A PROPOSAL

Measuring the Impact of Urban Form on Property Values – A Hedonic Price Analysis

Version 1.0
May 18, 2007

Mi Diao

Department of Urban Studies and Planning
Massachusetts Institute of Technology
Cambridge, Massachusetts

TABLE OF CONTENTS

SECTION I. INTRODUCTION

Concerns over the impacts of urban sprawl in U.S. metropolitan areas have led to increased advocacy for alternative urban development patterns. In 1993, a group of architects founded the Congress for the New Urbanism dedicated to "creating buildings, neighborhoods, and regions that provide a high quality of life for all residents, while protecting the natural environment" (NU 2002). Nowadays, new urbanism has been considered as one important tool to fight against urban sprawl and has attracted increasing interests among urban planners, architects, and developers. However, some critical issues about new urbanism are still under debate or not adequately addressed. One such question is whether good urban form can be capitalized in the market. If good urban form provides higher quality of life, the monetary value of this effect will be reflected in the property price, in addition to the value of other features such as structural attributes and neighborhood characteristics. The proposed study is to examine the impacts of urban form on property values using hedonic price analysis.

SECTION II. REVIEW OF LITERATURE

A. Literature Relevant to Urban Form Measures

In previous studies, many researchers generated various quantitative urban form measures. Their approaches vary in spatial resolution but generally employ a unit of analysis that is too coarse to keep sufficient spatial details of urban form. In order to examine development patterns and trends in Sacramento, California, and five comparable cities, Wassmer (2000) calculated share of metropolitan population, shares of employment, retail sales, farmland, poverty rates, income level, employment rate and commuting time for the central city, the central county, and the urbanized area respectively. He concluded that Sacramento exhibits “a high degree of relative decentralization,” and displays “a high level of negative metropolitan outcomes generally associated with a high degree of urban sprawl”. Galster et al (2001) defined urban sprawl based on eight distinct dimensions of land use patterns: density, continuity, concentration, clustering, centrality, nuclearity, mixed uses, and proximity. They compared measures for six dimensions and rank the level of urban sprawl for 13 urbanized areas. With the development of spatially disaggregate data and Geographic Information Systems (GIS) tools, more recent studies are conducted at more disaggregate level to improve the precision of measures. To explore the relationship between urban form and property values, Song and Knaap (2003) used residential neighborhood level data to develop five dimensions of urban form measures: street design and circulation system, density, land use mix, accessibility, and pedestrian access. Using a similar set of measures, Song (2005) compared three study areas to determine how well their urban development patterns meet smart growth principles. Rajamani et al. (2003) used neighborhood level urban form measures to assess the impacts of urban form on non-work trip mode choice.

B. Literature Relevant to Hedonic Price Analysis and Property Values

Hedonic price analysis is a method of estimating demand or price for a good that does not have a traditional economic market (Lipscomb 2007). It decomposes the item being researched into its constituent characteristics, and estimates the value of each characteristic. In the case of housing, which is a heterogeneous bundled good, researchers has been utilizing the hedonic pricing analysis to assess the implicit prices for a variety of characteristics associated with the property, such as structure components, environmental factors, public services, urban form, etc.

In the environmental economics field, Kim, Phipps and Anselin (2003) developed a spatial-econometric hedonic housing price model to estimate the marginal value of improvements in SO2 and NOX concentrations for the Seoul metropolitan area. They found that SO2 pollution levels have a significant impact on housing prices while NOX pollution doesn't. Hedonic price models are also used to examine the capitalization of public services. Edel and Sclar (1974) and King (1977) indicated the need to include measures of public service and taxes paid in the hedonic price analysis. Using data from Chicago for 1987-1991, Downes and Zabel (2002) estimated the impacts of school characteristics on house prices. They assigned to each house the school-level data for the closest school, and found that school outputs, i.e., test scores, have significant impact on house price, while school inputs, i.e., per-pupil expenditures don't.

