The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values

Elena G. Irwin and Nancy E. Bockstael

Open space is often cited as a primary attractor of urban and suburban residents to exurban areas located just beyond the metropolitan fringe. Included in the rural amenities afforded by open space are scenic views, recreational opportunities, and an absence of the disamenities associated with development, such as traffic congestion and air pollution. Several willingness-to-pay studies have demonstrated the positive amenity value of open space (Halstead; Beasley, Workman, and Williams), but evidence of the value of open space from revealed preference methods—namely, from hedonic models—is limited and mixed. In this article, we consider the issue of estimating open space spillovers using a hedonic pricing model with residential property sales and offer an explanation for why the positive amenity value of open space effects, even if it exists, may not always be empirically detected.

The estimation problems that we consider are ones of identification, which arise in a hedonic residential price model when the open space land is privately held and developable.[1] Under these circumstances, land parcels counted as open space are part of the market for residential land and therefore subject to the same economic forces that determine a location's residential value. Two identification problems arise as a result. The first is the standard type of econometric identification problem due to endogenous explanatory variables. The problem arises in testing whether the residential value of parcel i is affected by whether a neighboring parcel j is developed. The likelihood that parcel j is developed is a function of its value in residential use, which is itself a function of whether i is developed. The second problem arises because the factors that cause parcels to be more or less valuable in residential use are spatially correlated. If any of these spatially correlated explanatory variables are omitted from the estimation, then any variable measuring surrounding open space will be correlated with the error term. In both instances, biased coefficient estimates will result.

Although we focus specifically on open space in this article, identification of land use spillovers is a broader issue within hedonic modeling that applies to any measure of land use externalities in which the particular land use being measured is influenced by the residential housing market. The identification issues that arise are variants of the "reflection problem" and the problem of "correlated unobservables" discussed in the literature on agent interaction models (Manski; Brock and Durlauf).[2] This problem has not been, to our knowledge, addressed in the hedonics literature.

In this article we demonstrate the expected biases that result from the two identification problems and offer an empirical example. Using sales data on suburban and exurban residential properties in central Maryland, we find that using a simple instrumental variables technique, in which the instruments are carefully chosen to address both correlation problems, the estimated coefficient on open space increases.

This paper was presented at the ASSA winter meetings (New Orleans, LA, January 2001). Papers in these sessions are not subjected to the journal's standard refereeing process.

Elena G. Irwin is assistant professor, Department of Agricultural, Environmental, and Development Economics, Ohio State University, and Nancy E. Bockstael is professor, Department of Agricultural and Resource Economics, University of Maryland.

Support for this research was provided by a NCERQA-99-STAR-L1 grant under the Water and Watersheds program jointly sponsored by EPA and NSF.

[1] For this reason, this problem is most relevant to land in urban–rural fringe areas, where a significant amount of privately held open space land exists.

[2] These models originated with research on social interactions, in which the concern is in identifying the effects of social norms and peer pressures on individual choices. It has also been shown to arise in the context of land use externalities that lead to interaction effects among neighboring landowners making land use conversion decisions (Irwin and Bockstael).

Amer. J. Agr. Econ. 83(3) (August 2001): 698–704

Empirical Evidence of Open Space Effects

Surprisingly few published hedonic studies attempt to estimate the amenity effects that different types of surrounding open space have on residential property values. The findings from the few studies that attempt to include types of surrounding land use are inconclusive. Li and Brown estimated the converse of open space effects, density of surrounding residential development and proximity to non-residential uses, using sales of single-family homes in suburban towns in the Boston metropolitan area. They found that the surrounding residential density is insignificant, as is the externality effect associated with the proximity to commercial land use, although proximity to an industrial area has a negative and significant effect. Geoghegan, Wainger, and Bockstael examined the effect of surrounding agricultural and forested lands on the value of residential exurban land in a central Maryland region. Differing effects are found, depending on the size of the neighborhood considered: within a tenth of a kilometer radius, the proportion of open space has a positive and significant effect on land values, but within a larger one kilometer buffer has a negative and significant effect.

