Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach

Article *in* Real Estate Economics · February 2003

DOI: 10.1046/j.1080-8620.2003.00079.x · Source: RePEc

3 authors:

Tim Fik
University of Florida

David C Ling
University of Florida

Gordon Mulligan
The University of Arizona

2003 V31 4: pp. 623–646
REAL ESTATE
ECONOMICS

Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach

Timothy J. Fik,* David C. Ling** and Gordon F. Mulligan***

The absolute location of each real estate parcel in an urban housing market has a unique location-value signature. Accessibility indices, distant gradients and locational dummies cannot fully account for the influence of absolute location on the market price of housing because there are an indeterminable number of externalities (local and nonlocal) influencing a given property at a given location. Furthermore, the degree to which externalities affect real estate values is not only unique at each location but highly variable over space. Hence, absolute location must be viewed as interactive with other determinants of housing value. We present an interactive variables approach and test its ability to explain price variations in an urban residential housing market. The statistical evidence suggests that the value of location, as embodied in the selling price of housing units, may not be separable from other determinants of value. It is recommended that housing valuation models, therefore, be specified to allow site, structural and other independent attributes to interact with absolute location—$\{x, y\}$ coordinates—when accounting for intraurban variation in the market price of residential housing. This approach is especially useful when estimating the value of housing for geographic areas where very little is known a priori about the neighborhoods or submarkets.

Determining the underlying value of location in urban land and housing markets is inherently problematic because real estate values are simultaneously affected by a variety of site, structural and locational attributes. The most influential site and structural attributes typically are observable and can be easily included in house price regressions. Information about a property's location attributes is much more difficult to observe and quantify because numerous external effects (positive and negative) act upon a parcel of land at a given location. Moreover, these effects are reflected in a parcel's value.

We refer to the composite external effect at a given point in space as the *location-value signature* (*LVS*) of that parcel. Each parcel's relative location, and

*Department of Geography, University of Florida, Gainesville, FL 32611 or fik@geog.ufl.edu.
**Department of Finance, Insurance and Real Estate, University of Florida, Gainesville, FL 32611 or ling@ufl.edu.
***Department of Geography & Regional Development, University of Arizona, Tucson, AZ 85721 or mulligan@email.arizona.edu.

therefore its *LVS*, is unique with respect to influential nodes, districts, axes, corridors, competitive land uses and externalities found within an urban system. Thus, a parcel's *LVS* plays a role in the determination of the market value of a house residing on that parcel. This paper develops a hedonic pricing model that allows a parcel's unique *LVS* to be estimated.

Our proposed model is based on two fundamental assumptions. First, because locational characteristics are capitalized into the price of land, models for improved properties that incorporate location variables as additive terms, as is standard practice, are likely misspecified because the price of land (*i.e.*, the coefficient on lot size) is held spatially constant across the urban area. Other structural attributes, such as age and size of dwelling unit, also often display identifiable spatial patterns. The "spatial regularities" of the patterns strongly suggest that the hedonic price of the attributes also vary spatially. Thus, a more appropriate specification is to interact location variables with variables such as lot size, age and dwelling size. Unlike additive specifications, interactive specifications allow these attribute prices to vary spatially.

Second, although the search for alternative modeling frameworks has produced a rich literature on hedonic pricing models, few have directly incorporated parcel-specific locational information in a nonarbitrary form—such as the use of Cartesian $\{x, y\}$ coordinates.[1] Yet, there is a need to explore the full potential of these frameworks given that (1) no two parcels or properties occupy the same location in space, and (2) housing, land, locational accessibility and composite external effects are jointly purchased.

The inclusion of $\{x,y\}$ coordinates in an interactive capacity should have a significant impact on a model's ability to explain house price variation. To test this assertion, a series of models is estimated using a semi-log specification (specifying the logarithm of selling price as a function of various site and location variables). We first estimate a fully interactive model that contains no location variables. This "aspatial model" serves as a benchmark for subsequent models. In our second specification, all property characteristics variables, including lot size, are interacted with a set of discrete location dummy variables. This specification is theoretically preferable to an additive specification because it allows characteristics prices to vary spatially. Unfortunately, however, this specification is itself limited because land prices are averaged over discrete geographic boundaries. This is a concern given the theoretical support for smoothed, *continuous*, spatial patterns in land prices.

In our third interactive specification, lot size, age and dwelling size are interacted with a polynomial expansion of the property's $\{x, y\}$ coordinates to control for

[1] Recent notable exceptions include Clapp (2001) and Pavlov (2000).

the absolute location of the single-family home. The discrete location dummy variables are excluded from this specification. If we differentiate price with respect to location x (or y), we obtain the (geographic) slope of house price in the vicinity of some location $\{x, y\}$, taken along the x (or y) axis. Thus, it is possible to simulate the "price surface" for the city. On the other hand, if we differentiate price with respect to an attribute (say dwelling size), we obtain a location-specific rate of change in price with respect to a unit change in the attribute (*e.g.*, one extra square foot of size). A consequence of this is that the implicit prices for attributes are not globally constant throughout the city; instead, these implicit attribute prices vary continuously throughout, rising in some locations and falling in others.

The use of absolute location in hedonic price modeling is especially attractive when the researcher does not have adequate prior information on the boundaries of "neighborhoods" within the city. It is widely understood, for example, that zip codes and even census tracts can contain housing that is strikingly heterogeneous. More precise neighborhood delineation usually requires the help and expertise of knowledgeable local market participants, such as the property tax assessor and residential realtors. Even then, however, discrete definitions of neighborhoods may mask significant value discontinuities. For example, perhaps numerous households in an otherwise similar neighborhood back up to a tree-filled conservation area or park, while other homes have an unobstructed view of above ground power lines. These positive and negative externalities are averaged out when location dummy variables are employed, even when neighborhoods are well defined by those possessing local knowledge. However, these externalities, as well as other influential streets, corridors and land uses, are detected and controlled for with the use of $\{x, y\}$ coordinates (assuming, of course, that an adequate number of sales have occurred). This is a significant advantage to modelers who do not possess local market knowledge and who do not have the time or resources to invest in acquiring this information (such as those applying automated valuation system, or AVMs). This would include mortgage lenders attempting to value the properties that collateralize a portfolio of mortgages originated in numerous cities and states.

In our fourth (and preferred) specification, we include *both* absolute location and a set of location dummy variables. Unlike our third specification, this approach does not assume a single location-value price surface. Rather, this approach allows for multiple location-value surfaces to be specified based on prior knowledge of distinct submarkets.

Two recent articles have incorporated absolute location in the form of $\{x, y\}$ coordinates. Clapp (2001) develops a semi-parametric hedonic approach for estimating house value at any point in space. Similar to our approach, a significant strength of his model is that it does not require prior knowledge of neighborhood boundaries. A limitation of his model, however, is that the coefficients on other housing characteristics are held constant over space. Pavlov (2000) uses $\{x, y\}$ coordinates and allows the coefficients of the model to take on various functional forms (linear and quadratic), thus allowing characteristic variables to vary over space. Unlike Pavlov (2000), however, our preferred specification combines the use of $\{x, y\}$ coordinates and location dummies in a single parametric structure. In addition, our models are estimated using higher-ordered polynomials with all possible interactive combinations examined and tested using a stepwise regression procedure.

