

**ANISTROPIC AUTOCORRELATION
IN HOUSE PRICES**

by

Kevin Gillen

Thomas Thibodeau

Susan Wachter

Working Paper #383 **Anisotropic Autocorrelation in House Prices**

Kevin Gillen
Real Estate Department
The Wharton School
University of Pennsylvania
314 Lauder-Fischer Hall
256 South 37th Street
Philadelphia, PA 19104.6330
Phone: 215.898.4818
Fax: 215.573.5261
Email: gillenk@wharton.upenn.edu

Thomas Thibodeau*
Professor of Real Estate
Cox School of Business
Southern Methodist University
Dallas, TX 75275-0333
Phone: 214.768.3613
Fax: 214.768.4099
Email: tthibode@mail.cox.smu.edu

Susan Wachter

Assistant Secretary Designee for
Policy Development and Research
US Department of Housing and Urban Development
451 Seventh Street SW
Suite 8100
Washington, DC 20410
Phone: 202.708.1600
Fax: 202.619.8000

*Contact Author

October 2001

Abstract

This paper examines *anisotropic* spatial autocorrelation in single-family house prices and in hedonic house price equation residuals using a spherical semivariogram and transactions data for one county in the Philadelphia, Pennsylvania MSA. *Isotropic* semivariograms model spatial relationships as a function of the distance separating properties in space. *Anisotropic* semivariograms model spatial relationships as a function of both the distance and the direction separating observations in space.

The goals of this paper are: (1) to determine whether there is spatial autocorrelation in the hedonic residuals; and (2) to empirically examine the validity of the isotropy assumption. We estimate the parameters of spherical semivariograms for house prices and for hedonic house price equation residuals for 21 housing submarkets within Montgomery County, Pennsylvania. These housing submarkets are constructed by dividing the entire county into 21 groupings of economically similar adjacent census tracts. Census tracts are grouped according to 1990 census tract median house prices and according to characteristics of the housing stock. We fit the residuals of each submarket hedonic house price equation to both *isotropic* and *anisotropic* spherical semivariograms. We find evidence of spatial autocorrelation in the hedonic residuals in

spite of a very elaborate hedonic specification. Additionally, we have determined that, in some submarkets, the spatial autocorrelation in the hedonic residuals is anisotropic rather than isotropic. The empirical results suggest that the spatial autocorrelation in single-family house prices and in hedonic house price equation residuals is *anisotropic* in submarkets where most residents commute to a regional or local Central Business District (CBD).

We acknowledge Professor Tony Smith, John Green of REALIST, and Paul Amos for their assistance with this paper.

1. Introduction

House prices are spatially autocorrelated because properties in close proximity tend to have similar structural characteristics (e.g. square feet of building/living area, dwelling age, and design features). This is a natural consequence of the fact that spatially proximate properties tend to be developed at about the same time. In addition, properties within the same neighborhood share important neighborhood amenities (e.g. neighborhood properties have access to the same public schools and are served by the same municipal police and fire departments). Finally, house prices are likely to be spatially autocorrelated in neighborhoods where residents follow similar commuting patterns.

House price models attempt to explain spatial and/or temporal variation in house prices. These models are frequently used to mark residential property values to market. Hedonic models relate house prices to characteristics of the lot, the structure, and the neighborhood (Gillingham [1975], Goodman [1978], Thibodeau [1989, 1992, 1996], and many others). Repeat sales models measure changes in house prices as the average rate of appreciation for properties that have sold at least twice and have not undergone major structural changes between sales dates (Bailey, Muth and Nourse [1963], Case and Shiller [1987, 1989]). Hybrid models combine hedonic and repeat sales specifications to obtain more efficient parameter estimates (Case and Quigley [1991], Quigley [1995], and Hill, Knight and Sirmans [1997]). Assessed value models estimate price indices using information obtained from property tax assessment departments (Clapp and Giaccotto [1992]).

The parameters of hedonic house price equations are typically estimated using Ordinary Least Squares (OLS). This estimation procedure assumes the residuals are independently and identically distributed with zero mean, a constant variance, and zero covariance. When the residuals are spatially autocorrelated, the assumption of a zero covariance is violated and OLS yields inefficient parameter estimates. More accurate parameter estimates can be obtained by explicitly modeling the spatial autocorrelation. Modeling spatial relationships in hedonic house price equations can also significantly improve market value prediction accuracy.

The residuals produced by house price models may be spatially autocorrelated for three reasons. First, proximity externalities influence the market values of nearby properties in similar ways. Thibodeau [1990] demonstrated that high-rise office buildings reduce the market value of nearby homes by as much as 15%. Information on the determinants of proximity externalities in the single-family market is difficult and costly to obtain. Second, other information on important structural and neighborhood characteristics are not readily available and are excluded from empirical house price specifications. For example, data on the quality of local public schools and on area crime rates are difficult and costly to obtain, particularly on a large (e.g. national) scale. In addition, information on neighborhood socioeconomic and demographic characteristics are typically obtained from the U.S. Bureau of the Census. This information is collected only once every ten years. When hedonic equations are used to model house prices, the residuals will contain information on these unobserved housing characteristics. Third, even in ideal situations where all housing characteristic information is available, it is difficult to select the "correct" model specification. For example, it is difficult to model how public school quality gets capitalized into the price of single-family properties. Model mis-specification may also contribute to spatially autocorrelated hedonic house price equation residuals.