Hedonic price analysis is also widely applied to explore the relationship between urban form and property values. The implicit marginal prices of various urban form components, such as land use mix (Cao and Cory 1981; Song and Knaap 2004), street pattern (Guttery 2002), amenities (Shultz and King 2001; Benson et al 1998), proximity to transit stations and commercial centers (Bowes and Ihlanfeldt 2001; Song and Knaap 2004), etc., have been evaluated by various researchers. Cao and Cory (1981) showed that increasing industrial, commercial, multi-family and public land uses tends to increasing surrounding home values. They thus concluded that an optimal land use mix should be sought in locating economic activities into neighborhoods. Guttery (2002) examined the sales price of 1,672 houses located in the Greater Dallas-Fort Worth-Denton metroplex and found negative impacts from having rear-entry alleyways. Benson et al (1998) estimated the value of the view amenity in single-family residential real estate markets of Bellingham, Washington. Results from a hedonic price model suggest that willingness to pay for the view amenity is quite high. Shultz and King (2001) provided empirical evidence that proximity to the large protected natural areas, golf courses, and Class II wildlife habitats, as well as the percentage of vacant and commercial land use, positively influences housing values. Bowes and Ihlanfeldt (2001) looked into both the direct and the indirect effects of transit stations on the attractiveness of nearby neighborhoods. They found that stations located away from downtown have positive impacts on property values, while stations in low-income neighborhoods or close to downtown generate negative externalities to nearby properties. Song and Knaap (2004) incorporated quantitative measures of mixed land uses for neighborhoods in Washington County, OR into hedonic analysis, and concluded that housing prices increase with their proximity to public parks or neighborhood commercial centers. In another comprehensive study with neighborhood level urban form measures, Song and Knaap (2003) found that some features of new urbanism are capitalized into property values such as more connective street networks, more streets, shorter cul-de-sacs, smaller block size, better pedestrian accessibility to commercial uses, more evenly distributed mixed land uses and proximity to light rail stations. While features like higher density, containing more commercial, multifamily, and public use (relative to single-family uses), and containing major transportation arterials are not attractive to property buyers.

All together, considerable research suggests that (1) hedonic price analysis is a powerful tool in assessing various components of housing value; and (2) certain elements of urban form can be capitalized into property values. However, some potential problems associated with hedonic price analysis have not been fully addressed: (1) the spatial autocorrelation problem associated with the use of spatial data; and (2) the self-selection bias associated with the use of transaction based housing price data. The proposed study attempts to integrate advanced statistical techniques into the hedonic price models to get more rigorous estimates of the impact of urban form on property values in the market.

SECTION III. DATA

The main datasets used in the proposed study include Boston Sales of Residential Property Data, Boston Parcel Data, Census demographic data, land use data, and transportation data[1].

A. Boston Sales of Residential Property Data

Sales of Residential Property Data is a transaction-based dataset, containing parcel ID, sales price, address, sales date, living area, architectural style, and property type of each residential sale in the City of Boston. The data source is Suffolk County Registry of

[1] Another transaction based housing price dataset from Warren Group may also be available for this study. It records all the housing transactions in the Boston metropolitan area. The expansion of study area from City of Boston to the Boston metropolitan area could increase the variations of urban form measures, hence improve the outcomes of the proposed study.

Deeds and City of Boston Assessing Department.

B. Boson Parcel Data

Boston Parcel Data from the City of Boston Assessing Department provide a rich set of information for over 100,000 parcels, including parcel ID, parcel size, assessing value, tax, etc. It can be used to generate asset-specific variables in the hedonic price analysis.

C. Other Data Source

Land use data, Census demographic data 2000, and Census Transportation Planning Package (CTPP) data 2000 are also collected to generate related hedonic variables.

SECTION IV. METHODOLOGY

In the proposed study, we are to evaluate the virtues of urban form. Using Geographic Information Systems (GIS) tools, we will develop quantitative measures of urban form at neighborhood level, and incorporate those measures in hedonic price analysis. Spatial econometric techniques will be applied to deal with the spatial ‘spillover’ effects. To address the self-selection bias in the hedonic model, we will use a two-step estimate based on the classic Heckman procedure. In essence, our research will decompose urban form into its component parts and estimate the implicit prices associated with each part in the market place.

A. Urban Forum Measures

This study will decompose urban form into several components based on previous literature and quantify each component using GIS tools. Five major components commonly advocated by the new urbanism design will be calculated: residential density, land use mix, accessibility, local street pattern, and pedestrian walkability.

1) Residential Density

Density is an important indicator of urban form. A rich literature indicates that urban sprawl is characterized by low density single-family housing development, which leads to lower efficiency of land consumption and infrastructure investment, longer vehicle miles traveled and loss of sense of community. Two density measures will be offered in the proposed study: density of single-family house and population density.

2) Land Use Mix

Land use mix measures the degree to which land uses are mixed and balanced within the neighborhood. Greater mix of uses is believed to facilitate walking and biking, reduce vehicle trips generated and vehicle miles traveled, and enhance urban aesthetics. The commonly used measure of land use mix is land use entropy.

$$\text{Land use mix entropy} = -\sum_{j=1}^{J} \frac{P_j * \ln(P_j)}{\ln(J)} \quad (1)$$

where J is the total number of land use categories. P_j is the proportion of land in the jth land use category. A value of 0 means the land in the neighborhood is exclusively dedicated to a single use, while a value of 1 indicates perfect mixing of the *J* land use categories.

