$

where d is the Euclidean distance between two points. Accordingly, a covariogram can be derived from the corresponding semivariogram.

We used the exponential and spherical semivariogram models. The exponential semivariogram model and the spherical semivariogram model, respectively, are:

Exponential

$$
\gamma_{Exp}(d;\theta) = \begin{cases} 0 & \text{if } d = 0 \\ \theta_0 + \theta_1 \left\{ 1 - \exp\left(-\dfrac{d}{\theta_2} \right) \right\} & \text{if } 0 < d \end{cases}
$$

, (13)

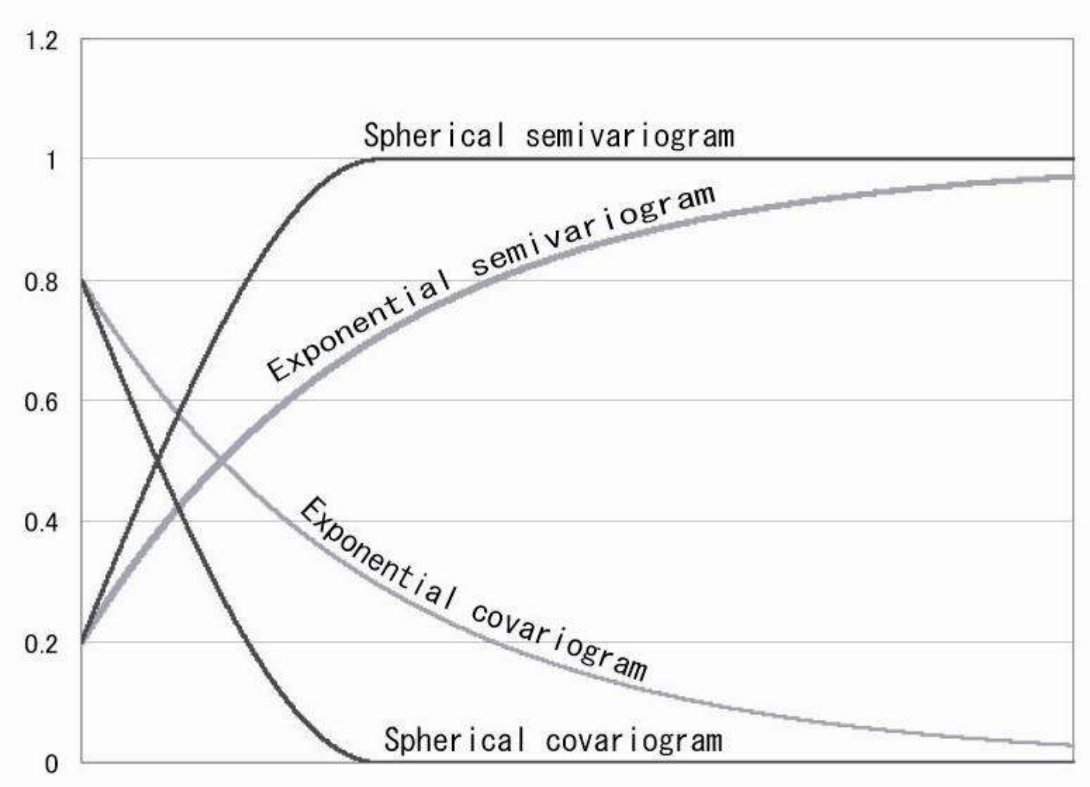
Spherical

$$
\gamma_{Sph}(d;\theta) = \begin{cases} 0 & \text{if } d = 0 \\ \theta_0 + \theta_1 \left\{ \dfrac{3}{2}\left(\dfrac{d}{\theta_2} \right) - \dfrac{1}{2}\left(\dfrac{d}{\theta_2} \right)^3 \right\} & \text{if } 0 < d \le \theta_2 \\ \theta_0 + \theta_1 & \text{if } d \ge \theta_2 \end{cases}
$$

where θ are parameters to be estimated. The θ_0 is called the *nugget*, $\theta_0 + \theta_1$ is called the *sill*, and θ_2 is called the *range*. The spherical semivariogram model actually reaches the specified sill, $\theta_0 + \theta$, at the specified range, θ_2; that is straightforward "cut-off" as for the boundary distance in Eq.(3), *h*, whereas the exponential approaches the sill asymptotically. Especially, when $d = 3 \times \theta_2$, exponential semivariance reaches 95% of the sill value; this distance is called the *practical range*. The estimated parameters for the semivariogram are then used to compute the associated covariogram. Figure 1 presents the basic form of the semivariograms and the covariograms of the two models.