Two additional articles look at the problem in a European setting, where land use controls are much stronger and therefore market effects much weaker. Cheshire and Sheppard estimated the effects of publicly and privately held open space on residential property values using data from two England towns. They found mixed results, depending on the amount of open space amenities in the two towns. Tyrvainen and Miettinen used the sales of houses from a semi-rural area in Finland to estimate the effects of surrounding forested lands on residential property values. Distance to the nearest small area of forest is found to be negative and significant and the presence of a forest view from the housing unit is found to have a positive influence. However, the relative amount of forested area within the housing neighborhood is not significant. None of these articles addressed the potential correlation of the variables measuring the open space spillovers with the hedonic model's error and most ignored the possibility of a spatially autocorrelation.[3]

The Identification Problem

To illustrate the issues associated with identification, assume that the value in residential use of parcel i in period t is given by

$$(1) \quad R_{it} = g(X_{it}, P_{it}) + \varepsilon_{it}$$

where X_{it} is a vector of observed explanatory variables associated with parcel i, P_{it} is the percent of the land around parcel i in privately held and developable open space in period t, and ε_{it} is the unobserved term which is likely both spatially and temporally correlated. The identification problems arise from the fact that the amount of surrounding open space, P_{it}, is a function of the past workings of the land market. The amount of open space in period t is inversely related to D_{it}, the amount of neighboring land that has been developed by period t. D_{it} will be a function of past economic signals, including all factors that have affected the value of surrounding parcels in residential use and all variables that have affected their costs of conversion from an open space to a developed use (including the opportunity cost of the land in an alternative use) up to time t. The concept is a temporally cumulative one because development is effectively irreversible.

One way to characterize the problem is to adopt the standard result from the monocentric model that exogenous increases in population shift the bid-rent gradient outward. We assume that in the presence of constant growth pressures, the value of land in a developed use, net of its opportunity costs, can only increase with time. If this is true, then the probability that land is found in the developed state as of time t will be increasing in R_{jt}, the value in residential use of a typical parcel in the neighborhood surrounding parcel i, and decreasing in W_{jt}, a vector of variables that affect the cost of converting that parcel. This assumption implies two things. First, if land was first worth developing at time $t - s$, then it will still be worth developing at time t. Second, the amount of land of

[3] Of these studies, only Tyrvainen and Miettinen test for the potential presence of spatial dependence in the data. They fail to reject the null hypothesis that the errors are random.

this sort that is found in the developed state will be increasing in $t-s$. D_{it} becomes a function of the values of R and W at time t[4] for all i's neighbors (where the set of neighbors is denoted N_i), and P_{it} can be defined as a different function of the same variables

$$(2) \quad P_{it} = h(R_{jt}, W_{jt}, \forall j \in N_i) + \eta_{jt}$$

Taking equation (1) and (2) together, we can now set out the identification problems. To expedite the presentation of the endogeneity problem, we specify a system of linear equations and assume that that parcel i has one neighbor, j, and j's only neighbor is i. The system of equations can be characterized by:

$$(3a) \quad R_{it} = \alpha X_{it} + \beta P_{it} + \varepsilon_{it}$$

$$(3b) \quad P_{it} = \gamma R_{jt} + \delta W_{jt} + \eta_{jt}$$

$$(3c) \quad R_{jt} = \alpha X_{jt} + \beta P_{jt} + \varepsilon_{jt}$$

$$(3d) \quad P_{jt} = \gamma R_{it} + \delta W_{it} + \eta_{it}$$

where R_{it} is the value of parcel i in residential use; P_{it} is a measure of open space around parcel i and, in this restricted example, is simply a measure of the likelihood that parcel j is developed. Also, assume that the variables in X_{it} are factors that increase the value of the parcel in residential use and variables in W_{it} are factors that increase the costs of developing parcel i. Therefore, α and δ are assumed to be positive; γ is negative; and we are usually interested in testing whether β is positive and significant.

