AN EMPIRICAL STUDY OF ATTITUDES TOWARDS GREEN URBAN DEVELOPMENT

A Dissertation
Presented to
The Academic Faculty

by

Lin-Han Chiang Hsieh

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Public Policy

Georgia Institute of Technology
December, 2013

AN EMPIRICAL STUDY OF ATTITUDES TOWARDS GREEN URBAN DEVELOPMENT

Approved by:

Dr. Daniel Matisoff, Advisor
School of Public Policy
Georgia Institute of Technology

Dr. Erik Johnson
School of Economics
Georgia Institute of Technology

Dr. Michael Elliott
School of Public Policy/School of City and Regional Planning
Georgia Institute of Technology

Dr. Douglas Noonan
School of Public and Environmental Affairs
Indiana University-Purdue University Indianapolis

Dr. Robert Kirkman
School of Public Policy
Georgia Institute of Technology

Date Approved: May 6, 2013

ACKNOWLEDGEMENTS

I wish to thank Dr. Doug Noonan, my previous advisor, for helping me develop the main theme of this dissertation. I also wish to thank my committee members, including my advisor, Dr. Dan Matisoff, for their great effort of guiding me through the whole process. I will not be able to have this page published without their generous help.

I would like to thank the Brook Byers Institute for Sustainable Systems, for their support through the U.S. NSF grant 0836046: EFRI – RESIN Project: Sustainable Infrastructures for Energy and Water Supply. I am also thankful to Fulton County Assessor Office for their generous provision of data.

Finally, I would like to thank my family, TSA at Georgia Tech, and students in School of Public Policy, for their support in many ways. Studying abroad is never easy, but you definitely make it much easier for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This study focuses on how spatial circumstances affect property owners' preference toward sustainable urban development, in the form of three-essays. In the first essay, property owners' preference toward the concept of compact development is identified. Compact development is an increasingly popular concept that includes multiple aspects, such as mixed land use, high density, and pedestrian/transit-friendly options. Previous hedonic literature on the comprehensive effect of compact development is limited. Also, spatial dependence in the data, something likely endemic to compact development, has not yet been thoroughly addressed. This study uses a spatial fixed-effect model, a spatial-autoregressive model with auto-regressive disturbances (SARAR), and a spatial fixed-effect SARAR model to determine the price effect of "compactness" in a major U.S. metropolitan area. By analyzing of 47,000 sales records in Fulton County over a decade, this study indicates that home buyers prefer to have smaller, more diffuse greenspace nearby, rather than a large, concentrated greenspace at a longer walking distance. High parcel density and diverse land use is consistently disvalued, and the premium on accessing public transportation is not identified among all models. No specific trend over time has been observed, despite the recession starting in 2008. Finally, a comprehensive index of compactness shows relatively high willingness-to-pay for compact development.

The second essay tests the spatial spillover of signaling within the pursuit of LEED certification. The benefit of pursuing green building certification mainly comes from two aspects: the cost-effectiveness from energy efficiency and the signaling consideration, including the premium on property values, benefits from a better reputation, morality

values, or purely pride. By analyzing all new constructions that received LEED certification from 2000 to 2012 (LEED-NC v2.0 to v2.2) in the U.S., this study tries to identify the size of the signaling effects, and spillover of signaling, as building owners pursue LEED certification. The results show that the signaling effect affects decision making in pursuing LEED certification, especially at scores around thresholds. The size of signaling effects differs among different owner types and different certificate levels. For the Gold level or below, government and non-profit-organization owners value signaling more than do profit-seeking firms. At the Platinum level, there is no significant difference among owner types. This study also finds that the signaling effect clusters spatially for government and profit-seeking firms. Finally, the results show that the cluster of signaling is independent from the cluster of LEED buildings, indicating that mechanisms behind the cluster of signaling are different from those of LEED constructions.

The third essay tests the distance effect on the support for Atlanta BeltLine. Atlanta BeltLine, a large urban redevelopment project currently underway in the center of Atlanta, transforms 22 miles of historical railroad corridors into parks, trails, pedestrian-friendly transit areas, and affordable housing. This study aims to determine the distance effect on the support of Atlanta BeltLine and whether the implement of Tax Increment Financing (TIF) affects the support. The contributions of this exercise are twofold. First, it demonstrates the risks and remedies to missing spatial data by solving the technical problem of missing precise spatial location values. Second, it tests underlying reasons why distance can help explain the level of support that Atlanta BeltLine has received, with striking implications for theories like the Homevoter hypothesis. Survey data used in this study was conducted in summer 2009, about three years after the declaration of the project. The support by both homeowners and renters significantly declines as distance from the BeltLine increases. However, when residents' tendency to use BeltLine parks and transits is entered as a variable, the distance effect disappears. By indicating that the distance effect comes from homeowners' and renters' the accessibility to BeltLine amenities, the result rejects the homevoter hypothesis, which holds that property value increment is the main mechanism behind support. The results also show that whether or not a homeowner or renter is a parent in City of Atlanta affects a person's support of the BeltLine. These results lead to the conclusion that the concern of TIF affecting future school quality hampers the support of the project.

Chapter 1

Introduction

This study focuses on how spatial circumstances affect property owners' preference toward sustainable urban development. Spatial circumstances, such as the quality of the neighboring environment, the distance to certain new amenities, or even the behaviors of other neighbors are usually critical variables affecting the success of sustainable urban developments. For example, the depreciation of high-density living patterns in the U.S. is usually cited/given as one of the main reasons why compact development is not popular in this country (Downs, 2005). Thus, understanding the causal relationship between spatial circumstances and property owners' attitudes is critical to the design and implementation of sustainable urban developments.

Since this study focuses on spatial factors that affect spatially sensitive problems, analytical tools that are suitable for spatial analysis are also critical. Due to the rapid growth of computation ability, methods that were previously difficult to apply have become feasible. This study aims to introduce spatial techniques that have not yet been applied to the field of sustainable urban development. Three empirical essays regarding attitudes towards compact development factors, LEED certification, and the Atlanta BeltLine project, are used as vehicles to address the problem between spatial circumstances and sustainable urban development.

In the first essay, a hedonic price analysis is applied to determine the price effect of compact attributes on the residential housing market. In contrast to the current sprawl city growth pattern, compact development/smart growth is described as the urban development with more mixed land use, higher density, and more pedestrian/transit-friendliness. Numerous previous hedonic studies have addressed the price effect of one or several of these compact attributes above (Irwin & Bockstael, 2001; Li & Brown, 1980; Song & Knaap, 2003), but few of them considered the concept of 'compactness' as a whole or analyze these factors at the same time. Also, since the residential housing market is highly spatially-sensitive, spatial autocorrelation is a critical issue affecting the validity of the model. This study addresses these two issues by analyzing the price effects of compact attributes with spatial regression models, including a spatial fixed-effect model, a spatial autoregressive model with spatial autoregressive disturbances (SARAR), and a spatial fixed-effect SARAR model. The data includes all the sales records of detached-single family houses from 2000 to 2010 in Fulton County, Georgia. By using data at the center of Atlanta, this study provides an interesting view of homebuyers' preferences on compact attributes in a relatively sprawling city.

The second essay examines how the scores of the LEED certification system are affected by different strategies, types of building owner, and possibly spatial spillover. LEED applies a scoring system to classify the levels of certificate. While determining the optimal scores and certificate level, building owners basically have two strategies they can follow: a performance-based strategy or a signaling strategy. If building owners purely follow the strategy of signaling, scores are expected to be right on the threshold levels because additional credits that are not enough for the next level are useless. On the other hand, if building owners purely follow a performance-improving strategy, the scores should be distributed smoothly, since buildings should achieve scores that optimize the energy performance. Also, the behaviors for different owner types, such as governments, non-profit organizations and profit firms, are expected to be different, due to their different goals. This study aims to identify the magnitude of signaling for different types of owners. Finally, this study also tries to identify the spatial spillover for signaling to see whether local competition affects how owners decide which scores/levels to achieve. LEED scoring data from USGBC will be analyzed with both statistical and GIS tools.

The last essay uses survey data on the Atlanta BeltLine project as a vehicle to examine effects of distance and tax increment financing (TIF) on the support of urban redevelopment activities. Atlanta BeltLine, which is a project of re-developing historical rail corridors into pedestrian-friendly rail lines and parks, is expected to boost property values in the nearby neighborhoods. The project is also an example of TIF, where the project cost is generally borne by future tax revenue increases caused by the project itself. On the one hand, homeowners are expected to be more supportive of Atlanta BeltLine as being closer to the BeltLine districts, due to the increased accessibility, or property value. This is an indirect test of the homevoter hypothesis, which holds that homeowners politically support policies/projects that increase their property values. On the other hand, since this re-distribution of tax revenue will possibly lower the proportion of future expenditure on public school, homeowners within the Atlanta BeltLine district, especially those with children, might be less likely to support this project. Survey data on Atlanta BeltLine is used in this study to test the homevoter hypothesis.

The main objective of this thesis project is to enhance the understanding of property owners' attitudes toward sustainable urban development projects. Also, by largely applying GIS and spatial econometric tools, this project aims to further explore the inference of behavior in the spatial perspective. Finally, the result of this study is expected to contribute to future implementation of sustainable urban development projects.

References

Downs, A. (2005). Smart Growth: Why We Discuss It More than We Do It. *Journal of the American Planning Association, 71*(4), 367-378. doi: 10.1080/01944360508976707

Irwin, E. G., & Bockstael, N. E. (2001). The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values. *American Journal of Agricultural Economics, 83*(3), 698-704.

Li, M. M., & Brown, H. J. (1980). Micro-Neighborhood Externalities and Hedonic Housing Prices. *Land Economics, 56*(2), 125-141.

Song, Y., & Knaap, G.-J. (2003). New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics, 54*(2), 218-238.

Chapter 2

Unpacking Compact Development: A Housing Hedonic Price Analysis in Atlanta

Abstract

Compact development is an increasingly popular concept that includes multiple aspects, such as mixed land use, high density, and pedestrian/transit-friendly options. Previous hedonic literature on the comprehensive effect of compact development is limited. Also, spatial dependence in the data, something likely endemic to compact development, has not yet been thoroughly addressed. This study uses a spatial fixed-effect model, a spatial-autoregressive model with auto-regressive disturbances (SARAR), and a spatial fixed-effect SARAR model to determine the price effect of "compactness" in a major U.S. metropolitan area. By analyzing of 47,000 sales records in Fulton County over a decade, this study indicates that home buyers prefer to have smaller, more diffuse greenspace nearby, rather than a large, concentrated greenspace at a longer walking distance. High parcel density and diverse land use is consistently disvalued, and the premium on accessing public transportation is not identified among all models. No specific trend over time has been observed, despite the recession starting in 2008. Finally, a comprehensive index of compactness shows relatively high willingness-to-pay for compact development.

Introduction

Compact development is a growing concept all over the world (Chen et al., 2008; Holden & Norland, 2005; Tang & Yiu, 2010), including in the U.S. (Daniels, 2001). Compact development/smart growth does not have particular patterns or definitions. Contrary to sprawl growth, which is basically the business-as-usual pattern, compact development generally refers to growth with higher density, more mixed uses, and more pedestrian-/transit-friendly design (Burton, 2000; R. H. Ewing, 2008). Not all compact cities have all these characteristics. For instance, Amsterdam, Portland, and Tokyo appear to be very different, though they are all considered to be “compact”. This is no surprise given compact development’s multidimensional nature. This study aims to identify preference for different aspects of compact development as revealed through market prices. Some previous hedonic price studies have discussed the price effects of one or a few of these factors. This study provides a broad test on these characteristics all embedded in the same analysis so that they are directly comparable. In this way, we can better understand the different dimensions of urban compactness to appreciate their relative contributions, at least in the case of a major (and arguably prototypical) U.S. city.

This study conducts a hedonic price analysis to determine the implicit price of compactness. Sales records of detached single family houses in Fulton County, which is the center of the Atlanta metropolitan area, will be used. Compared to cities that already have some compact-development related policies, Atlanta is relatively sprawling and decentralized (Levine & Frank, 2007). The results of this study show that the preferences of homebuyers in Atlanta for compact characteristics are not altogether different from those of homebuyers in more compact cities. Also, the result of an inquiry of residents’ preferences of compact development in Atlanta has several possible policy applications. First, the price premium of compactness can be used in calculating the potential benefit of applying compact development projects in the future. Furthermore, the preferences determined in this study can also be applied in other projection models, such as agent-based models, to see the predicted policy impact of compact development after decades of implementation.

Literature review

There is no generally accepted definition for smart growth/ compact development. A list of definitions used in various studies is summarized below.

Table 2.1 Definitions of smart growth/ compact development

Source	Definition
Burton (2000)	Compact city: ● Mixed use ● Relatively high density ● Efficient public transportation ● Significant walking and cycling
Downs (2005)	Smart growth principles: ● Limiting outward extension of new development ● Raising residential density ● Providing for more mixed land uses and pedestrian-friendly layouts ● Loading the public costs of new development onto its consumers ● Emphasizing public transit ● Revitalizing older existing neighborhoods

Table 2.1 Continue

Source	Definition
Emerine et al. (2006)	Smart growth principles: ● Mix land uses ● Take advantage of compact building design ● Create a range of housing opportunities and choices ● Create walkable neighborhoods ● Foster distinctive, attractive communities with a strong sense of place ● Preserve open space, farmland, and critical environmental areas ● Strengthen and direct development towards existing communities ● Provide a variety of transportation choices ● Make development decisions predictable, fair, and cost effective ● Encourage community and stakeholder collaboration in development decisions
Ewing et al. (2007)	Compact development: ● Median to high densities ● Mixed uses ● Centered development ● Interconnected streets Pedestrian and transit-friendly design

Among these definitions and principles, there are three factors in common: mixed land use, high density, and pedestrian-/transit-friendliness. Estimating the price effect of these three factors can be done using the diversity of neighboring land use, parcel density, the accessibility to public transportation, and greenspace availability, which are used as variables in this study. Few hedonic studies have addressed the impact of mixed land use, the effect of which is not consistent across studies. By using different indices to identify mixed land use, Moorhouse and Smith (1994), Song and Knaap (2003), Koster & Rouwendal (2012) indicate that more mixed land use increases property values. However, Geoghegan et al. (1997), who used entropy indices, did not find diversity significant. Also, Matthews and Turnbull (2007) use the proportion of non-residential use to identify mixed land use and reported that the effect is sensitive to model and method selection. This study applies an entropy index to represent the diversity of land use, similar to methods used in Yeh and Li (2001), Song and Knaap (2003), and Matthews and Turnbull (2007).

Residential density has been controlled for in a few hedonic studies as part of neighborhood characteristic control, but it is seldom discussed as a main theme (Li & Brown, 1980; Matthews & Turnbull, 2007; Shultz & King, 2001; Song & Knaap, 2003). In these studies, the effect of residential density on property value is always significant and negative. In this study, the total parcel number within an 800-meter distance is counted as parcel density in the neighborhood. Also, the population density at the block group level is controlled in the empirical model.

Accessibility to public transportation is directly related to the characteristic of transit-friendliness. Previous literature has been mixed about the impact of railway stations on property value (Bowes & Ihlanfeldt, 2001; Ryan, 2005; Song & Knaap, 2003). The effect ranges from negative to insignificant to positive. Debrezion et al. (2007) conducted a meta-analysis on these three studies, and indicate that the variations of effect are (can be?) attributed to the nature of the data, the particular spatial characteristics, the temporal effects, and the methodology. They concluded that a 4.2% premium on residential property values is average if a local station is within a 400-meter range (1/4 miles).

Accessibility to greenspace is another compact characteristic in this study for two reasons. First, greenspace is mentioned in most of the literature in Table 1, in terms of open space accessibility and neighborhood walkability. Second, numerous studies find an effect of greenspace on the housing market (Anderson & West, 2006; Cho et al., 2008; Geoghegan, 2002; Morancho, 2003). After reviewing a significant number of hedonic price analyses in the U.S. from early 1980s to 2000s, Crompton (2005) concluded that these studies “unequivocally support the contention that parks and open space contribute to increasing proximate property values.” Despite the difficulty in comparing the sizes of price effects across these different studies, Crompton boldly suggests that parks have a positive impact of 20% on proximate property value. Another meta-analysis done by Brander and Koetse (2011) confirms the positive price effect of open space. Using population density as a proxy to crowdedness, they also suggest that controlling for population density could be very important to identify the marginal value of park proximity.

The compact characteristics included in this study are all inherently spatial. One critical issue of estimating these spatially correlated variables in a hedonic housing price analysis is endogeneity or nonspherical errors. As Irwin and Bockstael (2001) have stated, there are two possible endogeneity problems while measuring the effect of open space on residential property values. First, privately owned open space is endogenous. Since the open space is privately owned and has its own value, the value of the open space is affected by itself, which causes an endogeneity problem. This problem is mitigated in this study because all the parks in the dataset are public ones. Second, unobserved spatial correlation could be another source of endogeneity. This is more critical in this study. For example, it is reasonable to expect that properties with parks nearby share other "neighborhood quality" factors regarding their location (Irwin and Bockstael, 2001). If neighborhood quality is imperfectly observed and appears in the error term, estimators from OLS may be biased. Evidence supports this concern about spatial dependence in hedonics models in several studies (e.g., Dubin and Sung, (1990), Can (1990), Basu and Thibodeau, (1998).

Several different approaches have been applied to address this problem, including an instrumental variable approach (IV) (Irwin, 2002), a stochastic approach (Tse, 2002), and moving window approaches (Páez et al., 2008). In the current study, spatial fixed-effect models and spatial econometric approaches are applied. Estimating hedonics models with spatial fixed effects is a popular approach due to its tractability. Anderson and West (2006) introduced fixed-effect model to control for observed and unobserved neighborhood characteristics. They found that in the Minneapolis-St. Paul metropolitan area, the value of proximity to open space is higher in neighborhoods that are dense, near the central business district, high-income, and high crime. We also apply a fixed-effect model in this study as a comparison to OLS and spatial regression models.

Another alternative to deal with the endogeneity issue is using spatial econometrics. Anselin (2001) developed spatial lag and spatial error models to consistently estimate models using spatial dependence. This approach has been applied to hedonic house price models (Anselin & Le Gallo, 2006), but has been rarely used to estimate both the greenspace and compactness effects on the housing market. This is another main contribution of this study. Both spatial lag and spatial error models are estimated, and the spatial error model is proved to fit these data better.

To sum up, there are two gaps in previous studies. First, many studies have estimated the price effect of one or only a few compactness attributes, but few studies have investigated the complex effect of compact development embedded in the same model. Second, new approaches to address the spatial autocorrelation issue have rarely been applied to this topic. This study aims lessen these two gaps by estimating the effect of compactness in the Fulton County housing market, using both a fixed-effect model and a spatial econometric to eliminate the endogeneity caused by spatial autocorrelation.