Several researchers have developed hedonic house price models that examine spatial autocorrelation in house prices and in hedonic house price model residuals. Dubin [1988] assumed the residual correlation between properties is a negative exponential function of the distance between them and estimated hedonic parameters using a maximum likelihood procedure suggested by Mardia and Marshall [1984]. Can [1992] incorporated spatially lagged values of

house prices in the hedonic specification. The absolute influence that nearby properties have on value is determined using an exogenously specified weighting matrix. Pace and Gilley [1997] model spatial dependence in house prices using a simultaneous autoregressive (SAR) model. SAR models explain variation in house prices as a function of property characteristics and spatially weighted hedonic residuals of comparable properties. They derive the SAR model by combining the OLS and grid estimator and then compare the OLS, grid, and SAR estimates of market value. Basu and Thibodeau [1998] examine spatial autocorrelation in Dallas house prices using a semi-log hedonic house price equation and a spherical autocorrelation function with data for over 5,000 transactions of homes sold between 1991:4 and 1993:1. Properties are geocoded and assigned to separate housing submarkets within metropolitan Dallas. Hedonic and spherical autocorrelation parameters are estimated separately for each submarket using estimated generalized least squares (EGLS). They conclude that house prices are spatially autocorrelated throughout metropolitan Dallas but that hedonic house price equation residuals are spatially autocorrelated in about half of the submarkets examined.

Most research on spatial autocorrelation in house prices has assumed that the correlation structure is *isotropic*--a function of only the distance between properties. The direction separating properties is ignored. Spatial data is *anisotropic* when spatial autocorrelation is a function of *both* the distance and the direction separating points in space. *Anisotropic* semivariograms have been examined by Journel and Huijbregts [1978], Oden and Sokal [1986], Isaaks and Srivastava [1989] and Simon [1997]. Simon [1997], for example, provides an exact test for *anisotropic* spatial autocorrelation. The exact tests are obtained by projecting the two dimensional spatial observations onto a single axis making an angle θ with the east-west axis. Spatial autocorrelations are then replaced by the conventional product-moment correlation coefficient.

House prices and hedonic house price equation residuals may exhibit *anisotropic* spatial autocorrelation. Residential location theory suggests that housing consumers tradeoff housing and commuting costs when selecting a residence. To reduce commuting costs, residential properties are developed initially around major transportation arteries. This pattern of residential development may result in stronger spatial autocorrelation in house prices (and in hedonic house

price equation residuals) along major transportation arteries, and in the direction of the Central Business District (CBD).

This paper examines *anisotropic* spatial autocorrelation for Montgomery County single-family transactions. We fit the parameters of a spherical autocorrelation function to the empirical (or sample) semivariogram for two directions in each of 21 housing submarkets in suburban Philadelphia. Our results suggest that the spatial autocorrelation in transaction prices and in hedonic house price equation residuals is *anisotropic* for some housing submarkets.

2. Specification

2.1 *The Hedonic House Price Specification*

We model the relationship between house prices and housing characteristics using a semi-logarithmic functional form. This specification regresses the log of transaction prices on a linear combination of (possibly transformed) housing characteristics. The semi-log functional form is given by:

$$V = e^{X\beta + \varepsilon} \quad (1)$$

where V is property value, X is a vector of (possibly transformed) housing characteristics, β is a vector of unknown hedonic coefficients, and ε is the residual. When the residual variance is constant and the residuals are spatially uncorrelated, ordinary least squares (OLS) yields best, linear, unbiased estimators of the parameters in the transformed equation:

$$Z = \log V = X\beta + \varepsilon, \quad (2)$$

where $\varepsilon \sim N(0, \sigma^2 I)$ so that $Z \sim N(X\beta, \sigma^2 I)$. OLS yields estimated coefficients

$$b = (X^T X)^{-1} X^T Z, \quad \text{where } b \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (3)$$

2.2 *Modeling Spatial Autocorrelation*

When the residuals are spatially autocorrelated, $E \{ \varepsilon \varepsilon' \} = \Omega$, a matrix with non-zero off-diagonal elements. In this situation, β can be estimated with the generalized least squares (GLS) estimator $B = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Z$. The empirical challenge is to estimate the elements of Ω .

Let $s_i = (a_i, b_i)$ denote the location of property i (e.g. a_i denotes the longitude and b_i the latitude for property i). Let $\xi(s_i)$ denote the hedonic price equation residual for a property located at s_i . If the stochastic process is weakly stationary, the *covariogram* for the distribution of residuals is $C(s_i - s_j) = \text{Cov} \{ \xi(s_i), \xi(s_j) \}$ for all (s_i, s_j) . Note that $C(0)$ is the (assumed constant) variance for the residual distribution. The *semivariogram* of the process is:

$$\gamma(s_i - s_j) = 0.5 \text{Var} \{ \xi(s_i) - \xi(s_j) \} = C(0) - C(s_i - s_j). \quad (4)$$

Let \mathbf{h} denote the (Euclidean) distance separating locations s_i and s_j . Clearly, $\gamma(-\mathbf{h}) = \gamma(\mathbf{h})$ and theoretically $\gamma(\mathbf{0}) = 0$. Empirically however, $\gamma(\mathbf{h})$ is sometimes discontinuous near the origin and $\gamma(\mathbf{h}) \rightarrow \theta_0 > 0$, as $\mathbf{h} \rightarrow 0$. The discontinuity, θ_0 , is labeled the *nugget*.