Our results can be summarized as follows. First, interactive models that incorporate absolute location outperform the aspatial or discrete-spatial models. Second, although the use of both discrete location dummies and absolute location $\{x, y\}$ produces the most accurate predictions of selling price, models that rely solely on $\{x, y\}$ coordinates performed extremely well. This finding supports the contention that detailed knowledge of an urban real estate market (and its various neighborhoods and submarkets) is not necessary when estimating market value. The use of Cartesian $\{x, y\}$ coordinates in an interactive model is a quick and inexpensive way to obtain reasonable estimates of urban residential housing values.

Before proceeding with the specification of our variable interaction model, we briefly review in the next section the noninteractive hedonic approaches commonly found in the house valuation literature. The third section describes our variable interaction model, and the fourth presents a description of our test market and data. The fifth section describes our empirical models and discusses the results, while the final section contains some concluding comments.

Noninteractive Hedonic Pricing Models

Hedonic house price equations typically regress price on structural characteristics (such as square footage and age) and location attributes. Location variables are of two general types: those that explicitly depict a spatial pattern (*e.g.*, distance gradients from a central urban node or potential externality) and those that indicate discrete differences in locations. In terms of the latter, dummy variables that denote parcel location by neighborhood, submarket or other areal designation have been commonly employed in hedonic models.

Location Dummies

The use of location dummies operates under the assumption that parcels in a given locality (neighborhood, census tract, zip code area, etc.) are subjected to the same general localized conditions and, therefore, share the same situational characteristics or externalities (including the effects of pollution, congestion, crime, environmental quality and amenities). This assumption, however, is limiting for at least two reasons: (1) the size of these areal units may vary dramatically over space, and (2) the processes which determine the intensity and impact of specific external effects can be highly variable even over short distances. Moreover, a property on the edge of a given areal unit may have more in common with properties in an adjoining unit/neighborhood than with properties in the same areal unit.

Goodman and Thibodeau (1998) develop a hierarchical modeling technique to identify housing submarket boundaries. The hierarchical model assumes that all homes within a spatially concentrated area share amenities associated with the property's location. Consequently, the housing characteristics that determine a property's market value are nested in a hierarchy—properties within neighborhoods, neighborhoods within school zones, school zones within municipalities and so on. Goodman and Thibodeau (2003) extend their earlier analysis by comparing the prediction accuracy of their hierarchical model for determining submarkets to the prediction accuracy of hedonic models that use zip code districts and census tracts to delineate submarkets. The authors find that their approach produces similar results, in terms of predictive accuracy, to zip code and census tract delineations of submarkets, but with the benefit of reduced variance. Note that Goodman and Thibodeau's (1998) approach derives, rather than imposes, a set of housing submarkets for a metropolitan area. This is similar, at least in part, to our approach in that we allow the data to characterize and measure the value of a parcel's location.

Distance Gradients

In contrast to areal units, distance gradients express an explicit spatial pattern. In fact, the use of a distance gradient *forces* the estimation of a uniform slope in all directions from the node. However, the effects of employment centers and externalities on property values may vary with direction from a given node. For example, property values may decrease (or increase) differently to the east of an urban node than in other directions. In addition, many spatial effects on property prices have an influence over only small geographic areas, suggesting the effects may diminish within relatively short distances (Palmquist 1992). In short, distance-based measures, with respect to a predetermined and limited number of locations or nodes, are inadequate representations of composite external effects as the importance or influence of any one given location or node to households will vary over space.[2]

[2] See Gillen, Thibodeau and Wachter (2001) for an extended discussion of directional spatial autocorrelation.

The Externality Literature

The "externality" literature shares some commonalities with the house price literature, such as the reliance on distance gradients, dummy variables or a combination of the two, to depict the locational characteristics of properties in reference to external factors. More sophisticated externality studies have included "accessibility" indices to measure employment potential or the influence of variations in population size. These types of approaches have incorporated various additive measures of locational differences, commonly specified as

$$P_i = \mathrm{f}(\boldsymbol{S}_i, L_i, D_i, LD_i, A_i), \tag{1}$$

where P_i is the transaction price of property i, $\boldsymbol{S}_i$ represents a vector of structural characteristics (*e.g.*, age, square footage, number of bedrooms and baths, etc.), L_i represents lot size, D_i represents the distance from property i to a central urban node or externality, LD_i represents a locational dummy variable to identify and differentiate parcel location by neighborhood and A_i is a locational accessibility measure.

Locational accessibility is commonly specified as a gravity-type index; namely,

$$A_i = \sum_{j=1}^{j} W_j / d_{ij}^{\lambda}, \tag{2}$$

in which W_j represents the proportional size of $j = 1, \ldots, J$ population or employment centers, d_{ij}^{λ} represents the distance between property i and center j and λ is an appropriate distance-decay exponent. Note, however, that the use of this type of measure raises several concerns. First, the functional form of this index is not easily justified, as other equally plausible forms do exist, and, second, this specification only describes the "potential" for interaction (Hansen 1959, Ingram 1971, Weibull 1976, 1980). Note also that the set of J population or employment enters and the value of λ must be predetermined, adding unnecessary complications.

Employing differing versions of Equation (1), researchers have produced evidence that land and housing values are significantly affected by local environmental quality (Kiel 1995, Kiel and McClain 1995), by neighboring nonresidential land uses (Dewees 1976, Li and Brown 1980, Nelson 1980) and by proximity to negative externalities (Guy, Hysom and Ruth 1985, Kohlhase 1991, Michaels and Smith 1990, Reichert, Small and Mohanty 1992).

Implications of Noninteractive Models and Spatial Statistics

From a strictly theoretical perspective, the standard additive hedonic price model assumes buyers evaluate structural characteristics and location characteristics separately when they purchase a home. Consequently, location represents a premium that is *added* to the price of the home—independent of the level or value of other structural and locational characteristics.[3] However, this assumption is inconsistent with established urban theory that posits the marginal value of housing characteristics vary spatially within the city. Thus, additive hedonic models fail in part to capture the full spatial properties of the housing data.

In recent years, especially with the advent of Geographic Information System (GIS) software packages, a variety of spatial statistical methods have been developed to help uncover localized trends in housing prices. Examples include simultaneous auto-regression (SAR), conditional auto-regression (CAR) and generalized least squares based upon an estimate of the covariance matrix (kriging).[4] Certainly, the use of proxies such as the average price of housing in surrounding areas or a spatial moving average of price may help to uncover localized trends in prices. Moreover, the use of spatial estimators may eliminate the problem of spatially autocorrelated error terms and thereby allow correct statistical inferences to be drawn regarding site and structural variables.[5]

Spatial estimators, however, may pose computational problems for large data sets, and they do not permit the estimation of a given parcel's *LVS*. Said differently, by using the error structure, spatial estimators attempt to *control* for each parcel's unobserved *LVS* that is not explicitly modeled in the hedonic regression. However, such estimators do not allow us to produce unit-specific estimates of location value and attribute-specific contributions to that value. The variable interaction approach includes *unit-specific* information on location and site attributes on the right-hand side of the hedonic house price equation.

---

[3] That is, "additive" regression models assume that house value is determined, in part, by X_1 *plus* X_2, not X_1 *times* X_2. If X_1 is lot size and X_2 is a location dummy variable, then an additive specification does not allow the impact of lot size on value to depend on location (or any other variable). With an interactive specification, lot size is multiplied by the location variable, thereby allowing the effects of lot size on house value to vary with location.

[4] See, for example, Pace and Barry (1997), Pace and Gilley (1997), Basu and Thibodeau (1998), Pace *et al.* (1998) and Dubin, Pace and Thibodeau (1999).