Models

This study uses a typical hedonic price model to describe the variation in sales prices:

$$\ln(\text{sales price})_i = (compactness)_i\beta_1 + X_i\beta_2 + u_i \qquad (1)$$

where β_1 and β_2 are estimators; *X* is a vector of other control variables, including property characteristics, neighboring amenities, location variables, and time-related variables. *u* is the error term including unobservable spatial factors, and the suffix *i* denotes individual sales record. Since compactness measurements (like distance to the nearest park) and the error term are both correlated with unobservable spatial characteristics, this spatial autocorrelation problem is likely to bias all estimators in the model.

To mitigate the spatial autocorrelation problem, we use a spatial fixed-effect model here. In order to explain this fixed-effect model, we first re-write equation (1) as

$\ln(\mathrm{sp})_{ij} = (compact)_{ij}\beta_1 + X_{ij}\beta_2 + u_{ij}$,Wwhere j denotes the census tract wherethat the property i is located in. The spatial fixed-effect model can be written as:

$$\left[\ln(sp)_{ij} - \overline{\ln(sp)_j}\right] = \left[(compact)_{ij} - \overline{(compact)_j}\right]\beta_1 + \left(X_{ij} - \overline{X}_j\right)\beta_2 + \left(u_{ij} - \overline{u}_j\right),$$

where $\overline{\ln(sp)_j}$ denotes the average logged sales price in census tract j, $\overline{(comapct)_j}$ represents the averaged measurement of compactness in census tract j, and so on. Because the means is subtracted at the tract level, the fixed-effect model deals only with the variation between individuals within census tract j, and wipes out any differences between each census tract j. Theoretically, the fixed-effect model can eliminate all the spatial correlation problems above the census tract level. Any spatial correlation within each census tract will remain in this fixed-effect model.

An alternative model, the spatial-autoregressive model with spatial-autoregressive disturbances (SARAR), is applied to further address the spatial autocorrelation problem:

$$\ln(\mathrm{sp})_i = \lambda W[\ln(sp)_i] + (compact)_i\beta_1 + X_i\beta_2 + u_i$$

$$\mathrm{u}_i = \rho W u_i + \varepsilon_i,$$

where W denotes a spatial weights matrix[1], and λ and ρ are estimators. The term $\lambda W[\ln(sp)_i]$ is supposed to capture the spatial dependence within the lagged sales

[1]In this study, inverted-distance spatial weights matrices are used because the concept of distance is critical. Inverted distance that is smaller than 1/1500 m is truncated for simplicity.

price, $\ln(sp)_i$. The term $\rho W u_i$ is used to absorb the spatial dependence within the error term. The effectiveness of SARAR regarding spatial autocorrelation depends on the definition of "neighbors", i.e. the selection of spatial weight matrix W. In this study, an inverse-distance weight matrix is used, which implies that the influence of neighbors weakens with distance. Under this setting, the SARAR model emphasizes the bias created by spatial autocorrelation at the neighborhood level.

As a spatial fixed-effect model works only for spatial autocorrelation above the tract level, and the SARAR model focuses on that at the neighborhood level, combining these two methods is expected to eliminate most of the spatial-autocorrelation problem. The combined spatial fixed-effect SARAR model (SFEAR) can be written as follows:

$$\begin{aligned}\left[\ln(sp)_{ij} - \overline{\ln(sp)_j}\right] \\ &= \lambda W\left[\ln(sp)_{ij} - \overline{\ln(sp)_j}\right] + \left[(compact)_{ij} - \overline{(compact)_j}\right]\beta_1 \\ &+ \left(X_{ij} - \overline{X_j}\right)\beta_2 + \left(u_{ij} - \overline{u_j}\right)\end{aligned}$$

$$(\mathrm{u}_{ij} - \overline{u_j}) = \rho W(\mathrm{u}_{ij} - \overline{u_j}) + \varepsilon_{ij} - \overline{\varepsilon_j},$$

where all the terms in the SARAR model are subtracted by corresponding tract-level means.

Finally, the fact that the data in this research spans a long period of time raises the concern that later sales might have influences on previous ones in the model. These counterfactual effects over-identify spatial correlations and, thus, likely make spatial autocorrelation become too statistically significant. This might not have been too serious an issue in previous hedonic price studies because their data usually ranged within one or two years (Anselin & Le Gallo, 2006; Basu & Thibodeau, 1998; Won Kim et al., 2003). However, pooling a decade of data might present a problem (see., e.g., Kuminoff et al., 2010). To address this concern, this study further introduces a temporal fixed-effect into the SFEAR model. This temporal-SFEAR model (TSFEAR) works just as a SFEAR model does, and can be written as

$$\begin{aligned}\left[\ln(sp)_{ijt} - \overline{\ln(sp)_{Jt}}\right] \\ &= \lambda W\left[\ln(sp)_{ijt} - \overline{\ln(sp)_{Jt}}\right] + \left[(compact)_{ijt} - \overline{(compact)_{Jt}}\right]\beta_1 \\ &+ \left(X_{ijt} - \overline{X_{Jt}}\right)\beta_2 + \left(u_{ijt} - \overline{u_{Jt}}\right)\end{aligned}$$

$$(\mathrm{u}_{ijt} - \overline{u_{Jt}}) = \rho W(\mathrm{u}_{ijt} - \overline{u_{Jt}}) + \varepsilon_{ijt} - \overline{\varepsilon_{Jt}},$$

where suffix *t* refers to the year of sale. Since TSFEAR controls for both the temporal and spatial fixed-effect, it examines only variations within a tract in a certain year. In other words, all the observations are compared as if they were in the same tract and sold in the same year. However, this powerful method raises another potential drawback of over-controlling. Dividing 45,000 observations into more than 1,500 groups (per tract per year) leaves each group with less than thirty observations, which might limit the precision of group means.

Data

This study includes all the sales of detached-single family houses in Fulton County from 2000 to 2010. The data used in this study is a combination of four sources: sales records, assessor’s appraisal data, parcel maps, and block group level census data.

The sales data is originally from the Fulton County Board of Assessors, from 2000 to 2010. To really capture the price effect in a real market, this study includes only arms-length sales. Sales that are excluded are those involving related individuals or corporations, multiple jurisdictions, foreclosures, banks as the seller/buyer, land contracts, conveying to adjoining property, burned/razed after sale, a deed of gift, a trade/exchange, portfolio sales, sales including multiple parcels, or any sales that are less than $5,000,. Immergluck (2009) applies a very similar rule to the same dataset. In order to further control the impact of market fluctuation on price, the current paper introduces the Case-Schiller home price index for the Atlanta metropolitan area to adjust the sale prices. Case and Shiller create the indices based on a weighted repeat-sales method (Case & Shiller, 1989). Adjusting sales price with this index should eliminate market-wide shifts in price over time.

Assessor's data are also from the Fulton County Board of Assessors, from 2000 to 2011. This dataset contains numerous property characteristics and neighborhood variables such as the build year, number of stories, lot size, indoor living area, number of rooms, construction style, type of external wall, location (CBD/major strip/secondary strip), and front (major road/secondary street). Merging assessor's data with sales data provides a basic dataset for hedonic price analysis. Some assessor's data have many missing values, especially for stories, living area, bedrooms, family rooms, bathrooms, and basements. The conditional-mean-imputation method is used to deal with this missing Xs issue (Little, 1992).

The block-group-level demographic data is from the 2000 Census. Factors such as median household income and population density serve as proxies for neighborhood quality. Parcel maps of Fulton County are used to calculate all the spatial-related variables such as distance to the nearest park, number of parks nearby, parcel density, and the diversity of land use. The phrase "neighborhood" or "nearby" is defined as an 800-meter-radius buffer circle. For each parcel, if there are 2 parks and 670 other parcels within its 800-meter buffer, the number of parks nearby is counted as 2, and the parcel density would be 670. 800 meters, i.e. a half mile, is selected because it takes less than ten minutes to walk through. This concept is borrowed from the calculation of Walk Score (http://www.walkscore.com/professional).

The diversity of land use is measured with an entropy index, following the method of Yeh and Li (2001):

$$diversity = -\sum_i^n P_i \log\left(\frac{1}{P_i}\right)/\log(n),$$

where P_i refers to the share of a certain type of land use, referring to property classes determined by the assessor's office, within each half-mile buffer. In this study, land uses are categorized into four classes: residential, commercial, industrial, and public use. Thus, *n* is four in this case. Under this definition, diversity ranges from zero to one. In the case of four categories uniformly distributed, the diversity score is one. In the case that there is only one type of land use class in the buffer, diversity equals zero. The average diversity score in this sample is 0.17, given that nearly ninety percent of the parcels are coded as residential use.

Since parcel boundaries can change over time (the number of parcels in Fulton County has increased by 30% in the past decade), parcel maps from 2002, 2006, and 2010 are used to capture the different neighboring parcels in different sales years.

Generally, the closest available map is selected. For sales that happened before 2004, the 2002 parcel map is used to obtain all the spatial variables. For sales between 2005 and 2008, the map of 2006 is used. For sales after 2009, the map of 2010 is used.

The description of variables and descriptive statistics are listed in Table 2.2. The dependent variable is logged sales price after adjustment. The measurement of compactness can be divided into three parts: greenspace, neighboring parcel characteristics, and transit. Greenspace variables include distance to the nearest park, the number of parks nearby, and the total park area nearby. Parcel density and the diversity of land use are used to identify the effect of neighboring parcel characteristics. The distance to the closest Marta railway station is used to reflect the effect of public transit. Factors including spatial relationships, property characteristics, locations, and time-related variables are used to control for other effects in all models.

Table 2.2 Descriptive Statistics

Variable description	Obs	Mean	Std. Dev.
Dependent variable			
Sales price	44743	266541	259199
Green space variables			
Distance to the closest park (km)	44743	1.656	1.572
Number of parks within the 0.8km buffer	44743	1.362	2.335
Sum of park area within the buffer (acre)	44743	225	1122
Neighboring parcels variables			
Number of parcels within the buffer	44743	668	481
Diversity score	44743	0.174	0.169
Transit variables			
Distance to the closest Marta station (km)	44743	14.486	11.158
Other spatial variables			
Distance to CBD (km)	44743	33.06	20.48
Distance to highway (km)	44743	6.372	6.393
Distance to closet hydro (km)	44743	0.763	0.733
Ddistance to ATL airport (km)	44743	25.74	16.61
Important location variables			
Median household income in BG	44743	73807	43712
Population density in BG	44743	2541	2094
Property characteristics			
Lot size (acre)	44743	0.40	0.419
Imputed stories	44743	1.485	0.403
Imputed indoor area (sq. ft.)	44743	7.542	0.415
Age of house	44743	23.46	26.47
Imputed number of bedrooms	44743	3.297	0.787
Imputed number of family rooms	44743	0.518	0.438
Imputed number of fixed bathrooms	44743	2.192	0.825
Basement: full	44743	0.190	0.393
Topo:level	44743	0.613	0.487
Topo:above street	44743	0.144	0.351
Topo:below street	44743	0.031	0.174

* Statistics of other control variables (location and time-related dummies) are listed in Appendix

Results

To identify the price effect of compact indicators, this study introduces a series of regression models, including OLS, spatial fixed-effect (SFE), SARAR, spatial fixed-effect SARAR (SFEAR), and temporal-spatial fixed-effect SARAR (TSFEAR) models. The results of six compact indicators for each model are listed in Table 2.3. All the variables listed in Appendix A1 are included in models as control variables.

Though the magnitudes of coefficients vary among models, the direction of effects and significances is generally consistent. However, the result for the SFEAR and TSFEAR models is preferred in this study, regarding controlling for spatial autocorrelation and over-identification. OLS serves as a nice benchmark against prior work, but it does not have any control for spatial autocorrelation. The spatial fixed-effect model eliminates spatial autocorrelation above the tract level, but the autocorrelation at the neighborhood level remains. Maping local indicators of spatial association (LISA) is a good practice to help further identify the problem of spatial autocorrelation within OLS and spatial fixed-effect models (Anselin 1995). Further details of the LISA maps and the results are given in Appendix 2.3. Combining the concern from econometric theory and the visual proof from LISA maps, the estimators generated by OLS and spatial fixed-effect models are viewed as inconsistent. As mentioned previously, the SARAR model is supposed to pick up most spatial autocorrelation with spatial lagged and error terms in the model. However, due to the design of the spatial weight matrix, SARAR is the most effective at eliminating autocorrelation at the neighborhood level. Thus, none of the above models addresses the problem thoroughly.

As a combination of SFE and SARAR, the SFEAR model is designed to eliminate spatial autocorrelation at both the above-tract level and the neighborhood level. As mentioned previously, TSFEAR further addresses the over-identification problem by introducing a temporal fixed-effect into the SFEAR model. However, since TSFEAR also introduces the concern of over-controlling as well, these two models are more appropriate to look at side-by-side, instead of relying on either one of them alone.

Table 2.3 Result of Compact Development Variables

	OLS	SFE	SARAR	SFEAR	TSFEAR
Obs.	44,743	44,743	44,743	44,743	44,743
Groups		161		161	1,513
Dist. to park (logged)	-0.022 ***	-0.006 **	-0.018 ***	-0.007 **	-0.004
Number of park	0.014 ***	0.009 ***	0.009 ***	0.007 ***	0.008 ***
Tot. park area (/10,000 acre)	-0.025	0.020	0.002	0.010	0.013
Parcel density (/100 pcls)	-0.006 ***	-0.009 ***	-0.004 ***	-0.006 ***	-0.007 ***
Land Use diversity	-0.007	-0.118 ***	-0.111 ***	-0.109 ***	-0.136 ***
Dist. to Marta (logged)	-0.040 ***	-0.012	-0.082 ***	-0.014	-0.012
Lambda			0.011 ***	2.609 ***	-0.004
rho			7.956 ***	1.282 ***	5.613 ***

*p < .10. **p < .05. ***p < .01.

After the number of parks nearby and the summation of park area are controlled for, distance to the closest park has a significant negative effect in the SFEAR model, but not in the TSFEAR model. Since both sales price and the distance to a park are logged, the interpretation of the coefficient can be viewed as elasticity. If all the other variables are held constant, doubling the distance of the closest park (from 1 to 2 kilometers, on average) drops the property value by 0.7 percent. The negative effect confirms that having a park nearby is a premium to property value, a result that is consistent with previous studies (Morancho, 2003; Song & Knaap, 2003). Another issue about the price effect of park distance is the inconsistency of statistical significance between the SFEAR and TSFEAR models. One way to explain the insignificance of the TSFEAR model is that the significance of the SFEAR model comes mainly from the over-identification between years. The effect of park distance matters only when later observations are able to influence previous sales, which is prohibited in the TSFEAR model. On the other hand, the price effect of park distance could be insignificant for the TSFEAR model because of over-controlling. Dividing observations into too many groups raises the standard errors, thus erasing the significance. The mechanism behind this difference is worth further investigation, but it is fair to state that the price effect of park distance is not as significant as other indicators, which are consistently significant for both models.

Having one more park nearby increases the sales price by 0.7 percent if all the other variables hold constant. Since the summation of total park area is also controlled in the model, it is more precise to describe this park number effect as “dividing one existing park into two.” In this sense, the effect is even stronger than it appears since there is no new park area created. The effect is statistically significant for both the SFEAR and TSFEAR models. This result implies that comparing to distance to the nearest park to the number of parks nearby has a more significant positive effect on property values. On the other hand, in both models, summation of the park area nearby does not have a significant premium on housing price. The evidence of homebuyers’ preferences for larger park area is only weak after controlling for park distance and park numbers.

The combined results of the three park variables suggest that homebuyers tend to prefer having parks nearby and that more parks within a walking distance are better. After controlling for the distance and number of parks, the size of total park area is relatively not as critical. However, if increasing the number of parks also implies increasing park acreage, the effect of total park acreage can be viewed as absorbed by the effect of park numbers. In other words, from the standpoint of stimulating property values, creating more small parks in the neighborhood is more efficient than developing a large park site.

Parcel density is defined as the number of parcels within the 800-meter-radius buffer, representing the walking-distance of the neighborhood. In both models, the effect is weak but significant. Holding all other variables constant, having an additional 100 parcels within an 800-meter buffer will decrease the housing price by only 0.6 percent. Given the fact that the average parcel density in this sample is 670, the effect of adding parcels is relatively trivial. Put another way, the 100 additional parcels raise the number of parcels in the buffer by 15%, but lower the price by only 0.6%.

The effect of land use diversity is also negatively significant in both models. This result is not consistent with Moorhouse and Smith (1994), Song and Knaap (2003) orKoster & Rouwendal (2012), who indicate that diversity is a premium to housing prices. The inconsistency is possibly due to different sample characteristics. While Moorhouse and Smith focused on the row houses in Boston, Song and Knaap analyzed the New Urbanism area in Portland, and Koster & Rouwendal (2012) studied the Rotterdam City Region in the Netherlands. Mixed land use is expected to be valued more highly in these relatively dense areas. On the other hand, the sample of this study is limited to detached single-family houses in the metro Atlanta area, where the preference for homogeneous land use is not surprising. Thus, this result implies that homebuyers' preferences on mixed land use are diverse and are highly sensitive to specific areas and housing types.

The distance to the nearest MARTA rail-transit station is a proxy for the accessibility to public transportation, measured in logged distance (kilometers). The effect of public transportation is surprisingly not significant in both models. One possible explanation is that the nearest MARTA station is so far for most observations, that observations within a census tract share a very similar distance. Since SFEAR and TSFEAR measure only how different an observation is from the tract mean, this lack of variation in distance results in the insignificance in coefficient. The significant and negative coefficient of distance to MARTA for the SARAR model, which refers to a positive attitude about accessing public transportation, serves as evidence of this explanation. Since the SARAR model does not involve the concept of demeaning at the tract level, the effect stays. The comparison of results for different models indicates that homebuyers in Fulton County value the accessibility to public transportation positively. However, this preference is not significant at the neighborhood level, since the distance to the nearest station does usually not vary enough to show any significance within a neighborhood.

Instead of controlling for the variation in homeowner attitudes among multiple years (by TSFEAR), this study also tries to identify the variation over time. By implementing the SFEAR model for individual years and comparing the coefficients, we can identify the trend of preference change over time can, if there is any. Table 2.4 shows the result of compact indicators for the SFEAR model, including year-by-year coefficients. For each compact indicator, eleven samples, including the pooled full sample and ten year-by-year subsamples, are run separately. Results of 2010 sales are not reported because the sample size is too small. They are still included in the pooled model, however.

Table 2.4 Result of Compact Development Variables for SFEAR Model

	Obs.	dist. to park (logged)	Number of park	Tot. park area (/10,000 acre)	Parcel density (/100 pcls)	Land Use diversity	dist. to Marta (logged)
Pooled	44,743	-0.007 **	0.007 ***	0.010	-0.006 ***	-0.109 ***	-0.014
00	3,889	-0.017	0.023 ***	-0.021	-0.011 ***	-0.290 ***	0.046
01	4,210	-0.015	0.019 ***	0.146 *	-0.001	-0.212 ***	-0.036
02	5,174	-0.002	0.003	-0.065	-0.005 *	-0.145 **	-0.003
03	5,208	-0.007	0.012 *	-0.006	-0.005 *	-0.103 *	-0.035
04	6,095	-0.027 ***	-0.01	0.208 ***	-0.008 ***	-0.111 ***	0.046
05	4,744	0.005	0.006	0.061	-0.011 ***	-0.166 ***	0.018
06	4,156	-0.002	0.003	0.052	-0.003 *	-0.076	-0.069 ***
07	4,481	0.007	-0.006	-0.052	-0.002	-0.115 **	-0.016
08	3,983	-0.001	0.003	-0.014	-0.005 **	-0.061	-0.133 ***
09	2,767	-0.005	0.002	-0.026	-0.002	-0.082	-0.090 *

*p < .10. **p < .05. ***p < .01.

