Measuring The Benefits of Air Quality Improvement:

A Spatial Hedonic Approach

Chong Won Kim, Tim Phipps, and Luc Anselin

Paper prepared for presentation at the AAEA annual meetings, Salt Lake City, August, 1998. Authors are Economist, Korea Research Institute for Human Settlements, Seoul, Korea; Professor, Agricultural and Resource Economics, West Virginia University; and Professor and Director, Regional Research Institute, West Virginia University, Morgantown, WV, 26505. Comments are welcome. Address e-mail to Tim Phipps at tphipps@wvu.edu.

Measuring The Benefits of Air Quality Improvement:

A Spatial Hedonic Approach

Authors: Chong Won Kim, Tim Phipps, and Luc Anselin[i]

The primary objective of this paper is to measure the marginal value of air quality improvement in Seoul, Korea by combining spatial econometric methods with a hedonic housing price model. Increasingly, spatial dependence and spatial heterogeneity are recognized as factors that may affect the efficiency and consistency of hedonic model estimates. The current literature shows that hedonic housing price models have generally neglected the issue of spatial dependence of housing data. Neglect of spatial considerations in econometric models may lead to serious errors in the interpretation of regression diagnostics (Anselin and Bera).

Can (1990, 1992) and Dubin (1988, 1992) considered the spatial nature of hedonic housing data in estimating the hedonic housing price model. Can considered the spatial lag with spatial expansion model and Dubin considered spatial autocorrelation using a geostatistical approach. None have jointly considered spatial econometric and environmental factors.

Seoul was chosen for two reasons. First, the United Nations Environment Programme and the World Health Organization (1994) have reported that Seoul has serious problems with sulfur dioxide and lesser, though increasing, problems with nitrogen oxides. Since ambient air quality patterns in Seoul vary in a systematic spatial pattern, the city provides a good laboratory for testing the spatial hedonic model. Second, the authors had access to air quality monitoring data for 20 stations in Seoul and an extensive geo-referenced survey of 1121 households.

We first describe the data, then develop the spatial hedonic models and finally present the estimation results.

Data

A sample of the Seoul housing market was taken by the Korea Research Institute for Human Settlement (KRIHS) in 1993. The survey involved in-person interviews of a random sample of owner and renter-occupied dwellings. The sample was designed to cover all 22 districts in the city of Seoul with three to five subdistricts chosen for sampling within each district (the sample covers 78 subdistricts). Originally 1560 households were selected for the survey but after screening for missing observations, 1121 observations (609 for owner occupied households, 512 for renter households) were selected this study. This paper only reports the results for owners.

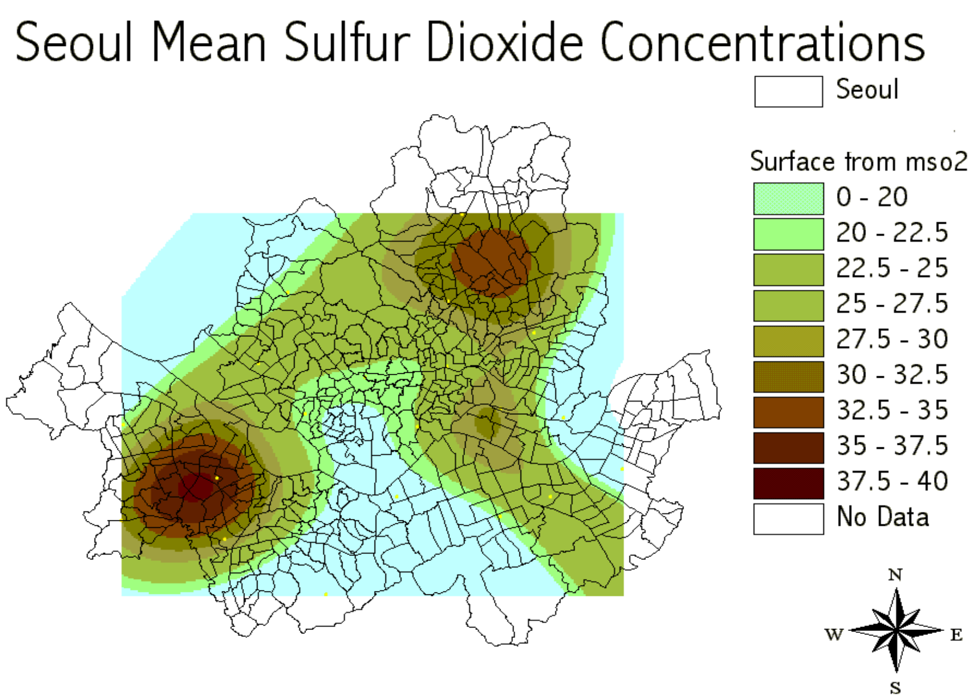
The survey collected data on housing price, house and neighborhood characteristics, and socio-economic data of the household (Table 1). The housing prices are based on respondent estimates. These estimates can be considered reasonably accurate given the institutional characteristics of the Seoul market. The Korean government has adopted a system of posted land prices for property tax collection. The posted land prices are reviewed and revised to reflect current land market prices based on sample areas across the whole country every two years.

