Parametric and Semiparametric Estimation of Land Value*

Heonsoo Park† and **Su Youn Jyoung**‡

4 July 2002

Abstract

Due to the large spatial variations of land prices in polycentric urban spatial structure, a traditional parametric approach has limits with respect to the functional forms of land prices, multicollinearity and misspecification problems. On the contrary, a nonparametric approach deals with the complicated curvature of land prices and the other two problems. A semiparametric approach offers important advantages for the estimation of land prices. It combines the benefits of both parametric and nonparametric approaches. In this paper, we use both parametric and semiparametric approaches to estimate land prices from Gangnam Gu, Seoul. In the parametric portion of the model, we include standard land characteristics such as zoning, land shape, school quality, among others. In the nonparametric portion of the model, we include some measures of accessibility such as distance to the city center, distance from the nearest arterial street, subway, or park. Applying the semiparametric approach to parcel data resulted in large improvements in estimation. Specifically, the R^2 went from 0.667 using the traditional parametric approach to 0.886 using the semiparametric model. In addition, the root mean squares of errors (RMSE) went from 0.117 under the parametric model to 0.069 under the semiparametric model, a reduction of 41.03%.

*Paper prepared for presentation at the AREUEA/AsRES International Meeting, Seoul, Korea, 3-6 July 2002. The paper is draft only, not for quotation or citation, but comments are welcome. Address e-mail to Heonsoo Park at heonsoo@cau.ac.kr.

†Department of Urban and Regional Planning, Chung-Ang University, 40-1 Nae-ri Daeduk-myun, Ansung, Kyeonggi 456-756, Korea. Email: heonsoo@cau.anyang.ac.kr

‡Economist, Korea Appraisal Research Institute, Email: jsy1970@kapaet.co.kr

1 Introduction

Accurate measurement of residential real estate price is an issue of great concern to landowners, real estate appraisers, economic analysts, the primary and secondary mortgage markets, producers of housing services, consumers of owner-occupied housing, policymakers, and others. Unfortunately there remain important shortcomings with the techniques commonly used to represent real estate price.

In general, land prices are affected by a number of land characteristics. Parcels of urban land differ from each other, and their prices vary accordingly. Of necessity, values will differ with respect to location. They will differ as well because of factors such as parcel size, shape, zoning, school qualities, and others.

In order to analyze land prices and examine the impact of differential access to transportation or changes in such access, for example, it is necessary to control for these various characteristics. One method for doing so is to estimate a hedonic equation, which attempts to explain variation in land prices using property physical and location characteristics. That is, the coefficients in a hedonic model yield information on the implicit value of the individual characteristics of the land (Rosen, 1974).

The parameters of hedonic land-price equations are typically estimated using ordinary least squares (OLS). This parametric estimation procedure assumes that the residuals are independently and identically distributed with zero mean, a constant variance, and zero covariance. However, the residuals produced by hedonic price equations may be spatially autocorrelated for several reasons. First, neighborhoods tend to be developed at the same time so neighborhood properties have similar structural characteristics. Second, neighborhood residential properties share location amenities. For example, the same policies and fire departments protect area residents and neighborhood children have access to the same public schools. Third, proximity externalities influence the market values of nearby properties in similar ways. Thibodeau (1990) demonstrated that high-rise office buildings reduce the market value of nearby homes by as much as 15 percent. Information on the determinants of proximity externalities in urban land market is difficult and costly to obtain. Consequently, these variables are frequently omitted from empirical land-price specifications. When hedonic equations are used to model land prices, the residuals will contain information on these unobserved land characteristics. Fourth, even in ideal situations where all land characteristic information is available, it is difficult to select the "correct" model specification.

For example, it is difficult to model how access to subway stations or parks gets capitalized into the price of properties. Model misspecification may also contribute to spatially autocorrelated hedonic land-price equation residuals (Gillen *et al.*, 2001).