The magnitudes and significances of indicators for the pooled model are generally the average effects of ten year-by-year models. There is no specific trend observed behind the randomly fluctuated preference change over time. It is worth noting that the recession that started in 2008 has not seemed to dramatically affect homebuyers' preferences on compact attributes. The magnitudes of effects after 2008 are still within the range of previous variation, though the significances disappear for most indicators. The only significant change after 2008 occurs in the preference for public transportation.

The accessibility to MARTA is positively valued as the distance effect is strongly negative. However, there is no evidence that this trend is related to the recession.

Finally, after the price effects of individual compact indicators are determined, a series of tests is performed to identify the comprehensive effect of all six compact indicators. First, a joint significance test is performed to test the hypothesis that these six variables have an influence on housing price. A χ^2-statistic of 124.5 for the pooled SFEAR model strongly rejects the null hypothesis that these six variables are not correlated with housing prices. Furthermore, this current study introduces a compactness index, which refers to the product of multiple compact indicators, into both the SFEAR and TSFEAR models. The compactness index can be written as:

$$\begin{aligned} Compactness\ Index &= (dist.to\ park\ index) \times (num.of\ parks\ index) \\ &\times (parcel\ density\ index) \times (diversity) \times (dist.to\ MARTA\ index) \end{aligned}$$

where the descriptions of indices for single variables are listed in Table 2.5.

Table 2.5 Descriptions of Indices for compact indicators

Index Name	Scale	Transformation	Descriptions
Dist. to park index	0 to 1	$= \frac{\log(1/_{Dist\ to\ park})}{Max(Dist\ to\ park)}$	1. Invert and nature log the distance. 2. Distances larger than 0.8 km are marked as 0. 3. Scale down to 0 to 1.
Num. of parks index	0 to 1	$= \frac{(Num\ of\ parks)}{Max(Num\ of\ parks)}$	1. Scale down to 0 to 1.
Parcel density index	0 to 1	$= \frac{(Num\ of\ parks)}{Max(Num\ of\ parks)}$	1. Scale down to 0 to 1.
Diversity index	0 to 1	-	
Dist. to MARTA index	0 to 1	$= \frac{\log(1/_{Dist\ to\ MARTA})}{Max(Dist\ to\ MARTA)}$	1. Invert and nature log the distance. 2. Distances larger than 0.8 km are marked as 0. 3. Scale down to 0 to 1.

Constructing the compact index requires that the indicators be preprocessed at different level so that they are comparable to each other. The distance to the nearest park is first inverted and nature logged so that larger numbers refer to being closer to parks. This process transfers the sign of the coefficient to be consistent with other indicators. For example, “preferring compactness” is now positive. Secondly, all distances larger than 0.8 kilometers are dropped from the index, and the inverse distance is assigned as zero, because this compact index is designed to reflect only the compactness at the neighborhood level. Finally, this indicator is scaled down to zero-one scale to be comparable. The summation of park area is not included in the compact index since it is not statistically significant in almost all the models. The number of parks and parcel density only need the pre-process of scaling down. Diversity needs no pre-process, because it is at the zero-one scale by definition. The distance to nearest MARTA station is processed simply as the distance to the nearest park. Since all five indicators now scale from zero to one, the compact index, which is the product of all five, is also under a zero-one scale. The results of compactness index for SFEAR model are listed as Table 2.6.

Table 2.6 Results of compactness index

Obs: 44,743

Groups: 161

	Coef.	Std. err.	P>|z|
Compactness Index (with dist. to MARTA)	2.367	16.044	0.883
Compactness Index (without dist. to MARTA)	10.133	1.763	0.000

The result shows that this compactness index is not statistically significant for the SFEAR model. However, after the distance to MARTA is omitted, the four-indicator index is strongly positive and significant. The coefficient is difficult to interpret since all four indicators are multiplied and scaled, but it strongly suggests that as a whole, these four dimensions (distance to park, number of parks, density, and diversity) compose a sort of *compact development* indicator that helps to explain housing prices in Atlanta.

Robustness

Several robustness checks are performed for the purpose of clarity. First, there are concerns about the result that diversity is a disamenity for homebuyers in Atlanta. One might consider industrial land use to be the main driving force of the negative preference. However, after industrial land use is dropped from the formation of the diversity index[2], the result is nearly identical, thus rejecting the hypothesis. Another concern is that the negative sign shown in this study is only an averaged effect at the margin. The price effect of diversity might differ for different types of diversity, or differ from neighborhood to neighborhood. Taking advantage of the numerous available observations, this study performs sensitivity checks by dividing observations into subsamples in multiple cases. In the first case, observations are categorized into four groups based on the median sales price in the neighborhood. The hypothesis is that diversity is more preferred in neighborhoods with higher prices. The result shows that the price effect of diversity is insignificant for the lowest-price group, while it is negatively significant in all other three groups. Thus, the hypothesis is rejected.

In the second case, observations are divided into four groups, based on diversity level. The theory is that homebuyers might value different levels of diversity in different ways. The result shows that diversity is positively valued only if the diversity score is less than 0.035. This result indicates that diversity is an amenity in the case of homogeneous environments, while the preference disappears dramatically when the diversity score rises. However, the average effect still shows that at the margin, diversity is generally disvalued for homebuyers in Atlanta.

[2] The diversity index without industrial use is only constructed from three land use type: residential, commercial, and public use. The highest score remains as one (one-third of parcels for each type), but the average score is slightly lower, since now only a small fraction of land use types (about 1.5 percent) is not included in the calculation.

Another concern regarding the interpretation/accuracy of the data is the imputation process while dealing with the missing value issue. The concern is that the imputation might generate additional noises that affect the result. Without any imputation, the sample size drops from 44,743 to 25,583. Most of the results are reasonably similar with and without imputation, including the magnitude of coefficients and significances. The only exception is diversity. Diversity is not statistically significant in the case of no imputation.

The covariance of estimators among three park-related variables, and between parcel density and population density, is determined to check for multicollinearity. The highest correlation is between the distance to park and the number of parks nearby, which is only 0.47. Collinearity does not seem to be a serious problem in this case.

Conclusion

Measures commonly embedded in the notion of compact development explain housing price variation in a major U.S. city over the past decade. The role of compactness indicators, however, may not be exactly as anticipated regarding the magnitudes of the effects and sometimes even the direction of the effects. For instance, total park area on its own plays a minimal role in explaining market prices. Unpacking the composite concept of compact development reveals a richness that defies simplistic characterization and also raises some interesting questions.

This study shows that homebuyers seem to care much more about being closer to their closest nearby park and having more parks within walking distance. Interpretation of these park variables should be done carefully. Extending this preference from parks to urban greenspace is fair because the effects of other urban greenspace, such as privately owned greenspace, should also be absorbed into this park effect. But following the same logic, it can be argued that the effect of other unobserved neighborhood amenities, like neighborhood walkability, may also be included in these park effects. It may be that these park variables do not just represent the preference to public parks, but also reflect the comprehensive effects of urban greenspace and other neighborhood amenities.

Homebuyers in metro Atlanta generally disvalue dense neighborhoods with mixed land use at the margin, though low-level of diversity is actually preffered. The result also shows that the premium of accessing public transportation does not exist in the pooled model, but a slight increase in appreciation of transit access seems to have begun during the recession. This trend might imply a potential benefit of expending the public transportation system, even in a relatively sprawling city like Atlanta. However, the concept of compactness is generally valued by homeowners, based on the result of a comprehensive compactness index. But, it cannot yet be claimed that people are starting to prefer to live compactly now. Future research with data that controls more for property quality is needed to further identify the cause of this phenomenon.

It should be emphasized that the relative contributions of these different dimensions of compactness to home values are based on an equilibrium analysis and represent marginal implicit prices. The estimates tell us how, given Atlanta's urban conditions, a marginal increase in one attribute compares to another. This is not to say that more diffuse parks rather than larger consolidated parks, for instance, will always result in higher home prices. On the margins they do, but we might expect diminishing returns for compactness in any of these dimensions – and if promoting one dimension leads to more gains than promoting another, then relative magnitudes might yet change as the status quo is redefined. Adding another park, for example, may boost property values more than adding another transit line when parks are relatively scarce. This could reverse as parks become more abundant. Compact development is still fundamentally a balancing act along multiple dimensions.

This study has several policy applications. First, at the margins at least, people prefer to have more small greenspace within walking distance rather than a single large park. This is information worth considering in the decision making process of park allocation. Since small parks in the neighborhood stimulate housing prices, city governments should also have incentives to invest in small public parks in lower quality neighborhoods, and expect growing property value increases. Secondly, though mixed land use is disvalued at the margin, it has a positive pricing effect when the diversity level is low. This result implies that there might be an optimum level of mixed land use, at least for this sample. Determining this optimum level would be valuable for future land use design. This study also shows that compactness at the neighborhood level, as a comprehensive concept, has a positive impact on housing price. Given the fact that pursuing compact development or city sustainability can actually increase property value and, presumably, tax revenue, compact growth should be discussed more at the municipal level. The conclusions of this study are even more interesting, considering this study takes place in a sprawling city like Atlanta.

References

Anderson, S. T., & West, S. E. (2006). Open space, residential property values, and spatial context. *Regional Science and Urban Economics, 36*(6), 773-789.

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis, 27*(2), 93-115.

Anselin, L. (2001). Spatial Effects in Econometric Practice in Environmental and Resource Economics. [Article]. *American Journal of Agricultural Economics, 83*(3).

Anselin, L., & Le Gallo, J. (2006). Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects. [Article]. *Spatial Economic Analysis, 1*(1), 31-52.

Basu, S., & Thibodeau, T. G. (1998). Analysis of Spatial Autocorrelation in House Prices. *The Journal of Real Estate Finance and Economics, 17*(1), 61-85.

Bowes, D. R., & Ihlanfeldt, K. R. (2001). Identifying the Impacts of Rail Transit Stations on Residential Property Values. *Journal of Urban Economics, 50*(1), 1-25. doi: 10.1006/juec.2001.2214

Brander, L. M., & Koetse, M. J. (2011). The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results. *Journal of Environmental Management, 92*(10), 2763-2773. doi: 10.1016/j.jenvman.2011.06.019

Burton, E. (2000). The Compact City: Just or Just Compact? A Preliminary Analysis. *Urban Studies, 37*(11), 1969-2006. doi: 10.1080/00420980050162184

Can, A. (1990). The Measurement of Neighborhood Dynamics in Urban House Prices. *Economic Geography, 66*(3), 254-272.

Case, K. E., & Shiller, R. J. (1989). The Efficiency of the Market for Single-Family Homes. *The American Economic Review, 79*(1), 125-137.

Chen, H., Jia, B., & Lau, S. S. Y. (2008). Sustainable urban form for Chinese compact cities: Challenges of a rapid urbanized economy. *Habitat International, 32*(1), 28-40.

Cho, S.-H., Poudyal, N. C., & Roberts, R. K. (2008). Spatial Analysis of the Amenity Value of Green Open Space. *Ecological Economics, 66*(2-3), 403-416.

Daniels, T. (2001). Smart Growth: A New American Approach to Regional Planning. [Article]. *Planning Practice & Research, 16*(3/4), 271-279.

Downs, A. (2005). Smart Growth: Why We Discuss It More than We Do It. *Journal of the American Planning Association, 71*(4), 367-378. doi: 10.1080/01944360508976707

Dubin, R. A., & Sung, C.-H. (1990). Specification of hedonic regressions: Non-nested tests on measures of neighborhood quality. *Journal of Urban Economics, 27*(1), 97-110.

Emerine, D., Shenot, C., Bailey, M. K., Sobel, L., & Susman, M. (2006). This Is Smart Growth: Smart Growth Network.

Ewing, R., Bartholomew, K., Winkelman, S., Walters, J., & Chen, D. (2007). Growing Cooler: The Evidence on Urban Development and Climate Change: Urban Land Institute.

Ewing, R. H. (2008). Characteristics, Causes, and Effects of Sprawl: A Literature Review
Urban Ecology. In J. M. Marzluff, E. Shulenberger, W. Endlicher, M. Alberti, G. Bradley, C. Ryan, U. Simon & C. ZumBrunnen (Eds.), (pp. 519-535): Springer US.

Geoghegan, J. (2002). The value of open spaces in residential land use. *Land Use Policy, 19*(1), 91-98. doi: 10.1016/s0264-8377(01)00040-0

Geoghegan, J., Wainger, L. A., & Bockstael, N. E. (1997). Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics, 23*(3), 251-264. doi: 10.1016/s0921-8009(97)00583-1

Holden, E., & Norland, I. T. (2005). Three challenges for the compact city as a sustainable urban form: Household consumption of energy and transport in eight residential areas in the greater Oslo region. [Article]. *Urban Studies, 42*(12), 2145-2166.

Immergluck, D. (2009). Large Redevelopment Initiatives, Housing Values and Gentrification: The Case of the Atlanta Beltline. *Urban Studies, 46*(8), 1723-1745.

Irwin, E. G. (2002). The Effects of Open Space on Residential Property Values. *Land Economics, 78*(4), 465-480.

Irwin, E. G., & Bockstael, N. E. (2001). The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values. *American Journal of Agricultural Economics, 83*(3), 698-704.

Koster, H. R. A., & Rouwendal, J. (2012). The Impact of Mixed Land Use on Residential Property Values. *Journal of Regional Science, 52*(5), 733-761. doi: 10.1111/j.1467-9787.2012.00776.x

Levine, J., & Frank, L. (2007). Transportation and land-use preferences and residents' neighborhood choices: the sufficiency of compact development in the Atlanta region. *Transportation, 34*(2), 255-274.

Li, M. M., & Brown, H. J. (1980). Micro-Neighborhood Externalities and Hedonic Housing Prices. *Land Economics, 56*(2), 125-141.

Little, R. J. A. (1992). Regression With Missing X's: A Review. *Journal of the American Statistical Association, 87*(420), 1227-1237.

Matthews, J., & Turnbull, G. (2007). Neighborhood Street Layout and Property Value: The Interaction of Accessibility and Land Use Mix. *The Journal of Real Estate Finance and Economics, 35*(2), 111-141. doi: 10.1007/s11146-007-9035-9

Moorhouse, J. C., & Smith, M. S. (1994). The Market for Residential Architecture: 19th Century Row Houses in Boston ´ s South End. *Journal of Urban Economics, 35*(3), 267-277. doi: 10.1006/juec.1994.1016

Morancho, A. B. (2003). A hedonic valuation of urban green areas. *Landscape and Urban Planning, 66*(1), 35-41.

Páez, A., Fei Long, & Farber, S. (2008). Moving Window Approaches for Hedonic Price Estimation: An Empirical Comparison of Modelling Techniques. *Urban Studies, 45*(8), 1565-1581. doi: 10.1177/0042098008091491

Ryan, S. (2005). The Value of Access to Highways and Light Rail Transit: Evidence for Industrial and Office Firms. *Urban Studies, 42*(4), 751-764. doi: 10.1080/00420980500060350

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

Song, Y., & Knaap, G.-J. (2003). New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics, 54*(2), 218-238.

Tang, B.-s., & Yiu, C. Y. (2010). Space and scale: A study of development intensity and housing price in Hong Kong. *Landscape and Urban Planning, 96*(3), 172-182.

Tse, R. Y. C. (2002). Estimating Neighbourhood Effects in House Prices: Towards a New Hedonic Model Approach. *Urban Studies, 39*(7), 1165-1180.

Won Kim, C., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of Environmental Economics and Management, 45*(1), 24-39. doi: Doi: 10.1016/s0095-0696(02)00013-x

Yeh, A. G.-O., & Li, X. (2001). Measurement and Monitoring of Urban Sprawl in a Rapidly Growing Region Using Entropy. *Photogrammetric Engineering & Remote Sensing, 67*(1), 83-90.

Chapter 3

Spatial Spillover on Signaling-Effect of LEED Certification

Abstract

The benefit of pursuing green building certification mainly comes from two aspects: the cost-effectiveness from energy efficiency and the signaling consideration, including the premium on property values, benefits from a better reputation, morality values, or purely pride. By analyzing all new constructions that received LEED certification from 2000 to 2012 (LEED-NC v2.0 to v2.2) in the U.S., this study tries to identify the size of the signaling effects, and spillover of signaling, as building owners pursue LEED certification. The results show that the signaling effect affects decision making in pursuing LEED certification, especially at scores around thresholds. The size of signaling effects differs among different owner types and different certificate levels. For the Gold level or below, government and non-profit-organization owners value signaling more than do profit-seeking firms. At the Platinum level, there is no significant difference among owner types. This study also finds that the signaling effect clusters spatially for government and profit-seeking firms. Finally, the results show that the cluster of signaling is independent from the cluster of LEED buildings, indicating that mechanisms behind the cluster of signaling are different from those of LEED constructions.

Introduction

This study examines the existence of spatial spillover in the signaling effect of LEED (Leadership in Energy and Environmental Design) certification. Because the program to label green buildings is a voluntary one, LEED participants are expected to be motivated mainly by factors outside of pure energy effectiveness. Some of these factors could be competition with local rivals or a desire to show social or environmental responsibility. In this study, the term “signaling” is used to represent this part of motivation. The first goal of this study is to identify the magnitude of the effect of signaling. This study is further interested in whether the signaling effect is spatially clustering. While several previous studies have already indicated that LEED buildings cluster spatially (Kahn & Vaughn, 2009; Kok et al., 2011), the spatial spillover within the signaling effect of LEED buildings has not yet been fully addressed. This study aims to demonstrate the spatial spillover of signaling in LEED buildings and compare this spillover with the spatial clustering of LEED buildings.

In theory, LEED certification serves as a signal to mitigate market failure caused by information asymmetry. Without a LEED certificate, building owners' investment in the sustainability perspective cannot be seen, thus hinders them from being compensated by green marketing/advertising, premium on property values, fulfilling legal obligation, showing social responsibility, or even just by expressing their pure pride. Most building owners' strategy is a mixture of both cost-effectiveness in energy efficiency and signaling—they pursue LEED certification for reasons of both energy-saving and signaling. In this study, all the reasons other than cost-effectiveness in energy efficiency are viewed as "signaling". The first part of this study uses the kernel density function to identify the proportion of signaling. The results show that signaling is a main reason for building owners to pursue LEED certification, especially for buildings achieving scores that are just above critical scores to the next certification level. The results also show that the level of signaling differs by owner types. Buildings owned by governments and non-profit organizations are more likely to "signal" than buildings owned by profit-seeking firms.