[5] In typical hedonic specifications, the spatial dependence of site and structural attributes, coupled with incomplete information, make spatial dependence of the regression residuals almost inevitable (Dubin, Pace and Thibodeau 1999). Ignoring spatial autocorrelation leads to a serious violation of the assumptions underlying ordinary least squares regression. In particular, OLS parameter estimates will be inefficient and will produce incorrect confidence intervals for estimated parameters and predicted values. In addition, standard tests of statistical significance will yield inaccurate conclusions. For an extensive review of the issues involved in estimating models with spatially autocorrelated error terms, see Dubin (1998).

Interactive Variables Approach Using $\{x, y\}$ Coordinates

Despite the importance of both absolute and relative location, few hedonic studies have utilized geocoded measures of locational accessibility. An early exception was Jackson's (1979) study of the Milwaukee housing market. Jackson uses census tract rents to derive a continuous measure of location value over space, while isolating the influence of locational accessibility on the price of land and housing. This was accomplished by specifying lot size as an interactive variable within a polynomial expression of location (as defined in terms of a unit's Cartesian $\{x, y\}$ coordinates). Effectively incorporating *all* general locational effects within one model, Jackson specifies

$$P_i = \mathrm{f}(\boldsymbol{S}_i, [L_i * A_i]), \tag{3}$$

where accessibility $A_i = g_i^j\{x_i, y_i\}$ is defined for location $\{x_i, y_i\}$ and $g_i^j\{x_i, y_i\}$ is a jth order polynomial expansion of $\{x, y\}$ coordinates, which are linked "interactively" to lot size (L_i). The pth order polynomial for the interactive terms $[L_i * A_i]$ may be expressed as

$$\Omega_i = \alpha + \sum_{j=0}^{p}\sum_{k=0}^{p} \beta_{jk}\big[L_i * \{x_i^j * y_j^k\}\big], \tag{4}$$

where α is the regression constant, the $\beta_{j,k}$s are the corresponding regression (slope) coefficients for the interactive terms included in the model, under the constraint that $j + k \leq p$. Note that the number of variables (m) in A_i increases dramatically with polynomial order. An increase from p to p', where $p' = p + 1$, would add $p' + 1$ additional terms to the model. A third-order polynomial expansion would produce $m = 9$ location variables ($x, x^2, x^3, y, y^2, y^3, x * y, x^2 * y$ and $x * y^2$), whereas a fourth-order expansion would yield $m = 14$ location variables. If the order of the function A_i were known, then the level of accessibility at a given location could be evaluated with respect to the spatial distribution of prominent nodes and localized externalities (employment centers, retail shopping outlets, landfills, hazardous sites, etc.) through the estimated coefficients associated with the m accessibility variables. Note that a pth order polynomial expression will account for $(p - 1)^2$ extrema (*i.e.*, structural shifts in the *LVS*) over space.

Individually, the estimated coefficients on the interaction of lot size and the polynomial expansion of $\{x_i, y_i\}$ are uninterpretable. When considered collectively, however, the coefficients for the expression $[L_i * A_i]$ yield an estimate of the unit price of land at any given location $\{x_i, y_i\}$. If only one value surface exists in the city, then the estimated regression coefficients (along with a specified lot size) can be used to estimate $\boldsymbol{\Omega}_i$. The value of $\boldsymbol{\Omega}_i$ will vary over space. In theory, the *LVS* of any ith parcel may be approximated by subtracting $\boldsymbol{\Omega}_i$ from the average observed $\boldsymbol{\Omega}$ value. Equation (4) reveals that structural characteristics ($\boldsymbol{S}_i$) are considered spatially constant, while lot size interacts with a parcel's absolute location. This formulation allows the price of land to vary spatially in a continuous manner. Given that each parcel has a unique set of $\{x, y\}$ coordinates, its corresponding location-value signature must also be unique. In addition, the interactive relationship between lot size and location reduces the likelihood of encountering multicollinear relations among the explanatory variables in the model—a condition that is common to trend surface models using higher-order polynomials (see Johnston 1978).

Overview of Test Market and Data

The city of Tucson was one of America's fastest growing cities during the 1980s and 1990s. Since 1980, the population of the city proper has grown from 330,000 to 487,000, while the metropolitan area has increased from 530,000 to 850,000 persons. Home to the University of Arizona—its largest single employer—the city also boasts a large and vibrant manufacturing and export base. Employment growth in high-technology sectors has been most impressive, with the number of high-tech jobs doubling over the past decade. The expansion of the manufacturing sector and linked industry has been noncentralized, with firms and supporting industry locating throughout the Tucson metropolitan area. The proliferation of industry and employment opportunities, along with a booming recreation and tourist industry and a sizable retirement community, have contributed greatly to the growth of Tucson's economy. This growth has also translated into a highly active urban housing market in terms of both the construction of new units and the sale of existing stock.

Although Tucson has a relatively weak central business district in comparison to most U.S. cities, it does display similar patterns of growth at the urban fringe. Older housing stock is typically located in the central portions of the city, whereas newer housing is typically found scattered around the city's peripheral locations. Analysis of the housing stock reveals a conelike age gradient that declines with distance from the city center. Tucson's housing market continues to expand rapidly to the north. The sprawling urban environment and transportation system reinforces the city's reliance on private transportation. In general, Tucson's housing-price gradient rises from low-valued housing in the south to high-valued housing as one moves north (as defined in terms of selling price per square foot). The housing-price gradient is less pronounced along the east-west corridor of the city, which is mostly associated with average-sized and average-price housing.

The housing data analyzed in this paper are drawn from a series of nine contiguous Multiple Listing Service (MLS) divisions that, as a whole, closely approximate Tucson's urbanized area in its geographical extent. These real estate divisions, which are perceived by real estate agents as being different

Table 1 ■ Selected descriptive statistics for residential housing units sold in 1998 in Tucson, AZ; $n = 2{,}971$ observations.

Variable (units)	Mean	Std. Dev.	Minimum	Maximum
SPRICE ($)	127,187	64,814	25,000	800,000
LOT (sq ft)	11,200	8,952	1,526	43,542
AGE	22.4	14.9	0	50
SQFT (sq ft)	1,717	520	1,000	4,400
X (x-coordinate)	807.25	29.43	717.01	863.72
Y (y-coordinate)	462.55	28.88	403.75	537.04

urban housing submarkets, are identified by dummy variables in the analysis that follows.[6]

An initial database was assembled for 3,703 geocoded housing units that were sold during 1998 in the nine MLS districts. From this we deleted observations with more than one acre of land, that were larger than 4,500 square feet or that exceeded 50 years of age. This left us with a final sample of 2,971 sales observations.

Presumably, most analysts fail to include lot size as an explanatory variable because the data are not available. Fortunately, the database contains unit-specific information on lot size (*LOT*), as well as age of dwelling in years (*AGE*) and square footage (*SQFT*). The absolute locations of housing units are represented by Cartesian $\{x, y\}$ coordinates, derived from latitude and longitude, referenced against the southwestern-most observation: $\{x_{min}, y_{min}\}$. Selected descriptive statistics are summarized in Table 1. The average selling price is $127,187 with a range of $25,000 to $800,000. Mean lot size and house size are 11,200 sq. ft. and 1,717 sq. ft., respectively. We view the significant variation in lot size in our sample as a strength of our data set. Lot size is certainly spatially autocorrelated. Thus, the inclusion of variable lot sizes provides valuable spatial information when interacted with locational attributes. The average age is 22 years with a standard deviation of 15 years. Overall, the sample contains significant heterogeneity in housing characteristics.

