# Using Spatial Econometrics to Assess the Impact of Swine Production on Residential Property Values

Jungik Kim
Ph.D. candidate
Department of Urban Planning
University of Illinois

Peter Goldsmith
Assistant Professor
NSRL Endowed Fellow in Agricultural Strategy
Department of Agricultural and Consumer Economics
University of Illinois

Michael Thomas
Assistant Professor
Department of Agribusiness
Florida A&M University

Presented as a Selected Paper at the annual meeting of the American Agricultural Economics Association, Denver, July, 2004

Contact author:
Peter Goldsmith
Email: pgoldsmi@uiuc.edu
Telephone: (217) 333-5131
Fax: (217) 333-5538

## Abstract

A spatial hedonic model is developed to assess monetary harm of confined animal feeding operations (CAFOs) on property values, taking explicitly spatial dependence in property values into account. Spatial autocorrelation was found in the form of spatial lag dependence, not spatial error dependence. When spatial lag dependence is explicitly taken into account, on average the impact is reduced by 18%. The magnitude of the spatial autoregressive parameter was about 0.2 for the 1-mile distance band, meaning one-fifth of the house value could be explained by the values of the neighboring houses.

Key Words: Spatial hedonics, spatial autocorrelation, spatial lag dependence, spatial error dependence, confined animal feeding operations, CAFO

## Introduction

Hedonic models in housing and real estate have traditionally used housing physical attributes and locational characteristics variables to estimate how a certain housing attribute marginally contributes to housing price. Examples of physical attributes variable include floor area, the number of rooms, the number of bathrooms, or house age. For locational characteristics, two kinds of measures are commonly employed: an accessibility measure such as distance to the Central Business District (CBD) and a neighborhood indicator measure such as median household income. Housing values, for example, would be negatively associated with distance to CBD due to increased transportation cost. Similarly, median household income can be an indicator for neighborhood quality. Hedonic models are also used to measure the negative externality of a noxious facility such as a landfill or a hazardous waste site (see Kohlhase, Kiel and Zabel, and Hite et al. among others). Distance to such a facility is additionally used as a locational variable to examine how the negative impacts of the facility are a function of distance.

Hedonic models using locational variables, however, do not fully account for spatial autocorrelation in housing prices. Spatial autocorrelation refers to the cluster of similar values in space (Anselin and Bera). Housing prices are spatially autocorrelated because neighborhood residential properties share location amenities such as public services (Dubin, Basu and Thibodeau). The boundaries of neighborhoods are not always clear-cut and the objective measures of neighborhood quality are often difficult to obtain. The residuals from hedonic models are likely to be spatially autocorrelated. OLS estimates then will be unbiased but inefficient (Dubin, Pace, and Thibodeau). Spatial autocorrelation also arises from a mismatch in spatial scale (Anselin and Bera). Two situations can be discerned. First, using an indicator variable for neighborhood quality, such as median income, will induce a spatial mismatch, as the data is only available on the census tract or block-group level. The spatial scale of the census tract or block-group level is larger than that of an individual house, resulting in positive spatial autocorrelation. Similarly, in the case of a noxious facility and its surrounding properties, the negative impacts of the facility tend to spill across houses. The spatial scale of the impact is larger than an individual house and causes spatial autocorrelation. In the presence of spatial autocorrelation, there is a loss of independent observations. The inclusion of a spatially lagged variable in the hedonic model addresses this loss of information (Anselin and Bera).

The issue of spatial autocorrelation has received little attention in hedonic studies in rural settings. The purpose of this article is to advance the methodology of rural hedonic studies by developing a spatial hedonic model that explicitly accounts for spatial autocorrelation in housing prices, in the assessment of the monetary harm of confined animal feeding operations (CAFOs) on rural property values. To date, none of the existing studies on the impact of swine operations on property values has addressed the issue of spatial autocorrelation in housing prices.[1] This article will contribute to the literature by providing a benchmark for future hedonic studies of rural land use.

## Literature Review

Spatial autocorrelation between spatial objects can be formally defined by the moment condition as follows (Anselin and Bera, p. 241):

(1) $$\text{cov}[y_i, y_j] = E[y_i y_j] - E[y_i] \cdot E[y_j] \neq 0, \quad \text{for } i \neq j$$

where $i$, $j$ refer to individual locations and $y_i$ and $y_j$ denote the value of a random variable at that location. The covariance structure becomes spatial when nonzero $i$, $j$ pairs are interpreted "in terms of spatial structure, spatial interaction or the spatial arrangement of the observations" (Anselin, 2001a, p. 312).

Two approaches, direct and indirect representation, are suggested in the literature to model the covariance structure (Anselin, 2001a). Direct representation, adopted from a geostatistical approach, assumes a continuous surface and expresses spatial interaction as a continuous function of distance. In contrast, an indirect representation, or lattice approach presumes that the spatial effect is a function of the interaction among discrete spatial objects. Although direct representation has been applied to real estate markets, an indirect approach is more suited to economic studies dealings with discrete objects in space, such as counties or regions (Anselin and Bera). Moreover, specifications used in direct representation have estimation and identification problems. For example, the nuisance parameter is not identified under the null hypothesis, resulting in a singular information matrix (Anselin, 2001b).