Figure 1

Semivariograms and covariograms



For the experimental procedure, the first step is to estimate semivariogram parameters using OLS residuals. The second step is to compute the corresponding covariogram. The covariogram then leads to Ω, and β can be estimated using the GLS estimator $(X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}P$. Using the estimated GLS coefficients, residuals can be recomputed, and the recomputed residuals are used to reestimate the new GLS estimator. With *K* estimated coefficients in the hedonic specification, this procedure iterates until

$$\sqrt{\sum_{k=1}^{K}(\hat{\beta}_{new.k}-\hat{\beta}_{old.k})^2/K} \fallingdotseq 0. \tag{14}$$

3. Data

3.1 Urban parks in Setagaya ward

Setagaya ward is situated on the southwestern edge of Tokyo's 23 wards. The population is approximately 810,000, and this ward is the largest among Tokyo's 23 wards, with the 14th highest population density (Population Census 2006). The population trend is toward growth. Because the location of Setagaya is very convenient in terms of transportation to access the business and commercial districts of Tokyo, it has become one of the most desirable residential zones.

In accordance with Setagaya ward's master plan, which emphasizes the role of urban parks and green spaces in the residential environment, various types of open space amenities exist, including urban parks, green spaces, and agricultural spaces. Table 1 shows a rough breakdown of the urban parks provided in Setagaya by legal definition. In addition to the legal category, parks can be classified by their area size. The size classification is considered to be more rational for estimating the effect of urban parks on property prices than the legal classification because in the study of urban park amenities, the size classification more precisely captures individuals' own perceptions of different types of parks. To construct urban park variables for the estimation based on the size classification, river terraces, green belts, and parks with an area < 500 m^2 were excluded for simplicity. With regard to the rest of the parks, parks with an area > 2 ha were separated from other parks as a particular variable. This is because, as noted previously, such large parks that offer several types of facilities to the people of an extensive region may have specific roles, and also because 2 ha is a threshold in classifying parks by size according to the *Toshikouen-hou* (the urban park law in Japan). The rest of the 241 parks with an area within 500 m^2 to 2 ha were simply grouped by quartile. Table 2 shows a summary of the reclassification of parks. The number of small parks was larger by a clear margin than the number of large parks. This phenomenon may be considered a reflection of the planning policy for urban parks in many municipalities in Japan in which the main policy objective has been to increase the total amount of open space per capita, regardless of the area size of the individual open space.

Table 1

Urban parks in Setagaya by legal classification

Class	Number of parks	Average size (m^2)
Ward parks	344	3454.88
Ward yards	136	1016.99
Subtotal	480	2764.14
Metropolitan parks	4	238740.19
Total	484	4714.36

Source: Setagaya Ward Statistics; Summary of Setagaya ward administration, Ch. 7

Table 2

Reclassification of Urban parks in Setagaya by size

Size	Number of parks	Average size (m^2)
500 m^2 to first quartile value	60	600.03
First quartile value to median	60	871.12
Median to third quartile value	60	1385.86
Third quartile value to 2 ha	61	4315.84
> 2 ha	17	99755.36
Total[a]	258	8257.86

[a]Note: River terraces, greenbelts, and parks with an area of 0 to 500 m^2 were excluded.