Written in this form, it is clear that (3a)–(3d) is a system of simultaneous equations. Solving this system simultaneously, R_{it}, R_{jt}, P_{it}, and P_{jt} can be defined as functions of X_{it}, X_{jt}, W_{it}, and W_{jt} as well as of a weighted sum of all the errors. If it is parcel i that is in the sales data set used to estimate the hedonic model, then we will be interested in knowing how P_{it} is correlated with ε_{it}. Solving the system explicitly yields the following expression

$$(4) \quad P_{it} = \frac{\alpha\gamma X_j}{1-\gamma^2\beta^2} + \frac{\alpha\beta\gamma^2 X_i}{1-\gamma^2\beta^2} + \frac{\delta W_j}{1-\gamma^2\beta^2} + \frac{\beta\delta\gamma W_i}{1-\gamma^2\beta^2} + \frac{\gamma\varepsilon_j + \beta\gamma^2\varepsilon_i + \eta_j + \beta\gamma\eta_i}{1-\gamma^2\beta^2}$$

The coefficient on ε_i, $\beta\gamma^2$, will be positive and, in a stable system, the denominator will also be positive, so P_{it} can be expected to be positively correlated with ε_i. The usual procedure in hedonic analysis of estimating (3a) by OLS will ignore the fact that P_{it} is endogenous and correlated in this way with ε_{it}. If this were the only problem, we would expect the OLS estimated coefficient on P_{it} to be biased upward.[5]

The second of our problems can be illustrated by substituting (3c) and (3b) into (3a) to get

$$(5) \quad R_{it} = g\{X_{it}, P_{it}[R_{jt}(X_{jt}, P_{jt}, \varepsilon_{jt}), W_{jt}, \eta_{jt}]\} + \varepsilon_{it}$$

Irrespective of the endogeneity issue, P_{it} will be correlated with ε_{it} if the unobserved variables affecting residential value (ε_{it} and ε_{jt}) are spatially correlated. Given that typical variables affecting the value of a parcel in residential use will include distance to employment centers, public services, school quality, proximity to locally undesirable land uses, we have every reason to expect the omitted variables embedded in ε_{it} to be correlated with P_{it}. The correlation here is negative: as the value of a location in residential use increases, the amount of surrounding open space will tend to decrease. This second source of correlation will bias the OLS estimate in a negative direction.

Estimation of Open Space Effects

To begin investigating this problem, we compare results from ordinary least squares (OLS) and instrumental variables (IV) approaches to estimation of the hedonic

[4] Strictly speaking, D_{it} may be more accurately described as being a function of R_j and W_j through periods $t-1$, because it takes time to develop. However, practically speaking this is irrelevant since most of the variables in R and W are either time invariant (like soil quality or distance to CBD) or are highly correlated over time.

[5] This holds for the case considered here, in which each parcel only has one neighbor. In the more general case, in which parcel i has many neighbors and each neighbor interacts with parcel i as well as other neighbors, the interaction effects can become very complex. In this case, the direction of the bias may be indeterminate.

pricing model in (3a). The data consist of arms-length, single transactions of owner-occupied residential properties within four Maryland counties in the suburban and exurban regions around Washington, D.C. The 55,799 sales occurred between January 1995 and December 1999. The data are from the Maryland Office of Assessment and Taxation and are made available in geocoded format by the Maryland Office of Planning through Maryland Property View, a GIS data product. Additional spatial variables were generated using ArcInfo GIS software.