Direct representation focuses on improving spatially autocorrelated residuals by directly estimating the covariance structure (Dubin, Basu and Thibodeau). Spatial covariance is expressed as a smooth decay function of the distance between observations. For example, Dubin specified the covariance structure as:

(2) $$K_{ij} = b_1 \exp(-d_{ij} / b_2)$$

where $d_{ij}$ is the distance between the $i$ and $j$ th observations and $b_1$ and $b_2$ are parameters to be estimated. Basu and Thibodeau used a semivariogram to estimate the covariance structure as:

(3) $$\gamma(s_i - s_j) = 0.5Var\{\xi(s_i) - \xi(s_j)\} = C(0) - C(s_i - s_j)$$

where $s_i$ denotes location of an observation $i$ and $\xi(s_i)$ denotes the hedonic residual for an observation at $s_i$.[2]

In indirect representation, the specification of a spatial process indirectly determines a covariance structure which requires constructing a relevant spatial weights matrix. In indirect representation, two types of spatial dependence, spatial lag dependence and error dependence, are conventionally distinguished (Anselin, 1988). Spatial lag dependence arises from spatial interaction between economic agents such as counties or states, or mismatch in spatial scale. Spatial lag dependence is modeled by incorporating a spatially lagged dependent variable in hedonic models, analogous to the inclusion of a serially autoregressive term for the dependent variable ($y_{t-1}$) in a time-series context. The inclusion of a spatial lagged dependent variable has two implications (Anselin and Bera). First, OLS estimates will be biased and inconsistent because a spatially lagged dependent variable is endogenous and always correlated with the error term. Second, the simultaneity must be explicitly accounted for, either in a maximum likelihood estimation framework or by using a proper set of instrumental variables.

On the other hand, spatial error dependence arises from noise in the model and can be specified as a spatial process for the disturbance term. The effect of a spatial residual autocorrelation on the OLS estimator is analogous to time-series results. The OLS estimates will be unbiased, but inefficient. More efficient estimators are obtained by specifying the error covariance implied by the spatial process.

*Spatial Weights Matrix*

The specification of a covariance structure in indirect representation requires constructing a relevant spatial weights matrix. The choice of spatial weights is made empirically in many applications, as there is very little theoretical guidance in the choice of spatial weights (Anselin, 2002).

A spatial weights matrix, usually denoted by *W*, specifies neighborhood sets for each observation as nonzero elements. In each row *i*, a nonzero element, $w_{ij}$, defines *j* as being a neighbor of *i*. So $w_{ij} = 1$ when $i$ and $j$ are neighbors, and $w_{ij} = 0$ otherwise. Establishing the neighborhood set in this fashion reflects the range of spatial interaction. In the case of the housing market, for example, nonzero pairs in a spatial weights matrix would indicate the extent of the externality effect on neighboring houses. By convention, an observation is not a neighbor to itself, so that the diagonal elements are zero ( $w_{ii} = 0$ ). In most cases, the spatial weights matrix is row standardized so that weights across rows are summed to one, which amounts to averaging of the neighboring values and allowing for spatial smoothing. Standardizing weights also makes the spatial parameters comparable between models.

There are different ways to define neighborhoods and a weights matrix.[3] Conventionally, neighbors refer to locations adjacent to each other sharing common boundaries or vertexes, so, for example, a county in a regular grid will have four neighbors (rook or bishop criterion) or eight neighbors (queen criterion) (Anselin, 2002). For example, in Figure 1, the five observations in a regular grid have one or four neighbors that share borders (rook criterion). The left side of Table 1 is a corresponding 5 × 5 spatial weights matrix and the right side shows a standardized weights matrix.[4]

For an irregular grid, the number of neighbors will depend on the shape of the grid. This notion of neighbors based on contiguity cannot be applied to rural housing markets because houses are not contiguous and can be separated by great distances. Constructing a weights matrix in rural areas based on contiguity will yield many "islands," or observations with no connections, and make the analysis of spatial autocorrelation irrelevant.

Neighbors can also be defined as the locations within a given distance. For the type of distance, a physical distance-band is commonly used (Can; Can and Megbolugbe; Pace and Gilley; Kim, Phipps, and Anselin). The number of neighbors, however, will vary if the sizes of spatial units differ. A distance-band weights matrix is not feasible for rural studies since lot sizes vary greatly in rural areas. Building a weights matrix on a distance band will produce an uneven number of neighbors from rural clusters (hamlets) or a small number of neighbors for larger lots (farms).

Alternatively, *k*-nearest neighbors for each house can be employed to address unconnected houses in a contiguity base or an uneven number of neighbors in a distance-band case (Can and Megbolugbe; Pace et al.). Choosing *k*-nearest neighbors ensures a constant number of neighbors for each house. The idea of *k*-nearest neighbors corresponds to the practice of the "comparable-sales" approach employed in residential real estate appraisal (Can and Megbolugbe).

*Empirical Studies*

Several hedonic studies in real estate and housing markets have applied an indirect representation of the covariance structure among spatial objects when addressing spatial autocorrelation. Yet, no study examines the issue of spatial autocorrelation in housing prices in rural settings, much less where CAFOs are concerned.

Incorporating a spatially lagged dependent variable and spatial parametric drift to housing prices in Columbus, Ohio, offered better explanation of the variations than traditional hedonic price models (Can). Similarly, incorporating a spatially lagged dependent variable to consider the house price index in Miami, Florida not only increased the explanatory power of the model, reflected in a higher $R^2$, but also addressed to some extent the problem of omitted housing structure variables (Can and Megbolugbe).

In contrast to these studies specifying a spatially lagged dependent variable, a spatially autocorrelated error term was specified through a weighted average of the errors on nearby properties for housing data in Boston (Pace and Gilley). The weight was given to each census tract and set as the distance between two tracts relative to all other tracts. The estimation results showed that modeling spatial dependence of the errors was significantly beneficial, resulting in a 44% reduction of the errors relative to the OLS.

A lattice perspective has also been used in the environmental economics literature. The OLS overestimated the effect of air quality on housing prices in Seoul, Korea in the presence of spatial lag dependence (Kim, Phipps, and Anselin). On the other hand, a spatial autoregressive (SAR) error model did not change the empirical results much in estimating the demand for air quality in the South Coast Air Basin counties (Beron et al.).

## Study Area and Data

Craven County located in southeastern North Carolina is chosen as a study area for three reasons: first, geographically coded real estate (N = 25,684 housing values) and swine industry data (N = 26 farms and 85,000 pigs) are available; second, the farms are located in one of the most significant swine regions nationally, with the farm size ranging from 600 to 12,600 hogs, and third, land use is heterogeneous where neither agriculture nor non-agriculture rural residents dominate.