If hedonic residuals are spatially autocorrelated, the assumption of a zero covariance is violated. The resulting parameter estimates will be inefficient and will produce incorrect confidence intervals for estimated parameters and for predicted values. In addition, the standard tests used to determine the statistical significance of property characteristics assume uncorrelated residuals. Spatial autocorrelation in hedonic residuals violate these assumptions and the standard statistical tests will yield inaccurate conclusions. Hedonic residuals that are positively and spatially autocorrelated will underestimate the population residual variance and the resulting t-statistics will be biased upwards. More accurate parameter estimates can be obtained by explicitly modeling the spatial autocorrelation. Modeling spatial relationships in hedonic land-price equations can also significantly improve the accuracy of market-value predictions (Basu and Thibodeau, 1998).

In a standard parametric hedonic model of land, spatial effects are captured through multiple dummy variables or a high-order polynomial in space. The dummy variable approach unrealistically implies discrete changes in location surfaces and produces imprecise results for local price surfaces with few sales. Polynomials become complex and unwieldy when location surfaces of land price are highly nonlinear (McMillen and Thorsnes, 2000).

Nonparametric approaches have distinct advantages over more conventional hedonic analyses of land markets. Nonparametric estimators impose little structure on the parametric specification of the hedonic models and allow the data to determine the appropriate degree of nonlinearity. Therefore, nonlinearities in location variables are easy to model using nonparametric approaches. High-order polynomials could produce the same results but are more cumbersome when the function is highly nonlinear. Nonparametric estimators are less likely to produce unrealistic edge effects than are polynomials,[1] and are, in general, more accurate at points away from the mean than are parametric approaches.

However, there are several primary disadvantages to nonparametric estimation. First, hypothesis tests are more difficult to construct than in traditional parametric estimation. Second, nonparametric estimators require large sample sizes to be accurate when the number of explana-

[1] For example, McDonald and Bowman (1979) find that the relationship between land values and distance from the Chicago city center is best described by a fourth-order polynomial. Similar results have been found in other studies of Chicago land values, such as McMillen (1996). These estimates typically imply an unrealistic upturn in land values at the edge of the area covered in the data set.

tory variables is large. Third, dummy variables are usually omitted from full nonparametric estimation, and recovering their effects is difficult (Horowitz and Härdle, 1996). However, dummy variables are common in studies of residential land market, and are occasionally the focus of the study. Fourth, kernel estimation is subject to asymptotic bias, making confidence intervals inexact.

Semiparametric estimation combines the benefits of parametric and nonparametric approaches. In residential land studies, for example, the effects of lot size and most dummy variables can be modeled with reasonable precision using a standard parametric specification, but location effects are expected to be highly nonlinear. In a semiparametric specification, efficiency is improved by specifying a parametric portion of the model for those characteristics whose effects are not expected to be highly nonlinear. Location effects are modeled nonparametrically (McMillen and Thorsnes, 2000, pp. 202-203).

In this paper, we use both parametric and semiparametric approaches to estimate land price in Gangnam Gu, Seoul. To contral spatial variations of land values, we include some standard measures of accessibility, such as distance from the city center, distance from the nearest arterial street, distance from the nearest subway station, and distane from the nearest park in hedonic land price equations and estimate them nonparametrically. The analysis of the paper yields a general conclusion that the semiparametric estimator yields more accurate land price than parametric specifications. In addition, applying the semiparametric approach results in large improvements in estimation.

The remainder of the paper consists of four sections. The next section introduces the formal framework of land price models. In Section 3, we briefly describe the data and then present statistical results in Section 4. Section 5 concludes.

2 Models

As is standard in the hedonic price literature, the dependent variable is the natural log of land price,[2] which we denote by y. The explanatory variables can be divided into two groups. The first group, which is represented by the vector $\mathbf{x}$, includes structural characteristics of land and dummy variables. Myriad hedonic price studies suggest that these variables can be assumed to enter the

[2]Since the price schedule represents an equilibrium locus, it cannot be interpreted as representing either the demand or the supply of those characteristics. It is the result of the interaction of both sides of the market. For this reason, the functional form for the hedonic equation is not determined theoretically. Rather, it is determined empirically. Of all the common functional forms we tried, the semi-logarithmic form provided the best fit to our data. Empirical issues in hedonic estimation are discussed in Palmquist (1991).

hedonic equation as the linear function $\beta'\mathbf{x}$. The other group of variables, which is represented by the vector $\mathbf{z}$, includes distance from the CBD, distance from the nearest subway station, and other continous location measures. The relationship between y and $\mathbf{z}$ may be highly nonlinear. We expect prices to rise near CBD and subway stations. But prices may not change uniformly with distances from these sites. Further, there are missing variables such as parks and major streets that have local effects on land prices. These missing variables will affect the estimated relationship between land prices and $\mathbf{z}$ if they are correlated with the distance variables.