The second part of this study aims to identify whether the signaling effect is spatially clustered. Theoretically, spatial spillover comes from at least three sources. First, local competition might enhance the pursuit of LEED certification. Building owners such as governments, schools, and firms who compete with their local rivals are likely to pursue the same or higher level of LEED certificates as their competitors. Second, spatial spillover could also come from knowledge spillover. The benefit gained from pursuing LEED is usually hidden information. The building owners who adopted LEED, or the contractors who are in charge of the application, do not need to report the benefit. Being spatially closer to these sources that hold this information is likely to increase the odds of pursuing LEED. Finally, legal requirements or incentives to adopt LEED within jurisdictions will also enhance clustering. In this study, the spatial spillover is measured by Moran's I, an index of spatial correlation, and local indicators of spatial association (LISA) maps. The results confirm the hypothesis that the signaling effect of LEED buildings clusters spatially, independent of the clustering of LEED buildings.

Literature Review

LEED, or any other green-building certification system, can possibly benefit building owners in several ways. First, LEED certified buildings can be expected to have better energy performance, which will significantly lower their maintenance cost. In fact, LEED certified buildings consume 25 to 30 percent less energy than conventional buildings, on average (Kats, 2003; Turner & Frankel, 2008). However, more than a quarter of LEED buildings actually consume more energy than their conventional counterparts (Newsham et al., 2009; Turner & Frankel, 2008). Moreover, even though the construction cost for green buildings is not significantly higher than that for non-green buildings (Morris & Matthiessen, 2007), the commissioning cost for LEED, which ranges from 1 to 6 percent of construction costs (D'Antonio, 2007; Mills et al., 2004), is another burden to building owners. The fact that a significant portion of LEED buildings bear extra costs without even enjoying energy savings implies that there are other reasons for the pursuit of LEED certification.

Another possible monetary driving force of LEED certification is the increment of property value, including premium sales prices, rental rates, and occupancy rates. Eichholtz et al. (2010) analyzed all Energy Star rated/ LEED certified US office buildings and found that, controlling for quality and location, a “green rating” increases rents by 3 percent and selling prices by 16 percent,. They further indicate that the relative premium for green buildings is higher in lower-cost regions or in less expensive parts of metropolitan areas. Fuerst & McAllister (2011a) confirm the effects of green-building labeling on office rents and real estate prices. The sources of this premium may be reduced utility costs, improved productivity (lower staff turnover, absenteeism), or an enhanced reputation (Fombrun & Shanley, 1990; Fuerst & McAllister, 2011b). The empirical evidence for reasons why owners seek LEED certification is either difficult to identify (reduced utility costs) or nearly intangible (improved productivity and enhanced reputation). However, if energy efficiency is the only benefit, energy intensive industries would have significantly higher incentives to pay the price. This conflicts with Eichholtz et al.’s study (2009) that oil, banking industries, government agencies, and non-profit organizations are top-ranking green office renters. This implies that labeling/signaling at least plays a role in the premium of rents and selling prices. In other words, signaling factors into the decision of building owners who pursue LEED certification. The first contribution of this study attempts to identify the how large a role signaling plays in the decision to apply for LEED certification.

The weight of signaling in the decision-making process of applying LEED certification varies by owner type. For government agencies and non-profit organizations, signaling is thought to play a more important role than it does for profit-maximizing firms because governments and non-profits are more eager to show that this is “the right thing to do” (Wood, 1991), and because they are less concerned about the profit-maximizing principal (Eichholtz, et al., 2009). For profit firms, signaling is expected to pay back in terms of gaining reputation, or so called Corporate Social Responsibility (CSR) so that firms can charge premium prices and attract better employees and investors (Eichholtz, et al., 2010; Milgrom & Roberts, 1986; Turban & Greening., 1997). Since CSR is difficult to identify, it is not surprising that there is not much positive empirical evidence on the relationship between CSR and profitability (Aupperle et al., 1985; Sen & Bhattacharya, 2001). However, CSR does benefit a firm at least in economic theory (Amacher et al., 2004; Lyon & Maxwell, 2008).

Some previous studies have indicated that LEED certified buildings cluster spatially. After analyzing political choices and the number of LEED buildings by zip code , Kahn and Vaughn (2009) indicated that LEED buildings tend to cluster in environmentalist communities. Kok et al. (2011) came a similar conclusion by looking at the diffusion of both Energy Star and LEED rated buildings. If LEED buildings cluster as environmentalism clusters, signaling should also cluster geographically, because signaling is defined as the reasons to go green other than efficiency, which is even more environmentalism than LEED certification per se. This study aims to contribute to this issue by showing that signaling also clusters spatially.

Data

The data used in this study is publicly open on the USGBC website. The dataset includes scores achieved and certificate levels as well as characteristics for each building, such as the project name, address, version of LEED, date of certification, owner type, project type, and building size. In order to keep the scoring system consistent and comparable, this study includes only buildings with LEED New Construction (LEED NC) version 2.0 to 2.2. In this scoring system, a score of 26 to 32 earns a certificate, a score of 33 to 38 earns a silver certificate, a score of 39 to 51 earns a gold certificate, and a score over 52 earns a platinum certificate. Figure 3.1 shows the histogram of frequencies on each score at the national level.

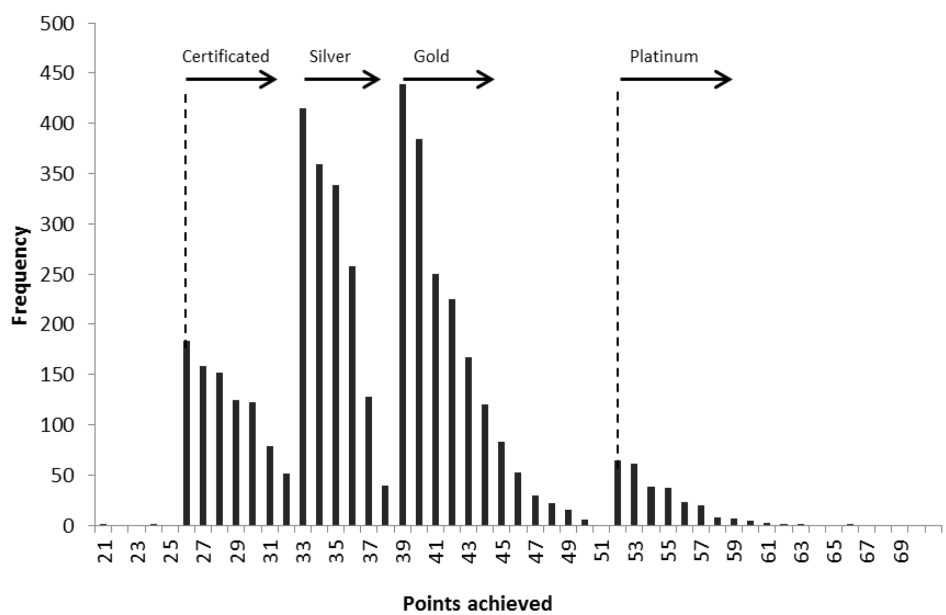


Figure 3.1 Histogram of LEED scores for US certified new-constructions

As shown in the figure, data scores crowd just above each threshold score, while only a relatively low number fall just below. This result implies the existence of a signaling effect. Without a signaling effect, i.e. with only performance-based thinking, building owners would achieve the most cost-effectiveness scores, without even considering which certificate level they would achieve. The high density at threshold scores implies that building owners tend to either earn a couple more credits to “upgrade” to the next certificate level, or stop making improvements since a couple more credits will not bring them to the next level. The high frequency of scores just above threshold scores seems to indicate that building owners sometimes overshoot when they aim at threshold scores. For a building that is only one-credit-shy of the next certificate level, the best improvement available might be worth at least three points. The building might end up with two points higher than the threshold, with a lot of signaling-based strategy mixed in its decision-making process. Thus, scores just above of thresholds should also be included as the indication of signaling.

Different types of owners are expected to use different strategies in pursuing LEED certification. To identify these differences, this study divides data scores into three types of owners: government, non-profit organizations, and profit-seeking firms. For simplicity, individual owners, and other owner types, which contribute to less than 10 percent of the data, are not included in this study.

Methodology

In the first stage, this study aims to identify the proportion of signaling in the pursuit of LEED certification. In this process, it uses the kernel density function, which generates smooth conditional expectations based on input-distributions (Silverman, 1981). The identification of signaling can be summarized as four steps. First, for each owner type, a histogram of scores achieved at the national level is generated. Similar to Figure 1, scores just above each threshold are expected to be closely grouped, and only a very few scores fall just below the thresholds. In other words, this study assumes that the signaling effect relocates buildings from scores just below thresholds to scores above, and thus locally skews the distribution. Second, all the scores that are around thresholds, both below and above, which are considered to be affected by the signaling effect, are dropped from the histogram. The proper definition of scores "around" the thresholds is discussed in later sections. The idea is that by removing scores that are highly affected by the signaling effect, the remaining scores are distributed as if there were no signaling effect. The problem with this pure performance-based distribution is that the densities of the dropped "signaling scores" are missing. Thus, in the third step, the kernel density estimation is used to generate a new distribution based on the histogram without the signaling scores so that the counterfactual densities of the signaling scores can be estimated as if there were no signaling. Finally, the "signaling factor" for each threshold score *i* can be identified as

$$(Signaling\ factor)_i = \frac{(Original\ density)_i - (Counterfactual\ density)_i}{(Original\ density)_i}$$

The identification of signaling scores is a judgment call based on the distributions of scores. Dropping fewer signaling scores keeps more information for the generation of kernel density, but signaling effects in these over-shot scores will still be included; thus, the measured signaling effect is likely to be underestimated. Dropping too many signaling scores will cover the signaling in these over-shot scores more thoroughly, but might leave less information for the generation of kernel density, thus biasing the estimates.

After identifying the intensity of the signaling effect for each signaling score, the second stage of the study then turns to the identification of the spatial dependence of the signaling effect. This study first assigns signaling factors to corresponding LEED certified buildings, based on owner types and the width (point spread) of thresholds. Note that only signaling factors of scores just above threshold scores, i.e. scores at the critical level and above, are assigned in this process. Signaling factors on the left-hand-side of thresholds, whose negative sign represents the other side of signaling clusters, are not included in the spatial dependence analysis to avoid the concern of double counting. They are coded as zeros, as are scores outside of the thresholds. The global Moran's Index (Moran's I), which is an index that reflects the level of spatial correlation (Moran, 1950), is then calculated to identify the spatial dependence of signaling factors. Moran's I is defined as

$$Moran's\ I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

where n is the number of spatial units, ω_{ij} denotes the element of spatial weights between *i* and *j*, *X* is the variable of interest, which is the signaling factor in this case, and $\bar{X}$ is the mean of *X*. Spatial weight ω reflects how "neighbors" are defined. It can be a dummy showing whether two units are directly contiguous, or it can be a value of inverse distance/squared inverse distance, reflecting the gradient of influence with distance. In this study, the method of inverse distance is used to calculate the global Moran's I.

Contiguity is not applicable here because nearly no LEED buildings are adjacent to each other. In normal cases, Moran's I ranges from -1 to 1, where -1 represents the case of perfectly dispersed, 1 illustrates perfectly clustered, and zero means purely random. However, in some extreme cases, such as when the variable *X* is strongly skewed or when ω_{ij} is extremely small, Moran's I can exceed the bounds of -1 and 1.

One concern regarding interpreting the spatial dependence of the signaling factor as the spatial cluster of signaling is that the cluster of signaling factors might be mainly due to the gathering of LEED certified buildings per se. Figure 2 illustrates the locations of all 4,472 LEED certified new constructions used in this study. Without any further data processing, it is apparent that LEED buildings in the US cluster in several major metropolitan areas. Since LEED buildings cluster for certain, it can be argued that a proportion of a subset of LEED certifications (signaling factors of buildings at threshold scores) is also spatially dependent. To avoid this potential issue, this study aggregates data at the zip code level and defines an index of ***Signaling Density (SD)*** as:

$$SD_j = \frac{\sum[(Signaling\ factor)_i \times (LEED\ buildings)_{ij}]}{(LEED\ buildings)_j,}$$

where *i* denotes the threshold score, *j* denotes the individual zip code area, $(LEED\ buildings)_{ij}$ refers to the number of LEED buildings with a score of *i* in zip code *j*, and $(LEED\ buildings)_j$ refers to the total number of LEED buildings in a zip code *j*. For each zip code *j*, SD_j is the summation of all signaling components divided by the total number of LEED buildings. By definition, since all of the negative signaling factors are excluded, SD ranges from zero to one. For any zip code in which all LEED buildings are at a certain signaling score whose signaling factor is one, the signaling density of this zip code is one[3]. For a zip code that does not have any LEED building but still achieves a signaling score, the signaling density is zero[4]. Given that the signaling densities for most zip codes are zero, the average signaling density is only 0.020, with a maximum of 0.899. Applying the signaling density avoids the concern that the cluster of signaling corresponds with the gathering of LEED buildings. In other words, with the signaling density, the spatial dependence discovered in this study is not disturbed by the spatial clustering of LEED buildings.

This study aggregates data to the zip code level to provide a denominator in signaling density (counts of LEED buildings) to hide the effect of LEED clustering. Aggregating data at too small of a level does not really diminish the effect of LEED clustering because there will be only one or a few LEED buildings in each zone (if there is even any), which limits the use of applying signaling density. On the other hand, aggregating data at too large of a level provides less interesting results. For example, aggregating data at the county level would provide results similar to Figure 3.2 since counties with non-zero SD are expected to cluster around metropolises.

[3] Since signaling factor will only equals to one in an extreme case, where the counterfactual density generated by kernel density function is zero, the signaling density is very unlikely to be as large as one.
[4] For zip codes that do not have any LEED buildings at all, the signaling density is also assigned to be zero.

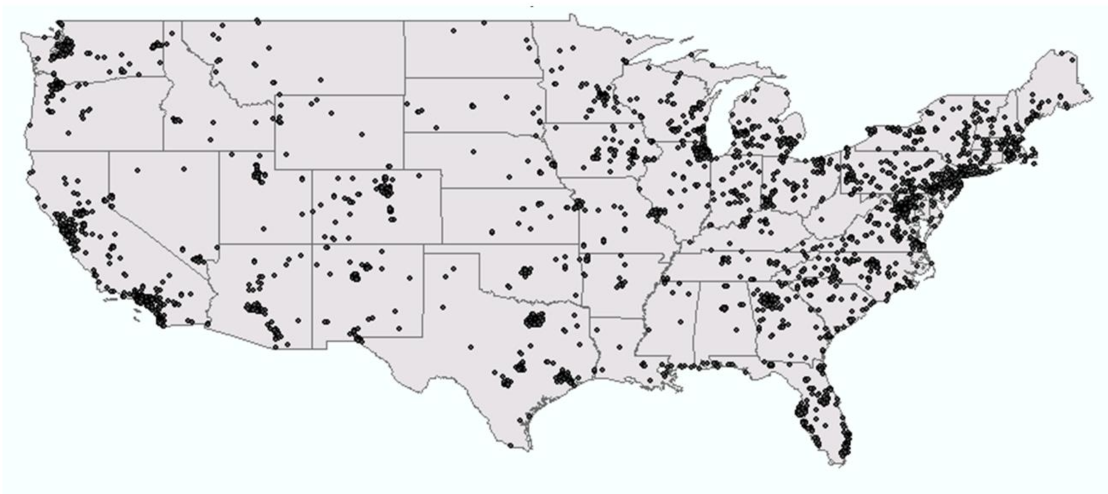


Figure 3.2 Distribution of LEED-NC certified buildings in the US

In order to further illustrate the spatial dependence of signaling, this study introduces the local indicators of spatial association (LISA) map, which plots the local Moran's I[5] statistics for each zip code area, to show the distribution of spatial dependence in several selected metropolitan areas. This study uses the first-order-contiguity method to generate spatial weight matrices in order to focus on the spatial spillover between adjacent zip code areas.

Result and Discussion

1. Identifying Signaling Factors

The first part of this study identifies the "signaling factor" for each score at the national level. As mentioned in the method section, the signaling factor reflects the

[5] See (Anselin, 1995).

weight of signaling in the decision-making process. A signaling factor of one means that buildings achieving this specific score have done so for the purpose of signaling, and a factor of zero represents that buildings at this score have pursued a performance-based strategy. Table 3.3 summarizes the results of the signaling factor under different conditions. For each owner type, three scenarios of dropping signaling scores are used: (1) Only the critical scores of each certificate level, i.e. scores of 26, 33, 39, and 52, are dropped; (2) Three scores, the critical score, the one above, and the one below, are dropped for each certificate level; (3) Five scores, including the critical score, two scores above and two below, are dropped for each level. The score-dropping scenario stops at two scores-below for two reasons. First, signaling is expected to occur for those buildings whose most cost-effective scores are just a few of credits shy of the next certificate level. When building owners believe signaling will payoff somehow, they will pursue the next level, regardless of the fact that getting extra credits is not efficient from the performance-based perspective. In this regard, signaling should move buildings from only a few credits shy to exactly at the next threshold, because these owners have no incentive to further exacerbate their already-not-efficient status. In practice, it is possible that the cheapest improvement merits more than one credit, thus making these buildings "overshoot" the goal and end up with a few credits higher than the threshold. However, it is not likely that these signaling building owners overshoot to the extent that they end up with three credits higher than the threshold. Secondly, dropping more than five points for each threshold makes it difficult to generate the counterfactual kernel estimates with the limited remaining information.

At each critical score, the identified signaling factor is generally higher than dropping more signaling scores in the counterfactual distribution generating process. For example, the signaling factor at 26 points for government-owned buildings grows from 0.62 to 0.90, when drop more scores in the process. By definition, a signaling factor of 0.90 means that 90 percent of the buildings in this category achieve this score only for signaling reasons. In other words, it can be inferred that 90 percent of the (reason for the) decision is due to signaling for each building in this category. However, there is not enough evidence indicating that dropping more credits/points leads to a more accurate estimate for signaling. On the one hand, dropping more signaling-like scores leaves a purer performance-based distribution in the system, which generates more accurate estimates of counterfactual density without signaling, thus identifying the signaling effect more clearer. On the other hand, dropping more scores leaves less information to cover more counterfactual estimates, thus making the results more likely to be biased. There is no reason to believe any dropping scenario is superior to any other. The bottom line is that, the signaling effect explains at least half of the reason for pursuing LEED certification for all three groups of building owners, under any score-dropping scenario.