Three phases of data collection are involved in this study. These data include: (1) assessed property values including location and description information, (2) general neighborhood indicators, and (3) hog operation and location.

Data on assessed property values are available from the Craven County GIS Website. Craven County completed a countywide revaluation of 50,000 individual parcels as of January 1, 2002 based on a 8 year total revaluation cycle mandated by North Carolina General Statutes. The County updated new values for all properties in January 2003.

For the purpose of analysis, rural houses are identified from the Craven parcels map with the following procedures: first, since the Craven parcels map includes every parcel including residential, commercial, agricultural, and open space, only parcels whose building type and land use are residential are deemed to be residential parcels with houses. Second, only houses with at least one bedroom and bathroom including mobile homes are selected. Third, houses in urban areas ($n$ = 12,799) are excluded as they, because of their higher population density, may swamp the data compared to the sparsely populated rural areas. As a result, 1,100 urban houses (9%) were lost within 2 mile of the nearest farm.[5] Additionally, 91 rural houses that are located in the same blocks with urban houses were deemed to be urban and excluded from the sample (Kim). Houses in the two townships (Township 5 and 6) with no hog farms are also excluded from the sample, as the minimum distance from the two townships to the nearest farm is about 7.25 and 19 miles, respectively. Houses with lot size over 10 acres are considered outliers and excluded to avoid houses with farm or timber tracts, following Palmquist, Roka, and Vukina.[6]

Information on the general neighborhood indicators is available from the 2000 Census. The Census Bureau reports the median household income and average commuting time to work by the census block-group level. The spatial information on the census block-groups is available in the form of the Census 2000 TIGER (Topologically Integrated Geographic Encoding and Referencing System) Shapefiles.

Data on the 26 hog operations are acquired from the North Carolina Department of Environment and Natural Resources, Division of Water Resources, and Craven County Appraisal Office. North Carolina hog data include: farm number, farm name, design capacity, steady-state live weight, owners name, and mailing address. The steady-state live weight represents the collective weight of all animals at a facility and is a more accurate means of size comparisons.[7] The steady-state live weight is divided by an average hog weight (135 lb.) to get an average number of animal "units" for an operation. Alternatively, the actual number of hogs could be used to compare farm sizes. But data on the number of hogs by different types such as sows, nurseries, or piglets are not available. Using steady-state live weight can account for different types of hogs on a farm and is considered as more representative of the hogs than the actual number of hogs.[8] The locations of the farms are identified by comparing owners' names and parcel IDs from Craven County Appraisal Office's records.

## Model Specification

The estimation of spatial hedonic model is preceded by diagnostics of spatial autocorrelation in the hedonic OLS model[9]. To that end, the Lagrange multiplier (LM) tests for spatial lag dependence and spatial error dependence are used (Anselin, 1988). The result of the LM test determines spatial lag or spatial error dependence. Spatial lag dependence is modeled by incorporating a spatially lagged dependent variable ($Wy$) into the model. Specifically, spatial lag model is specified as:

(4) $$Y = \rho W \cdot Y + X\beta + \varepsilon$$

where $\rho$ is the spatial autoregressive parameter, $W$ is the $k$-nearest neighbor weights matrix, $X$ is a vector of independent variables as above, and $\varepsilon$ is an error term.

Alternatively, a spatial error model can be specified with respect to the disturbance term. The most common specification is a spatial autoregressive (SAR) process in the error terms:

(5) $$Y = X\beta + \varepsilon$$

(6) $$\varepsilon = \lambda W\varepsilon + \xi$$

where $\varepsilon$ is an error term, $\lambda$ is the spatial autoregressive coefficient for the error lag $W\varepsilon$, and $\xi$ is an uncorrelated and homoskedastic error term.

Specific to our context, spatial lag dependence model is specified as:

(7) $$\begin{aligned}\text{VTF} = {} & \beta_0 + \rho\, W \cdot \text{VTF} + \beta_1 \text{BASEAREA} + \beta_2 \text{ROOM} + \beta_3 \text{BATHROOM} \\ & + \beta_4 \text{LOTSIZE} + \beta_5 \text{AGE} + \beta_6 \text{INCOME} + \beta_7 \text{DCBD} + \beta_8 \text{DOPEN} \\ & + \beta_9 \text{DSCHOOL} + \beta_{10} \text{HOG\_D} + \beta_{11} \text{SIZE}\end{aligned}$$

where VTF is Box-Cox transformed assessed property values, $\rho$ is the spatial autoregressive parameter, $W$ is the $k$-nearest neighbor weights matrix, BASEAREA is the base area of a house, ROOM is the number of rooms, BATHROOM is the number of bathrooms, LOTSIZE is lot size, AGE is house age, INCOME is median household income by census block groups, DCBD is the distance to CBD, DOPEN is the distance to nearest open space, DSCHOOL is the distance to the nearest school, HOG_D is the number of hogs in the nearest farm divided by the distance, and SIZE is a dummy variable for farm size (= 1 if greater than 2,500 head).

On the other hand, if the LM test points to spatial error dependence, the spatial autoregressive (SAR) error model is specified as:

$$(8) \qquad \text{VTF} = \beta_0 + \beta_1\,\text{BASEAREA} + \beta_2\,\text{ROOM} + \beta_3\,\text{BATHROOM} + \beta_4\,\text{LOTSIZE} + \beta_5\,\text{AGE} + \beta_6\,\text{INCOME} + \beta_7\,\text{DCBD} + \beta_8\,\text{DOPEN} + \beta_9\,\text{DSCHOOL} + \beta_{10}\,\text{HOG\_D} + \beta_{11}\,\text{SIZE}$$

where all notations remain the same as in spatial lag dependence model. The SAR error model achieves more efficient estimates by specifying the error covariance (Equation 6).