Table 1: Data

Variable	Definition
PRVAL	Property value of owner occupied house
TFLSP	Total floor space of house
NMRMS	Number of rooms
NMBATH	Number of bath rooms
HSAGE	Age of house.

DFUEL	Variable equals one if heated by oil or gas, zero if heated by briquette.
DHOUS	Variable equals one if housing type is house, else zero
DINCOM	Variable equals one if neighborhood incomes are high or high-middle, else zero
ACSHPT	Accessibility to the nearest hospital (time)
ACSSCH	Accessibility to the school of junior or high school (time)
ACSSUB	Accessibility to the nearest subway station (time)
ACSPRK	Accessibility to the nearest park or swimming pool (time)
SO2	Sulfur dioxide gas (SO_2) levels (unit: ppb).

Air pollution is monitored daily at 20 stations in Seoul. We used spline interpolation to impute the ambient air quality at each of the 78 residential subdistricts covered by the housing survey. We assumed that the air quality within each subdistrict was the same. Figures 1 shows the distribution of SO_2 for Seoul. SO_2 is most heavily concentrated near the industrial sections in the southwest and northeast regions.



SPATIAL HEDONIC MODELS

There are two basic types of spatial econometric model: the spatial lag model and the spatial error model. While the two are related, each has a different economic interpretation.

The Spatial Lag Model

The spatial lag hedonic model is analogous to an autoregressive time series model. The difference is that in a time series model past observations of the dependent variable partially explain current observations. In the hedonic spatial lag model, nearby observations of housing prices partially explain local housing price. The spatial lag model is an appropriate tool when capturing neighborhood spillover effects. That is, this model assumes that the spatially weighted sum of neighborhood housing prices is as an explanatory variable in housing price formation [i.e., $P_1 = \rho\ (w_{11}P_1 + w_{12}P_2 + w_{13}P_3 + \cdots + w_{1n}P_n) + X\beta + \varepsilon_1$], where w_{ij} is the spatial weight that links observation i and j]. This relationship is in accord with the standard real estate appraisal process of using comparable sales prices in forming an appraisal price. A general spatial lag hedonic housing price model can be written as follows:

$$P = \rho WP + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon \qquad (1)$$

where P is the housing price, W is a row-standardized spatial weight matrix, X_1 is the vector of structural characteristics, X_2 is the vector of neighborhood characteristics, X_3 is the environmental quality variable, and error terms are assumed as $\varepsilon \sim N(0, \sigma^2 I)$.

OLS estimators are biased and inconsistent if the spatial lag model is the correct model specification. In this case, maximum likelihood estimation (ML) and instrumental variable (IV) estimation are unbiased estimators. For ML estimation, we need to assume

$\varepsilon \sim N(0, \sigma^2 I)$. The normality assumption of the error term can be tested using the lagrange multiplier (LM) test against spatial error dependence in the presence of partially lagged dependent variables (Anselin 1988).

The Spatial Error Model

When spatial dependence is present in the error term, the spatial error model is appropriate. The hedonic spatial error model is:

$$P = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon \qquad (2)$$

$$\varepsilon = \lambda W_2 \varepsilon + \mu, \quad \mu \sim N(0, \sigma^2 I)$$

where λ is the spatial lag operator and W_2 is the spatial weight matrix.