3) Accessibility

Accessibility measures the ease to reach amenities and service opportunities. Three measures of accessibility will be calculated in the proposed study: distance to retail centers, distance to transit station, and distance to a public park.

4) Local Street Pattern

Literature indicates that urban sprawl type of development contains too many winding streets and cul-de-sacs, over-sized block, and poorly connected road network compared to the traditional grid-type street pattern, which leads to longer vehicle miles, worse air quality, less walking and biking activities. To show the differences between urban sprawl and traditional block pattern, we will use two indicators: average block parameter and total length of cul-de-sacs.

5) Pedestrian Walkability

The amenity of pedestrian infrastructure may include lighting, planting, sidewalk pavement, sidewalk width, curb situation, speed limit, etc. But most of them are not well recorded. Two factors will be included in the proposed study: average sidewalk width and the density of intersections along a path.

B. Hedonic Price Model

Hedonic regression assumes that sales price or rent of a property is a function of structural characteristics (S), neighborhood characteristics (N), location characteristics (L), contract characteristics (C), and the time rent or value is observed (T) (Malpezzi 2002), as is shown in Equation (2).

$$sale\_price = f(S, N, L, C, T) \qquad (2)$$

A standard hedonic price model can be specified as:

$$\ln(sale\_price_i) = \beta_i x_i + \varepsilon_i \qquad (3)$$

Where x_i is a vector of asset-specific characteristics of the properties (the hedonic variables), ε is normally distributed mean zero random error.

In the proposed study, the hedonic variables will cover five subsets of factors.

1) Structural Characteristics

Structural characteristics of a house – such as living area, lot size, architectural style, etc. – are widely known to affect its value.

2) Urban Form Characteristics

The urban form measures are discussed in the preceding section. Estimating their impacts is the major objective of the proposed study.

3) Socioeconomic Characteristics

Residents care about the socioeconomic status of their neighborhoods. To capture this effect, we include measures of racial composition and median household income.

4) Location Characteristics

Location characteristics include the public service level and accessibility to employment centers. Public service level can be measured with mean SAT score and Property tax rate. Accessibility to employment center can be measured with distance to the Central Business District (CBD).

5) Sales Date

To capture the overall housing market fluctuation, the quarter when the property was sold will be included in the model.

C. Spatial Econometrics

In spatial data analyses, when a value observed in one location depends on the values at neighboring locations, there is a spatial autocorrelation. Since the price of a property may be influenced by the characteristics of its neighboring properties, the proposed study also need to take spatial effects into consideration. Spatial data may show spatial autocorrelation in the variables or error terms. Accordingly, this study will address the spatial autocorrelation with spatial-lag model and spatial-error model developed by Luc Anselin (1999).

In the spatial lag case, the dependent variable y in place i is affected by the independent variables in both place i and its neighboring areas. With the existence of spatial lag, the assumption of uncorrelated error terms of Ordinary Least Square (OLS) estimation is violated; in addition, the assumption of independent observations is also violated. As a result, the OLS estimates are biased and inefficient. Spatial error means the error terms across spatial units are correlated. With the presence of spatial error, the assumption of uncorrelated error terms of OLS is violated. Therefore, the OLS estimates are inefficient.

In the spatial lag model, 'spillover' means the values of independent variables in one location will affect the values of dependent variable in adjacent areas (Anselin 1999).The spatial lag can be addressed by adding an additional regressor in the form of a spatially lagged dependent variable to the regression, as is shown in Equation (4).

$$Y = \rho W_y + \beta X + \varepsilon \tag{4}$$

where *Wy* is the spatial lag variable, ρ is the autoregressive coefficient. A spatial lag of a specified variable is computed by taking the weighted average of surrounding spatial units. The weights can take different forms, for example contiguity based weights, distance based weights, and K-nearest neighbor weights (Anselin 2003). The existence and magnitude of 'spillover' effects are indicated by the estimated value of the coefficient for the spatial lag variable (autoregressive coefficient).

A spatial-error model can be seen as a special case of a regression with a non-spherical error term. The off-diagonal elements of the covariance matrix express the structure of spatial dependence. The spatial error model can be estimated by:

$$Y = \beta X + \varepsilon \quad \varepsilon = \lambda W_\varepsilon + u \tag{5}$$

where u is an error term that meets the OLS assumptions, W_ε is the weighted average of error terms in neighboring areas. In the case of housing price, the spatial error model assumes that the spillover occurs indirectly through spatial correlation in the error terms for neighboring properties. That is, the independent variables have only local effects, but factors missing from the model specification are spatially correlated.