Observations may eventually become spatially uncorrelated as the distance between them increases. When this happens, the semivariogram stops increasing beyond some threshold and becomes constant. That is, $\gamma(\mathbf{h}) \rightarrow C^*$, as $\mathbf{h} \rightarrow \infty$. This limiting value, C^* , is called the *sill* of the semivariogram. The *range* of the semivariogram is the value \mathbf{h}_0 such that $\gamma(\mathbf{h}_0) = C^*$. So the range of a semivariogram is the distance beyond which observations are spatially uncorrelated. Finally, a semivariogram is *isotropic* if $\gamma(s_i - s_j)$ is a function of only the distance between s_i and s_j , $\|s_i - s_j\|$, and not the direction separating s_i and s_j . Spatial data is *anisotropic* when spatial autocorrelation is a function of *both* the distance and the direction separating points in space. *Geometric anisotropy* occurs when the range varies with direction but the sill is constant. *Zonal anisotropy* occurs when the sill varies with direction but the range is constant (Isaaks and Srivastava [1989]). Mixed models are models where both the sill and the range vary with the direction separating spatial observations.

The empirical (or sample) semivariogram examines how the spatial autocorrelation between observations changes as the distance between observations increases. The method of moments estimator for an empirical semivariogram (Matheron, [1963]) is:

$$g(\mathbf{h}) = \frac{\sum_{N(\mathbf{h})} [\xi(s_i) - \xi(s_j)]^2}{2|N(\mathbf{h})|} \quad (5)$$

where the average is taken over $N(\mathbf{h}) = \{(s_i, s_j) : s_i - s_j = \mathbf{h}\}$ and $|N(\mathbf{h})|$ is the distinct number of pairs in $N(\mathbf{h})$. For irregularly spaced data (e.g. single-family properties), $N(\mathbf{h})$ is modified so that $N(\mathbf{h}) = \{(s_i, s_j) : s_i - s_j \in T(\mathbf{h})\}$, where $T(\mathbf{h})$ is a tolerance region around \mathbf{h} . Figure 1 plots the points of the *isotropic* empirical (or sample) semivariogram for the log of transaction prices in the Ambler submarket, an area in the north-central region of Montgomery County. A point in the empirical semivariogram (e.g. a dot in Figure 1) is the difference between the variance and covariance in (the log of) house prices computed for properties within a given tolerance region. The first point to the right of the vertical axis measures the spatial autocorrelation for properties within +/- 200 meters of the separation distance $h = 250$ meters. The statistic is computed without regard for the direction separating properties. The points for the empirical semivariogram are computed for 40 values of h ranging from $h=250$ meters to $h=6,500$ meters. The tolerance ranges or bins are constructed so that each bin has approximately the same number of property pairs.

<Figure 1 About Here>

The next challenge is to fit a functional form to the points of the empirical semivariogram. Three popular *isotropic* semivariograms used to empirically examine spatial relationships are the spherical, exponential, and Gaussian semivariograms. (Cressie [1993] provides the functional forms for these semivariograms. Additional references on modeling spatial autocorrelation include Bailey and Gatrell [1995], Cliff and Ord [1973], Isaaks and Srivastava [1989], Ripley [1981], and Dubin, Pace and Thibodeau [1998]). The spherical semivariogram model has a finite range while the exponential and Gaussian semivariograms asymptotically approach a limiting value. The functional form for the spherical model is given by Cressie [1993]:

$$\gamma(\mathbf{h}; \theta) = \begin{cases} 0, & \mathbf{h} = 0 \\ \theta_0 + \theta_1 \left\{ 1.5 \left(\frac{\|\mathbf{h}\|}{\theta_2} \right) - 0.5 \left(\frac{\|\mathbf{h}\|}{\theta_2} \right)^3 \right\}, & 0 < \|\mathbf{h}\| \leq \theta_2 \\ \theta_0 + \theta_1, & \|\mathbf{h}\| \geq \theta_2 \end{cases} \quad (6)$$

The *nugget* for the spherical semivariogram is θ_0 , the *sill* is $\theta_0 + \theta_1$, and the *range* is θ_2 .

The parameters of the spherical semivariogram are estimated using nonlinear least squares. The three parameters of the spherical semivariogram model can be fit to the empirical semivariogram, $g(\mathbf{h})$, by minimizing the nonlinear function:

$$S(\theta) = \sum_{k=1}^K [g(\mathbf{h}(k)) - \gamma(\mathbf{h}(k), \theta)]^2 \quad (7)$$

with respect to the semivariogram parameters θ . The sequence $\mathbf{h}(1), \dots, \mathbf{h}(K)$ denotes the separation distances for which the sample semivariogram $g(\mathbf{h})$ are computed. Figure 1 also plots a spherical semivariogram fitted to the points of the empirical semivariogram for the Ambler submarket. The fitted spherical semivariogram is discontinuous at the origin (with an estimated *nugget* = 0.0747) and increases with separation distance h . The fitted spherical semivariogram becomes horizontal for $h > 4.38$ km, so house prices for properties separated by more than 4.38 km are spatially uncorrelated. If Ambler house prices were spatially uncorrelated, then the fitted semivariogram would be horizontal at all separation distances.