The identification problems outlined above occur when surrounding open space is privately held and developable. However, some open space is either owned outright by the public sector or the public sector owns its development rights. We distinguish among these three types of surrounding open space: (1) privately owned open space that is developable, including land in crops and pasture and privately owned forests (%OPENPR1); (2) privately owned open space that is protected from development, including agricultural easements and privately owned conservation areas (%OPENCON); and (3) publicly owned open space (%OPENPUB), including land owned by the federal, state, or county governments. Clearly the first type of open space is endogenously determined. The third type of land is arguably exogenous. Much of the public open space in this geographical area has been held by the public sector for many years and early decisions to purchase public lands were likely driven by a variety of motivations that may have little to do with today's real estate market. The second type of open space is problematic. Presumably the motivations for enrolling a parcel in an easement trust or an agricultural preservation program may be related to the market value of the land, even though these programs are often designed to reward the owner in a way commensurate with the difference between market and use value. To allow for either possibility, we estimated the model first by treating the surrounding amount of this type of open space as a separate exogenous variable (%OPENCON) and second by aggregating it with the endogenous open space variable. This aggregated variable is called %OPENPR2.[6]

[6] Each of these neighborhood variables measures the proportion of open space within a 400 meter radius of a parcel's centroid.

A variety of housing characteristics are included in the hedonic to capture variations over the housing stock. These include an index variable that rates the grade of the dwelling unit on a scale of 0–9, with 9 being the highest grade (DWLLGR), a dummy variable indicating whether the dwelling is a detached unit (DWLLTYP), number of full baths (BATHS_FU), number of half baths (BATHS_HA), the square feet of the structure in logs (LN_AREA), the footprint of the house in logs (LN_FTPRT), the age of the house in logs (LN_HAGE), and the year of the sale (SALEYR). In addition, lot size is hypothesized to influence residential values and is included in log form (LN_LSIZE). Locational attributes of the parcel that are hypothesized to influence price include the parcel's commuting distance to the two major urban centers in the region: Washington, DC (LN_DCDIST) and Baltimore (LN_BADIST). Both these variables are measured via the major roads network and are included in the model in log form. Several neighborhood variables, taken from the 1990 Census of Population, are hypothesized to influence price by providing signals about the historical nature of the neighborhood. The population density in the Census block group in which the parcel centroid is located (LNPOPDEN) is a proxy for congestion and is hypothesized to have a negative effect on residential values. In addition, the socioeconomic make-up of households within the block group is hypothesized to influence housing prices: median household income (LNMHHINC), the percent of the population with education beyond high school (%EDUHS+), and percent of the population that is African American (%BLKPOP). As the level of public services and goods and tax rates vary at the county level, county dummies for three of the four counties are included (CA, CH, and HO).

Table 1 reports the results from the OLS regression for a double log formulation of the hedonic equation and for the two different specifications of the endogenous open space variable. The model explains 70% of the variation in the log of sales price and most variables are highly significant, based on the OLS standard errors, and of the expected sign. Price is found to be increasing in all of the structural attributes and increasing at a decreasing rate in lot size, age of the house, and size of the house. Price decreases with