Empirical Models and Estimates

As a benchmark, we first estimate the following aspatial, but fully interactive, semi-log house price model:

[6] For additional information on the Tucson land market and a map of the realty divisions, see Mulligan, Franklin and Esparza (2002).

$$\ln P_i = \sigma_0 + \sum_{j=1}^{m} \alpha_j S_{j,i} + \sum_{j=1}^{m} \sum_{k=1}^{m} \Phi_{j,k} \left(S_{j,i}^{v} S_{k,i}^{w} \right) + \varepsilon_i \tag{5}$$

for $i = 1, \ldots, n$ observations, $j = 1, \ldots, m$ structural or site attributes $\boldsymbol{S}$ (including lot size) and ε is a random normal error term. Note that $k > j$ for interactive terms $(S_{j,i}S_{k,i})$ and $v + w \leq p$ (where p is the implied order of the model). This specification (Model 1) is consistent with the approach taken by Grether and Mieszkowski (1974) who interacted lot size with numerous neighborhood characteristics. A theoretical justification for our interactive model is contained in Parsons (1990).[7]

Equation (5) is estimated with a stepwise regression procedure. Stepwise regression is used to reduce the likelihood of inflated standard errors due to multicollinearity, a problem that is commonly observed in the estimation of spatial trend surface models (Johnston 1978). A stepwise introduction of variables to the model has the added advantage of ensuring that redundancy is minimized (by blocking the inclusion of collinear regressors). Only those variables offering significant explanatory power are retained and included in the final model. Stepwise regression eliminates the possibility of having collinear variables test insignificant when they are actually adding significant explanatory power to the model on an individual basis; a feature that is desirable to avoid estimation problems. It also simplifies the process of making cross-model comparisons. In addition, as several of the models contain a potentially large number of independent variables (particularly the models involving higher-ordered polynomial expressions and interactive combinations thereof), the use of stepwise regression sidesteps the issue of simply increasing model performance through the inclusion of more independent variables, thus avoiding spurious explanation.

Initial runs of each model, which included all of the interactions, were tested to identify an acceptable polynomial or model order—one that provided the greatest number of statistically significant coefficients and the highest adjusted R-square value. Sensitivity analyses of model order suggested the use of a fourth-order model as a maximum; that is, where the sum of the implied exponents of variable interactions is less than or equal to 4.

The results of the "aspatial model" are presented in Table 2. This model produced six explanatory variables: *SQFT*, *AGE*, *LOT*, AGE^2, LOT^2 and $SQFT^2$, which accounted for much of the variation in the selling prices of the 2,971 units

[7] Predicted prices recovered from taking $e^{[predicted \ln(price)]}$ may produce biased predictions for the expected transaction price (see Goldberger 1968 and Thibodeau 1989). This issue is addressed below.

Table 2 ■ Model 1: Interactive aspatial model; Dependent variable = Log of selling price.

Independent Variables	Estimated Coefficient	Standard Error	Standardized Coefficient	t-Stat.	Significance
Constant	10.668	0.036		291.98	0.000
SQFT	6.827E-04	3.56E-05	0.875	19.1	0.000
AGE	−2.196E-02	9.92E-04	−0.806	−22.1	0.000
LOT	2.841E-05	2.17E-06	0.627	13.1	0.000
AGE^2	3.131E-04	1.94E-05	0.576	16.1	0.000
LOT^2	−4.013E-10	4.87E-11	−0.382	−8.2	0.000
$SQFT^2$	−4.703E-08	8.01E-09	−0.252	−5.5	0.000

Parcel characteristics are lot size (*LOT*), age of dwelling in years (*AGE*) and square footage of improved space (*SQFT*).
Sample size = 2,971, adjusted R-square = 0.776, standard error of estimate = 0.1923121, degrees of freedom = 2,964, F = 1,711 (significance 0.000) and the inclusion/exclusion criteria are $p \leq 0.10$, $p > 0.10$. To test for the spatial independence of the estimated error terms, the residuals were regressed (using stepwise regression) on the independent variables. The following coefficient estimates are all significant at the 0.10 significance level: $x * y$, y^2, y^3, x^2, $y^2 * x$ and x. Thus, the null hypothesis of spatial independence in error terms is rejected.

sold during the study period. The model had an adjusted R-square of 0.776 and a standard error equal to 0.192. Note, however, that a trend-surface analysis of the error terms from this model revealed evidence of spatial autocorrelation. The variation in error was largely explained by a series of location variables, namely, $(x * y)$, y^2, x^2, $(y^2 * x)$ and x in descending order of significance (at the 90% confidence or higher). Violating the assumption of an independent error structure suggests that the model may be underspecified (*i.e.*, lacking adequate spatial information).

In our second interactive specification, dummy variables were added to control for the location of the property among nine MLS districts. Because interactive specifications can quickly generate a large number of regressors, we sought to minimize the number of location dummies. To accomplish this, we pretested the explanatory power of the nine MLS district dummy variables. Only three MLS districts proved to have significant explanatory power: *NDUM* (the north district), *NEDUM* (the northeast district) and *EDUM* (the eastern end of the east–west corridor district). The *NDUM* and *NEDUM* districts are characterized by up-scale housing associated with high amenity and/or mountain-view locations near or within the foothills region of the city. For the sake of parsimony, the remaining six districts are aggregated into a fourth location designation that serves as the base or default case.

We then estimated the following semi-log price model:

$$\ln P_i = \sigma_0 + \sum_{j=1}^{m} \alpha_j S_{j,i} + \sum_{r=1}^{d} \tau_r D_{r,i} + \sum_{r=1}^{d}\sum_{j=1}^{m} \delta_{r,j}^{q} (D_{r,i} S_{j,i})$$

$$+ \sum_{j=1}^{m}\sum_{k=1}^{m} \Phi_{j,k}\left(S_{j,i}^{v} S_{k,i}^{w}\right) + \sum_{d=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{m} \phi_{r,j,k}\left[D_{d,i}\left(S_{j,i}^{v} S_{k,i}^{w}\right)\right] + \varepsilon_i, \quad (6)$$

where $d = 1, \ldots, 3$ location dummies $\boldsymbol{D}$ (for four separate submarkets). As previously discussed, this specification is preferable to an additive specification because it allows the price of land and other attributes to vary across locations. However, a limitation is that it averages land prices over discrete geographic boundaries—in this case over our four MLS districts.

The results from this second regression estimation are presented in Table 3. The F-statistic indicates the model has significant explanatory power. This is also

Table 3 ■ Model 2: Contains dummy variables for MLS districts; Dependent variable = Log of selling price.