To estimate the standard errors of the semivariogram parameter estimates, we assume the semivariogram residuals are independently and identically distributed with mean zero and variance σ^2 , then

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^K (G_i - \gamma_i)^2}{K} \quad \text{where: } \begin{cases} G_i = \text{value of the empirical semivariogram at } D_i \\ \gamma_i = \text{value of the spherical semivariogram at } D_i \end{cases} \quad (8)$$

and the nonlinear least squares estimator $\hat{\theta}$

is approximately normally distributed with mean θ and covariance matrix (Maddala [1977])

$$\hat{\sigma}^2 [F(\hat{\theta})^T F(\hat{\theta})]^{-1} \quad (9)$$

$F(\theta)$ is an $N \times 3$ gradient matrix created by evaluating the partial derivatives of (6) with respect to θ_0 , θ_1 , and θ_2 at each of the K categorical distances D_i :

$$\begin{aligned} \mathbf{F}(\theta) &= \begin{bmatrix} \frac{\partial \gamma}{\partial \theta_0} & \frac{\partial \gamma}{\partial \theta_1} & \frac{\partial \gamma}{\partial \theta_2} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \left\{ \frac{3}{2} \left(\frac{D_i}{\theta_2} \right) - \frac{1}{2} \left(\frac{D_i}{\theta_2} \right)^3 \right\} & \theta_1 \left\{ -\frac{3}{2} \left(\frac{D_i}{\theta_2^2} \right) + \frac{3}{2} \left(\frac{D_i^3}{\theta_2^4} \right) \right\} \end{bmatrix} \text{ for } i=1, 2, \dots, K \end{bmatrix} \quad (10)$$

3. The Data

This paper examines spatial autocorrelation in house prices and in hedonic house price equation residuals using data for 21,562 transactions of single-family homes sold between 1995:Q1 and 1998:Q3 in Montgomery County, Pennsylvania. Montgomery County lies directly to the northwest of Philadelphia. The county contains many homes that pre-date the American Revolution. The eastern half of the county is a collection of various bedroom communities of educated professionals. It includes portions of Philadelphia's famous "Main Line", a sequence of affluent communities to the west of Philadelphia dating back to the 19th century. The region includes several of the "Seven Sisters" schools (Bryn Mawr, Villanova, Swarthmore, etc.). The western half of the county is predominately rural, containing several farming communities that date to colonial times, including a strong Amish presence. Map 1 illustrates the location of Montgomery County, PA.

<Map 1 About Here>

Realist, a private data vendor, provided 21,562 transactions of Montgomery County single-family properties. The typical Montgomery County single-family transaction had about 2,030 square feet of building area. The mean transaction price over the 1995-1998 period was \$179,823, or \$88.61 per square foot. While a few homes were built prior to the American Revolution, the average age for a Montgomery County property sold during the 1995-1998 period is 38.2 years.

3.1 The Submarkets

Twenty-one submarkets were defined by clustering contiguous census tracts with similar housing characteristics. Within each submarket, we would ideally prefer a distribution of house prices with a constant average unit price and a small variance, but the non-uniform spatial distribution of properties constrained this possibility. That is, if we defined the geographic submarkets such that each submarket could be classified as relatively "high-priced", "mid-priced", or "low-priced", there would be several submarkets with very few observations in them. So, for example, we were forced to occasionally include several low-priced tracts in an otherwise mid-priced submarket. The result was 21 distinct geographic housing submarkets that generally conformed with known municipal and geographic boundaries. Map 2 provides the submarket boundaries for the 21 Montgomery County submarkets.

<Map 2 About Here>

Table 1 contains descriptive statistics for transactions in the 21 submarkets. The median submarket transaction price ranged from \$108,000 for Pottstown to \$325,000 for the Main Line (West) submarket. The mean per square foot price ranged from \$73.33 for Pottstown to \$126.80 for the Main Line (West) submarket. The oldest properties are located in the eastern part of the county and the average age declines for submarkets in the western portion of the county. For example, the oldest properties are located in the Main Line (East) submarket (with an average age of 60 years) while the newest stock is in the Salford submarket (with an average age of 20 years).

<Table 1 About Here>

Later in the paper we provide semivariogram parameter estimates for both *isotropic* and *anisotropic* spherical semivariograms for each of the 21 submarkets. In addition, we provide detailed hedonic results for three representative submarkets: the Main Line (East) submarket, the Norristown submarket, and the Ambler submarket. A brief characterization of these three submarkets follows.

The Main Line is one of the oldest and wealthiest suburbs of the Philadelphia MSA. Named for a still-active commuter train line that dates back to the 19th century, the Main Line is a contiguous sequence of affluent suburbs that spans three counties, beginning at Philadelphia's westernmost city line. To normalize the relative size of the submarkets, we divided the segment of the Main Line that is within Montgomery County it into two distinct submarkets: Main Line (East) and Main Line (West). The Main Line's housing stock can be characterized by *global homogeneity*, but *local heterogeneity*: uniformly high-priced, but the value of any particular property can still differ from its neighbor substantially (e.g. \$600,000 v. \$400,000). The 1996 estimated average median census tract household income for the Main Line (East) submarket is \$136,256 and 40.7% of the submarket population has a college degree. The submarket is approximately 7.5 km by 9.0 km and has an area of 27,400 m². A typical property in the Main Line (East) submarket has 2,488 square feet of building area and sold for \$277,604, or \$111.66 per square foot. This area has some of the oldest homes in the Philadelphia area, with 25% of the transactions built prior to 1925. There are 1,399 transactions for the Main Line (East) submarket.

The Norristown submarket includes the edge city of Norristown and surrounding tracts. Norristown is an old industrial mill town dating back to the American Revolution that was built on the banks of the Schuylkill river for access to waterpower. The housing stock in the city proper contains many traditionally working-class row homes and cottages in high-density urban-like settings that are typical of neighborhoods in larger Northeastern cities. Although Norristown's industrial base is no longer extant, and the city has become steadily enveloped in Philadelphia's spreading suburban sprawl, Norristown persists as the area of Montgomery county with the lowest median income and highest percentage of minority residents. The 1996 estimated average median census tract household income for this submarket is \$63,391 and 16.2% of the

population has a college degree. The housing stock of Norristown can be characterized by both *global* and *local homogeneity*: uniformly low-priced neighborhoods of similar rowhomes and small homes. The Norristown submarket is approximately 9.8 km by 9.1 km and has an area of 35,139 m². The area residents are primarily blue collar. The average transaction price in Norristown was \$128,247, or \$81.67 per square foot. Transactions had an average of 1,639 square feet of building area and were 46 years old at the time of sale. There are 997 Norristown transactions.