## Estimation Results

The spatial lag model was estimated by taking all houses within certain distance from the nearest farm, starting from 0.5 mile up to 3 mile with a quarter mile increment. The estimated results are reported in Table 3, 4, and 5 for the three distance bands between 0.75 mile, 1 mile, and 1.25 mile.[10]

There was a strong presence of spatial lag dependence. The robust LM test was highly significant for spatial lag dependence ($p$ < 0.00), whereas it was not for spatial error dependence. The Robust LM test accounts for the presence of misspecification of spatial lag or spatial error dependence.[11] The presence of strong lag dependence was consistent across the 3- to 9-nearest neighbor weights matrix. All k {3, 5, 7, 9} weight matrix models were not significantly different from each other therefore only the diagnostic results and the estimation of spatial lag model from the 3-nearest neighbor weights matrix are reported.[12]

Two sets of spatial lag models, i.e., the maximum likelihood (ML) and the spatial two-stage least squares (2SLS) estimation, were estimated for the three distance bands. The ML estimation assumes normality. In contrast, the 2SLS estimation, using spatially lagged explanatory variables as instruments, is robust to nonnormality and consistent, but not necessarily efficient. Given the nonnormality found in the OLS estimates with significant Jarque-Bera test on normality ($p < 0.05$), the 2SLS estimates seemed more appropriate than the ML estimation. In addition, the 2SLS-robust estimation can be an alternative to the ML estimation, when heteroskedasticity is present. In fact, strong evidence of heteroskedasticity was present as indicated by significant White test ($p < 0.05$).

All housing physical variables (BASEAREA, ROOM, BATHROOM, LOTSIZE, and AGE) were significant at a 1% significance level and showed the expected signs. The locational variables (INCOME, DCBD, DOPEN, and DSCHOOL) were significant at 5%. INCOME was positive, meaning that property values are positively associated with household income. The negative DCBD meant that property values declined if they were located further away from the CBD, due to a lack of accessibility. The positive DOPEN and DSCHOOL suggested that the proximity to an open space and a school was not considered as an amenity in rural areas but instead may be viewed as a lack of accessibility.[13] The two variables related to hog impact (HOG_D and SIZE) were significant at 10% with a negative HOG_D and a positive SIZE. The negative HOG_D suggested that hog farms caused negative effects on property values. The positive SIZE implied that house values were higher if they had large farms in the neighborhood, suggesting for scale economies in abatement of environmental impact of hog farms.

Consistent with the results of the robust LM test, the spatial autoregressive parameter ($\rho$) in the ML estimation was positive and highly significant for all three distance bands ($p < 0.00$). Compared with the OLS estimates, all variables retained their signs and significance, but the magnitude of the coefficients decreased in the ML estimates, as the spatial lag effect was then incorporated into the model. This suggests that the OLS estimates will be biased if spatial lag dependence is not accounted for. The magnitude of the spatial autoregressive parameter captured the extent to which a house value at one location was related to its neighbors. For example, $\rho$ was about 0.2 for the 1-mile distance band, meaning one-fifth of the house value could be explained by the values of the neighboring houses. After the spatial lag effect was introduced, there was no remaining spatial error dependence, as the LM test was no longer significant for error dependence. As in the OLS estimation, however, heteroskedasticity was still present in the ML estimation. Subsequently, the 2SLS and 2SLS-robust estimations were carried out to address nonnormality and heteroskedasticity. Compared with the ML estimates, the change in the coefficients was more pronounced in the 2SLS-robust estimation than in the 2SLS estimation, given that the 2SLS-robust estimation took heteroskedasticity into account. All variables retained their signs and significance in the 2SLS and 2SLS-robust estimations.[14]

Accounting for spatial autocorrelation in housing prices significantly changed the magnitude of the impact of hog farms on property values. In the OLS estimates, property values declined per hog by -\$0.51 at 0.75 mile, -\$0.68 at 1 mile, and -\$0.53 at 1.25 mile. When spatial autocorrelation was taken into account in the form of spatial lag dependence, the negative impact on property loss was mitigated on average by 18%. Specifically, in spatial lag model property value losses decreased 4 cents (8%) to -\$0.47 at 0.75 mile, 16 cents (24%) to -\$0.52 at 1 mile, and 11cents (21%) to -\$0.42 at 1.25 mile. Thus, the impact on the value of the median house (\$63,520) 1 mile from a swine facility with 10,000 head was -\$5,200, or 8%.

## Conclusion

This article examined spatial autocorrelation in housing prices in the assessment of negative externality of hog farms on surrounding rural property values. Spatial autocorrelation in housing price is inherent in hedonic models using locational variables such as median household income by census block-group or distance to the facility under study. Substantial spatial autocorrelation was found in OLS estimates. A spatial lag dependence model was developed to correct for spatial autocorrelation as well as nonnormality and heteroskedasticity. The results from the 2SLS-robust estimation showed that when spatial autocorrelation was explicitly accounted for, the negative impact of hog farms on property loss was mitigated on average by 18%.

The findings of this article provide prescription for future hedonic studies in rural settings. The conventional use of sale prices as the dependent variable may overlook spatial autocorrelation in rural property values because they do not represent all rural properties. Turnover is low in rural areas and some rural houses may have never been sold. Predicting missing sale prices can be an alternative using a geostatistical approach (see Dubin). It remains to be seen how realistic such prediction can be if price is made to be a function of distance. Assessed property values were used for this research to remedy this problem creating a more spatially contiguous dataset.

## Endnotes.

[1] There are nine existing studies on the impact of swine operations on property values. See Kim, Goldsmith, and Thomas for an overview.
[2] The covariogram for the distribution of residuals is defined as $C(s_i - s_j) = Cov\{\xi(s_i), \xi(s_j)\}$ for all ($s_i$, $s_j$). The semicovariogram, dividing the covariogram by 2, is more commonly used to describe how spatial dependence changes as the distance between two observations increases. Typically, the semicovariogram curve is upward sloping, leveling off at some distance (the sill). See Cressie for technical details.
[3] The weights also can be based on economic distance or general metric such as a social network structure. See Anselin and Bera, and Anselin (2002) for details.
[4] For the further exposition of a spatial weights matrix, see Anselin (1988, 2002).