Table 3.3 Results of signaling factors at the national level

	Government			Non-profit			Profit		
Score	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
24			-			-			-
25		-	-		-	-		-	-
26	0.62	0.79	0.90	0.59	0.72	0.80	0.52	0.64	0.66
27		0.32	0.44		0.48	0.57		0.27	0.33
28			-0.03			0.18			0.23
31			-0.61			-0.39			-0.92
32		-0.57	-0.68		-0.87	-0.52		-1.26	-1.60
33	0.48	0.72	0.88	0.51	0.67	0.78	0.52	0.57	0.60
34		0.46	0.69		0.56	0.75		0.11	0.42
35			0.48			0.54			0.60
37			-0.33			-0.30			-0.10
38		-1.36	-1.34		-3.49	-1.78		-8.76	-6.36
39	0.56	0.77	0.84	0.53	0.64	0.71	0.48	0.60	0.58
40		0.51	0.59		0.48	0.54		0.59	0.56
41			-0.11			0.02			0.14
50			-0.72			-			-
51		-	-		-	-		-	-
52	0.60	0.81	0.88	0.50	0.65	0.69	0.63	0.79	0.75
53		0.52	0.55		0.56	0.60		0.65	0.67
54			-0.26			0.01			0.25

Note: the signaling factor is marked as '-' if the original density at that score is 0.

The signaling factors on the left-hand-side of thresholds are all negative, if not invalid, due to the absence of actual density. This "negative signaling," which results from the actual density being smaller than the counterfactual one, supports the first-half of the hypothesis that buildings just below the thresholds are attracted by the benefits brought by signaling and thus move to a score just above the thresholds.

The other side of the hypothesis is supported by the mostly-positive signaling factors just above the thresholds. Only 3 above-threshold signaling factors are negative, and all of them are in the government category and are two-scores higher than thresholds. These negative signaling factors just above thresholds imply that the signaling effect actually “drags” government buildings back to the thresholds. Thus, the actual densities are lower than expected. This result implies that government buildings tend to pursue a strong signaling strategy around thresholds and are less likely to seek higher scores once the “goal” certificate level is achieved. This result could be due partly to the legislative mandates that require government building a to “go green” (May & Koski, 2007), which requires that government buildings be certified even though it might not be cost-effective. However, this reason does not explain the signaling at the silver and gold levels. This result is consistent with the hypothesis that government buildings put more emphasis on signaling strategy than profit-seeking firms do because governments tend to care more about showing that they are “doing the right thing” and care less about the cost-effectiveness of LEED pursuit.

Comparing signaling factors among building owners provides even more direct evidence for the above hypothesis. For Gold level or under, the signaling factors at thresholds (scores of 26, 33, and 39) for government and non-profit-organization buildings are significantly higher than are the signaling factors for profit firms, but this significant difference disappears at the highest certificate level, Platinum. This result implies that profit-seeking firms do not value signaling as much as governments and non-profit organizations at lower certificate levels. However, for the Platinum level, profit-seeking firms seem to recognize that signaling provides much higher benefits such as a better reputation, a higher premium on property values, or a more competitive market position, thus making them downplay the energy-performance-based strategy and "signal" as much as governments and non-profit organizations do.

2. Spatial Dependence of Signaling

The second part of this study applies the results in the first part to the inquiry of spatial spillover of signaling. First, a signaling factor is assigned to every LEED-NC building that falls into the owner type categories and threshold definitions. As mentioned in the previous section, only signaling factors that are just above thresholds are introduced in this part. Buildings with scores on the left-hand-side of critical values, and scores outside of thresholds, are coded as zero. The global Moran's Indices, which reflect the level of spatial dependence, are then calculated based on signaling factors and are listed in Table 3.4.

The first column shows the global Moran's I's for all owner types. In the first three rows, global Moran's I is calculated based on signaling factors under different threshold definitions. As the thresholds expand from only critical scores (signaling factor 1) to two credits above critical values (signaling factor 3), both the strength and the significance of spatial dependence increases. This result is consistent with the hypothesis that the signaling effect clusters spatially. The first part of the results indicates that a larger threshold area identifies more signaling effect. Better identification of signaling would likely lead to clearer spatial dependence, if signaling does cluster spatially.

Table 3.4 Spatial Dependence (global Moran's I) of Signaling Factors

	Threshold scores	All types	Governments	Non-profits	Profits
Signaling factor1	**26, 33, 39, 52**	0.101 **	0.090 *	0.282 *	0.285
Signaling factor2	**26**, 27, **33**, 34, **39**, 40, **52**, 53	0.288 ***	0.297 ***	0.283 *	0.545 ***
Signaling factor3	**26**, 27, 28, **33**, 34, 25, **39**, 40, 41, **52**, 53, 54	0.396 ***	0.380 ***	0.230	0.558 ***
Signaling dummy1	**26, 33, 39, 52**	0.082 *	0.066	0.270	0.299 *
Signaling dummy2	**26**, 27, **33**, 34, **39**, 40, **52**, 53	0.332 ***	0.342 ***	0.263	0.467 **
Signaling dummy3	**26**, 27, 28, **33**, 34, 25, **39**, 40, 41, **52**, 53, 54	0.337 ***	0.349 ***	0.159	0.424 **

*p < .10. **p < .05. ***p < .01.

As a comparison, row 2 to 4 indicates the global Moran's Indices for signaling dummies. Instead of applying specific signaling factors, signaling dummies is defined as treating LEED buildings at threshold scores as one, and others as zero. Thus, the Moran's I of signaling dummy represents the spatial dependence of LEED buildings at threshold scores. In general, the spatial dependence of the building itself is not as strong as that of signaling, though the difference is limited. This result serves as good evidence that even without the identification of a signaling factor, LEED buildings at threshold scores per se, which tend to signal more, are spatially dependent. Moreover, the slightly higher spatial dependence of a signaling factor implies that the signaling factor is a better estimation of signaling than signaling dummies.

The following three columns divide LEED buildings into three owner types to compare clustering patterns among different owners. The result shows that the signaling of government buildings and that of profit-seeking firms are spatially dependent, while the signaling of non-profit organizations shows only a weak significance. This result is consistent with the hypothesis that the main mechanism behind the spatial dependence of signaling is local competition. The spatial dependence could be the result of local governments competing with each other for the leadership in going green (Sharpe, 1970), or it could come from local firms using signaling to compete for market share. Non-profit organizations are less likely to compete locally, partly because of the nature of non-profits. Since this study does not divide non-profit organizations into categories, the possible competition within categories cannot be identified.

Between two significantly clustering owner types, the spatial dependence of government buildings is generally not as strong as that of profit-seeking firms. This is not surprising because the competition among government agencies is not supposed to be stronger than that among for-profit firms. Another possible factor is regulation mandates. An increasing number of states require public(government) facilities to be constructed to be green buildings (May & Koski, 2007). These requirements are likely to mitigate the influence of signaling because seeking LEED certification is not a voluntary decision based on signaling/performance anymore. Since these legislative mandates apply only to public facilities, the already-weaker spatial dependence for government buildings can be further weakened when it is compared to that of private firms.

To avoid the concern that signaling clusters only as LEED buildings do, this study introduces the concept of signaling density. Since the data is aggregated at the zip code level and the total signaling factors in the zip code are divided by the total number of LEED constructions, the signaling density is able to identify a pure spatial dependence independent from the cluster of LEED constructions. Table 5 lists the global Moran's Indices of signaling densities. Since the data is aggregated at the zip code level, it is possible to define spatial weight ω by identifying the contiguity of zip code boundaries, as well as by using the inverse distance of zip code centroids. The two right-hand columns in Table 3.5 show the Moran's Indices based on these two methods. Though the numbers differ, the trend is consistent between methods.

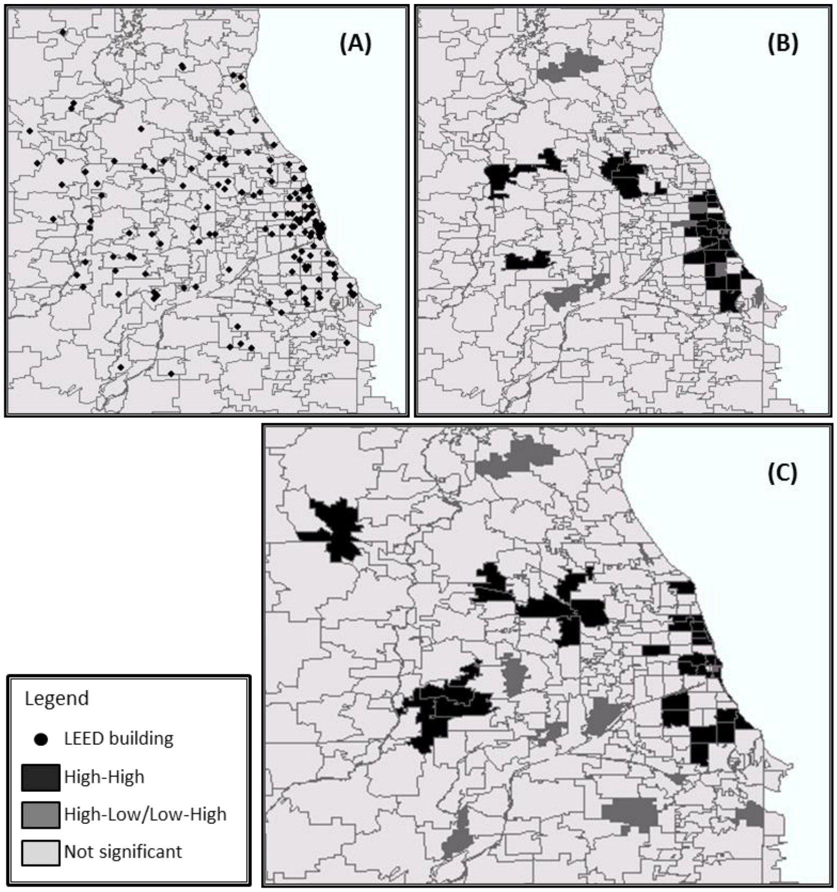
Table 3.5 The Spatial dependence (Global Moran's I) of Signaling Density

	Threshold scores	**Contiguity**	**Inverse Distance**
LEED counts	All scores	0.178 ***	0.056 ***
Signaling density 1	**26, 33, 39, 52**	0.058 ***	0.020 ***
Signaling density 2	**26**, 27, **33**, 34, **39**, 40, **52**, 53	0.090 ***	0.033 ***
Signaling density 3	**26**, 27, 28, **33**, 34, 25, **39**, 40, 41, **52**, 53, 54	0.103 ***	0.039 ***

*p < .10. **p < .05. ***p < .01.

As shown in Table 3.5, signaling density spatially clusters at the zip code level. Since signaling density is not correlated with the number of LEED buildings in each zip code, this study concludes that signaling is spatially dependent, even after controlling for the cluster of LEED certifications. While the spatial dependence of signaling density increases as thresholds expand, which follows the same trend in Table 4, the clustering of LEED buildings is much stronger than that of signaling densities. This result is as expected because LEED certifications as a whole have more reasons than signaling to cluster in space. One such reason is that LEED certified buildings are just more likely to cluster in metropolitan areas.

Finally, to further investigate whether signaling clusters in the same pattern as LEED certifications do, this study uses the LISA maps for Chicago, the metropolitan area that has most LEED certified new constructions in the US. In Figure 3.6, map (A) shows the distribution of LEED buildings in Chicago. Map (B) and (C) are the LISA maps for LEED building counts and signaling density, respectively. A LISA map illustrates the local Moran's I for specific variables. A significant and positive local Moran's I value indicates that a zip code with high LEED counts is adjacent to more high-count zip codes than random, which is illustrated in black on the map. A significant but negative local Moran's I,s colored as dark-gray, refers to either a low-count zip code adjacent to more high-count areas, or vise versa. Light-gray areas refer to an insignificant local Moran's I, which indicates that the spatial distribution of high/low LEED counts in adjacent areas is close to random. Similarly, map (C) illustrates the spatial dependence of signaling density by local Moran's I.



Map (A): Distribution of LEED building
Map (B): Spatial dependence of LEED building counts
Map (C): Spatial dependence of Signaling Index

Figure 3.6 Spatial Dependence maps for Chicago

Figure 3.6 shows the spatial dependence of LEED buildings and signaling density in Chicago. A symbol in promoting energy efficiency, Chicago leads the US with 156 LEED-NC certifications. In map (A), LEED certified buildings clearly cluster in the central business district (CBD). The same trend can also be observed in map (B), where high-LEED-count zip codes surrounded by other high-LEED-counts (High-High blocks) cluster mostly in the CBD area. This is not direct evidence t that buildings in the CBD are more likely to pursue LEED certification because this study does not control for building density in each zip code. However, it does indicate that LEED certifications tend to cluster in central, high-density areas. The result shown in map (C) needs less clarification because signaling density inherently controls for LEED density. map (C) illustrates the spatial dependence of signaling density with clustering high-high zip codes. Moreover, the map clearly shows that signaling clusters not only in CBD, but also in suburban areas. One possible explanation stems from Eichholtz et al. (2010), who indicate that the relative premium for green buildings is higher in lower-cost regions or in less expensive parts of metropolitan areas. Since the premium for "being green" is one of the mechanisms behind signaling, higher premiums create incentives to signal more in suburban areas.

Conclusion

The result of this study concludes that the signaling effect does affect the decision making in pursuing LEED certification, especially at scores around thresholds. The size of the signaling effect differs among different owner types and different certificate levels. For the Gold level or below, government and non-profit-organization owners signal more than profit-seeking firms. At the Platinum level, there is no significant difference among owner types. This result generally confirms the hypothesis that governments and non-profit organizations are more likely to signal because they care less about the cost effectiveness from an energy perspective. The difference disappears at the Platinum level because governments and non-profit organizations have to more careful consider higher initial construction costs, and firms value signaling more with higher property values or reputation benefits.

This study also finds that the spatial cluster of signaling varies among owner types. Governments and for-profit firms cluster significantly, while non-profit organizations do not. This result implicitly supports the hypothesis that the cluster of signaling is mainly from local competition since non-profit organizations are less likely than other owner types to be involved in competing situations. Moreover, the cluster of signaling is significant at the zip code level even after controlling for the cluster of LEED constructions, which further confirms that the mechanism behind signaling clustering is independent of that behind LEED-construction clustering.

The results of this study have some policy applications. First, better understanding the strategies behind the pursuit of LEED certification will help future regulation making. For example, subsidies might not be as effective for non-profit organizations as for profit firms because non-profit organizations do not consider the costs of pursuing LEED much, especially at thresholds. Second, the existence of spatial spillover in signaling can possibly shape policies to stimulate green buildings. By encouraging pilot projects locally, such as encouraging government buildings to achieve higher certificate levels, policy makers who push for the adoption of LEED may expect the number of LEED buildings to grow exponentially due to the spatial spillover.

There are several limitations to this study. First, the definition of signaling is too broad to identify any mechanism in this study. Generally, factors other than energy-efficient-oriented purposes, such as green marketing/advertising, premiums on property values, legal obligations, social responsibility, or even pure pride, are all considered as signaling-oriented rationales. Due to the limitation of data, there is no way to identify the single effect of any of those signaling-oriented rationales.

Second, a lack of information regarding building density limits the comparison between LEED density and signaling density. Currently, this study compares only the distribution of LEED certificate numbers and signaling density. Though it is enough to say that signaling clusters differently from LEED certificates, it would be interesting to see if signaling density also clusters differently from LEED density.

Finally, calculating the spatial dependence with a long-time-period data raises the concern of letting old buildings be affected by new buildings. The dataset includes certificates from 2000 to 2012. Calculating spatial dependence with all observations together appears to allow new certificate scores to affect existing scores, which is not actually possible. This causes the problem of over-identification, which could make the spatial dependence be overly significant. This issue must be kept in mind while interpreting the spatial cluster in LISA maps.

References

Amacher, G. S., Koskela, E., & Ollikainen, M. (2004). Environmental quality competition and eco-labeling. *Journal of Environmental Economics and Management, 47*(2), 284-306. doi: 10.1016/s0095-0696(03)00078-0

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis, 27*(2), 93-115.

Aupperle, K. E., Carroll, A. B., & Hatfield, J. D. (1985). An Empirical Examination of the Relationship between Corporate Social Responsibility and Profitability. *The Academy of Management Journal, 28*(2), 446-463. doi: 10.2307/256210

D'Antonio, P. C. (2007). *Costs and Benefits of Commissioning LEED-NC™ Buildings*. Paper presented at the National Conference on Building Commissioning, Chicago, IL.

Eichholtz, P., Kok, N., & Quigley, J. M. (2009). Why Do Companies Rent Green? Real Property and Corporate Social Responsibility. University of California, Berkeley Program on Housing and Urban Policy Working Paper W09-004.

Eichholtz, P., Kok, N., & Quigley, J. M. (2010). Doing Well by Doing Good? Green Office Buildings. *American Economic Review, 100*(5), 2492-2509.

Fombrun, C., & Shanley, M. (1990). What's in a Name? Reputation Building and Corporate Strategy. *The Academy of Management Journal, 33*(2), 233-258. doi: 10.2307/256324

Fuerst, F., & McAllister, P. (2011a). Green Noise or Green Value? Measuring the Effects of Environmental Certification on Office Values. *Real Estate Economics, 39*(1), 45-69. doi: 10.1111/j.1540-6229.2010.00286.x

Fuerst, F., & McAllister, P. (2011b). Eco-labeling in commercial office markets: Do LEED and Energy Star offices obtain multiple premiums? *Ecological Economics, 70*(6), 1220-1230. doi: 10.1016/j.ecolecon.2011.01.026

Kahn, M. E., & Vaughn, R. K. (2009). Green Market Geography: The Spatial Clustering of Hybrid Vehicles and LEED Registered Buildings. *9*(2), Article 2.

Kats, G. H. (2003). Green Building Costs and Financial Benefits Retrieved from http://www.dcaaia.com/images/firm/Kats-Green-Buildings-Cost.pdf

Kok, N., McGraw, M., & Quigley, J. M. (2011). The Diffusion of Energy Efficiency in Building. *American Economic Review: Papers & Proceedings, 101*(3), 77-82.

Lyon, T. P., & Maxwell, J. W. (2008). Corporate Social Responsibility and the Environment: A Theoretical Perspective. *Review of Environmental Economics and Policy, 2*(2), 240-260. doi: 10.1093/reep/ren004

May, P. J., & Koski, C. (2007). State Environmental Policies: Analyzing Green Building Mandates. *Review of Policy Research, 24*(1), 49-65. doi: 10.1111/j.1541-1338.2007.00267.x

Milgrom, P., & Roberts, J. (1986). Price and Advertising Signals of Product Quality. *Journal of Political Economy, 94*(4), 796-821. doi: 10.2307/1833203

Mills, E., Friedman, H., Powell, T., Bourassa, N., Claridge, D., Haasl, T., & Piette, M. A. (2004). The Cost-Effectiveness of Commercial-Buildings Commissioning, A Meta-Analysis of Existing Buildings and New Construction in the United States: Lawrence Berkeley National Laboratory, Portland Energy Conservation Inc., Energy Systems Laboratory, Texas A&M University.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika, 37*(1/2), 17-23. doi: 10.2307/2332142

Morris, P., & Matthiessen, L. F. (2007). Cost of Green Revisited: Reexamining the Feasibility and Cost Impact of Sustainable Design in the Light of Increased Market Adoption Retrieved from http://www.davislangdon.com/upload/images/publications/USA/The%20Cost%20of%20Green%20Revisited.pdf

Newsham, G. R., Mancini, S., & Birt, B. J. (2009). Do LEED-certified buildings save energy? Yes, but…. *Energy and Buildings, 41*(8), 897-905. doi: http://dx.doi.org/10.1016/j.enbuild.2009.03.014

Sen, S., & Bhattacharya, C. B. (2001). Does Doing Good Always Lead to Doing Better? Consumer Reactions to Corporate Social Responsibility. *Journal of Marketing Research, 38*(2), 225-243. doi: 10.2307/1558626

Sharpe, L. J. (1970). Theories and Values of Local Government. *Political Studies, 18*(2), 153-174. doi: 10.1111/j.1467-9248.1970.tb00867.x

Silverman, B. W. (1981). Using Kernel Density Estimates to Investigate Multimodality. *Journal of the Royal Statistical Society. Series B (Methodological), 43*(1), 97-99.