The hedonic-price function is assumed to have the following form:

$$y_i = \beta'\mathbf{x}_i + g(\mathbf{z}_i) + \varepsilon_i \tag{1}$$

where $i = 1, \ldots, n$. The function $g(\cdot)$ is assumed to be smooth and continuous. The error terms, ε_i, are assumed to be independent, but they may be heteroskedastic and do not have to have zero mean.

We use two approaches to estimate (1). The first is a standard parametric approach, where $g(\mathbf{z}_i)$ is the linear function $\delta'\mathbf{z}_i$. The second alternative is the semiparametric approach proposed by Robinson (1988).[3] In a parametric approach, the vector of coefficients, δ, is most easily estimated by regressing y on $\mathbf{x}$ and $\mathbf{z}$. Identical estimates are obtained from a three-stage procedure. In the first stage, regress y and each of the K variables in the vector $\mathbf{x}$ on $\mathbf{z}$, and form the predicted values $\hat{y}$ and $\hat{\mathbf{x}}_k$. Next, regress $y - \hat{y}$ on the vector of residuals $\mathbf{x}^*_k \equiv \mathbf{x}_k - \hat{\mathbf{x}}_k$ to estimate β. In the third step, estimate δ by regressing $y - \hat{\beta}'\mathbf{x}$ on $\mathbf{z}$. Intuitively, the first step purges y and $\mathbf{x}$ of the effects of $\mathbf{z}$, and the second stage estimates the independent effect of $\mathbf{x}$ on y. The third stage estimates the effect of $\mathbf{z}$ on the dependent variable after the effects of $\mathbf{z}$ have been removed (Thorsnes and McMillen, 1998).

The semiparametric approach is identical to the linear case, but a nonparametric estimator replaces OLS when $\mathbf{z}$ is involved in the calculations. Estimation proceeds as similar as in the parametric approach. In the first stage, use a nonparametric procedure to estimate $E(y_i|\mathbf{z}_i)$ for each of K explanatory variables in the vector $\mathbf{x}$. The predicted values are $\hat{y}_i$ and $\hat{\mathbf{x}}_{ki}$. Next, regress $y_i - \hat{y}_i$ on all $\mathbf{x}^*_{ki} \equiv \mathbf{x}_{ki} - \hat{\mathbf{x}}_{ki}$ to estimate β. The estimate is $\hat{\beta}$. In the third step, use the nonparametric procedure to estimate $(y_i - \beta'\mathbf{x}_i|\mathbf{z}_i)$ using $E(y_i - \hat{\beta}'\mathbf{x}_i|\mathbf{z}_i)$ as the dependent variable. The estimate is $\hat{g}(\mathbf{z}_i)$. We estimate σ^2 as $\hat{\sigma}^2 = 1/n \sum_{i=1}^{n}(y_i - \hat{\beta}'\mathbf{x}_i - \hat{g}(\mathbf{z}_i))^2$ and then get the covariance matrix

[3] Relevant applications of the semiparametric approach include Engle *et al*. (1986), Stock (1991), Anglin and Gencay (1996), Thorsnes and McMillen (1998), and Pavlov (2000).

for $\hat{\beta}$ is $\hat{\sigma}^2(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}$, where $\mathbf{X}^*$ is the matrix containing $\mathbf{x}^*_{ki}$.

Linearity is a special case of the semiparametric procedure. If $g(\cdot)$ is linear, the semiparametric approach provides consistent but inefficient estimates. The semiparametric approach provides consistent estimates when the function is nonlinear, where OLS estimates are inconsistent.