The Ambler submarket is the stereotypical postwar middle-class bedroom community of the American suburban landscape. It is also the housing submarket that is most similar in characteristics to the overall Montgomery housing market. Geographically, Ambler is the largest of the three submarkets profiled here, and unlike the other two submarkets, it contains no real geographically defined center. It is best described as a collection of similar suburban housing developments near to, but without any real relationship to, the township of Ambler. The 1996 estimated average median census tract household income for this submarket is \$98,255, and 29.8% of the population have a college degree. The housing stock of Ambler is best characterized as *globally heterogeneous*, but *locally homogenous*: clusters of architecturally similar and like-priced homes, but some relative variance between clusters of development due to the timing and nature of general suburban growth. The Ambler submarket is approximately 13.1 km by 11.7 km and has an area of 55,490 m². The average transaction price in Ambler was \$232,301, or \$94.92 per square foot. Transactions had an average of 2,446 square feet of building area and were 28 years old at the time of sale. There are 1,773 transactions for Ambler.

4. Empirical Specifications

4.1 The Empirical Hedonic House Price Specification

With a few exceptions, the empirical specification is standard. The house price specification relates the log of transaction price to various structural (dwelling size, dwelling age, number of stories, etc.) and location (distance to CBD) characteristics and includes dummy variables for sales date.

The continuous variables in the model are common to most traditional hedonic model specifications: log of building square footage, frontage, and log of number of stories. We also take the ratio of building square footage to lot square footage to measure the pricing of aesthetic proportionality of the property. As this variable becomes larger, it increasingly denotes a large house on a small lot. Hence the negative sign on this variable's coefficient denotes a discounting for the lack of yard space and/or privacy. Additionally, we also take the ratio of total number of rooms to building square footage to measure the effects of average room size in a property. In accordance with the Victorian aesthetic of the time, many homes in Montgomery County built during the 19th century contain a very tight partitioning of the building into a sequence of many rooms that would be considered awkwardly small by today's standards. The positive sign on its coefficient indicates that, for a fixed house size, today's consumers generally prefer fewer large rooms to numerous small rooms.

A set of dummy variables measure the effects of categorical housing characteristics. For example, a dummy variable is created for each unique value of qualitative variables, like exterior material and type of heating fuel. Additionally, dummy variables are created for discrete variables, like number of fireplaces or number of bathrooms to allow for nonlinear relationships between these structural characteristics and house prices. As a general rule, the category with the greatest number of observations associated with it serves as the omitted variable in the empirical specification. There are eight possible categories for exterior material. The dummy variable for aluminum exterior is omitted from the specification since it had the largest percentage of observations associated with it.

The empirical specification includes variables designed to measure the influence that housing vintage has on house price. Exhibit 1 illustrates the distribution for year of construction. Although the volume of construction is certainly increasing over time as the population of the region increased, the historical pattern of housing construction exhibits several cycles of development. Measuring from trough-to-trough, 6 distinct cycles of construction emerge, which are labeled: pre-1865, 1866-1887, 1888-1918, 1919-1945, 1946-1975, and 1976-1998. Each wave of new construction after 1865 is approximately 25 years in length, and all properties built during a particular cycle are characterized by a common set of hedonic characteristics: architectural style, exterior material, heating fuel, etc.

To capture the unique and nonlinear effects each cycle, we take the interaction of a property's age with a dummy variable that equals "1" if the property was constructed during that cycle. We do the same with age-squared and age-cubed. The general specification of the formula is:

$$\begin{aligned} \text{AGE}_{i,j} &= \text{age of the } i\text{th house} * (\text{dummy} = 1 \text{ if built during } j\text{th cycle,} = 0 \text{ otherwise)} \\ \text{AGESQ}_{i,j} &= \text{age - squared of the } i\text{th house} * (\text{dummy} = 1 \text{ if built during } j\text{th cycle,} = 0 \\ &\quad \text{otherwise)} \\ \text{AGECUBE}_{i,j} &= \text{age - cubed of the } i\text{th house} * (\text{dummy} = 1 \text{ if built during } j\text{th cycle,} = 0 \\ &\quad \text{otherwise), for } j = 1, \dots, 6; i = 1, \dots, 21,562. \end{aligned}$$

<Exhibit 1 About Here>

The specification also includes variables that measure the influence that distance to three centers of economic activity have on house price. The three centers of economic activity are the City of Philadelphia, the Montgomery County CBD measured at the King of Prussia and the distance to the submarket center of economic activity.

Finally, the specification includes dummy variables for sales quarter beginning with the first quarter of 1995 and ending with the second quarter of 1998. The omitted category is for properties sold at the end of the period (1998:3).

Although we impose the same specification on all submarkets, we did consider allowing the specification to vary across submarkets to be better tailored to variations in consumer preferences across these submarkets. But, we rejected this strategy since we reasoned that it would make us vulnerable to the accusation that we possibly tweaked the specification in order to induce anisotropy in the residuals. By imposing the same large and uniform specification on all submarkets, we reduce the probability of there being any spatial effects present in the residuals. Hence, any spatial autocorrelation in the residuals is not spurious. Reducing the number of variables in the specification increases the likelihood of the residuals being correlated with any omitted variables: surely an undesirable result.