Independent Variables	Estimated Coefficient	Standard Error	Standardized Coefficient	t-Stat.	Significance
Constant	10.822	0.061		178.7	0.000
SQFT	5.701E-04	5.78E-05	0.731	9.8	0.000
SQFT^2	−3.450E-08	1.30E-08	−0.185	−2.6	0.008
AGE	−3.069E-02	2.56E-03	−1.126	−12.0	0.000
AGE^2	3.676E-04	1.89E-05	0.676	19.4	0.000
AGE ∗ *SQFT*	6.032E-06	2.32E-06	0.378	2.6	0.009
AGE ∗ *SQFT*^{2}	−1.170E-09	5.71E-10	−0.171	−2.0	0.002
LOT	2.646E-05	2.03E-06	0.584	13.0	0.000
LOT^2	−4.261E-10	4.55E-11	−0.406	−9.4	0.000
NDUM	0.311	0.016	0.188	20.1	0.000
EDUM	0.252	0.043	0.219	5.9	0.000
EDUM ∗ *SQFT*	−1.217E-04	2.46E-05	−0.184	−4.9	0.000
NEDUM	0.154	0.014	0.098	10.9	0.000

Parcel characteristics are lot size (*LOT*), age of dwelling in years (*AGE*) and square footage of improved space (*SQFT*). Location dummies are *NDUM* (the north MLS district), *NEDUM* (the northeast MLS district) and *EDUM* (the east MLS district that runs across the eastern portion of the east–west corridor). Sample size = 2,971, adjusted R-square = 0.807, standard error of estimate = 0.1781, degrees of freedom = 2,958, $F = 1{,}039$ (significance 0.000) and the inclusion/exclusion criteria are $p \leq 0.10$, $p > 0.10$. To test for the spatial independence of the estimated error terms, the residuals were regressed (using stepwise regression) on the independent variables. The following coefficient estimates are all significant at the 0.10 significance level: $x * y$, y^2, y^3, x^2, $y^2 * x$ and x. Thus, the null hypothesis of spatial independence in error terms is rejected.

indicated by an adjusted R-squared of 0.807 and a relatively low standard error (0.178).

Seven site and structural variables are found to be significant: $SQFT$, $SQFT^2$, AGE, AGE^2, $(AGE * SQFT)$, $(AGE * SQFT^2)$, LOT and LOT^2, along with one interactive site-location variable: $(EDUM * SQFT)$ (slope-shifting effect). In addition, all three location dummies ($NDUM$, $EDUM$ and $NEDUM$) are found to have a positive and significant intercept-shifting impact on the selling price. Note that the higher premium attached to locating in the east (where $EDUM = 1$) is eroded as square footage goes up (as suggested by the negative coefficient associated with $[EDUM * SQFT]$). Note also that square footage ($SQFT$) is highly interactive with AGE. This suggests that square footage and age are inherently nonseparable as predictors of selling price within the Tucson market.

The results for Model 2 suggest that AGE—a variable that is often used as a proxy for housing quality—is an important interactive variable. It is interesting to note, however, that AGE does not directly "interact" with any location variables. This is because the age of a dwelling may already indicate a housing unit's location within the urban area—as older housing units tend to be more centralized, whereas new units tend to be associated with more peripheral locations.

Despite the inclusion of location dummies in this second model, a trend surface analysis of the residuals reveals a spatial dependence in the model's error structure. Error terms were found to be explained by the following location variables: $(x * y)$, y^3, x^2, $(y^2 * x)$ and x (all significant at the 90% confidence level or higher).

In our third interactive specification, square footage, lot size and age are interacted with a polynomial expansion of the property's $\{x, y\}$ coordinates *instead* of our discrete location dummy variables. More specifically, this semi-log model may be specified as

$$\ln P_i = \sigma_0 + \sum_{j=1}^{m} \alpha_j S_{j,i} + \sum_{j=1}^{m} \sum_{k=1}^{m} \Phi_{j,k}\left(S_{j,i}^{v} S_{k,i}^{W}\right) + \sum_{j=0}^{p} \sum_{k=0}^{p} \beta_{j,k}\left[S\left(x_i^j y_i^k\right)\right] + \varepsilon_i \tag{7}$$

for a specified pth-order interactive polynomial expression and/or pth-order model with unit-specific Cartesian coordinates $\{x_i, y_i\}$ as derived from latitude and longitude measures, where $j + k \leq p$. The inclusion of $\{x, y\}$ coordinates allow structural or site attributes (S) to interact with a unit's absolute location.

Table 4 ■ Model 3: Contains $\{x, y\}$ coordinates; Dependent variable = Log of selling price.

Independent Variables	Estimated Coefficient	Standard Error	Standardized Coefficient	t-Stat.	Significance
Constant	10.926	0.130	—	83.7	0.000
$SQFT * x^2 * y$	9.177E-12	8.39E-13	3.971	10.9	0.000
AGE	−2.682E-02	1.28E-03	−0.984	−20.9	0.000
$LOT * y^3$	−2.550E-12	2.20E-13	−6.326	−11.5	0.000
AGE^2	3.336E-04	1.74E-05	0.613	19.1	0.000
$SQFT^2$	−2.972E-08	8.01E-09	−0.159	−3.5	0.000
LOT^2	−4.253E-10	4.27E-11	−0.405	−9.9	0.000
$LOT * y^2$	1.832E-09	1.45E-10	9.368	12.5	0.000
$LOT * x$	−1.396E-07	1.20E-08	−2.513	−11.7	0.000
$SQFT * y^3$	−5.944E-12	6.26E-13	−1.043	−9.4	0.000
$SQFT * x^2$	−2.571E-09	2.94E-10	−2.281	−8.7	0.000
$AGE * SQFT$	2.470E-06	4.43E-07	0.155	5.5	0.000

Parcel characteristics are lot size (*LOT*), age of dwelling in years (*AGE*), square footage of improved space (*SQFT*) and absolute value $\{x, y\}$ coordinate(s). Sample size = 2,971, adjusted R-square = 0.834, standard error of estimate = 0.1654718, degrees of freedom = 2,958, $F = 1{,}242$ (significance 0.000) and the inclusion/exclusion criteria are $p \leq 0.10, p > 0.10$. To test for the spatial independence of the estimated error terms, the residuals were regressed (using stepwise regression) on the independent variables. None of the coefficient estimates are significant at the 0.10 significance level; thus, the null hypothesis of spatial independence in the error terms cannot be rejected.

Note that $\{\sigma, \alpha, \Phi, \phi$ and $\beta\}$ represents the set of regression coefficients to be estimated and, as before, ε is a random normal error term. This specification is included to assess the relative explanatory power of a model that relies exclusively on absolute location to model the location attributes of a property. The results are reported in Table 4.

Overall, these results suggest that site and location attributes are interrelated; consider the significance of $(SQFT * x^2 * y)$, $(SQFT * y^3)$, $(SQFT * x^2)$ and $(AGE * SQFT)$ (as older houses tend to be more centralized in this market). Once absolute location is accounted for through the inclusion of $\{x, y\}$ coordinates, site attributes tend to enter into the equation in quadratic form (*e.g.*, AGE, AGE^2, $(AGE * SQFT)$, $SQFT^2$ and LOT^2). The model boasts an adjusted R-square of 0.834 and a standard error of 0.165. Moreover, the error terms of the model are found to be spatially independent; that is, no spatial autocorrelation in error is detected when running a trend surface model on the residuals (at a 90% confidence level).