[5] There are 12 and 317 urban houses within 1 mile and 1.5 mile distance of the nearest farm, respectively.
[6] An additional outlier was identified and excluded from the sample. See Kim for details.

[7] Division of Water Resources website, "Important Facts about Lists of Animal Operations", ftp://h2o.enr.state.nc.us/pub/Non-Discharge/Animal%20Operations%20Info/
[8] Environmental Defense website, http://www.scorecard.org/env-releases/def/aw_wastes.html
[9] see Kim, Goldsmith, and Thomas for a complete discussion of the Box-Cox hedonic Swine Impact Model
[10] Results for closer and further distance bands (0.5 mile and beyond 1.75 mile up to 3 mile) are not reported, as the variable of interest HOG_D was not significant. The people who were living in close proximity, within a 0.5-mile distance, could be related to the hog farms and may be more tolerant of them, thereby having no negative effect on assessed values. HOG_D was also not significant for the distance band beyond 1.75 mile up to 3 mile. For two distance bands (1.5 and 1.75 mile), the results of the robust LM test indicated that both spatial lag and error dependence were present, preventing the estimation of the spatial lag or spatial error dependence model.
[11] In practice, the LM test statistic can be significant for both spatial lag and spatial error dependence, but the robust LM test statistics point to spatial lag or spatial error dependence.
[12] The number of nearest neighbors selected as a weights matrix in real estate studies ranges from 3 (Can and Megbolugbe) up to 15 nearest neighbors (Pace et al.). The use of 3-nearest neighbor weights matrix for this article conforms to the practice of using three sales comparables in real estate appraisal (Lusht).
[13] The expected sign for DOPEN was either positive or negative. It could be negative if people valued living near an open space in a more quiet environment, or it could be positive if living near open space was viewed as a lack of accessibility. The expected sign for DSCHOOL was negative, as distance was negatively associated with the degree of accessibility to the school.
[14] The only exception was HOG _D in the 2SLS-robust estimation for the 0.75-mile distance band. But given that HOG_D was only marginally insignificant ($p = 0.12$) in the case of the 0.75-mile distance band, the 2SLS-robust estimation would be the basis for estimating the marginal price in the spatial hedonic model.

## References


Anselin, L. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic, 1988.

_________. "Spatial Econometrics." *Companion to Theoretical Econometrics*. Badi Baltagi, ed., pp. 310-330. Oxford: Basil Blackwell, 2001a.

_________. "Rao's Score Test in Spatial Econometrics." Journal of Statistical Planning and Inference 97 (2001b): 113-139.

_________. "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." Agricultural Economics 27 (2002): 247-267.

Anselin, L., and A. K. Bera. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." *Handbook of Applied Economic Statistics*. A. Ullah, and D. E. A. Giles, eds., pp. 237-289. New York: Marcel Dekker, 1998.

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

Beron, K. J., Y. Hanson, J. Murdoch, and M. Thayer. "Hedonic Price Functions and Spatial Dependence: Implications for the Demand for Urban Air Quality." *Advances in Spatial Econometrics*. L. Anselin, R. Florax, and S. Rey, eds. Heidelberg: Springer-Verlag (in press).

Can, A. "The Measurement of Neighborhood Dynamics in Urban House Prices." *Economic Geography* 66 (1990):203-222.

Can, A., and I. Megbolugbe. "Spatial Dependence and House Price Index Construction," *Journal of Real Estate Finance and Economics* 14 (1997): 203-222.

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

Cropper, M. L., L. B. Deck, and K. E. McConnell. "On the Choice of Functional Form for Hedonic Function," *The Review of Economics and Statistics* 70 (1988): 668 - 675.

Davidson, R., and J. G. MacKinnon. *Estimation and inference in Econometrics*. Oxford: Oxford University Press, 1993.

Dubin, R. A. "Predicting House Prices Using Multiple Listings Data." *Journal of Real Estate Finance and Economics* 17 (1998): 35-59.

Dubin, R., R. K. Pace, and T. G. Thibodeau. "Spatial Autoregression Techniques for Real Estate Data." *Journal of Real Estate Literature* 7 (1999): 79-95.

Hite, D., W. Chern, F. Hitzhusen, and A. Randall. "Property-value Impacts of an Environmental Disamenity: The case of landfills." *Journal of Real Estate Finance and Economics* 22 (2001):185-202.

Kiel, K., and J. Zabel. "Estimating the Economic Benefits of Cleaning Up Superfund Sites: The Case of Woburn, Massachusetts." *Journal of Real Estate Finance and Economics* 22 (2001): 163-184.

Kim, J. "An Assessment of the Discommodity Effects of Swine Production on Rural Property Values: Spatial Analysis." PhD dissertation, University of Illinois at Urbana-Champaign, 2004

Kim, J., P. Goldsmith, and M. Thomas. "An Assessment of the Discommodity Effects of Swine Production on Rural Property Values. *Review of Agricultural Economics*. Submitted. June, 2004.

Kim, C. W., T. T. Phipps, and L. Anselin. "Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach," *Journal of Environmental Economics and Management* 45 (2003): 24-39.

Kohlhase, J. E. "The Impact of Toxic Waste Sites on Housing Value." *Journal of Urban Economics* 30 (1991):1-26.

Lusht, K. M. *Real Estate Valuation: Principles and Applications*. New York: McGraw-Hill, 2000.

Pace, R. K., and O. W. Gilley. "Using the Spatial Configuration of the Data to Improve Estimation." *Journal of Real Estate Finance and Economics* 14 (1997): 333-340.

Pace, R. K., R. Barry, J. M. Clapp, and M. Rodriquez. "Spatiotemporal Autoregressive Models of Neighborhood Effects," *Journal of Real Estate Finance and Economics* 17 (1988):5-13.