We use a kernel procedure for the nonparametric estimates.[4] As an example, consider estimating $E(y_i|\mathbf{z}_i)$. The estimate for observation i is:

$$\hat{y}_i = \frac{\sum_{j=1}^n k_h(z_i - z_j) y_j}{\sum_{j=1}^n k_h(z_i - z_j)} \tag{2}$$

where, apart from a constant term that cancels out in Equation (2)

$$k_h(z_i - z_j) = K\left(\frac{z_i - z_j}{h}\right) \tag{3}$$

where $K(\cdot)$ is the kernel function and h is the bandwidth. We use a stadard normal multivariate kernel for our estimates. The first step is to calculate the Choleski decomposition of the inverse of the sample covariance matrix for $\mathbf{z}$, which we denote by $\mathbf{S}$. Next, construct the new vector $\mathbf{v} \equiv \mathbf{Sz}$. the multivariate normal kernel is simply:

$$K\left(\frac{z_i - z_j}{h}\right) = \prod_{q=1}^{m} \phi\left(\frac{v_{qi} - v_{qj}}{h}\right) \tag{4}$$

where ϕ is the standard normal density function and m the number of elements in $\mathbf{z}$.

When constructing the estimate for observation i, the kernel estimator puts more weight on observations that have values of $\mathbf{z}_j$ that are close to $\mathbf{z}_i$. The kernel estimatr allows for interactions among the m variables that compose $\mathbf{z}$ because the multivariate normal density kernel takes into account the correlations among the variables. The bandwidth parameter controls the relative weight given to nearby observations. When h is small, nearby observations are given more weight relative to when h is large.

We use the method of cross validation to choose the value of h. The model is estimated for every observation in the data set. The cross validated estimates are obtained by omitting observation i and reestimating the entire model. The residual for observation i is $y_i - \hat{\beta}_i'\mathbf{x}_i - \hat{g}_i(\mathbf{z}_i)$, where $\hat{\beta}_i$ and $\hat{g}_i$ are the estimates obtained with observation i deleted. The bandwidth choice is the value of h that minimizes $n^{-1}\sum_{i=1}^n (y_i - \hat{\beta}_i'\mathbf{x}_i - \hat{g}_i(\mathbf{z}_i))^2$. A grid search leads to an optimal value of h. We then estimate the model again including observation i when constructing $\hat{y}_i$.

A pseudo-R^2 measures the goodness of fit. The pseudo-R^2 is the standard R^2 obtained by regressing $\hat{y}$ on y. The pseudo-R^2 is identical to the conventional R^2 in a linear parametric model.

[4] Useful reviews of kernel regression procedures include Härdle (1990) and Härdle and Lington (1994).

As with a conventional R^2, the pseudo-R^2 rises when the predicted values are closer to the actual values of y. We use a Hausman (1978) test to determine whether $g(\cdot)$ is statistically different from the null parametric specification. Let $\hat{\beta}_{sp}$ and $\hat{\beta}_{ols}$ denote the $(k-1) \times 1$ vectors of semiparametric and OLS coefficient estimates (the intercept is not included). Although the semiparametric estimates are inefficient when the parametric specification is correct, we should find that $\hat{\beta}_{sp}$ and $\hat{\beta}_{ols}$ are approximately equal because both sets of estimates are consistent. When the model is more nonlinear than the parametric specification, the semiparametric estimates are consistent but the OLS estimates are not, and thus $\hat{\beta}_{sp}$ and $\hat{\beta}_{ols}$ should be different. Letting $\hat{\mathbf{V}}_{sp}$ and $\hat{\mathbf{V}}_{ols}$ denote the covariance matrix estimates for $\hat{\beta}_{sp}$ and $\hat{\beta}_{ols}$, respectively, the test statistic is as follows:

$$H = (\hat{\beta}_{sp} - \hat{\beta}_{ols})'(\hat{\mathbf{V}}_{sp} - \hat{\mathbf{V}}_{ols})^{-1}(\hat{\beta}_{sp} - \hat{\beta}_{ols}) \tag{5}$$

where H is distributed χ^2 with $k-1$ degrees of freedom.

3 Data

The data set includes 471 residential parcels appraised in 1998 in Gangnam Gu, Seoul. These data are called as "comparables" because they are used later for automately evaluating each residential parcels ("subjects") in the same area. Concentrating on one land use reduces unobserved heterogeneity and permits us to focus in detail on the value-location relationship for a single use. Moreover, the focus on residential land helps us to eliminate the selection bias problems noted in Colwell and Sirmans (1993). Previous researchers (e.g., McMillen and McDonald, 1991; Wallace, 1988) find that selection bias is most prominent in manufacturing and commercial land uses, while selection-bias correction variables are insignificant for the single-family residential category.