Source: Summary of Setagaya ward administration, Ch. 7

3.2 Data used for analysis

Data for single-room dwellings were collected from the entire area of Setagaya ward, providing 2370 samples. Table 3 presents a list of the variables used and their definitions. The independent variable is the natural log of the monthly rental price of a single-room dwelling (RENT). The prices and characteristics of dwellings, including AGE (age of dwelling), WALK (time to walk from the nearest train station to the dwelling, in minutes), and AREA (area of the dwelling), were extracted from May through June 2007 from the website of a private real-estate office called *forrent* (http://www.forrent.jp).[6]

The average tax payment of a ward inhabitant per payer population in the block (TAX) and the number of businesses within the block area per hectare (BUSPL) were collected from Setagaya Ward Statistics 2006. Because the inhabitant tax is progressive, the TAX variable corresponds to the income level of each neighborhood. As both TAX and BUSPL increase, the rental price is also expected to increase.

Although the structural characteristics were collected in detail for all 2370 samples, only a block-level address could be obtained for each property from the real-estate website. Therefore, to reach a correspondence between block-level locational variables and a specific dwelling on a one-to-one basis, it is necessary to aggregate every structural data point of the dwellings into a block,[7] but a good hedonic approach must include data sets that are as disaggregated as possible. Nevertheless, Shultz and King (2001) determined that the aggregation of land-use data by alternative levels of census geography does not appear to seriously affect the overall quality of the marginal implicit estimates for open-space amenities.

Using the geographical information of the block-level address, the Euclidean distances to Shibuya station (SHIBUYA) and to particular types of parks from the block were calculated. Shibuya is one of the best-known districts in Japan because of its extreme vitality in terms of both culture and business. Moreover, Shibuya station is an important hub of many transportation systems for anyone living in Tokyo. Therefore, it is easily inferred that as the distance to Shibuya station increases, the rent price

6 The one possible problem is that the data may be biased from sample selection because dwellings for sale can be considered dead stock, whereas most good dwellings have already been sold. However, Jud and Seaks (1994) showed that ignoring the sample selection bias will only lead to a subtle error in estimates.

7 It is still possible to use both disaggregated structural variables and aggregated locational variables in one framework. However, this has been criticized as yielding invalid inferences, the so-called ecological fallacy problem (Anselin 2002). More fundamental problems occur when estimating the GLS estimators. However, the variance–covariance matrix must be positive, but in the situation when some observations share the same locational data, it will no longer be definitely positive

would decrease. The data on the size and the location for every park (except for some metropolitan parks) in Setagaya ward can be obtained from the Summary of Setagaya ward administration, Ch. 7. These data were grouped by area size based on the classification of Table 2. Because the effects of park amenities on properties would not be observable if the distance to the parks is too long to freely access them (except for versatile large parks), we need to define the boundary of an accessible distance. An easily accessible distance of 500 m for anyone on foot was considered to be suitable. This could have been more precisely estimated if it were based on an actual survey of visitors and their origin, means of transportation to reach the park, and the purpose of the visit. The variables for park amenities were defined as the number of parks accessible within a 500 m radius of the block and park sizes of 500 m^2 to the first quartile value (PARKQ1), first quartile value to the median (PARKQ2), median to the third quartile value (PARKQ3), third quartile value to 2 ha (PARKQ4), and > 2 ha (PARKL). Table 4 shows the descriptive statistics for all variables.

Table 3

Variables and their definitions

Variable	Definition
Structural variables	
RENT	Natural log of monthly rental price of a single-room dwelling (yen)
AGE	Age of dwelling (years)
WALK	Time to walk from the nearest train station to the dwelling (min)
AREA	Area of the dwelling (m^2)
Locational variables	
TAX	Average tax payment of a ward inhabitant per payer population in the block (yen)
BUSPL	Number of businesses within the block area per hectare
SHIBUYA	Distance from the block to Shibuya station (km)
PARKQ1	Number of parks accessible within a 500 m radius of the block; each park's size is 500 m^2 to the first quartile value

PARKQ2	Number of parks accessible within a 500 m radius of the block; each park's size is first quartile value to median
PARKQ3	Number of parks accessible within a 500 m radius of the block; each park's size is median to third quartile value
PARKQ4	Number of parks accessible within a 500 m radius of the block; each park's size is third quartile value to 2 ha.
PARKL	Number of parks accessible within a 500 m radius of the block; each park's size is > 2 ha.