Turban, D. B., & Greening., D. W. (1997). Corporate Social Performance and Organizational Attractiveness to Prospective Employees. *Academy of Management Journal, 40*(3), 658-672. doi: 10.2307/257057

Turner, C., & Frankel, M. (2008). Energy Performance of LEED® for New Construction Buildings: New Buildings Institute.

Wood, D. J. (1991). Corporate Social Performance Revisited. *The Academy of Management Review, 16*(4), 691-718. doi: 10.2307/258977

Chapter 4

The Closer the Better? The Distance Effect on the Support of Atlanta BeltLine

Abstract

Atlanta BeltLine, a large urban redevelopment project currently underway in the center of Atlanta, transforms 22 miles of historical railroad corridors into parks, trails, pedestrian-friendly transit areas, and affordable housing. This study aims to determine the distance effect on the support of Atlanta BeltLine and whether the implement of Tax Increment Financing (TIF) affects the support. The contributions of this exercise are twofold. First, it demonstrates the risks and remedies to missing spatial data by solving the technical problem of missing precise spatial location values. Second, it tests underlying reasons why distance can help explain the level of support that Atlanta BeltLine has received, with striking implications for theories like the Homevoter hypothesis. Survey data used in this study was conducted in summer 2009, about three years after the declaration of the project. The support by both homeowners and renters significantly declines as distance from the BeltLine increases. However, when residents' tendency to use BeltLine parks and transits is entered as a variable, the distance effect disappears. By indicating that the distance effect comes from homeowners' and renters' the accessibility to BeltLine amenities, the result rejects the homevoter hypothesis, which holds that property value increment is the main mechanism behind support. The results also show that whether or not a homeowner or renter is a parent in City of Atlanta affects a person's support of the BeltLine. These results lead to the conclusion that the concern of TIF affecting future school quality hampers the support of the project.

Introduction

This study uses Atlanta BeltLine as an example to determine the effect of distance on the support of urban redevelopment projects. First launched by the City of Atlanta, Atlanta BeltLine is a large urban redevelopment project taking place at the center of

Atlanta. The key concept of the project is to transform 22 miles of historic railroad corridors into pedestrian-friendly rail transit, multi-use trails, parks, and affordable housing. Being one of the largest urban reformation and mobility projects currently undertaken in the United States, Atlanta BeltLine has been characterized as a local project that will be able to “transform the city” (Atlanta Development Authority, 2005). Thus, given the potential impact of Atlanta BeltLine and the expectation of residents, it is fair to include all people in metro Atlanta area into the scope of this study.

The effect of distance could be a mix of several mechanisms. First, the accessibility to new parks, trails, and public transits, is surely an amenity. Being closer to green space or transit options is expected to be positively related to residents’ support, especially for those who regularly use these facilities. Those who do not directly use these amenities could still benefit from the premium on property values that the amenities are likely to cause. The homevoter hypothesis is a perfect theory to describe this mechanism. First developed by Fischel (2005), the homevoter hypothesis holds that homeowners politically support local governments actions that increase property values. In this case, it is expected that local homeowners will support Atlanta BeltLine, as long as the project increases their property values. The property value increment could be due to the increase of actual accessibility to amenities, or the expectation of “Atlanta being a better place.” Thus, it is reasonable to apply the homevoter hypothesis to the whole city. Since it is expected that being closer to the BeltLine creates a higher premium, homeowners will be more supportive, which makes the distance effect under this mechanism theoretically negative. This study implicitly tests the homevoter hypothesis by determining the distance effect on homeowner support, under the assumption that property value increment due to the BeltLine is correlated with the distance to it.

On the other hand, support for the Atlanta BeltLine could be hampered by the provision of affordable housing. According to Atlanta BeltLine Inc. (ABI), the BeltLine project aims to create 5,600 units of affordable housing over twenty-five years (Atlanta BeltLine, 2013). If the homevoter hypothesis is true, homeowners who live around the proposed BeltLine affordable housing should oppose the project since more nearby housing supply would be harmful to their property values. Neighborhoods also often oppose the introduction of nearby public housing based on a fear of increasing crime (Roncek et al., 1981). Though affordable housing in the BeltLine project has little similarity with public housing, it is arguable that the provision of par-price housing increases residents' concern about crime.

The idea of Tax Increment Financing (TIF) could be another important factor tthat might affect the distance effect regarding citizen support. TIF allows local governments to pay for a certain project by the future tax increment created by the project itself. Mainly funded by TIF, the Atlanta BeltLine basically is expected to pay for itself by the property tax increment collected in the Tax Allocation District (TAD) over the next 25 years. Since TIF blocks the use of future tax increment towards public education, the quality of public school is expected to be lower in the future if the population keeps growing. Thus, it is expected that homeowners with children, especially those who plan to send their kids to public schools, will be less likely to support Atlanta BeltLine. This negative effect of having kids might not be constant throughout different school districts. Theoretically, only public schools that contain TAD in their catchment zones are affected.

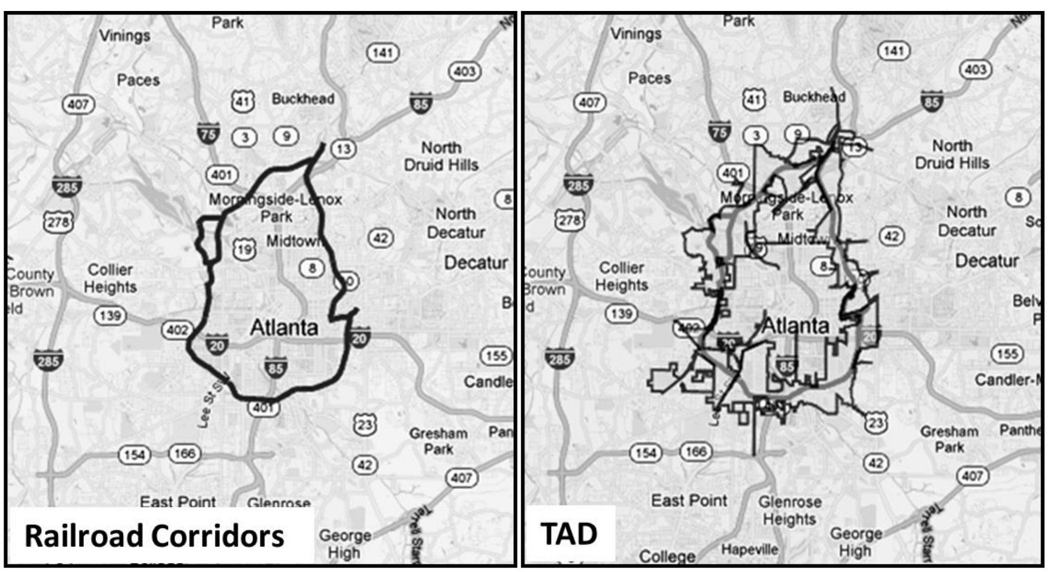
However, since Atlanta Public Schools could possibly adjust the budget distribution, all public schools within the City of Atlanta could be affected. Thus, it is expected that this negative impact of having kids on support could also occur to all parents within the City of Atlanta. In other words, the interaction term of jurisdiction dummies and having kids is expected to be negatively significant.

Data used in this study is from a survey conducted in summer 2009, asking about opinions and expectations regarding Atlanta BeltLine. The main difficulty of applying the survey data to this study is that only half of the respondents provided their actual addresses. To solve this missing-spatial-location problem, this study tries several imputation approaches, including utilizing zip-code centroids, population-weighted zip-code centroids, and two multiple-imputation methods. The results show that imputation is preferred to dropping observations with missing values in order to maintain a workable sample size. The selection of imputation methods does not significantly affect the results.

The result of this study indicates that the support for Atlanta BeltLine among homeowners decays along with distance with statistical significance. However, this study also shows that accessibility is the main mechanism behind the distance effect. This finding rejects the homevoter hypothesis. Also, the interaction terms of jurisdiction dummies and the number of children in a household have slightly significant effects on the support, showing that having kids in affected school zones does reduce homeowners' support for Atlanta BeltLine.

Background

Atlanta BeltLine is an urban redevelopment project undertaken in city of Atlanta that will transform 22-miles of historic railroad corridors and 45 connected neighborhoods into 22 miles of pedestrian-friendly rail transit, 33 miles of multi-use trails, 1,300 acres of parks, and 5,600 units of affordable housing. Stemming from a 1999 masters thesis by Georgia Tech student Ryan Gravel, the BeltLine project gained the support of the City of Atlanta in 2005 and became an official/active project after the creation of Atlanta BeltLine Inc. (ABI) in 2006. The Atlanta BeltLine Tax Allocation District (TAD) serves as the primary funding source. 6,500 acres of TAD is projected to generate 1.7 billion dollars of tax increment in a twenty-five year window, which is about sixty percent of the original estimated cost. The remainder of the cost is expected to be covered by local contributions and federal funds (Atlanta BeltLine, 2013). Figure 4.1 illustrates the location of the 22-mile railroad corridor and 6,500 acres of the Atlanta BeltLine TAD.



(Source: http://beltline.org)

Figure 4.1 The Location of Railroad Corridors and the BeltLine TAD

The Atlanta BeltLine project is expected to boost property values in the surrounding neighborhood, or even in a larger range. Immergluck (2009) conducted a hedonic housing price analysis for single-family house sales in Atlanta from 2000 to 2006 and found that after 2005, sales closer to the BeltLine TAD had a higher premium on sales price. The author claims that this premium is a result of both gentrification and local newspaper coverage of the Atlanta BeltLine project. Following Immergluck's conclusion, the price effect of the BeltLine should not be limited only to neighborhoods because the effect of local newspaper is not limited to neighborhoods surrounding the Atlanta BeltLine. In 2005 alone, more than 100 stories about the BeltLine appeared in *The Atlanta Journal Constitution*, the major principle daily paper in Atlanta (Immergluck, 2009). Influenced by the frequent coverage, the real estate market in the city was likely affected, thus making the price effect of the BeltLine project affect a much wider range, though the effect should weaken with distance.

The relationship between the support for the BeltLine and increases in property value can be predicted by the homevoter hypothesis. As mentioned previously, the homevoter hypothesis holds that homeowners politically support local governments actions that increase property values (Fischel, 2005). Developed by Fischel in 2005, the homevoter hypothesis is rarely tested by empirical studies. Brunner et al. (2001; 2003) analyzed the voting results of a school voucher in California and concluded that homeowners in neighborhoods with superior public schools were less likely to vote for the voucher because they worried about a property value decrease. Dehring et al. (2008) analyzed the results of a 2004 referendum in Arlington, Texas, concerning a publicly subsidized stadium to host the NFL Dallas Cowboys, and found that potential property value increment is positively correlated with the support rate in the referendum, which is consistent with the homevoter hypothesis.. The Atlanta BeltLine case provides an additional chance to test the homevoter hypothesis in an indirectly way. Under the assumption that the level of property value increment caused by Atlanta BeltLine is correlated with distance, which is suggested by Immergluck (2009), the homevoter hypothesis can be indirectly tested by showing that the distance from the BeltLine project is correlated with the support for it by local homeowners. However, the relationship between distance and support can still come from sources other than property value increments, such as direct use value.

The introduction of TIF is another important factor that can possibly affect the support of Atlanta BeltLine. As the most widely used local government program for financing economic development in the United States, TIF has the main advantage of bringing in no outside money and needing no new revenue-raising authority (Briffault, 2010; Man & Rosentraub, 1998). One concern about TIF that is related to the examination of the homevoter hypothesis is the impact of tax-reallocation on education expenditures. Weber (2003) analyzed TIF's impact on the finances of school districts in Cook County, Illinois, and revealed that municipal use of TIF depleted the property tax revenues of schools during the lifespan of the TIF. Since the quality of public school is proven to be one critical determinant of property values (Brasington, 1999; Haurin & Brasington, 1996), the fear of future decline of school quality can be interpreted as the fear of a drop in property value. If the homevoter hypothesis holds, homeowners within a TAD, or those who live outside a TAD but expect the project to lower their school qualities and decrease their property values, are less likely to favor Atlanta BeltLine, thus, showing that the support is influenced by the fear of a decline in school quality, which is indirect evidence that the support is affected by potential property value decline.

Data

Data used in this study is collected mainly from an online survey about Atlanta BeltLine, conducted in the summer of 2009. 37 questions were asked, including questions about participants' backgrounds, their opinions about the Atlanta region as it is and as it may be, and their attitudes and expectations about the BeltLine project. To avoid the warm glow effect, i.e. respondents tend to "pretend" they support the BeltLine in a BeltLine survey, the invitation letter indicated that it was an opinion survey for Atlanta area residents on the topic of "housing, green space, and transportation." A random sample was drawn from Survey Sampling International's (SSI) online panel, selecting adults in the Atlanta metropolitan area, with 60 percent from within the city of Atlanta. A response rate of 5 percent was reported, which is favorable compared with those of other web-based surveys. The spatial distribution of the 946 respondents is shown in Figure 4.2, and the descriptive statistics of key variables are summarized in Table 4.3.

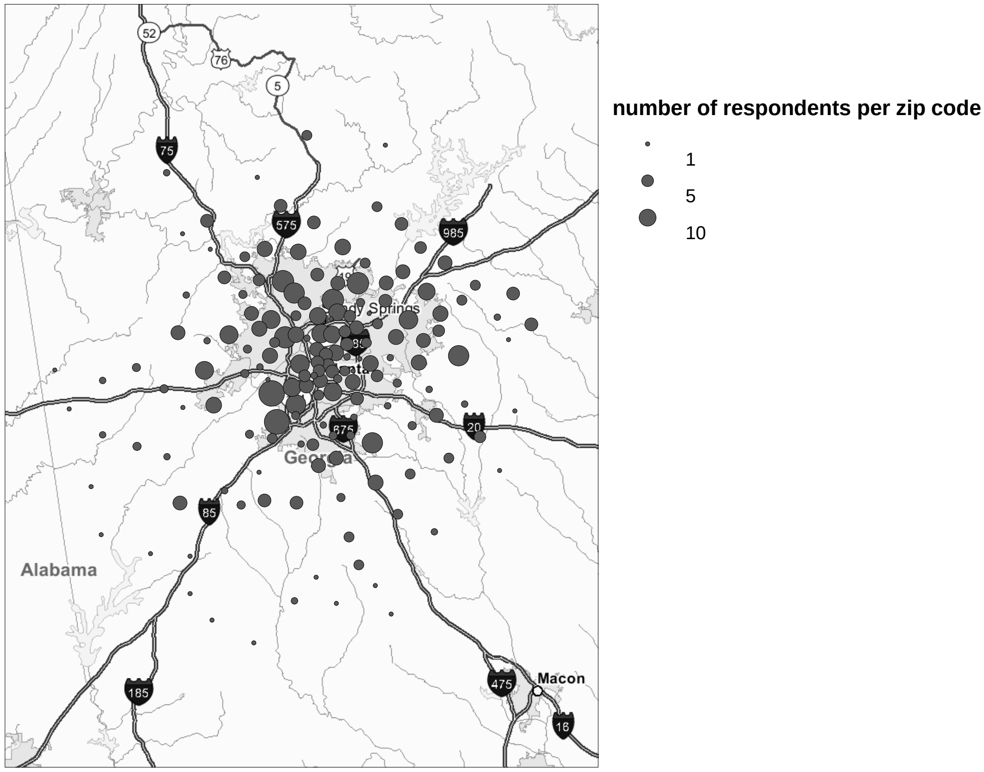


Figure 4.2 Spatial Distribution of Respondents

Table 4.3 Descriptive Statistics of Key Variables

Variable	Obs.	Mean	Sd.	Min.	Max
Support for BeltLine (-2= not support, 0= need info., 2= very support)	900	0.98	1.04	-2	2
BeltLine will transform ATL (0= not sure, 3= strongly agree)	816	1.98	0.81	0	3
I expect the BeltLine to be bigger than planned (-2= much smaller, 2= much bigger)	640	0.10	1.49	-2	2
My frequency of using BeltLine parks/trails (0= never, 3= several times per week)	880	0.95	0.92	0	3
My frequency of using BeltLine transits (0= never, 3= several times per week)	885	1.02	1.04	0	3
Will it get easier to transport in ATL in 5 yrs?	876	-0.68	0.58	-1	1
Hoped my nbhd would change	929	-0.50	0.73	-1	1
ATL quality of life improving in 5 yrs	838	-0.38	0.72	-1	1
How long lived at current home (midpoints)?	873	10.11	9.65	0.5	30
How long lived in ATL (midpoints)?	874	20.53	10.80	0.5	30
Household size (midpoint)	873	2.60	1.38	1	8
Autos owned	876	1.94	1.10	0	6
Commute minutes (round trip, 0 = no commute)	780	39.60	50	0	480
Household income (midpoint)	852	68,430	44,541	15,000	190,000
Age (midpoint)	870	49.44	13.81	21	75
Years of education (midpoint)	865	15.15	2.26	10	19

Respondents were generally supportive of the BeltLine. In a -2 to 2 scale, the mean is close to 1, which represents “(The BeltLine) is more good than bad”. Respondents also had a relatively strong belief that the BeltLine will transform Atlanta. This is consistent with my previous argument that the BeltLine is presented as a project to transform the whole city. On average, respondents were 50 years old, with a relatively high education level, had a household of 2.6 people, and had an annual household income of $68,000.

Since this study focuses on the relationship between support and the distance to the BeltLine TAD, the location of respondents must be accurate. At the end of the survey, respondents were asked if they wanted to receive a report when the survey was done. Half of them provided their actual mailing addresses for the report. For the other half, their locations are accurate only at the zip code level. One simple solution to getting accurate locations is by dropping all records without accurate addresses, but dropping these records will possibly cause two problems. First, dropping half of the records with accurate information just because of one missing variable is simply a waste from the perspective of efficiency. Secondly, and even more importantly, dropping records without actual addresses raises the concern of selection bias. It is fair to argue that people willing to receive reports care more about green space and transportation, thus are more likely to support projects like the BeltLine. To keep a large sample size, and to avoid selection bias, this study introduces four approaches to impute missing location. These approaches are described in detail in the Methodology section.