Table 1. Estimation Results

	OLS		IV	
Variable	Model 1	Model 2	Model 1	Model 2
Intercept	4.21876* (0.11223)	4.212852* (0.11222)	4.274455* (0.11359)	4.267121* (0.11355)
DWLLGR	0.161342* (0.00243)	0.16148* (0.00243)	0.161214* (0.00244)	0.161358* (0.00244)
DWLLTYP	0.166001* (0.00389)	0.166253* (0.00389)	0.16773* (0.00391)	0.167884* (0.00391)
BATHS_FU	0.067622* (0.00221)	0.06765* (0.00221)	0.067695* (0.00222)	0.067711* (0.00223)
BATHS_HA	0.045654* (0.00239)	0.045738* (0.00239)	0.045564* (0.0024)	0.045648* (0.0024)
LN_FTPRNT	0.102229* (0.0056)	0.102456* (0.0056)	0.101658* (0.00562)	0.101866* (0.00562)
LN_AREA	0.343946* (0.00503)	0.343869* (0.00503)	0.344668* (0.00505)	0.344598* (0.00505)
LN_LSIZE	0.031557* (0.00144)	0.031366* (0.00144)	0.029509* (0.00155)	0.029398* (0.00155)
LN_HAGE	−0.01979* (0.00104)	−0.01969* (0.00102)	−0.01955* (0.0009)	−0.01951* (0.00099)
SALEYR	0.022746* (0.00076)	0.022767* (0.00076)	0.022651* (0.00076)	0.022683* (0.00076)
LN_BADIST	0.058226* (0.00331)	0.058053* (0.00331)	0.0565* (0.00336)	0.056386* (0.00336)
LN_DCDIST	−0.06336* (0.00472)	−0.06267* (0.00472)	−0.06367* (0.00473)	−0.06308* (0.00472)
LN_MHHINC	0.121049* (0.00615)	0.120463* (0.00615)	0.117275* (0.00627)	0.116831* (0.00626)
LN_POPDEN	−0.03057* (0.00103)	−0.03077* (0.00103)	−0.0296* (0.00103)	−0.02985* (0.00103)
%BLKPOP	−0.05453* (0.00906)	−0.05547* (0.00906)	−0.05663* (0.00909)	−0.05745* (0.00909)
%EDUHS+	0.175214* (0.0097)	0.175675* (0.0097)	0.179333* (0.00976)	0.17959* (0.00976)
%OPENPR1	−0.00824 (0.00531)	—	0.063296** (0.02183)	—
%OPENPR2	—	−0.00695 (.005301)	—	0.062136** (.021639)
%OPENPUB	0.028427† (0.01362)	0.028375† (0.01362)	0.030577† (0.0135)	0.030092† (0.0135)
%OPENCON	0.213814** (0.07371)	—	0.210559** (0.07385)	—
CA	−0.28153* (0.00566)	−0.28106* (0.00566)	−0.28613* (0.00591)	−0.2857* (0.00591)
CH	−0.23875* (0.00533)	−0.23875* (0.00533)	−0.23708* (0.00537)	−0.23714* (0.00537)
HO	−0.14091* (0.00354)	−0.14088* (0.00354)	−0.14267* (0.00358)	−0.14264* (0.00358)
R^2	0.706	0.706	0.705	0.705

Note: *, **, and † indicate significance at the .001, .005 and .05 levels, respectively.

distance to Washington, D.C., which is consistent with the urban bid-rent model, but is increasing with distance to Baltimore. Given that the majority of our study area is located to the south of Baltimore, this is not too surprising. The area just south of Baltimore contains large industrial sites, as well as a large airport, both of which make this area less desirable as a residential location. The neighborhood variables all have the expected signs: price is increasing in the median income and the educational attainment of households within the parcel's block group and decreasing in population density and the proportion of the population that is African American.

The estimated effects of open space on residential property values are of primary interest. In both specifications, price is increasing in the proportion of surrounding land that is publicly owned. In addition, when the proportion of conservation land that is privately owned is treated as an exogenous variable, price is also found to be increasing in this variable. In both specifications, the proportion of privately owned and developable open space within a parcel's neighborhood is found to have a negative, but insignificant, effect on price. However, these results are doubtful because of the identification problems that we suspect exist.

To test empirically whether these problems exist, we conduct a Hausman endogeneity test and we test for spatial autocorrelation. A joint Hausman test supports the hypothesis that both the private open space measure and the privately owned conservation lands are endogenous and their coefficients biased and inconsistent. A Lagrange multiplier test supports our hypothesis that the residuals in the hedonic regression are spatially correlated.[7] For reasons given earlier, this spatial correlation in the errors will result in a biased and inconsistent coefficient on the private open space variable. It will also produce inefficient parameter estimates for other variables, but this inefficiency is not of major concern because of the large number of observations.[8]