Table 1 contains a brief description of the variables included in the model. The dependent variable, LPRICE, is the natural logarithm of the ratio of land price to total square meters. The AREA variable is used to analyze the relationship between land values and parcel sizes.

After determining the location of each parcel on a map, additional variables are added to the data set. Added variables include standard measures of accessibility: distance from the Seoul central business district or subcenters (CBD), distance from the nearest arterial street (ROAD), distance from the nearest subway station (SUBWAY), and distance from the nearest park (PARK). All distance variables enter the estimating equations in level form, so their coefficients are directly interpretable as gradients.[5]

[5]The simple log-level econometric model can be derived from theory when the housing production function is

Table 1: Variable name, descriptive statistics and definitions

Variable	Mean	Definition
LPRICE	14.3 (0.20)	Ratio of land price to square meters (in log)
AREA	5.92 (1.28)	parcel size (in log of square meters)
PARK	0.24(0.19)	Distance (in kilometers) to the nearest park
ROAD	0.16 (0.10)	Distance (in kilometers) to the nearest arterial road
SUBWAY	0.66 (0.38)	Distance (in kilometers) to the nearest subway station
CBD	9.05 (2.31)	Distance (in kilometers) to the nearest city center
LANDUSE	0.72	1 if parcel is used for the single-family detached, 0 otherwse
SLOPE	0.88	1 if parcel is flat, 0 otherwise
SHAPE1	0.30	1 if parcel is squared, 0 otherwise
SHAPE2	0.23	1 if parcel is long-squared, 0 otherwise
SHAPE3	0.30	1 if parcel is trapezoid, 0 otherwise
FACE1	0.07	1 if parcel is faced to one or more roads of class 1, 0 otherwise
FACE2	0.11	1 if parcel is faced to one or more roads of class 2, 0 otherwise
FACE3	0.57	1 if parcel is faced to one or more roads of class 3, 0 otherwise

Note: Roads are classified by their width. Class 1 has over 25 meters in width. Class 2 is in the middle, 12-25 meters in width. Class 3 is in the lower, 8-12 meters in width. Standard deviations are in parentheses and have been omitted for dummy variables.

The other variables are included to control for additional factors that play some role in the determination of land values. LANDUSE equals 1 if a parcel is used for the single-family detached and 0 otherwise, and SLOPE equals 1 if a parcel is flat and 0 otherwise. One would expect the price embodied in land values due to these characteristics to yield positive coefficients in each case.

Finally, we have included a vector of dummy variables, SHAPE and FACE, which capture the effect on land values of parcel's shape and the type of the road to which a parcel is faced, respectively. These variables capture the effect on land values of a parcel's dominant shape and the type of the road to which the parcel is faced.

The second column of Table 1 contains the descriptive statistics, where standard deviations have been ommitted for dummy variables.

4 Results

Estimation results are presented in Table 2. We present detailed results for the Gaussian kernel estimates only, but the results of other kernels are nearly idential. Cross validation indicates that $h = 0.29$ provides the best fit, and we use this value in all subsequent calculations with the Gaussian kernel. To simplify comparison of parametric and semiparametric estimates, we calculate

Cobb-Douglas and the price elasticity of housing demand is unitary (Mills and Hamilton, 1994).

Table 2: Estimation results

Variable	Parametric Estimates		Gaussian Semiparametric Estimates	
Constant	14.572	(0.058)***	–	
AREA	0.032	(0.008)***	0.033	(0.010)***
LANDUSE	-0.027	(0.014)**	0.028	(0.014)**
SLOPE	0.056	(0.018)***	0.031	(0.019)*
SHAPE1	0.036	(0.018)**	0.027	(0.018)*
SHAPE2	0.053	(0.018)***	0.042	(0.020)**
SHAPE3	0.074	(0.018)***	0.067	(0.018)***
FACE1	0.163	(0.040)***	0.188	(0.043)***
FACE2	0.084	(0.020)***	0.068	(0.025)**
FACE3	-0.119	(0.013)***	-0.133	(0.013)***
PARK	-0.151	(0.032)***	–	
ROAD	-0.195	(0.061)***	–	
SUBWAY	-0.053	(0.021)**	–	
CBD	-0.043	(0.003)***	–	
R^2	0.667		0.886	
RMSE	0.117		0.069	
Hausman test	–		32.34***	

Note: Standard deviations are in parentheses
* indicates significance at $\alpha \leq 0.10$.
** indicates significance at $\alpha \leq 0.05$.
*** indicates significance at $\alpha \leq 0.01$.

root mean squares errors (RMSE) for both parametric and semiparametric approaches.