**Appendix**

The derivation of the marginal price of the spatial hedonic model follows Kim, Phipps, and Anselin (pp. 34-35). First, define the spatial lag model with Box-Cox transformed as

(1) $$Y^{(\lambda)} = \rho W Y^{(\lambda)} + \beta X + \varepsilon$$

where $Y^{(\lambda)}$ is a $(n\times 1)$ column vector of transformed assessed values, $W$ is a $(n\times n)$ weights matrix, $X$ is a $(n\times k)$ matrix (where $k$ is the number of explanatory variables), $\beta$ is a $(k\times 1)$ column vector, and $\varepsilon$ is a $(n\times 1)$ column vector.

The reduced form is

(2) $$Y^{(\lambda)} = [I-\rho W]^{-1}\beta X + [I-\rho W]^{-1}\varepsilon$$

Let $\nu = [I-\rho W]^{-1}\varepsilon$ and $A = [I-\rho W]^{-1}$, then

(3) $$Y^{(\lambda)} = A\beta X + \nu$$

Equation (3) can be written as

(4) $$\begin{bmatrix} Y_1^{(\lambda)} \\ Y_2^{(\lambda)} \\ . \\ . \\ Y_n^{(\lambda)} \end{bmatrix} = \begin{bmatrix} a_{11}, a_{12}, \ldots\ldots, a_{1n} \\ a_{21}, a_{22}, \ldots\ldots, a_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}, \ldots\ldots\ldots\ldots, a_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_{11}, x_{12}, \ldots\ldots, x_{1n} \\ x_{21}, x_{22}, \ldots\ldots, x_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{n1}, \ldots\ldots\ldots\ldots, x_{nn} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} \nu_1 \\ \nu_2 \\ . \\ . \\ \nu_n \end{bmatrix}$$

Define $X_k$ as a column vector $(n\times 1)$ of one housing characteristic. The marginal price can then be derived by taking the derivative of both sides of Equation (4). First, the derivative of $Y^{(\lambda)}(n\times 1)$ with respect to $X_k$ is defined as follows:

(5) $$\frac{\partial Y^{(\lambda)}}{\partial X_k^{'}} = \begin{bmatrix} \partial Y_1^{(\lambda)}/\partial x_{1k}, \partial Y_1^{(\lambda)}/\partial x_{2k}, \ldots\ldots\ldots, \partial Y_1^{(\lambda)}/\partial x_{nk} \\ \partial Y_2^{(\lambda)}/\partial x_{1k}, \partial Y_2^{(\lambda)}/\partial x_{2k}, \ldots\ldots\ldots, \partial Y_2^{(\lambda)}/\partial x_{nk} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \partial Y_n^{(\lambda)}/\partial x_{1k}, \partial Y_n^{(\lambda)}/\partial x_{2k}, \ldots\ldots\ldots, \partial Y_n^{(\lambda)}/\partial x_{nk} \end{bmatrix}$$

$$= Y_k^{(\lambda-1)} \begin{bmatrix} \partial Y_1 / \partial x_{1k}, \partial Y_1 / \partial x_{2k}, \ldots\ldots, \partial Y_1 / \partial x_{nk} \\ \partial Y_2 / \partial x_{1k}, \partial Y_2 / \partial x_{2k}, \ldots\ldots, \partial Y_2 / \partial x_{nk} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \partial Y_n / \partial x_{1k}, \partial Y_n / \partial x_{2k}, \ldots\ldots, \partial Y_n / \partial x_{nk} \end{bmatrix} = Y_k^{(\lambda-1)} \frac{\partial Y}{\partial X_k^{'}}$$

On the other hand, the derivative on the right side of Equation (4) with respect to $X_k$ is

$$(6) \qquad \begin{bmatrix} \beta_k a_{11}, \beta_k a_{12}, \ldots\ldots, \beta_k a_{1n} \\ \beta_k a_{21}, \beta_k a_{22}, \ldots\ldots, \beta_k a_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \beta_k a_{n1}, \beta_k a_{n2}, \ldots\ldots, \beta_k a_{nn} \end{bmatrix} = \beta_k A = \beta_k [I - \rho W]^{-1}$$

It follows that

$$(7) \qquad \frac{\partial Y}{\partial X_k^{'}} = [I - \rho W]^{-1} \frac{\beta_k}{Y_k^{(\lambda-1)}}$$

The marginal price of the spatial lag hedonic model consists of two elements. The first element $[I - \rho W]^{-1}$ amounts to a spatial multiplier and can be expanded into an infinite series:

$$(12) \qquad [I - \rho W]^{-1} = I + \rho W + \rho^2 W^2 + \ldots\ldots$$

$$= \left(\frac{1}{1-\rho}\right) \text{ if } W \text{ is row-standardized and } |\rho| < 1$$

The presence of a spatial multiplier accounts for spatial spillover effects where a change in the property value at one location affects all other locations, whereas the degree of spillover gradually diminishes over space. In the case of hog farms, the spatial multiplier captures the spillover effects of an environmental impact caused by hog farms over *k*-nearest neighbors. The second term $\frac{\beta_k}{Y_k^{(\lambda-1)}}$ is due to Box-Cox transformation and reflects nonlinear relationship between property values and housing attributes.