The empirical hedonic house price specification is:

$$\begin{aligned}
 \text{Ln (TRANSACTION PRICE}_i) &= \beta_0 \\
 &+ \beta_1 * \text{DWELLING AGE} + \beta_2 * \text{AGESQ} + \beta_3 * \text{AGECUBE} \\
 &+ \beta_4 * (\text{DWELLING AGE} * \text{DEVELOPMENT CYCLE}) \\
 &+ \beta_5 * (\text{DWELLING AGE} * \text{DEVELOPMENT CYCLE})^2 \\
 &+ \beta_6 * (\text{DWELLING AGE} * \text{DEVELOPMENT CYCLE})^3 \\
 &+ \beta_7 * \text{LN(BUILDING SQUARE FOOTAGE)} \\
 &+ \beta_8 * (\text{BUILDING SF/LOT SQUARE FOOTAGE}) \\
 &+ \beta_9 * (\text{TOTAL ROOMS/BUILDING SF}) \\
 &+ \beta_{10} * \text{LN(NUMBER OF STORIES)} \\
 &+ \beta_{11} * \text{TYPE OF EXTERIOR} \\
 &+ \beta_{12} * \text{TYPE OF SPACE HEATING SYSTEM} \\
 &+ \beta_{13} * \text{TYPE OF HEATING FUEL} \\
 &+ \beta_{14} * \text{TYPE OF BASEMENT} \\
 &+ \beta_{15} * \text{NUMBER OF FIREPLACES} \\
 &+ \beta_{16} * \text{NUMBER OF BATHROOMS} \\
 &+ \beta_{17} * \text{GARAGE CAPACITY} \\
 &+ \beta_{18} * \text{FRONTAGE} \\
 &+ \beta_{19} * \text{CORNER LOT} \\
 &+ \beta_{20} * \text{SOURCE OF WATER SUPPLY} \\
 &+ \beta_{21} * \text{SEWER SYSTEM} \\
 &+ \sum_{i=1}^3 \phi_i * \text{DISTANCE TO CBD}_i \\
 &+ \sum_{j=1}^T \delta_j * \text{SOLD}_j + \xi_i,
 \end{aligned} \tag{10}$$

where

$DISTANCE_{1i}$ = Distance of the i th house to the Philadelphia CBD

$DISTANCE_{2i}$ = Distance of the i th house to the the Montgomery County CBD*

$DISTANCE_{3i}$ = Distance of the i th house to its submarket's CBD

* *King of Prussia*

and $SOLD_j = 1$ if property sold in quarter j and is zero otherwise; $j = 1995:1, 1995:2, \dots, 1998:2$.

There are a number of discrete categorical variables in the specification which necessitated the creation of a series of dummy variables. As a general rule, we always chose that value as the omitted category for which the number of observations was most numerous. The following chart gives the value of the omitted discrete variables used in the specification:

<Table 1a About Here>

The descriptive statistics for the housing characteristics included in the hedonic specification for the Main Line (East), Norristown, and Ambler submarkets are provided in Table 2.

<Table 2 About Here>

5. Estimation Results

5.1 House Price Spherical Semivariograms

We examine directional spatial autocorrelation in Montgomery County house prices (and in hedonic house price equation residuals) by fitting a spherical function to the empirical semivariogram computed using properties separated by a given direction. Our analysis examines

directional autocorrelation for two directions: north-south and east-west. Since few properties are located exactly north-south (or east-west) of one another, we compute the spatial autocorrelation for properties within a tolerance range of the desired direction. The tolerance region used here is $\pm 45\%$. Consequently, the points for the empirical semivariogram for property pairs separated by a northerly direction are computed for properties located $90^\circ \pm 45^\circ$. Similarly, the points for the empirical semivariogram for property pairs separated in an easterly direction are computed for properties located $0^\circ \pm 45^\circ$.

The following table presents the empirically estimated thetas and (approximate) t-scores of the spherical semivariogram for $\log(\text{house price})$ in the Ambler submarket. The measurement of θ_2 , the range, has been re-scaled to kilometers.

Estimated Semivariogram Parameters: Thetas and (t-scores)
--

Log (House Price)

Main Line (East) Submarket

Isotropic Spherical Semivariogram

θ_0	0.099003
(t-score)	(12.801)
θ_1	0.08001
(t-score)	(9.1672)
θ_2	4.4115
(t-score)	(5.7187)

Anisotropic Spherical Semivariogram

Properties North-South

θ_0	0.099888
(t-score)	(11.746)
θ_1	1.4489
(t-score)	(0.0084357)
θ_2	85.228
(t-score)	(0.0084211)

Properties East-West

θ_0	0.10862
(t-score)	(12.637)
θ_1	0.044479
(t-score)	(4.9164)
θ_2	3.3459
(t-score)	(3.1912)

The empirical semivariogram results clearly indicate that the spatial autocorrelation of house prices in this submarket is *anisotropic*. The values of the spherical semivariogram parameters for the *isotropic* semivariogram are statistically different from zero, indicating that house values are correlated up to a distance of approximately 4.4 km. However, the *anisotropic* semivariograms differ in their results: although the estimated semivariogram parameters are uniformly significant for the "east" semivariogram, only θ_0 is significant for the "north" semivariogram. This suggests that the Main Line (East) spatial autocorrelation in house prices exists solely along the east-west axis of the submarket. Figure 2 plots the *isotropic* and *anisotropic* semivariograms.