Most results are as expected. As is common in studes of urban land values, the R^2s are high at 0.667 for the parametric (OLS) model and 0.886 for the semiparametric model.[6] The semiparametric model shows strong improvements in goodness of fit relative to the parametric (OLS) model. For example, the root mean squares of errors (RMSE) went from 0.117 under the parametric model to 0.069 under the semiparametric model, a reduction of 41.03%.

Urban theory predicts that land values are a smooth, convex function of distance from the city center and other sites for which access is valuable (see, for example, Mills, 1972; Muth, 1969). Our empirical studies support this prediction. Land values decline by about 4.3% per kilometer with distance from the Seoul CBD or subcenters. They also decline by about 5.3% per kilometer with distance from the nearest subway station. Land values appear to decline rapidly with distance for the nearest arterial street and park by about 19.5% and 15.1%, respectively. Land values are

[6]The R^2 for the semiparametric estimator is calculated by regressing the predicted values on actual values.

higher in low slope parcels. Land values are highest in trapezoided parcels. They are also highest in parcels which are faced to one or more roads of class 1.

One of the interesting results is the parameter estimates of AREA. The positive parameter estimate implies a convex value-size relationship, which suggests that assembling costs of parcels cause small parcels to trade at a discount.[7] This results support the results of Colwell and Sirmans (1978), who find that the value-size relationship is convex for small parcels and then concave at larger values.

Another interesting result is the parameter estimate of LANDUSE. In a parametric approach, the sign of the parameter is negative and statistically significant, meaning that the land values are higher in the multi-family detached or apartment complex sites. On the contrary, in a semiparametric approach, the sign of the parameter is positive and statistically significant, meaning that the land values are higher in the single-family detached areas and coinciding more closely with most individuals' priors than OLS.

5 Conclusions

"Location, location, location" is a frequent explanation of land values. In real estate valuation literature, location is used as a proxy for numerous unobserved variables. The standard hedonic approach to property valuation assumes that the price of a land is a function of the values markets place on its physical and locatioal characteristics. One of the most significant drawbacks of hedonic valuation, however, is that we cannot observe or measure all relevant characteristics. This problem is typically addressed by using an artificial measure of location as a proxy for variations in the unobserved covariates.

In this paper, we use both parametric and semiparametric approaches to estimate land prices of parcels from Gangnam Gu, Seoul. In the parametric approach, we use several access variables and estimate them parametrically. In the semiparametric approach, on the contrary, we estimate those locational variables nonparametrically. Applying the semiparametric approach to parcel data resulted in large improvements in estimation. Specifically, the R^2 went from 0.667 using the traditional parametric approach to 0.886 using the semiparametric model. In addition, the root mean squares of errors (RMSE) went from 0.117 under the parametric model to 0.069 under the semiparametric model, a reduction of 41.03%.

[7]From Table 2, the parameter estimate of AREA is 0.032, which means $\partial\left(\frac{LPRICE}{AREA}\right)/\partial AREA$. Therefore, $\partial LPRICE/\partial AREA$ equals to 1.032.

It should be noted that the effects of some variables to land values are differ according to which approaches we use. For instance, in a parametric approach, the parameter estimate of LANDUSE is negatively and statistically significant, meaning that the land values are higher in the multi-family detached or apartment complex sites. In a semiparametric approach, its parameter estimate is changed to be positively and statistically significant, meaning that the land values are higher in the single-family detached areas and coinciding more closely with most individuals' priors than OLS.

In addition, according to Hausman test in Table 2, hedonic price functions of parcel are expected to be nonlinear, but theory offers little guidance on the form of the nonlinearity. Especially, choosing the appropriate functional form is particularly difficult for location variables because the appropriate form tends to be unique to the place. Therefore, semiparametric estimation offers important advantages for hedonic price function estimation. Parametric structure is maintained for the part of the function whose structure is known a priori, whereas variables whose effects are expected to be highly nonlinear enter the nonparametric part of the model. The semiparametric approach preserves degrees of freedom but acknowledges that the full functional form is not known with confidence.