Another critical issue regarding the distance to BeltLine TAD is whether the respondent is within the "donut" of BeltLine, in other words, living in a residence that has the Beltline on all sides of it. The distance effect within the donut can become blurry because being away from one side of the BeltLine necessarily means being closer to another side. Thus, the distance effect for this group of respondents is expected to be smaller. However, the sample size of this group might be too small to affect the comprehensive result. Only 7 (out of 459) respondents with actual addresses were within the donut. After including all missing-address respondents that were in zip codes adjacent to BeltLine TAD, the total possible donut respondents numbered only 20 (out of 854).

The maps of jurisdiction, including the zip code maps, the boundary of the BeltLine TAD, and the city of Atlanta, were obtained from the City of Atlanta GIS and other sources. The block group level census data is from GeoLytics, including block group level maps and population data based on the 2000 census.

Methodology

This study aims to identify the relationship between the support for BeltLine and the distance to the BeltLine TAD. Identifying how the distance and having-kids affect the the respondents' support for the BeltLine, will be done by constructing an OLS regression model:

$$Support = (distance)\beta_1 + (jurisdiction)\beta_2 + (kids)\beta_3 + (jurisdiction \times kids)\beta_4$$

$$+(renter)\beta_5 + (renter \times distance)\beta_6 + X\gamma + u,$$

where β and γ are estimators, X is a vector of other explanatory variables, and u is the error term. For simplicity of interpretation, the distance to the BeltLine TAD is logged. The *kids* variable represents the number of kids in the household, which is generated from the survey question of household size. As mentioned previously, since the concern about lowering future school quality affects the whole city, the jurisdiction dummy is defined as being in the City of Atlanta. The interaction term between the jurisdiction variable and the number of children is introduced to capture the additional concern of school quality for parents in Atlanta. The dummy and interaction variables for renters are used to identify the possibly different distance effects between homeowners and renters.

If the homevoter hypothesis holds, the distance effect between homeowners and renters should differ because property value increment has different meanings for these two groups.

In order to further identify the mechanism behind the distance effect, the study adds the accessibility to Beltline amenities to the model:

$$Support = (distance)\beta_1' + (accessibility)\beta_5 + (jurisdiction)\beta_2' + (kids)\beta_3' + (jurisdiction \times kids)\beta_4' + X\gamma' + u'$$

If the accessibility absorbs most of the significance of the distance effect, accessibility is proven to be the main mechanism behind the distance effect. Other mechanisms, including the property value increment and the concern about affordable housing, are less important in regard to residents' support. Two variables, the expected frequency of using BeltLine parks, and the expected frequency of using the transit, are used here to represent the accessibility of BeltLine amenities.

As mentioned previously, only half of the respondents provided their actual addresses. To expand the sample size, and to avoid selection bias, this study introduces four approaches to impute missing locations. First, zip code centroids are used to represent the locations of the no-address respondents. This approach has two significant shortcomings. First, assigning missing-address respondents to zip code centroids makes sense only when the population is uniformly distributed in zip codes. For zip codes that consisted of large non-residential area, such as a large park or other public facilities, using centroids is misleading. Second, assigning all missing-address respondents in a certain zip code to the centroid eliminates the potential explanatory power of distances on their different support levels. This is likely to introduce measurement error and weaken the significant level of distance coefficients.

The second approach uses population-weighted centroids. Instead of using geographic centroids, the population-weighted centroids can help avoid the first shortage mentioned in the previous paragraph. By overlapping the census-block population map and zip code map, the population distribution at the block group level within each zip code area can be captured. The population-weighted centroids can thus be generated, which is expected to be more accurate than geographic centroids, in the sense of introducing the consideration of population distribution. However, this approach does not help mitigate the problem of eliminating the explanatory power of distance since all missing-address respondents in a zip code are still assigned to the same location.

The third approach is multiple imputation. In theory, the imputation approach first regresses valid distances on all the other variables, including the support for the BeltLine, and utilizes the regression results to impute missing distances (Little, 1992). This approach generates a specific distance for each missing-address respondent, thus eliminating the problem of assigning missing values to the same location. As a result, the distance effect with this approach is expected to be more statistically significant than that with centroids approaches.

One concern about the imputation method is that regression coefficients are directly applied to the generation of missing distances, neglecting the fact that regression estimates are distributions, not constants. To fix this problem, this study introduces a multiple imputation approach as the third approach of generating missing distances. The concept of multiple imputation is similar to that of regular imputation, but instead of applying regression coefficients as constants, multiple imputation picks random numbers for variables in each imputing round, based on the distributions behind the regression coefficients and standard errors (Rubin, 1987). After this imputation process is repeated multiple times, a series of counterfactual distances for each missing-address respondent is generated. In this study, the imputation process is repeated 1,000 times. The average of these 1,000 imputed distances is used in the empirical model.

Finally, the fourth approach to filling missing distances applies a truncate regression method to the multiple imputation process. One problem with the imputation process is that the imputed distance might fall outside of the possible range, given the restriction of zip code boundaries. For each missing-address respondent, the possible distance to the BeltLine TAD is bounded by the shortest and longest distance from the zip code to the BeltLine TAD. To add this restriction to the multiple imputation process, this study introduces the method of truncate regression. By providing the lower and upper bounds for each missing distance, truncate regression allows the multiple imputation process to generate imputed distances that are within zip code boundaries.[6] Again, the imputation is repeated 1,000 times, and the mean is recorded as the imputed distance.

The distances generated by these four methods are introduced into the regression model. The estimated effects of distance are then compared with each other, and with the estimator generated by including respondents with actual addresses.

[6] Due to computation limitation, the upper bound of missing distance is generated by doubling the distance between lower bound and geographic centroid:

(Upper bound)=(lower bound)+2×(Distance between lower bound and the centroid)

The generation of jurisdiction variable is straightforward for respondents with actual locations. Dummy variables are generated with GIS tools, based on whether they are in the jurisdictions or not. This is a more complicated task for missing-address respondents since their actual locations are not known. In this study, the proportion of zip code areas within certain jurisdiction districts is used to generate the variable. For example, for missing-address respondents in a zip code that is not adjacent to the BeltLine TAD, their BeltLine TAD variable is coded as zero. For a zip code that that is half within the City of Atlanta, the Atlanta variable is coded as 0.5 for respondents with missing addresses in it.

Results

The regression results are listed in Table 4.4. Each column represents a specific approach to imputing missing distances. Column (0) lists results for actual-address respondents only. Column (1) locates missing addresses at their zip code geographical centroids; Column (2) uses population-weighted zip code centroids; Column (3) utilizes multiple imputations; Column (4) applies multiple imputations with truncate regressions. The dependent variable is a categorical variable of the support for the BeltLine, ranging from -2 to 2. The independent variables include logged distance, jurisdiction dummies (located in BeltLine TAD/ City of Atlanta), number of kids, interaction terms between the number of children and the jurisdiction dummies, and demographic characteristics of respondents, including logged household income, age, and years of education.

Table 4.4 Distance Effect on Support for Atlanta BeltLine

	(0)	(0)'	(1)	(2)	(3)	(4)
Imputation Method	**None**	**All to zipcode centroids**	**Zipcode Centroids**	**Pop-weighted Centroids**	**Multiple Imputation**	**MI: truncate regression**
Obs.	449	812	813	809	801	782
R-squared	0.057	0.031	0.040	0.035	0.046	0.044
log(distance) (km)	-0.16 **	-0.08	-0.14 **	-0.10 *	-0.17 ***	-0.15 **
In TAD	-0.46	-0.17	-0.46	-0.29	-0.38	-0.49
In City of Atlanta	0.06	0.20	0.03	0.14	0.15	0.04
Number of Kids	0.05	0.05	0.05 *	0.05	0.03	0.05
ATL*Kids	-0.34 ***	-0.36 ***	-0.35 ***	-0.33 ***	-0.30 ***	-0.37 ***
Renter	0.02	0.15	0.14	0.15	0.05	0.15
Renter*log_dist	-0.04	-0.04	-0.04	-0.04	-0.02	-0.05
Household income (log)	-0.10	-0.03	-0.04	-0.03	-0.03	-0.04
Age (yr)	-0.01 **	0.00	0.00	0.00	0.00	0.00
Education (yr)	0.01	0.00	0.00	0.00	0.00	0.00
Constant	2.92 ***	1.68 **	2.01 ***	1.80 ***	1.98 ***	2.12 ***
Kids Joint Test	***	***	***	***	***	***

*p < .10. **p < .05. ***p < .01.

The distance effect is consistently significant among all four imputation methods, as well as in the model without any imputation (column (0)). The same trend also applies to other variables, showing that imputation does not significantly affect the result, no matter what kind of imputation method is selected. In other words, imputation enlarges the sample size without disturbing the results. Thus, imputation should be a preferred solution to the missing-precise-location issue, rather than dropping observations with missing values. Note that imputation is only a preferred solution when at least some precise locations are known in the data. Column (0)' indicates an artificial case, as if there were no precise location recorded in the dataset. The distance to the BeltLine TAD is generated by assigning all observations to their corresponding zip code centroids. The insignificant distance effect in model (0)' shows that imputing too many missing values without having any actual data raises the issue/possibility of disturbing the result.

The selection of imputation methods does not seem to affect the results for either the magnitude of coefficients or the significances. The distance effects in all four imputation models are significant, and the coefficients range only from -0.10 to -0.17. Holding all the other variables constant, doubling the current distance to the BeltLine TAD will decrease respondents' support by only less than one-fifth category. However, given that the support variable have only five categories, and given that 73 percent of the homeowners in the sample are supportive of the BeltLine (reporting 1 or 2 in the survey), the distance effect on support is significantly strong.

Jurisdiction variables and the number of kids alone do not explain different attitudes towards the BeltLine in any of the models. The significances of jurisdiction variables are likely to be absorbed by the distance variables since they are just measuring the same thing in different ways. The variable of the number of children is not significant either. In theory, the number of children can affect the attitudes about the BeltLine in two ways. First, having more children could potentially create additional value of access to parks, trails, and even transit for parents because children usually love to play in such places. However, this additional support on the part of parents does not show in the result. Second, as mentioned previously, parents with kids are likely to worry that the implementation of the BeltLine TAD might hurt the future quality of public school, thus reducing their support for the project. This dis-support is sensitive to jurisdictions. Only parents in school zones affected by the BeltLine TAD need to worry about this. School catchment zones are difficult to include in the model because they differ by levels. Jurisdiction (City of Atlanta) is a good proxy for school catchment zones because if school quality is affected by the BeltLine, it is fair to argue that the impact will be eventually borne by all public schools in the city. Since both of these mechanisms interact with jurisdictions, it is expected that the number of children alone does not affect the support, after the interaction terms of jurisdictions and the number of children are included.

The interaction terms between jurisdictions and the number of children generally support the argument in the previous paragraph. The interaction terms between the City of Atlanta and the number of children show a strong and significant negative effect on the support. Holding all the other variables constant, having one additional child decreases the support for Atlanta BeltLine by one-third level for respondents in the City of Atlanta. After all, the joint significance of two child-related variables shows that the number of children does affect respondents' support for the BeltLine.

The two renter-related variables are not significant for any of the (four) models, showing that the distance effect for renters and homeowners is not significantly different. This result is inconsistent with the homevoter hypothesis. As the homevoter hypothesis holds, property value increment is the main mechanism behind the distance effect. In this case, renters should not be as supportive for the BeltLine as homeowners at the same close distance because renters will suffer from the property value increment by having to pay higher rents. There is also no evidence, for all imputation models, showing that the demographic characteristics of respondents affect their support for the BeltLine project. However, the demographic characteristics serve as a good set of control variables.

To further identify the mechanism behind the distance effect, the second set of models that includes accessibility variables is introduced in this study. The results are listed in Table 4.5. Both restricted models (without accessibility) and unrestricted models are listed in Table 4.5, for comparison. In order to further investigate the difference between homeowners and renters, the study treats these two groups as two subsamples. Though results from different imputation methods are similar to each other, the comparison focuses on results using multiple imputation with truncate regressions.

Table 4.5 Mechanisms behind the Distance Effect

	Restricted		**Unrestricted**	
	Homeowners	**Renters**	**Homeowners**	**Renters**
Obs.	526	208	526	207
R-squared	0.0301	0.0889	0.2081	0.2113
log(distance) (km)	-0.14 **	-0.23 ***	-0.02	-0.11
BL park usage			0.34 ***	0.12
BL transit usage			0.25 ***	0.24 ***
In TAD	-0.48	-0.36	-0.25	0.39
In City of Atlanta	0.04	-0.08	-0.05	-0.07
Number of Kids	0.08	0.00	0.04	-0.01
ATL*Kids	-0.39 **	-0.39 ***	-0.30 **	-0.32 **
Household income (log)	-0.05	0.01	-0.05	-0.01
Age (yr)	0.00	-0.01	0.00	0.00
Education (yr)	-0.01	0.00	0.00	-0.02
Constant	2.22 **	1.90	0.97	1.34
Kids Joint Test	**	**	**	**

*p < .10. **p < .05. ***p < .01.

The accessibility variables are generally strongly significant for both homeowners and renters. Also, the significance of the distance variable disappears after the inclusion of accessibility. This result strongly suggests that the main mechanism behind the distance effect is the accessibility to BeltLine amenities. In other words, property value increment is not likely to be significantly correlated with support for BeltLine, which rejects the homevoter hypothesis. There are several explanations for this surprising result. First, the logged distance to the BeltLine TAD might not be a good proxy for the price gradient caused by the project. Maybe the price effect is hidden in the interaction terms between the logged distance and any omitted neighborhood quality variables, which were not able to be tested in this study due to the limitations of the data. More importantly, given that Atlanta BeltLine is a mixed project that includes green space, transit, and affordable housing, the price gradient might not be as straightforward as a function of distance. Noonan (2012) provides some empirical evidence that the price impacts of BeltLine (driven by speculation) are not consistently positive according to a variety of hedonic price models. Second, this study tests its hypothesis based on survey responses, instead of actual votes. The relatively low response rate and potentially less deliberation in survey answers could possibly, though not likely, bias the result. Last but not least, residents might just not be rational or deliberative enough to consider their support for BeltLine outside of their direct use value. In this regard, homevoter hypothesis is rejected in the case of Atlanta BeltLine.

There is no significant difference between the results for homeowners and those for renters, except for the accessibility to parks. However, the insignificance of accessing parks is not enough to reject the dominance of accessibility since the distance effect still losses significance after the introduction of the accessibility to public transit. Thus, the similar attitudes between renters and homeowners also imply that the homevoter hypothesis should be rejected in the case of Atlanta BeltLine.

Finally, a robustness test that uses ordered logit models, instead of OLS, is performed. The results for ordered logit generally tell the same story as the results for the OLS models performed in this study.[7] This study uses OLS models for two main reasons. First, ordered logit presumes an S-shape relationship, while OLS tests the relationship linearly. Though the dependent variable (support for the BeltLine) is indeed a categorical variable ranging from -2 to 2, there is no particular reason to believe that the relationship between support and the distance is not a linear line. Besides, the coefficients of OLS models, especially interaction terms, can be interpreted more intuitively than those of ordered logit models. This facilitates the understanding of relationships between variables, and the relationship of interests.

Discussions and Conclusions

This study aims to determine the distance effect on the support of Atlanta BeltLine. The support significantly declines as distance from the BeltLine increases without the inclusion of accessibility. However, the inclusion of accessibility absorbs the significance of the distance effect, which leads to the conclusion that the direct usage value is the main mechanism behind the distance effect. This result rejects the homevoter hypothesis, given that property value increment is not directly related to the distance effect on support. Also, the result shows that being parents in the City of Atlanta does affect

[7] The result for ordered logit models are available upon request.

homeowner or renter support for the BeltLine. This result leads to the conclusion that the issue of TIF affecting future school quality does lessen the support for the project.

The most important policy application of this study is to disprove the impact of property value increment on support for Atlanta BeltLine. In this case, residents support Atlanta BeltLine because it provides local amenities, without considering whether this project benefits their housing price or not. This result serves as a possible guide to the direction of promotion for these kinds of urban redevelopment projects. However, the result should be interpreted with caution since respondents gave responses to the support and the usage expectation questions almost simultaneously. For example, although this study suggests that residents who use parks more frequently are more likely to support the BeltLine, it is inappropriate to claim that encouraging residents to use parks more stimulates their support for the program. The causal relationship remains vague in this case.

There are several concerns and limitations to this study, especially in the data processing. The first issue is the possible measurement error issue in the dependent variable, the support for the BeltLine. Respondents of the survey range all over the metro Atlanta area. It can be argued that some of them are too far fom the project to be exposed to the effect of Atlanta BeltLine— though the effective range of BeltLine is also arguable. If we somehow restrict the effect of BeltLine to a certain range, no matter where the TAD is, a 1-mile buffer, or a 5-mile buffer, respondents that are outside the effective area are considered not affected. One argument is that these distant responses are just random noise, which should not affect the estimates. On the other hand, another argument is that these distant respondents respond only because they care about the BeltLine, even though they are not really affected. In this case, the supports from these distant areas in the sample are expected to be too high, compared to the population. This self-selection problem might result in a too low estimate for the distance effect because the expected low support in distant areas is not as low as it should be. To sum up, the measurement error caused by distant respondents either does not affect the estimators or gives us a lower-bound estimate, depending on how the measurement error is interpreted.

Another measurement error issue comes from missing addresses. Half of the respondents refused to provide their addresses after they were asked if they wanted to receive the final report of the survey. This missing-address issue is not random because respondents who refuse to receive the report are likely to care less about the survey, thus should be less supportive of the BeltLine. Thus, it may be concluded that dropping people who offer less support will bias the estimator of distance toward zero since the effect is diluted. However, this bias is not significant in the results. The four approaches to impute these missing addresses also introduce potential problems. First, the centroids methods introduce additional random measurement errors, thus inflating the standard errors and lowering the significance level.

Measurement error in the variable of the number of children also needs further revision. First, using categorical family size to infer number of kids creates error. Again, this random error in independent variable is expected to increase standard error and biases the estimator toward zero. Also, not having control for the age of the children is another source of error. For respondents with older children, having children should not affect the distance effect since it can be assumed that these parents would care less about the potential negative influence of the BeltLine on school expenditure. Including this group of samples will amplify the effect of children, while they are wrongly expected to have less support when they really don't. On the other hand, failure to include respondents who plan to have children in the future will bias the child effect downward since this group's lower level of support should be due to their worry about possibly reduced future school expenditures, which this study fails to capture. If all the sources of error are taken together, the estimated effect of children in this study is expected be biased, but the direction is unclear.

Generally speaking, the possible sources of measurement errors are likely to either amplify the variances or even bias the estimators. As long as we understand the possible error sources and the consequences, we can at least have a sense of what the true values would likely be, even if the estimators are not perfectly accurate.

References

Atlanta BeltLine. (2013). Atlanta BeltLine Website, from http://beltline.org

Atlanta Development Authority. (2005). Atlanta BeltLine Redevelopment Plan: http://www.investatlanta.com/adaInitiatives/BeltLineRedevelopmentPlanA.jsp.