Table 4

Descriptive statistics

Variable	Mean (N = 244)	Standard deviation
Structural variables		
RENT	11.170	0.087
AGE	14.982	5.032
WALK	9.469	5.713
AREA	20.718	2.505
Locational variables		
TAX	270482.5	191703.1
BUSPL	5.097	5.246
SHIBUYA	8.047	3.004
PARKQ1	0.516	0.682
PARKQ2	0.496	0.711
PARKQ3	0.545	0.744
PARKQ4	0.389	0.629
PARKL	0.143	0.385

4. Results

Table 5 reports the results of the three specification tests. The spatial weight matrix used in the tests was the simple form in Eq. (3). The test based on the statistic $LM_{\lambda,\rho}$ reveals the presence of some kind of misspecification, with OLS at 90% significance. Subsequently, to determine the cause of the misspecification, two one-directional tests were conducted. According to the results, the null hypothesis of LM_{λ} was rejected significantly at 95%, whereas that of LM_{ρ} could not be rejected. This means that a spatial error process, rather than a spatial lag process, exists in the data. In other words, the main cause of the misspecification is the spatial autocorrelation in disturbances caused by omitted variables. This can be a positive reason to apply GLS type estimations to this data set, including the spatial error models and kriging models.

Table 5

Specification tests

	$LM_{\lambda.\rho}$	LM_{λ}	LM_{ρ}
LM statistic	5.185 *	4.827 **	0.340
Null hypothesis	H_0: $\rho = 0, \lambda = 0$	H_0: $\lambda = 0$	H_0: $\rho = 0$

** and * indicate significance at the 0.05 and 0.1 levels, respectively.

All estimation results of coefficients and parameters from the six models, i.e., the OLS, the spatial lag, the spatial error (simple), the spatial error (flexible), the exponential kriging, and the spherical kriging models, are reported in Table 6. The dependent variable is the natural log of the monthly rental price aggregated to block level. Consequently, the estimated coefficients and parameters are roughly within expectations and are critically interesting.

We first focus on the overall estimation results. Regarding the structural variables, we expected RENT to be positively related to AREA and negatively related to WALK and AGE. As a result, the estimated coefficients of these variables have the expected sign at a statistically significant level. Regarding locational variables, we expected RENT to be positively related to TAX and BUSPL and negatively related to SHIBUYA. In terms of the locational variables of park amenities, we expected the basic trend to be an improvement in the residential environment as the number of accessible parks increased. Thus, RENT was considered to be positively related to five park variables. As a result, the estimated coefficients of these, except for the park variables, are within the predictions. According to the results, all six models estimate that PARKQ1 is significantly negative, PARKQ3 is significantly positive, and all of the other park variables are not significantly related to RENT. For PARKQ1, it is difficult to determine a definite causality of the negative relation. One possible interpretation is that, as a general tendency, most such small parks occur in zones of clusters of low-rise dwellings, just like *empty lots,* rather than park amenities; this variable may play a role as an indicator of few dwelling amenities. Whether this is true, what seems certain is that a small park with an area of approximately 600 m^2 does not have a significant positive effect on the housing market. For the larger park variables PARKQ4 and PARKL, the value of such a park would not be considered to be capitalized into local housing markets *wholly* in terms of their versatile applications and facilities to attract many visitors globally.[8] Because the proximity to such parks might be an external economy and diseconomy at the same time for the neighborhood residents, we assumed that the coefficient is not significant. Eventually, at least from the point of view of the housing market, only medium-sized parks, i.e., PARKQ3, have a positive effect on the market. These results are reasonable, interesting, and have several policy implications in terms of urban park planning.

We then compared individual results of the six regression models. The first area that must be focused on is parameters of the spatial autoregression model. Without spatial autocorrelation, the autoregression parameters ρ and λ should not be significant. The results indicate that, as predicted, ρ is insignificant and λ is significant at the 95% level for the simple model and at the 99% level for the flexible model, meaning that a spatial autocorrelation in disturbances caused by omitted variables exists in this case. The likelihood ratios between the spatial error models and the OLS model show the validity of using the spatial error models at a certain significance.