The spatial econometric techniques discussed in this section are based on approaches proposed by Luc Anselin (1999). There are still other approaches that could be utilized in the proposed study, for example the Cokriging approach (Chica-Olmo 2007) and Geographically Weighted Regression approach (Fotheringham, Brunsdon, and Charlton 2002). It would be interesting to compare the results of various approaches in dealing with spatial autocorrelation.

D. Heckman's Two-Stage Estimates

The sample used to estimate the hedonic price model are transaction based, consisting only of properties sold in the market, where buyer's reservation prices are higher than seller's reservation prices. So they are not necessarily random draws from the population of properties. The possibility of sample selection bias arises when the unobserved property characteristics affecting the transaction sales propensity also influence the transaction price level. To deal with the self-selection bias, the proposed study will follow the method presented by Fisher, Geltner, and Pollakowski (2006), which is an application of the Heckman's two-stage estimates (Heckman 1979). The reservation price equations for both the buyer side and the seller side, and transaction price for the sold properties take the following forms:

$$RP_{it}^{b} = \sum \alpha_{j}^{b} X_{ijt} + \sum \beta_{t}^{b} Z_{t} + \varepsilon_{it}^{b} \tag{6}$$

$$RP_{it}^{s} = \sum \alpha_{j}^{s} X_{ijt} + \sum \beta_{t}^{s} Z_{t} + \varepsilon_{it}^{s} \tag{7}$$

$$P_{it} = \sum \alpha_j X_{ijt} + \sum \beta_t Z_t + (\varepsilon_{it} \big| RP_{it}^b \geq RP_{it}^s) \quad (8)$$

where X_{ijt} is a set of j asset-specific characteristics of the properties including the structural, urban form, socioeconomic, and location characteristics discussed in the preceding section, Z_t is a vector of zero/one time-dummy variables ($Z_t = 1$ in quarter t), RP_{it}^b (RP_{it}^s) is the natural logarithm of a buyer's (seller's) reservation price for property i as of time t, ε_{it}^b (ε_{it}^s) is normally distributed mean zero random errors, and $\alpha_j = \frac{1}{2}(\alpha_j^b + \alpha_j^s), \beta_t = \frac{1}{2}(\beta_t^b + \beta_t^s), and \varepsilon_{it} = \frac{1}{2}(\varepsilon_{it}^b + \varepsilon_{it}^s)$ (Fisher, Geltner, and Pollakowski 2006).

It should be noted that the error term in Equation (8) may have a nonzero mean because the observed transaction sample consists only of selected properties, i.e., properties whose buyer reservation price is higher than seller reservation price. When $E[(\varepsilon_{it} \big| RP_{it}^b \geq RP_{it}^s] \neq 0$, OLS regression using the observed data produces biased estimates of α and β. The standard treatment for the sample selection problem of this type is Heckman's two-step estimates (Heckman, 1976, 1979). Heckman's procedure first uses all the properties to model the possibility that a property is sold with a probit model. A property is sold when the reservation price of the buyer is higher than that of the seller. Use S_{it}^* to denote the difference between the two reservation prices.

$$\begin{aligned} S_{it}^* = RP_{it}^b - RP_{it}^s &= \sum (\alpha_j^b - \alpha_j^s) X_{ijt} + \sum (\beta_t^b - \beta_t^s) Z_t + (\varepsilon_{it}^b - \varepsilon_{it}^s) \\ &= \sum \varpi_j X_{ijt} + \sum \gamma_t Z_t + \eta_{it} \end{aligned} \quad (9)$$

The probability that a property i is sold can be modeled as:

$$\Pr[\text{Property } i \text{ is sold}] = \Phi(\sum \varpi_j X_{ijt} + \sum \gamma_t Z_t) \quad (10)$$

Where $\Phi(\ )$ is the cumulative density function of the normal distribution.

The results of the probit estimation can then be used to construct estimates of the inverse Mills ratio ($\hat{\lambda}_i$), an instrumental variable to control for self-selection:

$$\hat{\lambda}_i = \frac{\phi\left(\sum \varpi_j X_{ijt} + \sum \gamma_t Z\right)}{\Phi\left(\sum \varpi_j X_{ijt} + \sum \gamma_t Z\right)} \quad (11)$$

where ϕ and Φ denote the probability density function (*pdf*) and cumulative distribution function (*cdf*) of the standard univariate normal distribution, respectively. When adding the inverse Mills ratio $\hat{\lambda}_i$ into the hedonic price equation, the OLS estimates of α and β will be unbiased and consistent (Fisher, Geltner, and Pollakowski 2006).