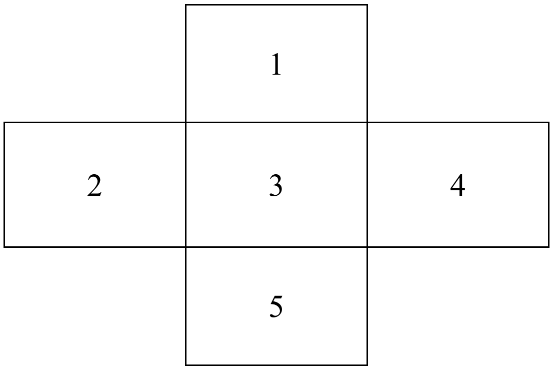


**Figure 1. Neighbors in Regular Grid**

**Table 1. Spatial Weights Matrix**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 1 | 1 | 0.25 | 0.25 | 0 | 0.25 | 0.25 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

**Table 2. Descriptive Statistics for Houses within 3 mile Distance (n = 2,155)**

| Variables | Definition | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| *Dependent Variable* | | | | | |
| VALUE | Assessed property values ($) | 81,862 | 60,892 | 6,260 | 628,710 |
| *Independent variable* | | | | | |
| BASEAREA | Base area (sq. ft) | 1,461 | 449 | 480 | 3,876 |
| ROOM | Number of rooms | 5.8 | 1.1 | 2 | 13 |
| BATHROOM | Number of bathrooms | 1.6 | 0.6 | 1 | 6 |
| LOTSIZE | Lot size (acres) | 1.4 | 1.6 | 0.05 | 10 |
| AGE | Age of house (yrs) | 32 | 25 | 1 | 173 |
| INCOME | Median household income by census block-group ($1,000) | 36.2 | 11.3 | 22.1 | 61.4 |
| DCDB | Distance to the central business district (miles) | 14.1 | 6.3 | 2.9 | 26.7 |
| DSCHOOL | Distance to the nearest school (miles) | 3.8 | 2.1 | 0.1 | 9.2 |
| DOPEN | Distance to the nearest open space (miles) | 0.2 | 0.2 | 0.001 | 1.3 |
| HOG_D | Number of hogs divided by distance to the nearest farm | 3,158 | 3,779 | 264 | 32,555 |
| SIZE | 1 if a farm is large (> 2,500 head) | 0.5 | 0.5 | 0 | 1 |

**Table 3. Estimated Spatial Hedonic Model for 0.75 Mile Distance Band**

| | Models | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | BoxCox (OLS) | | Spatial Lag (ML) | | Spatial Lag (2SLS) | | Spatial Lag (2SLS-Robust) | |
| ρ | | | 0.213 | *** | 0.227 | *** | 0.221 | *** |
| | | | (0.046) | | (0.058) | | (0.058) | |
| CONSTANT | 32.536 | *** | 21.831 | *** | 21.138 | *** | 19.788 | *** |
| | (4.930) | | (5.140) | | (5.559) | | (5.455) | |
| BASEAREA | 0.012 | *** | 0.012 | *** | 0.012 | *** | 0.011 | *** |
| | (0.001) | | (0.001) | | (0.001) | | (0.001) | |
| ROOM | 1.123 | ** | 1.253 | *** | 1.262 | *** | 1.305 | *** |
| | (0.507) | | (0.474) | | (0.486) | | (0.459) | |
| BATHROOM | 4.486 | *** | 3.937 | *** | 3.901 | *** | 4.549 | *** |
| | (1.129) | | (1.056) | | (1.091) | | (1.222) | |
| LOTSIZE | 1.701 | *** | 1.514 | *** | 1.502 | *** | 1.496 | *** |
| | (0.292) | | (0.275) | | (0.284) | | (0.358) | |
| AGE | -0.148 | *** | -0.150 | *** | -0.150 | *** | -0.141 | *** |
| | (0.024) | | (0.022) | | (0.023) | | (0.032) | |
| INCOME | 0.341 | *** | 0.239 | *** | 0.232 | *** | 0.247 | *** |
| | (0.067) | | (0.066) | | (0.070) | | (0.065) | |
| DCBD | -0.334 | *** | -0.264 | ** | -0.260 | ** | -0.247 | ** |
| | (0.126) | | (0.119) | | (0.122) | | (0.107) | |
| DOPEN | 8.589 | *** | 7.349 | *** | 7.269 | *** | 7.348 | *** |
| | (1.939) | | (1.840) | | (1.887) | | (1.665) | |
| DSCHOOL | 0.869 | *** | 0.693 | ** | 0.682 | ** | 0.693 | ** |
| | (0.304) | | (0.286) | | (0.294) | | (0.273) | |
| HOG_D | -0.00017 | * | -0.00015 | * | -0.00014 | * | -0.00012 | |
| | (0.0001) | | (0.0001) | | (0.0001) | | (0.0001) | |
| SIZE | 2.925 | ** | 2.123 | * | 2.071 | * | 1.532 | |
| | (1.254) | | (1.182) | | (1.220) | | (1.158) | |
| Jarque-Bera[1] | 8.51 | ** | | | | | | |
| White/BP[2] | 100.51 | ** | 46.11 | *** | | | | |
| LM-error | 8.92 | *** | 0.06 | | 0.09 | | | |
| Robust LM-error | 10.39 | | | | | | | |
| LM-lag | 11.07 | *** | | | | | | |
| Robust LM-lag | 10.71 | *** | | | | | | |

***Significant at 1%, ** significant at 5%, * significant at 10%

[1] Jarque-Bera test on normality of errors

[2] White or Breusch-Pagan test on heteroskedasticity

Note: LM tests and ML estimation are carried out with 3-nearest neighbor weight matrix.