8 The travel cost method would be the most appropriate to evaluate such a large park.

Table 6

Estimation results

Dependent Variable: *RENT*,

(t-value in parentheses)

Variables	OLS	Spatial lag model	Spatial error model: Simple	Spatial error model: Flexible
Intercept	11.0838 ***	11.0736 ***	11.0890 ***	11.0946 ***
	(330.283)	(330.221)	(322.601)	(327.483)
AGE	-0.0074 ***	-0.0074 ***	-0.0076 ***	-0.0078 ***
	(-10.683)	(-10.680)	(-11.174)	(-11.411)
WALK	-0.0024 ***	-0.0023 ***	-0.0024 ***	-0.0024 ***
	(-3.748)	(-3.640)	(-3.604)	(-3.511)
AREA	0.0156 ***	0.0156 ***	0.0158 ***	0.0156 ***
	(10.988)	(10.986)	(11.432)	(11.348)
TAX	5.2730E-08 ***	5.3775E-08 ***	4.5247E-08 **	4.4434E-08 **
	(2.896)	(2.956)	(2.395)	(2.326)
BUSPL	0.0019 ***	0.0018 ***	0.0015 **	0.0015 **
	(2.765)	(2.742)	(2.283)	(2.197)
SHIBUYA	-0.0158 ***	-0.0157 ***	-0.0163 ***	-0.0163 ***
	(-12.972)	(-12.845)	(-9.833)	(-10.103)
PARKQ1	-0.0116 **	-0.0120 **	-0.0107 **	-0.0116 **
	(-2.308)	(-2.402)	(-2.052)	(-2.192)
PARKQ2	0.0004	-3.5938E-05	0.0010	0.0010
	(0.092)	(-0.007)	(0.190)	(0.189)
PARKQ3	0.0103 **	0.0101 **	0.0111 **	0.0124 **
	(2.308)	(2.272)	(2.302)	(2.525)
PARKQ4	0.0035	0.0035	0.0044	0.0065
	(0.506)	(0.498)	(0.611)	(0.888)
PARKL	-0.0031	-0.0020	-0.0034	-0.0038
	(-0.275)	(-0.177)	(-0.296)	(-0.327)

Model parameters		ρ	0.0012		λ	0.6451	**	λ	0.0674	***
			(0.598)			(2.436)			(2.808)	
								h	0.9288 (km)	***
									(11.455)	
Adj. R^2	0.6705		0.6710			0.6810			0.6851	
Log-likelihood	392.047		392.226			394.650			395.965	

***, **, and * indicate significance at the 0.01, 0.05, and 0.1 levels, respectively.

Table 6 (cont)

Variables	Kriging model			
	Exponential		Spherical	
Intercept	11.1018	***	11.1035	***
	(318.429)		(325.535)	
AGE	-0.0079	***	-0.0079	***
	(-11.687)		(-11.700)	
WALK	-0.0026	***	-0.0025	***
	(-3.518)		(-3.455)	
AREA	0.0156	***	0.0155	***
	(11.605)		(11.526)	
TAX	3.7596E-08	*	3.7075E-08	*
	(1.922)		(1.888)	
BUSPL	0.0012	*	0.0012	*
	(1.829)		(1.872)	
SHIBUYA	-0.0166	***	-0.0165	***
	(-8.539)		(-9.251)	
PARKQ1	-0.0114	**	-0.0124	**
	(-2.116)		(-2.304)	
PARKQ2	0.0013		0.0015	
	(0.245)		(0.270)	
PARKQ3	0.0132	**	0.0137	***

		(2.546)		(2.650)
PARKQ4		0.0067		0.0075
		(0.891)		(1.000)
PARKL		-0.0046		-0.0047
		(-0.388)		(-0.391)
Model parameters	$\Theta_0+\theta_1$	0.0024	$\theta_0+\theta_1$	0.0024
	θ_0	0.0010	θ_0	0.0014
	$3\times\theta_2$	1.2870 (km)	θ_2	1.3175 (km)
Adj. R^2		-		-
Log-likelihood[a]		394.905		395.001

***, **, and * indicate significance at the 0.01, 0.05, and 0.1 levels, respectively.

[a]Note: To calculate the log likelihood of the kriging models, an additional assumption of normality is required.