The spatial error model implies that spatial interactions among the observations are the result of omitted variables that are spatially autocorrelated. The idea of this model is similar to the first order moving average process in time series models. While the error (e_t) in the time series model can be expressed as the weighted sum of uncorrelated and identically distributed random errors (μ_t, μ_{t-1} , $\mu_{t-2\ldots}$), the error term (ε) in the spatial error model is the sum of all weighted errors since the error terms can be written as $e = [I - lW]^{-1} m = [I + lW + (lW)^2 + (lW)^3 + \ldots] \cdot m$.

OLS estimators remain unbiased, but they are no longer efficient if the spatial error model is a valid model specification. The spatial error hedonic property value model can be estimated by ML estimation assuming normal error terms.

The normality assumption of the error term can be tested using the LM test against heteroskedasticity and spatial error dependence as in the case of the spatial lag model (Anselin 1988).

Creation of Spatial Weight Matrices

We assume that observations in each of the subdistricts are distributed uniformly from the district centroid. The distance between subdistricts is measured by the distance between the subdistrict centroids. The strength of spatial interaction between spatial units is generally inversely related to the distance. We examined a number of different specifications for the spatial weight matrix, including inverse distance, inverse distance raised to a power, and general contiguity matrices. Our final specification used a general contiguity matrix based on distance between subdistrict centroids:

$$w_{ij} = 1 \text{ if } d_{ij} \prec D_c$$
$$else\ w_{ij} = 0$$
$$where\ D_c \text{ is the critical distance}$$

Hypotheses and Results

Our qualitative hypotheses for the estimated variables are straightforward. For house characteristics we hypothesize that price is positively related to floor space, number of rooms, number of bathrooms, the fuel type dummy (indicating presence of a modern heating system), and the detached house dummy variable, and negatively related to age. For neighborhood characteristics we expect price to be positively related to neighborhood income and negatively related to each of the accessability variables, since they are measured in terms of travel time from the house to the facility. We expect SO_2 concentrations to be negatively related to housing price. We expect SO_2 to be a statistically significant factor because levels of sulfur dioxide have been stable over time, concentrations are publicly reported, and perhaps most importantly, sulfur dioxide is a visible pollutant.[ii]

We considered four functional forms: linear-linear, linear-log. log-linear, and log-log functional forms with nine different weight matrices. To conserve space, only the best fitting model, the log-linear form for the spatial lag model, is presented for three critical distance values: 2 km, 3 km, and 4 km. Results for the spatial error model are not presented because after estimating the spatial lag model, the LM test for spatial error dependence showed no significant error dependence in any model.

Maximum likelihood estimations of the log-linear functional form show that the signs of parameters for all the variables are as hypothesized (Table 2). All variables are significant at least at the 5% level except for two of the four accessibility variables (ACSPHT and ACSSUB).

The criteria for measures of goodness-of-fit also show that the model with a 4 km cut-off weight matrix fits best. There is no detectable heteroskedasticity problem in the spatial lag model since the Breusch-Pagan (spatial) test for heteroskedasticity can not reject the null hypothesis at the 5% level $(H_0 : E(e_t^2) = s^2)$.

Table 2 MLE for Log-Linear Spatial Lag Model

VARIABLE	2 km	3 km	4 km
W_PRVAL	0.294*** (0.049)	0.349*** (0.063)	0.473*** (0.068)
CONSTANT	12.73*** (0.948)	11.69*** (1.222)	9.321*** (1.315)
DHOUS	0.112*** (0.039)	0.125*** (0.039)	0.129*** (0.039)
DINCOM	0.163*** (0.04)	0.166*** (0.04)	0.156*** (0.04)
DFUEL	0.200*** (0.05)	0.197*** (0.05)	0.185*** (0.05)
TFLSP	0.0112*** (0.0008)	0.011*** (0.0008)	0.0112*** (0.0008)
NMRMS	0.0785*** (0.013)	0.078*** (0.013)	0.078*** (0.013)
NMBATH	0.082*** (0.030)	0.082*** (0.030)	0.0826*** (0.030)
HSAGE	-0.0049** (0.002)	-0.0052** (0.002)	-0.0055** (0.002)

ACSHPT	-0.0041* (0.002)	-0.0038* (0.002)	-0.0041* (0.002)
ACSSCH	-0.0061*** (0.002)	-0.0058*** (0.002)	-0.0057*** (0.002)
ACSSUB	-0.0026 (0.001)	-0.0028* (0.001)	-0.00268 (0.001)
ACSPRK	-0.0029* (0.001)	-0.0029* (0.001)	-0.0031** (0.001)
MSO2T	-0.0089*** (0.0026)	-0.0089*** (0.0026)	-0.007*** (0.0026)
R2 Sq.	0.6376	0.6320	0.6375
Corr.	0.6255	0.6236	0.6277
LIK AIC	-269.254	-272.361	-267.629
SC	628.274	634.487	625.023