Brasington, D. (1999). Which Measures of School Quality Does the Housing Market Value? *Journal of Real Estate Research, 18*(3), 395-413.

Briffault, R. (2010). The Most Popular Tool: Tax Increment Financing and the Political Economy of Local Government. *The University of Chicago Law Review, 77*(1), 65-95.

Brunner, E., & Sonstelie, J. (2003). Homeowners, property values, and the political economy of the school voucher. *Journal of Urban Economics, 54*(2), 239-257. doi: 10.1016/s0094-1190(03)00063-9

Brunner, E., Sonstelie, J., & Thayer, M. (2001). Capitalization and the Voucher: An Analysis of Precinct Returns from California's Proposition 174. *Journal of Urban Economics, 50*(3), 517-536. doi: 10.1006/juec.2001.2232

Dehring, C. A., Depken Ii, C. A., & Ward, M. R. (2008). A direct test of the homevoter hypothesis. *Journal of Urban Economics, 64*(1), 155-170. doi: 10.1016/j.jue.2007.11.001

Fischel, W. A. (2005). The homevoter hypothesis :how home values influence local government taxation, school finance, and land-use policies: Harvard University Press.

Haurin, D. R., & Brasington, D. (1996). School Quality and Real House Prices: Inter- and Intrametropolitan Effects. *Journal of Housing Economics, 5*(4), 351-368. doi: http://dx.doi.org/10.1006/jhec.1996.0018

Immergluck, D. (2009). Large Redevelopment Initiatives, Housing Values and Gentrification: The Case of the Atlanta Beltline. *Urban Studies, 46*(8), 1723-1745. doi: 10.1177/0042098009105500

Little, R. J. A. (1992). Regression With Missing X's: A Review. *Journal of the American Statistical Association, 87*(420), 1227-1237.

Man, J. Y., & Rosentraub, M. S. (1998). Tax Increment Financing: Municipal Adoption and Effects On Property Value Growth. *Public Finance Review, 26*(6), 523-547. doi: 10.1177/109114219802600601

Noonan, D. S. (2012). Amenities Tomorrow: A Greenbelt Project's Impacts over Space, Time. In H. S. Banzhaf (Ed.), *Markets for Land and Pollution: Implications for Environmental Justice* (pp. 170-196): Stanford University Press.

Roncek, D. W., Bell, R., & Francik, J. M. A. (1981). Housing Projects and Crime: Testing a Proximity Hypothesis. *Social Problems, 29*(2), 151-166. doi: 10.2307/800421

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc.

Weber, R. (2003). Equity and Entrepreneurialism: The Impact of Tax Increment Financing on School Finance. *Urban Affairs Review, 38*(5), 619-644. doi: 10.1177/1078087403038005001

Chapter 5

Conclusion

The results of all three essays indicate that property owners' attitudes towards sustainable urban development are sensitive to the spatial environment they face. From the methodology perspective, these results confirm the necessity of taking spatial issues into consideration while constructing models for urban redevelopment projects. From the policy application perspective, these results provide evidence of how neighboring actors and conditions affect property owners' preferences and behaviors. A detailed contribution for each essay from both a methodological and a policy application perspective is discussed in below.

The first essay introduces spatial modeling techniques, including a spatial fixed-effect model and a spatial autoregressive model with spatial autoregressive disturbances, into the conventional hedonic housing price analysis. The result shows that in the determining of price effects of neighboring environments, such as green space, density and diversity, the consideration of spatial autocorrelation is inevitable. The results also show that neglecting spatial autocorrelation between variables of interest and residuals leads to biased estimators and misleading interpretations. The result per se has a significant policy application. By realizing that homebuyers prefer to have multiple parks within a walkable distance, rather than having a single large park in the neighborhood, policy makers or city planners will be able to be more strategic in the placement and design of greenspace. For example, the City of Atlanta is targeting increasing publicly accessible parkland from current 7.5 Acres/1,000 residents to 10 Acres/1,000 residents in the future (City of Atlanta, 2009). In the sense of boosting property values with parkland and investing the additional property tax revenue to create more parks, the policy goal is too vague to be optimal. Instead, targeting to increase the number of parks, rather than the total park land per 1,000 residents might be a more effective goal to achieve.

The first contribution of the second essay from the methodology perspective is to quantify the proportion of signaling in building owners' decision-making process for which certificate level of LEED to target. The signaling factor generated in this study indicates that, at certificate levels lower than Gold, governmental and non-profit-organization owners are more driven by the signaling effect than are profitable firms,. This result is highly applicable to the policy-making process of promoting green building certification. For example, subsidies for the cost of applying for LEED certification are not effective in promoting LEED because a large number of LEED applicants are governments and non-profit organizations, and their decisions are not based purely on cost effectiveness.

The second part of the second essay further emphasizes the importance of spatial spillover on the role of signaling in the pursuit of LEED certification. From the policy application perspective, this result implies that the effect of pilot projects on promoting green building certification could be significant. By encouraging specific building owners to apply for LEED certification, the spatial spillover of signaling will significantly increase the likelihood of applications from their local competitors, thus boosting the certified rate of certain owner types in the area.

In the third essay, spatially sensitive issues such as accessibility and the worry of lowering future public school quality for parents in affected areas is proven to dominate respondents' support for the project. This essay does not involve advanced spatial techniques, but the result confirms the importance of spatial circumstances on residents' attitudes towards certain urban redevelopment projects. Also, by rejecting the homevoter hypothesis, this result indicates that residents do not always support projects that increase their property values. From the policy application perspective, the result illustrates that residents seem to care more about issues related to their everyday life, such as the enjoyment of amenities, and a concern for school quality, instead of their property value increment. The result can be applied not only to the ongoing BeltLine project, but also to any similar urban redevelopment projects.

To sum up, the three essays in this study emphasize the importance of spatial circumstances on the attitudes of property owners/residents for sustainable urban development. Several advanced spatial techniques that have not been applied to this field are introduced into this study. While these techniques are proven to be critical to urban environment studies, methods and techniques regarding spatial elements in this field need further development and investigation.

References

City of Atlanta. (2009). Project Greenspace Summary Report: Prepared by Wallace Roberts & Todd, LLC.

Appendix

Table A2.1 Full descriptive statistics of all variables

Variable description	Obs	Mean	Std. Dev.
Dependent variable			
Logged sales price	44743	12.212	0.726
Green space variables			
Logged distance to closest park (m)	44743	0.032	1.074
Number of parks within the 0.4km buffer	44743	1.362	2.335
Sum of park area within the buffer	44743	225	1122
Neighboring parcels variables			
Number of parcels within the buffer	44743	668	481
Diversity score	44743	0.174	0.169
Transit variables			
Logged distance to closest Marta station	44743	2.229	1.072
Other spatial variables			
Logged distance to CBD (m)	44743	3.220	0.836
Logged distance to highway	44743	1.208	1.271
Logged distance to closet hydro	44743	-0.677	0.958
Logged distance to ATL airport	44743	3.010	0.724
Property characteristics			
Logged lot size	44743	-1.192	0.710
Imputed stories	44743	1.485	0.403
Imputed indoor area (sq. ft.)	44743	7.542	0.415
Age of house	44743	23.46	26.47
Imputed number of bedrooms	44743	3.297	0.787
Imputed number of family rooms	44743	0.518	0.438
Imputed number of fixed bathrooms	44743	2.192	0.825
Basement: crawl	44743	0.168	0.374
Basement: part	44743	0.070	0.256
Basement: full	44743	0.190	0.393
extwall:frame	44743	0.271	0.444
extwall:brick	44743	0.004	0.066

Table A2.1 Continue

Variable description	Obs	Mean	Std. Dev.
extwall:masonry	44743	0.028	0.164
extwall:block	44743	0.046	0.210
extwall:stucco	44743	0.003	0.056
extwall:aluminum	44743	0.027	0.162
extwall:stone	44743	0.169	0.374
extwall:asbestos	44743	0.001	0.032
extwall:concrete	44743	0.025	0.156
Heat:none	44743	0.009	0.096
Heat:central	44743	0.022	0.145
Heat:central air	44743	0.064	0.244
Heat:heat pump	44743	0.479	0.500
Fuel:gas	44743	0.560	0.496
Fuel:electricity	44743	0.005	0.069
Fuel:oil	44743	0.000	0.014
Fuel:coal/wood	44743	0.001	0.031
Fuel:solar	44743	0.000	0.000
Fuel:none	44743	0.007	0.085
Topo:level	44743	0.613	0.487
Topo:above street	44743	0.144	0.351
Topo:below street	44743	0.031	0.174
Location variables			
Logged median household income in BG	44743	11.029	0.622
Logged population density in BG	44743	7.395	1.109
City of Atlanta	44743	0.345	0.475
City of Atlanta: midtown	44743	0.001	0.025
City of Atlanta: buckhead	44743	0.003	0.059
City: Alpharetta	44743	0.064	0.245
City: College Park	44743	0.009	0.096
City: East Point	44743	0.044	0.205
City: Fairburn	44743	0.034	0.182
City: Hapeville	44743	0.006	0.080
City: Mountain Park	44743	0.000	0.019
City: Palmetto	44743	0.008	0.087

Table A2.1 Continue

Variable description	Obs	Mean	Std. Dev.
City: Roswell	44743	0.080	0.272
Union City	44743	0.031	0.174
City: Unincorporated south Fulton county	44743	0.151	0.358
City: Milton	44743	0.048	0.215
City: Johns Creeks	44743	0.105	0.306
City: Sandy Spring	44743	0.071	0.257
Time-related variables			
Sales in 2000	44743	0.087	0.282
Sales in 2001	44743	0.094	0.292
Sales in 2002	44743	0.116	0.320
Sales in 2003	44743	0.116	0.321
Sales in 2004	44743	0.136	0.343
Sales in 2005	44743	0.106	0.308
Sales in 2006	44743	0.093	0.290
Sales in 2007	44743	0.100	0.300
Sales in 2008	44743	0.089	0.285
Sales in 2009	44743	0.062	0.241
Sales in 2010	44743	0.001	0.028
Sold in winter	44743	0.193	0.395
Sold in spring	44743	0.292	0.455
Sold in summer	44743	0.311	0.463
Sold in fall	44743	0.204	0.403

Table A2.2 Comparison of Results from different models

		dist. to park (log)				Num. of park			
	Obs.	OLS	SFE	SAR	SFEAR	OLS	SFE	SAR	SFEAR
Pooled	44,743	-0.022 ***	-0.006 **	-0.018 ***	-0.007 **	0.014 ***	0.009 ***	0.009 ***	0.007 ***
00	3,889	-0.031 ***	-0.015	-0.032 ***	-0.017	0.018 ***	0.022 ***	0.015 ***	0.023 ***
01	4,210	-0.017 *	-0.02 **	-0.020 **	-0.015	0.026 ***	0.017 ***	0.019 ***	0.019 ***
02	5,174	-0.005	0.003	-0.005	-0.002	0.015 ***	0.007	0.010 **	0.003
03	5,208	-0.014	-0.004	-0.012	-0.007	0.015 *	0.017 **	0.019 ***	0.012 *
04	6,095	-0.033 ***	-0.022 ***	-0.031 ***	-0.027 ***	-0.003	-0.008	0.00	-0.009
05	4,744	-0.018 **	-0.004	-0.010	0.005	0.017 ***	0.004	0.018 ***	0.006
06	4,156	-0.032 ***	-0.006	-0.018 **	-0.002	0.009 **	-5E-04	0.013 ***	0.003
07	4,481	-0.022 ***	0.023 ***	-0.023 ***	0.007	0.010 **	-0.001	0.008 **	-0.006
08	3,983	-0.017 *	0.001	-0.010	-9E-04	0.018 ***	0.008	0.011 **	0.003
09	2,767	-0.033 ***	-9E-04	-0.023 **	-0.005	0.006	-0.003	0.009 *	0.002

*p < .10. **p < .05. ***p < .01.

		Sum of park area (10,000 acre)				Parcel density (100 pcl.)			
	Obs.	OLS	SFE	SAR	SFEAR	OLS	SFE	SAR	SFEAR
Pooled	44,743	-0.025	0.020	0.00	0.01	-0.006 ***	-0.009 ***	-0.004 ***	-0.006 ***
00	3,889	-0.041	-0.051	-0.02	-0.02	-0.018 ***	-0.013 ***	-0.022 ***	-0.011 ***
01	4,210	0.062	0.146 *	0.18 **	0.15 *	-0.013 ***	-0.005	-0.011 ***	-0.001
02	5,174	-0.099 *	-0.055	0.00	-0.06	-0.010 ***	-0.013 ***	-0.012 ***	-0.005 *
03	5,208	-0.042	0.015	-0.09 *	-0.01	-0.012 ***	-0.013 ***	-0.009 ***	-0.005 *
04	6,095	0.116	0.241 ***	0.11 *	0.21 ***	-0.008 ***	-0.01 ***	-0.01 ***	-0.008 ***
05	4,744	0.098 *	0.097	0.10	0.06	-0.007 ***	-0.007 ***	-0.006 ***	-0.011 ***
06	4,156	-0.055	0.007	-0.06	0.05	-0.001	-0.002	0.00	-0.003 *
07	4,481	-0.107	-0.045	-0.09	-0.05	-0.001	-0.004 *	0.00	-0.002
08	3,983	-0.052	-0.036	-0.03	-0.01	-0.003	-0.008 ***	-0.004 **	-0.005 **
09	2,767	-0.081	-0.03	-0.06	-0.03	0.002	-0.005 **	0.00	-0.002

*p < .10. **p < .05. ***p < .01.

Table A2.2 Continue

	Obs.	Land Use Diversity OLS	SFE	SAR	SFEAR	Distance to Marta (log) OLS	SFE	SAR	SFEAR
Pooled	44,743	-0.007	-0.118 ***	-0.111 ***	-0.109 ***	-0.04 ***	-0.012	-0.082 ***	-0.014
00	3,889	-0.017	-0.31 ***	0.0161	-0.29 ***	-0.039 **	0.029	-0.032	0.046
01	4,210	0.080	-0.159 **	5E-06	-0.212 ***	-0.073 ***	-0.052 *	-0.097 ***	-0.036
02	5,174	0.048	-0.117 **	-0.032	-0.145 **	-0.089 ***	-0.008	-0.093 ***	-0.003
03	5,208	-0.085	-0.194 ***	-0.159 ***	-0.103 *	-0.055 **	-0.05	-0.042 **	-0.035
04	6,095	0.134 ***	-0.072 *	0.1043 **	-0.111 ***	-0.033 *	0.079 ***	-0.04 **	0.046
05	4,744	-0.179 ***	-0.215 ***	-0.24 ***	-0.166 ***	-0.022	0.027	-0.034 **	0.018
06	4,156	-0.074	-0.126 **	-0.139 ***	-0.076	0.003	-0.046 *	-0.071 ***	-0.069 ***
07	4,481	-0.104 **	-0.127 **	-0.084 *	-0.115 **	-0.009	0.022	-0.022	-0.016
08	3,983	-0.096 *	-0.145 ***	-0.086 *	-0.061	-0.074 ***	-0.104 ***	-0.078 ***	-0.133 ***
09	2,767	-0.227 ***	-0.085	-0.161 **	-0.082	0.031	-0.066	-0.067 **	-0.09 *

*p < .10. **p < .05. ***p < .01.

A2.3: Spatial correlation within residuals

The main cause of endogeneity in this study is the spatial autocorrelation. Theoretically, if spatial autocorrelation can be identified in the residuals of a certain model, then the estimators of spatial related variables are endogenous. A map of local indicators of spatial association (LISA) is a useful tool to identify the spatial correlation in the residual. Developed by Anselin (1995), a LISA map can demonstrate the level of spatial correlation for each variable compared to the average distribution. In the map below, dark dots denote high prices clustering with high prices, white dots refer to low-low clustering, and gray dots are non-significant. By looking at the LISA maps of residuals from pooled OLS and spatial fixed-effect models, the spatial correlation left in the residuals can be identified. As shown in the figure below, there is some spatial correlation left in the residuals of the pooled OLS model, and the correlation is significantly less in the spatial fixed-effect residuals. This result indicates that the spatial fixed-effect model effectively mitigates the endogeneity problem. However, the residual map of the SARAR model is not available due to calculation limitations, so a comparison between all three models is not available either. However, it is a good practice to test the existence of spatial correlation in OLS and spatial fixed-effect models.

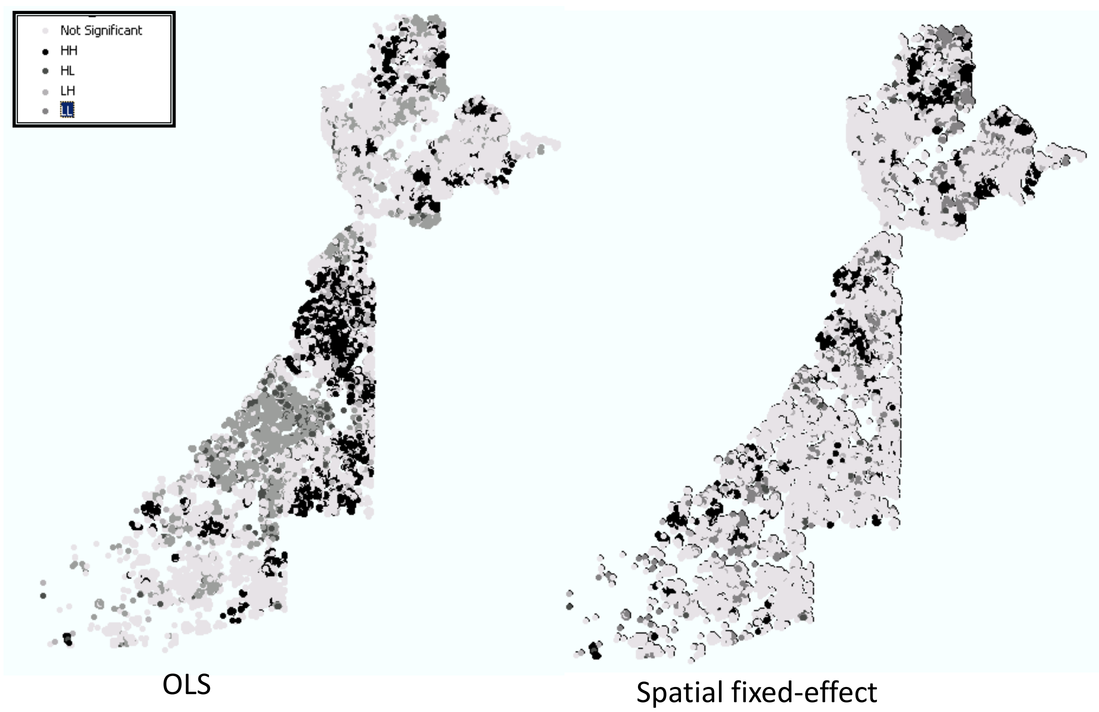


Figure A2.3 Comparison of Spatial Autocorrelation in Residuals