We could not find any significant difference between the spatial error models and the kriging models; the log-likelihood values were very close. However, we cannot compare the log-likelihood values of kriging models this simply without an additional distributional assumption. Figure 2.a and b presents the estimated exponential and spherical semivariograms, respectively, which show a clear spatial interdependent relationship. The estimated *h, practical range* for the exponential model, and *range* for the spherical model are 0.9288 km, 1.2870 km, and 1.3175 km. The kriging models detected wider spatial autocorrelation than the spatial error model. That is, the boundary of spatial correlation existed at a 0.9288–1.3175 km radius, on average, from the source.

Figure 2.a

Estimated exponential semivariogram

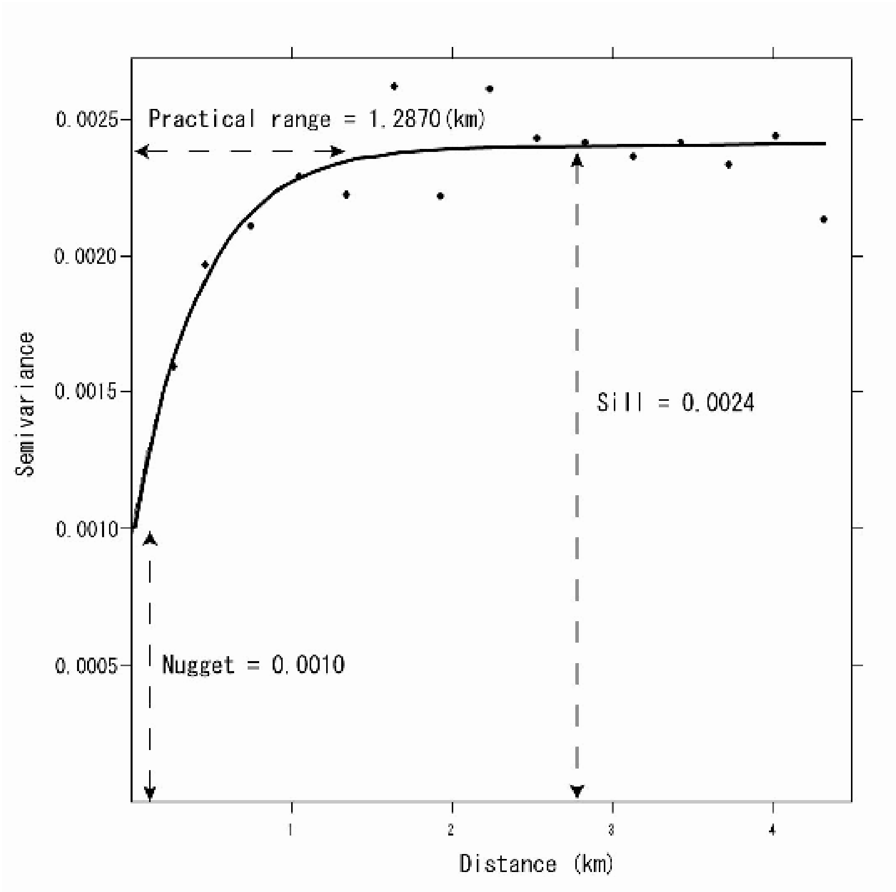
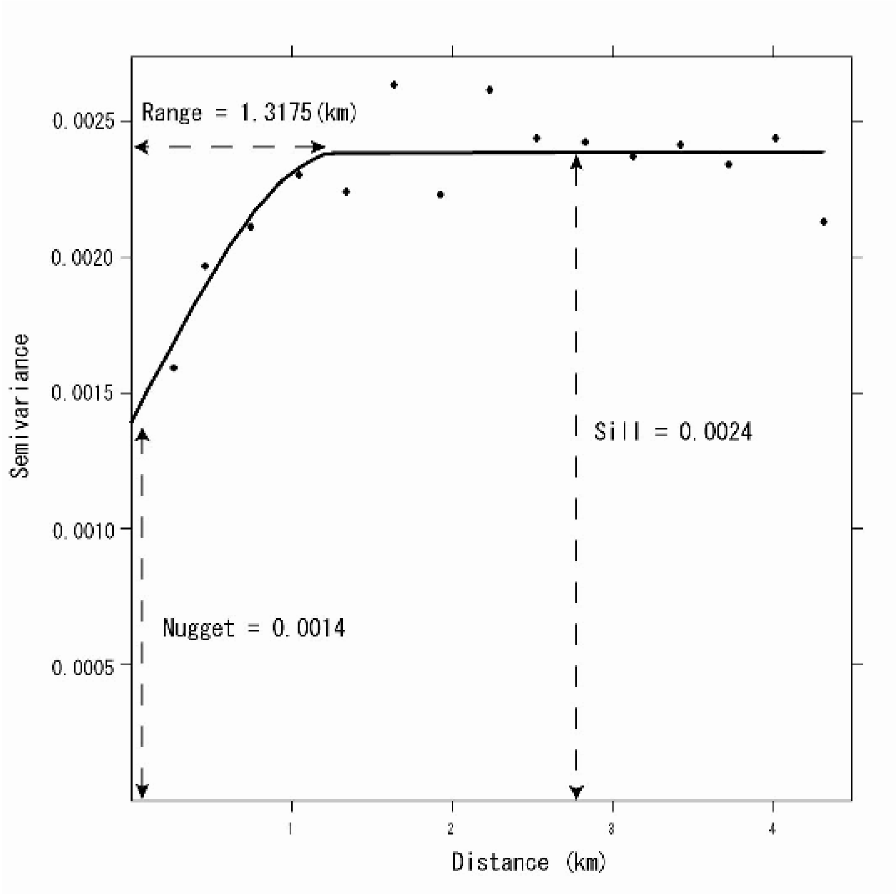