<Figure 2 about here>

First, note that the *isotropic* semivariogram lies exactly between the *anisotropic* semivariograms. This would suggest that the *isotropic* measurement of spatial dependence is an average of the *anisotropic* measurements. This *isotropic* semivariogram illustrates that (the log of) the Main Line house prices are spatially autocorrelated up to a range of about 4.4 kilometers. Beyond 4.4 kilometers, (the log of) the Main Line house prices are spatially uncorrelated. But, the *anisotropic* semivariogram for "east" indicates that Main Line house prices are spatially autocorrelated up to a distance of only 3.3 kilometers. At first glance, the non-converging *anisotropic* semivariogram for "north" would suggest that the spatial stochastic process of house prices is nonstationary, and the actual range of spatial autocorrelation is much greater than 4.4 kilometers. However, the insignificant t-scores for its parameters suggest just the opposite. In actuality, there is no spatial autocorrelation in house prices in a north-south direction. This spatial stochastic process is just a pure 'nugget' effect as house prices vary randomly from one property to the next with no covariance. All spatial autocorrelation occurs along the east-west axis.

Cities generally develop outwards from their centers. Consequently, we might *a priori* expect that the direction of greatest autocorrelation would be in the direction of the city's CBD. This is exactly true for this submarket: Philadelphia lies directly across the eastern edge of this submarket's border.

5.2 Hedonic Parameters

The parameters of the hedonic equation and the spherical semivariogram are estimated separately for the twenty-one submarkets. Table 3 provides the OLS regression statistics for the Main Line (East), Norristown, and Ambler submarkets. (Results for the other submarkets are available from the authors upon request).

<Table 3 About Here>

In general, dwelling size and age explain most of the variation in the log of transaction prices. t-statistics for the logarithm of square feet of building area coefficients range from 15.7 for Norristown to 23.9 for Ambler. (The estimated coefficient for the log of building area for the Montgomery County hedonic has a t-statistic of 78.6.) Estimated coefficients for dwelling age polynomials are jointly statistically significant for each submarket and document different depreciation patterns. The omitted type of exterior is aluminum, so the estimated coefficients measure differences from this type of exterior. The frame, masonry, stone, and stucco exteriors command premiums in all submarkets. Properties with electric heat and heat pumps receive discounts relative to properties with gas heat. A fireplace is worth between three and eight percent of the value of the property. The estimated coefficients for the ratio of building square feet to lot size are all negative and highly significant. Estimated coefficients for square feet of garage space are statistically significant with the expected influence in each of the submarkets. Corner lots are capitalized into all house prices as a discount. Additional rooms per square foot of building area commands premiums in Norristown and Ambler, but not in the Main Line. The distance of any given property to the Philadelphia CBD is strongly significant for the entire county, but is generally insignificant within any given submarket. However, the distance to the submarket's CBD is strongly significant for all submarkets, thus reflecting the importance of relative global and local effects. Finally, between 1995:1 and 1998:3, house prices have increased 5% for Montgomery County properties. Similarly, prices increased over 11% for the Main Line properties and over 7% for Ambler properties, but house prices have been essentially constant for Norristown homes.

5.3 Spherical Semivariograms for Hedonic Residuals

Table 4 provides the semivariogram parameter estimates for the residuals from the hedonic house price equations for all 21 submarkets. The top half of the table lists semivariogram parameters for the *isotropic* spherical semivariogram while the bottom half of the table presents the spherical semivariogram parameters for two *anisotropic* spherical semivariograms.

<Table 4 About Here>

Isotropic and *anisotropic* estimates of the nugget are statistically significant in each of the 21 submarkets. *Isotropic* estimates of θ_1 and θ_2 (the range) are statistically significant in 9 submarkets while *anisotropic* estimates of these parameters are significant in 11 submarkets. For 5 of these 11 submarkets, the estimated parameters for θ_1 and θ_2 are statistically in only one direction. In Pottstown, New Hanover, Bryn Athyn, and the Main Line (West) submarket, hedonic house price equation residuals are spatially correlated only for properties separated in a north-south direction. For Cheltenham, hedonic house price equation residuals are spatially correlated only for properties separated by an east-west direction.

The range of the *anisotropic* spatially autocorrelated residuals also varies for several submarkets that exhibit both north-south and east-west spatial autocorrelation. For example, the range of spatial autocorrelation in the Main Line (East) submarket for properties separated in a north-south direction is over one kilometer further than the range for properties separated by an east-west direction.

6. Conclusion

This paper examines *anisotropic* spatial autocorrelation in house prices and in hedonic house price equation residuals for twenty-one housing submarkets in suburban Philadelphia. We find that both house prices and hedonic house price equation residuals are spatially autocorrelated and the spatial autocorrelation changes with the direction separating properties for some submarkets.

Since most real estate development tends to spread outwards from existing city centers along major transportation arteries, it is reasonable to expect that the direction of greatest spatial

autocorrelation occurs in the direction of the CBD. Our empirical results are consistent with this hypothesis. For those submarkets where anisotropy obtains, several are inner-ring suburban bedroom communities that are close to Philadelphia. For two of the submarkets (Bryn Athyn, Cheltenham), the direction of spatial autocorrelation is directly towards the Philadelphia City Center. For a third submarket (Main Line West), the direction of anisotropy is towards a major transportation artery connecting the Philadelphia CBD with the County CBD, both of which are major employment centers for the residents of this submarket. For two other submarkets on the rural western edge of the county (Pottstown, New Hanover), the direction of spatial autocorrelation is greatest towards the Pottstown city center, which is the local CBD for this area of the county.