SECTION V. SUMMARY

In summary, the proposed study will estimate the impacts of urban form on property values with hedonic price analysis using transaction data of residential properties in the Boston area. Neighborhood level urban form measures will be calculated with GIS tools and integrated into the hedonic price model. Furthermore, we attempt to address some important issues not well addressed in the literature with advanced statistical techniques: (1) spatial econometric methods, in particular spatial lag model and spatial error model, will be used to deal with the spatial autocorrelation problem brought by the spatial data used in the analysis; (2) Heckman's two-step estimates will be used to fix the self-selection bias associated with transaction based property value data.

BIBLIOGRAPHY

Anselin, Luc. 1999. "Spatial Econometrics", Working Paper.

Anselin, Luc. 2003. "GeoDa 0.9 User's Guide."

Benson, Earl D. et al. 1998. "Pricing Residential Amenities: The Value of a View", *Journal of Real Estate Finance and Economics*, 16: 1, 55-73.

Bowes, D.R. and K.R. Ihlanfeldt. 2001. "Identifying the Impacts of Rail Transit Stations on Residential Property Values", *Journal of Urban Economics* 50, 1: 1-25.

Cao, T. V. and D. C. Cory. 1981. "Mixed Land Uses, Land-Use Externalities, and Residential Property Values: A Re-evaluation", *Annals of Regional Science* 16: 1-24.

Chica-Olmo, Jorge. 2007. "Prediction of Housing Location Price by a Multivariate Spatial Method: Cokriging", *Journal of Real Estate Research* 29, 1: 91-114.

Downes, Thomas A. and Jeffrey E. Zabel. 2002. "The impact of school characteristics on house prices: Chicago 1987–1991", *Journal of Urban Economics* 52: 1–25.

Edel, Matthew and Elliot Sclar. 1974. "Taxes, Spending and Property Values: Supply Adjustment in a Tiebout-OOates Model", *Journal of Political Economy* 82: 941-954.

Fisher, Jeff, David Geltner, and Henry Pollakowski. 2006. "A Quarterly Transactions-Based Index of Institutional Real Estate Investment Performance and Movements in Supply and Demand", Working Paper.

Fotheringham, A. Steward, Chris Brunsdon, and Martin Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, John & Sons, Incorporated.

Galster, George, et al. 2001. "Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept", *Housing Policy Debate* 12, 4: 681-717.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error", *Econometrica* 47, 1: 153-161.

Guttery, R. S. 2002. "The Effects of Subdivision Design on Housing Values: the Case of Alleyways", *Journal of Planning Education and Research* 23, 3: 265-273.

Kim, Chong, Tim T. Phipps and Luc Anselin. 2003. "Measuring the benefits of air quality improvement: a spatial hedonic approach", *Journal of Environmental Economics and Management* 45, 1: 24-39

King, A. Thomas. 1977. "Estimating Property Tax Capitalization: A Critical Comment", Journal of Political Economy 85, 2: 425-431.

Lipscomb, Clifford A. 2007. "An Alternative Spatial Hedonic Estimation Approach", Journal of Housing Research 15, 2: 143-160.

Malpezzi, Stephen. 2002. "Hedonic Pricing Models: A Selective and Applied Review", Working paper.

New Urbanism (NU). 2002. Creating Livable Neighborhoods, Available at http://www.newurbanism.org.

Rajamani, et al. 2003. "Assessing the Impacts of Urban Form Measures in Nonwrok Trip Mode Choices after Controlling for Demographic and Level-of-Service Effects", *Transportation Research Record* 1831: 158-165

Song, Yan and Gerrit J. Knaap. 2003. "New Urbanism and Housing Values: a Disaggregate Assessment", *Journal of Urban Economics* 54: 218-238.

Song, Yan and Gerrit J. Knaap. 2004. "Measuring the Effects of Mixed Land Uses on Housing Values", Regional Science and Urban Economics 34: 663-680.

Song, Yan. 2005. "Smart Growth and Urban Development Pattern: a Comparative Study", *International Regional Science Review* 28, 2: 239-265.

Shultz, Steven D. and David A. King, 2001. "The Use of Census Data for Hedonic Price Estimates of Open-Space Amenities and Land Use", *Journal of Real Estate Finance and Economics* 22, 2: 239-252.

Wassmer, R. 2000. "Urban Sprawl in a U.S. Metropolitan Area: Ways to Measure and a Comparison of the Sacramento to Similar Metropolitan Areas in California and the U.S.", Lincoln Institute of Land Policy Working Paper. https://www.lincolninst.edu/pubs/dl/114_Wassmer00.pdf.