Figure 2.b

Estimated spherical semivariogram



Finally, we considered estimated effect of park amenities on rental prices. The value of the marginal effect is toward an increase from the basic OLS model to the spatial models, except for the spatial lag model, although the statistical difference among the models is not clear. However, with the presence of autocorrelation in the error terms, the estimates of the OLS model are less efficient than those of the spatial error models and the kriging models. Therefore, the marginal value estimated by OLS may be highly *undervalued*.

5. Conclusions and Future Research

This study provides two main empirical findings. The first is that the effect of a neighborhood park's amenities on the rental price of a dwelling depends on the park's size. According to the estimation results, the number of accessible parks has a positive influence on rental prices of neighboring dwellings only if the size of the park is approximately 1400 m^2. This casts doubt on the validity of past urban park planning policy in which the primary concern was not for the size of individual open spaces, but solely the proportion of the aggregated open spaces. This suggests that when the amount of available open spaces to supply a residential zone is limited, it is better policy to allocate a few medium-sized parks than to allocate a lot of small parks to improve the environment of a residential area. The second finding is that, in terms of efficiency, spatial hedonic approaches are preferable to the basic OLS hedonic approach when spatial autocorrelation exists. Additionally, a significant difference between the spatial autoregression model and the kriging model was not observed in this analysis. Because these models are based on different theoretical mechanisms, we cannot simply infer which functional form is best. A possible method to obtain such information is through *cross-validation*, but the limitations in the number of available data mean that this we could not use this method in our analysis.

A further technical challenge is to transform spatial regression models into nonlinear models. To date, basic spatial models are linear, and their functional forms have not been compared with nonlinear models. The extension of spatial models to allow the use of more flexible functional forms should be the subject of further research.

The most powerful use of spatial models is in finding the locational relationship between objects that can be explicitly mathematically modeled. Therefore, such models can be used to predict the effects of proposed locational relationships that have not yet been implemented. In the case of our analysis, it would be possible to compare the current policy of urban park planning and a future proposal, which would be of great interest for future research.