Ideally, it would prove useful to have information on both the absolute and relative location of residential housing units. Building on the work of Jackson

(1979), Johnson and Ragas (1987), Waddell, Berry and Hoch (1993a, 1993b) and Smersh (1996), the model depicted by Equation (7) can be expanded to include a vector of dummy variables to differentiate the various geographic submarkets. These dummy variables may be used to account for local (intra-submarket) variations not easily represented by a single trend surface. Such an approach would prove useful when the various submarkets or neighborhoods are separated by physical barriers or boundaries (*e.g.*, river or highway) and the underlying trend surface has potential "discontinuities." Additional dummy variables may be included to assess the impact of infrastructure, land use patterns, zoning or the existence of specialized districts or corridors. As multiple forces or processes are simultaneously at work in the urban land and housing markets, locational differences in the $\beta_{j,k}$s of Equations (6) and (7) are to be expected. Thus, a more complete interactive variables framework is specified by modeling the log of selling price ($\ln P_i$) as a function of both absolute location (*i.e.*, $\{x, y\}$ coordinates) *and* discrete location dummies. More specifically, the semi-log version may be expressed as

$$\begin{aligned}\ln P_i &= \sigma_0 + \sum_{j=1}^{m} \alpha_j S_{j,i} + \sum_{r=1}^{d} \tau_r D_{r,i} + \sum_{r=1}^{d}\sum_{j=1}^{m} \delta_{r,j}\left(D_{r,i} S_{j,i}^{q}\right) \\ &\quad + \sum_{j=1}^{m}\sum_{k=1}^{m} \Phi_{j,k}\left(S_{j,i}^{v} S_{k,i}^{W}\right) + \sum_{d=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{m} \phi_{r,j,k}\left[D_{d,i}\left(S_{j,i}^{v} S_{k,i}^{W}\right)\right] \\ &\quad + \sum_{j=0}^{p}\sum_{k=0}^{p} \beta_{j,k}\left[S(x_i^j y_i^k)\right] + \sum_{d=1}^{r}\sum_{j=0}^{p}\sum_{k=0}^{p} \Theta_{r,j,k}\left\{D_{d,i}\left[S\left(x_i^j y_i^k\right)\right]\right\} + \varepsilon_i .\end{aligned} \tag{8}$$

In this specification, discontinuities or structural shifts in various portions of a price surface may be explained by interacting submarket dummies with absolute location. In a sense, this approach assumes that "multiple" location-value response surfaces may be present, as an outcome of location, market segmentation, and variability in the nature and scale of "localized externalities." Ultimately, a composite location-value response surface is a byproduct of these overlapping surfaces.

The output from this final model is presented in Table 5. As expected, this model outperforms all prior specifications. As with Model 3, the null hypothesis of spatially independent residuals could not be rejected at the 90% confidence level. The results indicate marked geographic differences in the selling prices of units in terms of square footage, especially in the northern and eastern sections of the city (a finding that is consistent with Model 2). Note that submarket dummies and absolute location interact in several ways as shown by the significance of $NDUM * x$ and $NDUM * x * y$.

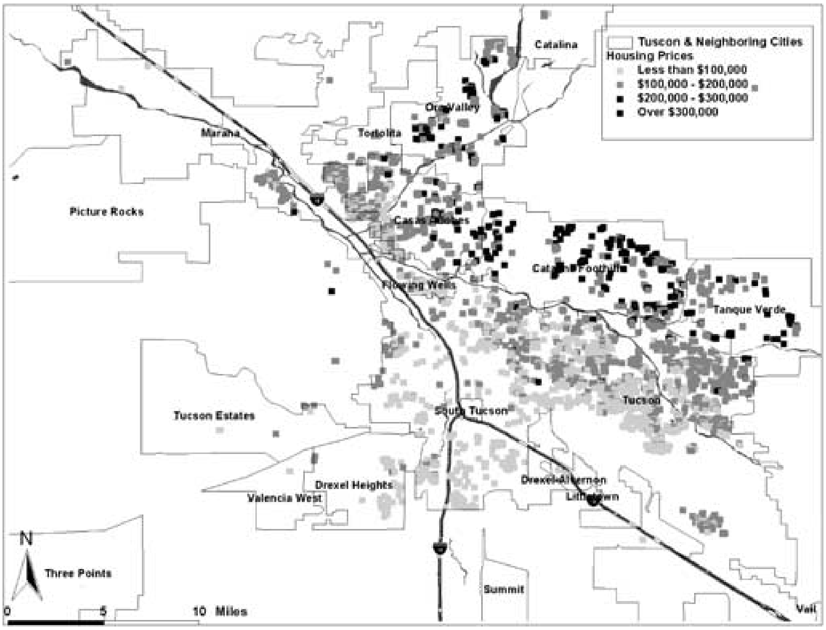
Table 5 ■ Model 4: Contains both $\{x, y\}$ coordinates and dummy variables for MLS districts; Dependent variable = Log of selling price.

Independent Variables	Estimated Coefficient	Standard Error	Standardized Coefficient	*t*-Stat.	Significance
Constant	10.366	0.128		81.1	0.000
$SQFT^2$	−5.932E-08	8.01E-09	−0.318	−7.4	0.000
AGE	−2.069E-02	8.63E-04	−0.759	−23.9	0.000
AGE^2	3.007E-04	1.65E-05	0.553	18.2	0.000
LOT^2	−4.268E-10	4.13E-11	−0.406	−10.3	0.000
$SQFT * x^2$	−8.827E-10	2.90E-10	−0.783	−3.0	0.002
$SQFT * y^3$	−2.659E-12	6.28E-13	−0.467	−4.2	0.000
$SQFT * x^2 * y$	4.985E-12	8.40E-13	2.157	5.9	0.000
$SQFT^2 * EDUM$	−2.231E-08	3.00E-09	−0.065	−7.4	0.000
$LOT * x$	−1.414E-07	1.20E-08	−2.544	−12.2	0.000
$LOT * y^2$	1.889E-09	1.41E-10	9.659	13.3	0.000
$LOT * y^3$	−2.665E-12	3.30E-11	−6.609	−12.4	0.000
$NDUM * x$	−6.583E-03	1.18E-03	−3.249	−5.5	0.000
$NDUM * x * y$	1.418E-05	2.48E-06	3.344	5.7	0.000
$y * x^2$	1.333E-09	4.56E-10	0.067	2.9	0.003

Parcel characteristics are lot size (*LOT*), age of dwelling in years (*AGE*), square footage of improved space (*SQFT*) and absolute value of the $\{x, y\}$ coordinate(s). Location dummies are *NDUM* (the north MLS district), *NEDUM* (the northeast district) and *EDUM* (the east–west corridor district). Sample size = 2,971, adjusted R-square = 0.844, standard error of estimate = 0.1604022, degrees of freedom = 2,956, $F = 1{,}147$ (significance 0.000) and the inclusion/exclusion criteria are $p \leq 0.10$, $p > 0.10$. To test for the spatial independence of the estimated error terms, the residuals were regressed (using stepwise regression) on the independent variables. None of the coefficient estimates are significant at the 0.10 significance level; thus, the null hypothesis of spatial independence in error terms cannot be rejected.

Although Models 1 and 2 are largely insensitive to price variations within geographic submarkets, the inclusion of the interaction terms in Model 4 helps to uncover the directional aspects of selling price within the various submarkets. Overall, the empirical evidence suggests that consumers are willing to pay a premium for locating in the relatively affluent areas of North and Northeast Tucson, with higher selling prices predicted along the y and x axes, particularly as x and y increase simultaneously (see Figure 1). Moreover, the results highlight the interactive nature of absolute location and housing attributes as they combine to explain variations in the selling price of housing in the Tucson market. Although the price of housing generally increases as one trends north or northeast, there is statistical evidence that intrasubmarket variations in the selling price of housing exist in this market. Note that of the three submarket dummies that tested significant in Model 4, none entered into the equation in an intercept-shifting capacity.

Figure 1 ■ Predicted prices for full sample of 2,971 transactions.