Especially interesting is the characterization of local versus global *anisotropic* effects. While the global pattern of development is outward from the Philadelphia CBD, local development also spreads away from local CBDs such as the county and submarket CBDs. The direction anisotropy is influenced by which CBD is the primary shopping and employment center for submarket residents. For bedroom communities of Philadelphia commuters, anisotropy obtains in the direction of the Philadelphia CBD. For submarkets where residents are equally likely to commute to either the MSA CBD or county CBD, anisotropy obtains in the direction of the transportation artery connecting these two CBDs. Finally, for the more rural and isolated submarkets where the local CBD is the primary retail and employment district, anisotropy obtains in the direction of this CBD, unaffected by the larger spread of development outward from Philadelphia. Further investigation into the estimation of the local v. global 'layers' of anisotropy is an area for further research.

It is our ultimate goal to evaluate the benefits of explicitly modeling anisotropic spatial dependence in the residuals of a hedonic estimation. However, we leave this to a future paper. The goal of this paper was first to identify if anisotropy obtains in the residuals, given a well-specified hedonic estimation. We leave the evaluation of the benefits obtained from explicitly modeling anisotropic spatial dependence in the residuals to a future paper. Future research will also perhaps allow the hedonic specification to vary in order to maximally exploit the presence of anisotropically autocorrelated residuals in the prediction of house values.

References

- Anselin, Luc. 1978. *Spatial Econometrics: Methods and Models* Dordrecht: Kluwer Academic.
- Bailey, Trevor C. and Anthony C. Gatrell. 1995. *Interactive Spatial Data Analysis* Addison Wesley Longman Limited. Essex, England.
- Bailey, M., R. Muth and H. Nourse. 1963. "A Regression Method for Real Estate Price Index Construction" *Journal of the American Statistical Association* 58 (304):933-942.
- Basu, Sabyasachi and Thomas Thibodeau. 1998. "Analysis of Spatial Autocorrelation in House Prices" *Journal of Real Estate Finance and Economics* 17 (1):61-85.
- Can, Ayse. 1992. "Specification and Estimation of Hedonic Housing Price Models" *Regional Science and Urban Economics* 22:453-477.
- Case, Bradford and John Quigley. 1991. "The Dynamics of Real Estate Prices" *Review of Economics and Statistics* 83:50-58.
- Case, K. and R. Shiller. 1987. "Prices of Single Family Homes Since 1970: New Indexes for Four Cities" National Bureau of Economic Research Working Paper No. 2393. Cambridge, MA.
- _____. 1989. "The Efficiency of the Market for Single-Family Homes" *American Economic Review* 79 (1):125-137.
- Clapp, J. and C. Giaccotto. 1992. "Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods" *Journal of the American Statistical Association* 87:300-306.
- Cliff, Andrew D. and J. K. Ord. 1973. *Spatial Autocorrelation* London: Pion Limited.
- Cressie, Noel A. 1993. *Statistics for Spatial Data* New York: John Wiley and Sons.
- Dubin, Robin A. 1988. "Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms" *Review of Economics and Statistics* 70:466-474.
- Dubin, Robin A., R. Kelley Pace and Thomas G. Thibodeau. 1998. "Spatial Autoregression Techniques for Real Estate Data" *The Journal of Real Estate Literature*
- Gillingham, Robert. 1975. "Place to Place Rent Comparisons" *Annals of Economic and Social Measurement* 4(1):153-174.

- Goodman, Allen C. 1978. "Hedonic Prices, Price Indices, and Housing Markets" *Journal of Urban Economics* 5(4):471-484.
- Hill, R. C., J.R. Knight and C.F. Sirmans. 1997. "Estimating Capital Asset Price Indexes" *The Review of Economics and Statistics* 79 (2):226-233.
- Isaaks, Edward H., and R. Mohan Srivastava. 1989. *An Introduction to Applied Geostatistics*. New York: Oxford University Press.
- Journel, A. G. and C. J. Huijbregts. 1978. *Mining Geostatistics* London: Academic Press.
- Maddala, G. S. 1977. *Econometrics*. New York: McGraw Hill.
- Mardia, K. V. and R. J. Marshall. 1984. "Maximum Likelihood Estimation Of Models for Residual Covariance in Spatial Regression" *Biometrika* 71:135-146.
- Matheron, G. 1963. "Principles of Geostatistics" *Economic Geology* 58:1246-1266.
- Oden, N. L. and R. R. Sokal. 1986. "Directional Autocorrelation: An Extension of Spatial Correlograms in Two Dimensions" *Systematic Zoology* 35:608-617.
- Pace, R. Kelley, and O.W. Gilley. 1997. "Using the Spatial Configuration of the Data to Improve Estimation" *Journal of the Real Estate Finance and Economics* 14 (3):333-340.
- Quigley, John. 1995. "A Simple Hybrid Model for Estimating Real Estate Indexes" *Journal of Housing Economics* 4(1):1-12.
- Ripley, Brian. 1981. *Spatial Statistics* New York: John Wiley and Sons.
- Simon, Gary. 1997. "An Angular Version of Spatial Correlations, with Exact Significance Tests" *Geographical Analysis* 29 (3):267-278.
- Thibodeau, Thomas G. 1989. "Housing Price Indexes from the 1974-83 SMSA Annual Housing Surveys" *Journal of the American Real Estate and Urban Association* 17:100-117.
- _____. 1990. "Estimating the Effect of High Rise Office Buildings on Residential Property Values" *Land Economics* 66 (4):402-408.
- _____. 1992. *Residential Real Estate Prices from the 1974-1983 Standard Metropolitan Statistical Area American Housing Survey* Mount Pleasant, Michigan: Blackstone Books: Studies in Urban and Resource Economics.
- _____. 1996. "House Price Indices from the 1984-1992 MSA American Housing Surveys" *Journal of Housing Research* 6(3):439-481.