References

Acharya G, Bennett L.L. (2001) Valuing Open Space and Land-Use Patterns in Urban Watersheds. J Real Estate Finance Econ 22, pp. 221-237.

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

Anselin L. (2002) Under the Hood Issues in the Specification and Interpretation of Spatial Regression Models. Agricultural Econ 27, pp. 247-267.

Anselin L, Bera A.K, Florax R, Yoon M.J. (1996) Simple Diagnostic Tests for Spatial Dependence. Regional Sci Urban Econ 26, pp.77-104.

Basu S, Thibodeau T.G. (1998) Analysis of Spatial Autocorrelation in House Prices. J Real Estate Finance Econ 17, pp. 61-85.

Cheshire P. and Sheppard S. (1995) On the Price of Land and the Value of Amenities. Economica 62, pp. 247-267.

Cho S.H, Bowker J.M, Park W.M. (2006) Measuring the Contribution of Water and Green Space Amenities to Housing Values: An Application and Comparison of Spatially Weighted Hedonic Models. J Agri Resource Econ 31, pp. 485-507.

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

Cressie N.A. (1993) Statistics for Spatial Data, New York: Wiley.

Dubin R.A. (1988) Estimation of Coefficients in the Presence of Spatially Autocorrelated Error Terms. The Rev Econ Stat 70, pp. 466-474.

Dubin R.A, Pace R.K, Thibodeau T.G. (1999) Spatial Autoregression Techniques for Real Estate Data. J Real Estate Lit 7, pp. 79-95.

Forrent. http://www.forrent.jp, Cited May through June 2007.

Florax R.J.G.M, Nijkamp P. (2003) Misspecification in Linear Spatial Regression Models. Tinbergen Institute Discussion Paper, TI 2003-081/3.

Geoghegan J. (2002) The Value of Open Spaces in Residential Land Use. Land Use Pol 19, pp. 91-98.

Irwin E.G. (2002) The Effects of Open Space on Residential Property Values. Land Econ 78, pp. 465-480.

Irwin E.G, Bockstael N.E. (2001) The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values. Amer J Agri Econ 83, pp. 698-704.

Jacobs J. (1961) The Death and Life of Great American Cities. Vintage Books, A Division of Random House, Inc., New York.

Jud G.D, Seaks T.G. (1994) Sample Selection Bias in Estimating Housing Sales Prices. J Real Estate Research 9, pp. 289-298.

Kim C.W, Phipps T.T, Anselin L. (2003) Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach. J Environ Econ Manage 45, pp. 24-39.

Leggett C.G, Bockstael N.E. (2000) Evidence of the Effects of Water Quality on Residential Land Prices. J Environ Econ Manage 39, pp. 121-144.

Lutzenhiser M, Netusil N.R. (2001) The Effect of Open Space on a Home's Sale Price. Contemporary Econ Pol 19, pp. 291-298.

Palmquist R.B. (1999) Hedonic Models, in .J.C.J.M. Van Den Bergh (Ed.), Handbook of Environmental and Resource Economics, Edward Elgar, pp. 765-776.

Palmquist R.B. (1992) Valuing Localized Externalities. J Urban Econ 32, pp. 40-44.

Paterson W.R, Boyle K.J. (2002) Out of Sight, Out of Mind? Using GIS to Incorporate Visibility in Hedonic Property Value Models. Land Econ 78, pp. 417-425.

Rosen S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. J Pol Econ 82, pp. 34-55.

Setagaya Ward Statistics 2006, http://www.city.setagaya.tokyo.jp, Cited June 2007.

Shultz S.D, King D.A. (2001) The Use of Census Data for Hedonic Price Estimates of Open-Space Amenities and Land Use. J Real Estate Finance Econ 22, pp. 239-252.

Smith V.K, Poulos C, Kim H. (2002) Treating Open Space as an Urban Amenity. Resource Energy Econ 24, pp. 107-129.

Summary of Setagaya ward administration, 2006, Ch.7. http://www.city.setagaya.tokyo.jp, Cited May through June 2007

Tobler W.R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. Econ Geography 46, pp. 234-240.