Note : ***: significant at 1 % ** : significant at 5 %
*: significant at 10 % ( ) : Standard Deviation

Economic Interpretation of the Spatial Lag Hedonic Model

To interpret the spatial-lag hedonic model, first express the model as:

$$P = [I - \rho W]^{-1} X\beta + \nu \qquad (3)$$

where P is an (n ×1) column vector, $[I-\rho W]^{-1}$ is an (n × n) inverse matrix, X is an (n × k) matrix, β is an (k×1) column vector, and ν is an (n × 1) column vector. Assume that A = $[I-\rho W]^{-1}$. Then (3) can be written as:

$$\begin{bmatrix} P_1 \\ P_2 \\ \cdot \\ \cdot \\ P_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ a_{n1} & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & x_{1k} \\ x_{21} & x_{22} & \cdot & \cdot & x_{2k} \\ \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{nk} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} + \begin{bmatrix} \nu_1 \\ \nu_2 \\ \cdot \\ \cdot \\ \nu_n \end{bmatrix} \qquad (4)$$

Define $\mathbf{x_k}$ as a column vector (n ×1) of one housing characteristic variable. Then the derivative of P(n×1) with respect to $\mathbf{x_k}'$ is defined as follows:

$$\frac{\partial P}{\partial x_k'} = \begin{bmatrix} \partial P_1/\partial x_{1k} & \partial P_1/\partial x_{2k} & \cdot & \cdot & \cdot & \partial P_1/\partial x_{nk} \\ \partial P_2/\partial x_{1k} & \partial P_2/\partial x_{2k} & \cdot & \cdot & \cdot & \partial P_2/\partial x_{nk} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \partial P_n/\partial x_{1k} & \partial P_n/\partial x_{2k} & \cdot & \cdot & \cdot & \partial P_n/\partial x_{nk} \end{bmatrix} \qquad (5)$$

This is the Jacobian matrix of P with respect to $\mathbf{x_k}'$. Based on the above definition, the marginal implicit price (marginal benefit) of the hedonic equation is derived as follows:

$$\frac{\partial P}{\partial x_k'} = \begin{bmatrix} \beta_k \alpha_{11} & \beta_k \alpha_{12} & \cdot & \cdot & \cdot & \beta_k \alpha_{1n} \\ \beta_k \alpha_{21} & \beta_k \alpha_{22} & \cdot & \cdot & \cdot & \beta_k \alpha_{2n} \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \beta_k \alpha_{n1} & \beta_k \alpha_{n2} & \cdot & \cdot & \cdot & \beta_k \alpha_{nn} \end{bmatrix} = \beta_k \cdot A = \beta_k \cdot [I - \rho W]^{-1} \qquad (6)$$

The marginal implicit price from a spatial error hedonic model or traditional linear hedonic model is the constant (β_k), but the marginal implicit price of the spatial lag hedonic model is $\beta_k \cdot [\text{I}-\rho\text{W}]^{-1}$.

The Jacobian matrix (5) can be interpreted as follows:

Focusing on the first row, the housing price of location "1" is not only affected by a marginal change of one housing characteristic (say air quality) of location "1" but also is affected by marginal changes of housing characteristics in the other locations (x_{ik}, i=1,2 ..,n). That is, the total impact of a change in air quality on housing price at location "1" is the sum of direct impacts ($\partial P_1 / \partial x_{1k}$) plus induced impacts ($\sum_{i=2}^{n} \partial P_1 / \partial x_{ik}$). The sum of each row of the inverse matrix of row-standardized spatial weights is 1/(1-ρ).

The spatial lag property value model can capture the induced effects of a neighborhood's housing characteristic change, since the weighted neighborhood housing prices are an explanatory variables of house "***i***". The traditional hedonic property value model cannot capture these induced effects of a neighborhood's housing characteristic change. Therefore, a traditional hedonic property value model may lead a biased or at least imprecise estimate of benefits of housing characteristic change if these induced effects are present. As seen from (6) the total effects of a marginal change in air quality at house "***i***" is $\frac{1}{1-r}$. Since $\frac{1}{1-r}$ is a sum of an infinite geometric progression

$(1+r+r^2+r^3+.......)$, the amplified effect of a housing characteristic change on

housing price can be called a "spatial multiplier".