The significance of the interactive variables yields strong statistical evidence that the underlying selling price of housing is not related simply to site attributes within various submarkets or a premium attached to its location within a given submarket. Instead, housing prices are simultaneously related to both site attributes and absolute location of the units in question. Although *AGE* is found, once again, to be an important explanatory variable, it is shown to interact exclusively with itself and not with other site characteristic variables once absolute location is introduced into the model (compare results of Models 2 and 3 with Model 4). Note that $\{x, y\}$ variables enter into Model 4 in a largely interactive sense. In other words, absolute location is shown to be interactive in some way with site attributes or location dummies. Once again, the results support the contention that site and structural attributes are nonseparable from location, given that the majority of explanatory variables in the model are of the interactive variety. Thus, the underlying location-value signatures of residential housing units in this market are unique given that (1) the value of each housing unit is inextricably linked to both physical attributes and its unique location and (2) the physical and locational attributes are not independent of one another. As would be expected, therefore, Model 4 outperforms all other specifications with an adjusted R-squared of 0.844 and a standard error of 0.1604.

Predictive Accuracy

To evaluate the prediction accuracy of the four alternative models, the sample of 2,971 transactions is separated into two subsamples: an estimation subsample and a prediction subsample. The estimation sample is a random sample of 2,471 transactions from the 2,971 total observations. These transactions are used to estimate the parameters of the alternative hedonic models. The remaining 500 transactions (*e.g.*, the prediction sample) are excluded from the estimation sample and are used to evaluate prediction accuracy for the alternative hedonic specifications. The prediction error is the actual minus the predicted selling price, where predicted price is based on recovered estimates from $e^{[predicted\,\ln(price)]}$. The same estimation and prediction subsamples are used for each of the four alternative specifications. Thus, variations in prediction accuracy cannot be attributed to differences in the underlying sample.

Our comparison of out-of-sample predictive power is contained in Table 6. Panel A contains the mean absolute error (MAE) and standard deviation for each model specification. The MAE is the average of the absolute value of the prediction errors. For the interactive aspatial model (Model 1), the holdout sample produced a MAE of \$20,210 and a standard deviation of \$27,430. The MAE is approximately 15.3% of the average selling price in the holdout sample. Results for the holdout sample for Model 2 (which includes location dummy variables) show a marginal improvement over the aspatial model. The MAE is \$18,943, or 14.3% of the average selling price for units in the holdout sample, and the standard deviation declines to \$25,319.

In Model 3, square footage, lot size and age are interacted with a polynomial expansion of the property's $\{x, y\}$ coordinates instead of discrete location dummies. Model 3's MAE of \$17,242 is 13.0% of the average selling price for the holdout sample. Finally, because it would seem useful to have information on both absolute and relative location, Model 4 includes both $\{x, y\}$ coordinates and discrete location dummies. The MAE for Model 4 is \$16,805, compared to \$18,943 for Model 2 and \$17,242 for Model 3. By incorporating information on both absolute and relative location, Model 4 is able to account for a greater percentage of the intraurban variation in selling price of housing and to reduce the standard error of the estimate.

The mean absolute error is a relative metric for comparing alternative hedonic specifications. As an alternative metric of prediction accuracy, we also

Table 6 ■ Comparison of out-of-sample predictive power.[a,b]

Panel A: Mean absolute error = Average selling price − Estimated selling price		
Interactive Model	MAE	Std. Dev.
Model 1: Aspatial	\$20,210	\$27,430
Model 2: Dummy variables only	\$18,943	\$25,319
Model 3: $\{x, y\}$ coordinates only	\$17,242	\$22,280
Model 4: Dummies and $\{x, y\}$ coordinates	\$16,805	\$21,912
Panel B: Percentage of predicted prices within 10% of actual price		
Interactive Model	Holdout Sample	Full Sample
Model 1: Aspatial	46.4%	49.2%
Model 2: Dummy variables only	54.6%	51.9%
Model 3: $\{x, y\}$ coordinates only	54.2%	56.8%
Model 4: Dummies and $\{x, y\}$ coordinates	65.0%	59.9%

[a]From the 2,971 total observations, 500 randomly selected observations were withheld. The model was then rerun with 2,471 observations. The estimated regression was then used to predict the selling prices of the properties in the holdout sample. The prediction error is the actual selling price minus the predicted. Predicted price is based on recovered estimates from $e^{[predicted \ln(price)]}$.

[b]Using the predicted price recovered from $e^{[predicted \ln(price)]}$ may produce a biased prediction for the expected transaction price (see Goldberger 1968 and Thibodeau 1989). An Equality of Means test was therefore run for each model for the estimation sample data. The mean of the observed mean transaction prices for the holdout sample is \$132,084.70. The means of the predicted prices for Models 1–4 are \$128,836.6, \$129,397.2, \$131,319.5 and 131,399.8, respectively. The observed and predicted means for Models 1 and 2 are significantly different from each other at the 95% confidence level; therefore, a correction (based on the standard deviation of the residuals) is required. However, the observed and predicted means for Models 3 and 4 are not significantly different from one another at the 95% confidence level. Thus, a correction factor to ensure unbiased estimation of price was not employed for Models 3 and 4 given the test results and the fact that the sample size is large enough for the asymptotic results on unbiasedness to be valid.

calculated the percentage of predicted prices that are within 10% of the observed transaction price.[8] These results are reported in Panel B of Table 6. Using our holdout sample of 500 transactions, Model 1's predicted prices are within 10% of the actual price in 46.4% of the cases. Adding dummy location variables to the aspatial model improves the prediction accuracy to 54.6%. Model 3, which employs $\{x, y\}$ coordinates instead of discrete location dummies, has a predictive accuracy of 54.2%. Finally, Model 4, which incorporates

[8] In a May 2002 presentation to the Homer Hoyt Institute about Freddie Mac's automated valuation models (AVMs), Michael Bradley of Freddie indicated that Freddie likes to have at least 50% of the predictions within 10% of the actual transaction price.

information on both absolute and relative location, has a predictive accuracy of 65.0%. The results produced by using the entire sample of 2,971 transactions are broadly similar. However, Model 3's predictive accuracy is now greater than Model 2's (56.8% *vs.* 51.9%). Model 4's predictive accuracy is 59.9%.

These results strongly suggest that the inclusion of information on the absolute location of residential housing may lead to more precise estimation of selling price, particularly as each unit's location is truly unique in any given market. Again, it is important to emphasize that the modeling approach applied here requires no prior knowledge of the local real estate market, just the $\{x, y\}$ coordinates of parcels themselves. This framework offers great utility to those modeling the value of residential housing units in a market (or a series of markets) with which they are unfamiliar.

Nonetheless, Model 3 does a relatively good job with fewer explanatory variables and no *a priori* specification real estate submarkets. Once again, this approach offers a distinct advantage to modelers or mortgage lenders who do not posses local knowledge of numerous and diverse real estate markets and submarkets across various regions. Moreover, in cases where expert knowledge is available, the inclusion of $\{x, y\}$ coordinates (in an interactive capacity) can be used to improve the predictive power of models that traditionally incorporate discrete location dummies.

Conclusions

Real estate valuation models have often underspecified the locational or situational characteristics when attempting to explain variations in the market price of residential housing. Although hedonic house price models frequently incorporate dummy variables to distinguish between locations or submarkets, such frameworks are generally unable to distinguish more refined spatial patterns or trends. Moreover, despite early advancements in hedonic modeling using distance gradients and locational accessibility, relatively few studies have fully explored the potential of using Cartesian coordinates (in an interactive fashion) to explain intraurban variations in the price of housing and to capture the composite impact of externalities on property values.