**Table 4. Estimated Spatial Hedonic Model for 1 mile Distance Band**

| | Models | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | BoxCox (OLS) | | Spatial Lag (ML) | | Spatial Lag (2SLS) | | Spatial Lag (2SLS-Robust) | |
| ρ | | | 0.187 | *** | 0.194 | *** | 0.188 | ** |
| | | | (0.034) | | (0.045) | | (0.048) | * |
| CONSTANT | 41.114 | *** | 28.475 | *** | 27.953 | *** | 27.313 | ** |
| | (4.695) | | (5.054) | | (5.474) | | (5.657) | * |
| BASEAREA | 0.016 | *** | 0.015 | *** | 0.015 | *** | 0.014 | ** |
| | (0.001) | | (0.001) | | (0.001) | | (0.001) | * |
| ROOM | 2.072 | *** | 2.063 | *** | 2.063 | *** | 2.293 | ** |
| | (0.482) | | (0.461) | | (0.467) | | (0.483) | * |
| BATHROOM | 6.807 | *** | 6.196 | *** | 6.171 | *** | 6.392 | ** |
| | (1.130) | | (1.086) | | (1.105) | | (1.260) | * |
| LOTSIZE | 2.203 | *** | 1.993 | *** | 1.984 | *** | 2.056 | ** |
| | (0.312) | | (0.300) | | (0.306) | | (0.366) | * |
| AGE | -0.248 | *** | -0.248 | *** | -0.248 | *** | -0.237 | ** |
| | (0.023) | | (0.022) | | (0.022) | | (0.030) | * |
| INCOME | 0.495 | *** | 0.385 | *** | 0.380 | *** | 0.393 | ** |
| | (0.067) | | (0.067) | | (0.070) | | (0.068) | * |
| DCBD | -0.386 | *** | -0.267 | ** | -0.263 | * | -0.273 | ** |
| | (0.140) | | (0.136) | | (0.139) | | (0.132) | |
| DOPEN | 8.533 | *** | 7.147 | *** | 7.089 | *** | 6.826 | ** |
| | (1.905) | | (1.845) | | (1.877) | | (1.735) | * |
| DSCHOOL | 0.789 | *** | 0.707 | *** | 0.703 | ** | 0.664 | ** |
| | (0.283) | | (0.272) | | (0.275) | | (0.276) | |
| HOG_D | -0.00033 | *** | -0.00024 | ** | -0.00024 | ** | -0.00021 | ** |
| | (0.0001) | | (0.0001) | | (0.0001) | | (0.0001) | |
| SIZE | 4.999 | *** | 3.208 | *** | 3.134 | ** | 2.703 | ** |
| | (1.223) | | (1.219) | | (1.262) | | (2.703) | |
| Jarque-Bera[1] | 15.22 | *** | | | | | | |
| White/BP[2] | 138.00 | *** | 82.34 | *** | | | | |
| LM-error | 15.14 | *** | 0.10 | | 0.006 | | | |
| Robust LM-error | 0.09 | | | | | | | |
| LM-lag | 30.84 | *** | | | | | | |
| Robust LM-lag | 15.80 | *** | | | | | | |

***Significant at 1%, ** significant at 5%, * significant at 10%

[1] Jarque-Bera test on normality of errors

[2] White or Breusch-Pagan test on heteroskedasticity

Note: LM tests and ML estimation are carried out with 3-nearest neighbor weight matrix.

**Table 5. Estimated Spatial Hedonic Model for 1.25 Mile Distance Band**

| Variable | BoxCox (OLS) | | Spatial Lag (ML) | | Spatial Lag (2SLS) | | Spatial Lag (2SLS-Robust) | |
|---|---|---|---|---|---|---|---|---|
| ρ | | | 0.167 | *** | 0.151 | *** | 0.144 | *** |
| | | | (0.029) | | (0.037) | | (0.039) | |
| CONSTANT | 38.524 | *** | 28.177 | *** | 29.365 | *** | 29.336 | *** |
| | (3.150) | | (3.539) | | (3.851) | | (4.010) | |
| BASEAREA | 0.012 | *** | 0.012 | *** | 0.012 | *** | 0.012 | *** |
| | (0.001) | | (0.001) | | (0.001) | | (0.001) | |
| ROOM | 1.488 | *** | 1.502 | *** | 1.504 | *** | 1.595 | *** |
| | (0.327) | | (0.316) | | (0.319) | | (0.339) | |
| BATHROOM | 5.085 | *** | 4.747 | *** | 4.796 | *** | 4.911 | *** |
| | (0.750) | | (0.726) | | (0.735) | | (0.794) | |
| LOTSIZE | 1.979 | *** | 1.794 | *** | 1.784 | *** | 1.812 | *** |
| | (0.218) | | (0.212) | | (0.216) | | (0.263) | |
| AGE | -0.192 | *** | -0.187 | *** | -0.188 | *** | -0.185 | *** |
| | (0.015) | | (0.014) | | (0.015) | | (0.019) | |
| INCOME | 0.397 | *** | 0.315 | *** | 0.318 | *** | 0.323 | *** |
| | (0.046) | | (0.046) | | (0.049) | | (0.048) | |
| DCBD | -0.298 | *** | -0.230 | ** | -0.243 | ** | -0.247 | *** |
| | (0.095) | | (0.093) | | (0.095) | | (0.091) | |
| DOPEN | 6.418 | *** | 5.542 | *** | 5.654 | *** | 5.813 | *** |
| | (1.275) | | (1.243) | | (1.259) | | (1.165) | |
| DSCHOOL | 0.424 | ** | 0.432 | ** | 0.432 | ** | 0.401 | ** |
| | (0.187) | | (0.181) | | (0.182) | | (0.180) | |
| HOG_D | -0.00021 | *** | -0.00016 | ** | -0.00017 | ** | -0.00014 | ** |
| | (0.0001) | | (0.0001) | | (0.0001) | | (0.0001) | |
| SIZE | 3.821 | *** | 2.790 | *** | 2.968 | *** | 2.877 | *** |
| | (0.780) | | (0.774) | | (0.795) | | (0.804) | |
| Jarque-Bera[1] | 8.87 | ** | | | | | | |
| White/BP[2] | 139.94 | *** | 79.43 | *** | | | | |
| LM-error | 24.06 | *** | 1.36 | | 0.95 | | | |
| Robust LM-error | 1.82 | | | | | | | |
| LM-lag | 36.09 | *** | | | | | | |
| Robust LM-lag | 13.85 | *** | | | | | | |

***Significant at 1%, ** significant at 5%, * significant at 10%

[1] Jarque-Bera test on normality of errors

[2] White or Breusch-Pagan test on heteroskedasticity

Note: LM tests and ML estimation are carried out with 3-nearest neighbor weight matrix.
Figure 1.