Marginal Benefit Estimation

The derivative of the hedonic price equation with respect to each explanatory variable is the marginal implicit price. This marginal implicit price can be interpreted as the marginal WTP assuming the housing market is in equilibrium. An important consideration is the fact that the estimated marginal benefits represent the capitalized rather than the annual value of the benefits of pollution abatement.

The elasticity of housing price from a given small change in air quality at the mean value for the log-linear functional form for model 1 is estimated as follows:

$$e_{\mathbf{x_k}} = b_{11} \cdot (\frac{1}{1-r}) \cdot \mathrm{SO}_2$$

$$= 0.0089 \cdot (\frac{1}{1-0.294}) \cdot 24.2$$

$= 0.31$, where b_{11} is coefficient of air pollution variable and r is the coefficient of spatial lag, and SO_2 is the mean value of air pollution.

The marginal benefits per household using $b_{11} \cdot [\mathbf{I} - r\mathbf{W}]^{-1}\mathbf{P}$ is about \$ 3,050 ~ \$3,325. This is the capitalized value of the benefits of pollution abatement in the housing price. Note that this is a partial measure that would not hold for a non marginal change in air pollution. Such a non marginal change would induce a new housing price equilibrium that would have to be calculated before benefits could be estimated.

Table 11 Marginal Implicit Prices (unit: million won)

	WTP per household	% of housing price	Elasticity
Model 1	2.46 ($ 3,075)	1.26	0.31
Model 2	2.66 ($ 3,325)	1.36	0.33
Model 3	2.44 ($ 3,050)	1.25	0.33

Note : $1= 800 won
Model 1 is with 2 km cut-off weights matrix,
Model 2 is with 3 km cut-off weights matrix,
Model 3 is with 4 km cut-off weights matrix.

SUMMARY AND CONCLUSIONS

Two conceptual models are considered for this study: the spatial lag hedonic property value model and the spatial error hedonic property value model. The regression diagnostics showed that the spatial lag model specification is valid for the housing market in Seoul.

Marginal WTP for a small change in air quality (4% improvement) is about $3,000 ~ $3,300(1.2%~1.5%) for owners. The value of air quality capitalized into the price of the house is the present value of air quality the owner expects to receive while living there.

The current study improves on past hedonic modeling efforts by measuring both the direct and induced effects of a change in a public good such as air quality. Therefore, the spatial lag hedonic model deals with neighborhood effects which cannot be captured by non spatial techniques. This contribution will increase the statistical efficiency of empirical hedonic models and the development of estimation methods. Future studies may compare the results of spatial and non-spatial models and extend the spatial lag model to allow the use of more flexible functional forms such as the Box-Cox transformation.

[i] Authors are Economist, Korea Research Institute for Human Settlements, Seoul, Korea; Professor, Agricultural and Resource Economics, West Virginia University; and Professor and Director, Regional Research Institute, West Virginia University, Morgantown, WV.

[ii] In the broader study that this study is drawn from, we hyothesize that nitrogen oxide levels would not be a significant factor since NOx is not a visible pollutant and because levels are increasing rapidly.

References

Anselin, Luc, 1988a, *Spatial Econometrics: Methods and Models*, Dordrecht: KluwerAcademic Publishers.

Anselin, Luc and Anil Bera (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In David Giles and Amman Ullah (eds) Handbook of Applied Economic Statistics, pp. 237-289. New York: Marcel Dekker.

Can, Ayse, 1990, " The Measurement of Neighborhood Dynamics in Urban House Prices," *Economic Geography* 66:254-272.

Can, Ayse., 1992, " Specification and Estimation of Hedonic Housing Price Models," *Regional Science and Urban Economics* 22:453-474.

Dubin, Robin A.,1988, "Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms," *The Review of Economics and Statistics* 70: 466-474.

Dubin, Robin A.,1992, "Spatial Autocorrelation and Neighborhood Quality," *Regional Science and Urban Economics* 22:432-452.

The United Nations Environment Programme and The World Health Organization, "Air Pollution in The World's Megacities," *Environment* 36: 4-37.