We argue that each property can be thought of as having a unique *location value signature* or *LVS* in relation to the sum total of all externalities which affect a given property/location. Accessibility indices and submarket dummies cannot fully account for the influence of absolute location on the market price of housing as there are an indeterminable number of externalities influencing a given property at its given location. Furthermore, the degree to which any one or more externalities affects real estate values is highly variable over space.

Hedonic models that do not directly incorporate absolute location (and potential interactions between explanatory variables and absolute location) will most likely fall short of explaining the true underlying impact location has on market price.

This paper develops several hedonic specifications that attempt to more fully capture the interactive components of location value. More specifically, we present an interactive variables approach and test its ability to explain price variations in urban residential housing. Relative to a noninteractive model, explanatory power is significantly increased. Moreover, the spatial dependence in error terms observed in the noninteractive specification is removed in our sample by the interactive variables approach. Future research should more closely examine the predictive accuracy of our variable interactions approach relative to alternative specifications (no lot size, linear lot size, lot size interacted with other housing characteristics).

Although we employ a cross-sectional hedonic model, with results pertaining to a given year, the analysis can easily be expanded to include an interactive time variable. The inclusion of a temporal variable could allow one to test both the stability of estimated coefficients and structural changes in the spatial distribution of the underlying location value signatures over time. Future research could focus on explaining relative shifts in intraurban price appreciation. Because the focus of this research is prediction accuracy, we have not calculated marginal lot valuations, age effects, and so forth, although this would be a useful extension. An additional avenue for future research would be to investigate the effects of adding $\{x, y\}$ coordinates to the age-related heteroskedasticity problem discussed by Goodman and Thibodeau (1997).

Continued advances in spatial modeling and the use of GIS will make it easier to compare and contrast alternative approaches in the development of location-sensitive hedonic models. The recent explosion of geocoded housing market data provides a wealth of opportunities for examining the relationships between location, externalities and property values both within and between geographic markets and submarkets. In our view, the variable interaction approach discussed here offers an interesting alternative for real estate analysts who wish to uncover the importance of "location, location, location" in an absolute sense without having prior expert knowledge of the geographic markets they are analyzing.

References

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

Clapp, J.M. 2001. A Semi Parametric Method for Estimating Local House Price Indices. Working paper. University of Connecticut: Storrs, Connecticut.

Dewees, D.N. 1976. The Effect of a Subway on Residential Property Values in Toronto. *Journal of Urban Economics* 3: 357–369.

Dubin, R.A. 1998. Spatial Autocorrelation: A Primer. *Journal of Housing Economics* 7: 304–327.

———, R.K. Pace and T.G. Thibodeau. 1999. Spatial Autoregression Techniques for Real Estate Data. *Journal of Real Estate Literature* 7: 79–95.

Gillen, K., T. Thibodeau and S. Wachter. 2001. Anisotropic Autocorrelation in House Prices. *Journal of Real Estate Finance and Economics* 23: 5–30.

Goldberger, A.S. 1968. The Interpretation and Estimation of Cobb–Douglass Functions. *Econometrics* 36(3): 464–472.

Goodman, A.C. and T.G. Thibodeau. 1997. Dwelling Age Heteroskedasticity in Hedonic House Price Equations: An Extension. *Journal of Housing Research* 8: 299–317.

——— and ———. 1998. Housing Market Segmentation. *Journal of Housing Economics* 7: 121–143.

——— and ———. 2003. Housing Market Segmentation and Hedonic Prediction Accuracy. *Journal of Housing Economics* 12(3): 181–201.

Grether, D.M. and P. Mieszowski. 1974. Determinants of Real Estate Value. *Journal of Urban Economics* 1: 127–146.

Guy, D.C., J.L. Hysom and S.R. Ruth. 1985. The Effect of Subsidized Housing on Values of Adjacent Housing. *AREUEA Journal* 13: 378–387.

Hansen, W. 1959. How Accessibility Shapes Land Use. *Journal of the American Institute of Planners* 25: 73–76.

Ingram, D.R. 1971. The Concept of Accessibility: A Search for an Operational Form. *Regional Studies* 5: 101–107.

Jackson, J.R. 1979. Intraurban Variation in the Price of Housing. *Journal of Urban Economics* 6: 464–479.

Johnson, M.S. and W.R. Ragas. 1987. CBD Land Values and Multiple Externalities. *Land Economics* 63: 337–347.

Johnston, R.J. 1978. *Multivariate Statistical Analysis in Geography*. Longman: London.

Kiel, K.A. 1995. Measuring the Impact of the Discovery and Cleaning of Identified Hazardous Waste Sites on House Values. *Land Economics* 74(4): 428–435.

——— and K.T. McClain. 1995. The Effects of an Incinerator Sitting on Housing Appreciation Rates. *Journal of Urban Economics* 37: 311–323.

Kohlhase, J.E. 1991. The Impact of Toxic Waste Sites on Housing Values. *Journal of Urban Economics* 30: 1–26.

Li, M.M. and H.J. Brown. 1980. Micro-Neighborhood Externalities and Hedonic Housing Prices. *Land Economics* 56: 125–141.

Michaels R.G. and V.K. Smith. 1990. Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites. *Journal of Urban Economics* 28: 223–242.

Mulligan, G.F., R. Franklin and A.X. Esparaza. 2002. Housing Prices in Tucson, Arizona. *Urban Geography* 23(5): 446–470.

Nelson, J.P. 1980. Airports and Property Values: A Survey of Recent Evidence. *Journal of Transportation Economics and Policy* 14: 37–52.

Pace, R.K. and R. Barry. 1997. Fast Spatial Estimation. *Applied Economic Letters* 4: 337–341.

———, ———, J.M. Clapp and M. Rodriguez. 1998. Spatiotemporal Autoregressive Models of Neighborhood Effects. *Journal of Real Estate Finance and Economics* 17: 15–33.

———. and O.W. Gilley. 1997. Using the Spatial Configuration of the Data to Improve Estimation. *Journal of Real Estate Finance and Economics* 14: 333–340.

Palmquist, R.B. 1992. Valuing Localized Externalities. *Journal of Urban Economics* 31: 59–68.

Parsons, G.R. 1990. Hedonic Prices and Public Goods: An Argument for Weighting Locational Attributes in Hedonic Regressions by Lot Size. *Journal of Urban Economics* 27: 308–321.

Pavlov, A.D. 2000. Space-Varying Regression Coefficient: A Semi-parametric Approach Applied to Real Estate Markets. *Real Estate Economics* 28: 249–283.

Reichert, A.K., M. Small and S. Mohanty. 1992. The Impact of Landfills on Residential Property Values. *Journal of Real Estate Research* 7: 297–314.

Smersh, G. 1996. Intraurban Variation in House Price Appreciation: A Case Study, Jacksonville, Florida 1980–1990. Ph.D. Dissertation. Department of Geography, University of Florida: Gainesville, FL.

Thibodeau, T. 1989. Housing Price Indexes from the 1974–83 SMSA Annual Housing Surveys. *AREUEA Journal* 17(1): 100–117.

Waddell, P., B.J.L. Berry and I. Hoch. 1993a. Housing Price Gradients: The Intersection of Space and Built Form. *Geographical Analysis* 25: 5–19.

———, ——— and ———. 1993b. Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location. *Journal of Real Estate Finance and Economic* 7: 117–141.

Weibull, J.W. 1976. An Axiomatic Approach to the Measurement of Accessibility. *Regional Science and Urban Economics* 6: 357–379.

———. 1980. On the Numerical Measurement of Accessibility. *Environment and Planning* 12: 53–67.