References

[1] Anglin, P.M. and R. Gencay (1996). Semiparametric Estimation of a Hedonic Price Function. *Journal of Applied Economics*, **11**: 633-648.

[2] Basu, S., and T. Thibodeau. (1998). "Analysis of Spatial Autocorrelation in House Prices," *Journal of Real Estate Finance and Economics*, **17:1**, 61-85.

[3] Colwell, P.F., and C.F. Sirmans. (1993). "A Comment on Zoning, Returns to Scale, and the Value of Undeveloped Land," *Review of Economics and Statistics*, **73**, 61-85.

[4] Colwell, P.F., and C.F. Sirmans. (1978). "Area, Time, and the Value of Urban Land," *Land Economics*, **54**, 514-519.

[5] Engle, R, C.W.J. Granger, J. Rice, and A. Weiss (1986). Semiparametric Estimates of the Relationship between Weather and Electricity Sales. *Journal of the American Statistical Association*, **81**, 310-320.

[6] Gillen, K., T. G. Thibodeau, and S. Wachter. (2001). "Anisotropic Autocorrelation in Housing Prices," *Journal of Real Estate Finance and Economics*, **23:1**, 5-30.

[7] Härdle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.

[8] Härdle, W. and D. Linton (1994). "Applied Nonparametric Methods," In R.F. Engle and D.L. McFadden, (eds.), *Handbook of Economics*, Vol. 4, New York : Elsevier.

[9] Hausman, J.A.(1978). "Specification Tests in Econometrics," *Econometrica*, **46**: 1251-1272.

[10] Horowitz, J.L., and W. Härdle (1996). "Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates," *Journal of the American Statistical Association*, **91**, 1632-1640.

[11] McDonald, J.F., and H.W. Bowman. (1979). "Land Value Functions: A Reevaluation," *Journal of Urban Economics*, **6**, 25-41.

[12] McMillen, D. P. (1996). "One Hundred Fifty Years of Land Values in Chicago: A Non-Parametric Approach," *Journal of Urban Economics*, **40**, 100-124.

[13] McMillen, D. P., and J.F. McDonald. (1991). "Urban Land Value Functions with Endogenous Zoning," *Journal of Urban Economics*, **29**, 14-27.

[14] McMillen, D. P., and P. Thorsnes. (2000). "Housing Prices and Information on Superfund Sites," *Advances in Econometrics*, **14**, 201-228.

[15] Mills, Edwin S. (1972). *Studies in the Structure of the Urban Economy*. Baltimore: Johns Hopkins University Press.

[16] Mills, Edwin S., and Bruce W. Hamilton. (1994). *Urban Economics*. New York: Harper Collins.

[17] Muth, Richard F. (1969). *Cities and Housing*. Chicago: University of Chicago Press.

[18] Palmquist, R.B. (1991). "Hedonic Methods," In J.B. Braden and C.D. Kolstad, (eds.), *Measuring the Demand for Environmental Quality*, pp. 77-120, North Holland.

[19] Pavlov, A.D. (2000). "Space-Varying Regression Coeficients: A Semi-parametric Approach Applied to Real Estate Markets," *Real Estate Economics*, **28:2**, 249-283.

[20] Robinson, P.M. (1998). "Root-N-Consistent Semiparmetric Regression," *Econometrica*, **56**: 931-954.

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

[22] Stock, J.H. (1991). "Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits," In W.A Barnett, J.Powell, and G.M. Tauchen, (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, New York: Cambridge University Press.

[23] Thibodeau, T. G. (1990). "Estimating the Effect of High Rise Office Buildings on Residential Property Values," *Land Economics*, **66:4**, 402-408.

[24] Thorsnes, P., and D.P. McMillen. (1998). "Land Value and Parcel Size: A Semiparametric Analysis," *Journal of Real Estate Finance and Economics*, **17:3**, 233-244.

[25] Wallace, Nancy E. (1998). "The Market Effects of Zoning Undeveloped Land: Does Zoning Follow the Market?" *Journal of Urban Economics* **23**, 307-326.