To address these problems, we adopt an instrumental variables approach. From equation (4), it is clear that P_{it} can be written as a function of the exogenous variables X_{it}, X_{jt}, W_{it}, and W_{jt}, where X_t is a vector of observed explanatory variables that influence residential value and W_t is a vector of variables influencing the costs of development. In identifying appropriate instruments for P_{it}, we seek variables that not only meet the standard criteria of being correlated with P_{it} and uncorrelated with the error, but also variables that will minimize potential multicollinearity problems. Neighboring parcels will share many of the same locational attributes (e.g., distance to major urban centers), which gives rise to a positive spatial association among many of the variables in X_t. Because P_{it} is a function of X_{jt}, multicollinearity may be a problem in estimating (3a). For this reason, W_{it} and W_{jt} are chosen as instruments for P_{it}. Specifically, we use variables indicating the parcel's slope and the soil's drainage ability, as well as whether the parcel has high quality soils—a proxy for the opportunity cost of development.

Table 1 also reports the results for both specifications of the endogenous open space variable using this instrumental variables model. The OLS and IV estimates are very similar with the exception of the coefficient associated with the private open space variable. Whereas the OLS estimate of this coefficient is negative, but insignificant, the IV estimate is positive and significant at the 1% level in both specifications. Also of interest are the other estimated open space coefficients. In the first specification, in which the proportion of privately own land in conservation is treated as an exogenous variable, this variable also remains positive and highly significant. Under both specifications, the proportion of publicly owned open space remains positive and significant at the 5% level.

Conclusions

Using hedonic models to test whether people value open space amenities is hampered by the fact that a parcel's land use is in part determined by its residential value. This fact generates two identification problems that lead to correlation between the surrounding open space measure and the error term in the hedonic model. In this article, we show that these two effects bias the estimated coefficient on the open space

[7] The Hausman statistic is 42.997, which is distributed as a chi-squared with two degrees of freedom and is significant at the 0.005 level. The LaGrange Multiplier statistic is 17.611, which is distributed as a chi-squared with 1 degree of freedom and is significant at the 0.005 level as well.

[8] Spatial autocorrelation will also bias the standard errors, so there is a (probably small) chance that some estimated coefficients are not significantly different from zero.

measure. An empirical attempt to obtain consistent parameter estimates using instrumental variables produces evidence of a positive and statistically significant effect. The results showed that the OLS estimation biases the estimated marginal value of open space downward. Identification strategies that break the correlation between the endogenous land use externality variables and the error term, such as the instrumental variables approach used in this paper, are necessary to test for the existence of these spillover effects on residential property values.

References

Beasley, S.D., W.G. Workman, and N.A. Williams. "Estimating Amenity Values of Urban Fringe Farmland: A Contingent Valuation Approach." *Growth and Change* 17(1986): 70–78.

Brock, W., and S. Durlauf. "Interactions-Based Models." *The Handbook of Econometrics 5*. J.J. Heckman, and E. Leamer, eds. Amsterdam: Elsevier Science, in press.

Cheshire, P., and S. Sheppard. "On the Price of Land and the Value of Amenities." *Economica* 62(1995):247–67.

Geoghegan, J., L.A. Wainger, and N.E. Bockstael. "Spatial Landscape Indices in a Hedonic Framework: An Ecological Economics Analysis Using GIS." *Ecol. Econ.* 23(1997):251–64.

Halstead, J. "Measuring the Nonmarket Value of Massachusetts Agricultural Land: A Case Study." *J. of Northeastern Agricultural Economic Council*, 13(1984):226–47.

Irwin, E., and N.E. Bockstael. "Interacting Agents, Spatial Externalities, and the Endogenous Evolution of Residential Land Use Pattern." Department of Agricultural, Environmental, and Development Economics Working Paper AEDE-WP-0010-01, Ohio State University.

Li, M., and H.J. Brown. "Micro-Neighborhood Externalities and Hedonic Housing Prices." *Land Econ.* 56(May 1980):125–40.

Manski, C. "Identification of Endogenous Social Effects: The Reflection Problem." *Rev. Econ. Studies* 60(1993):531–42.

Tyrvainen, L., and A. Miettinen. "Property Prices and Urban Forest Amenities." *J. Environ. Econ. Manag.* 39(2000):205–23.