Three Essays in Hedonic Housing Price

by

Sezhu Liu

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 13, 2014

Keywords: housing price, spatial autocorrelation, regression discontinuity, propensity score
green space, natural experiments, difference in difference

Copyright 2014 by Sezhu Liu

Approved by

Diane Hite, Chair, Professor of Agricultural Economics and Rural Sociology
Valentina Hartarska, Professor of Agricultural Economics and Rural Sociology
Denis Nadolnyak, Professor of Agricultural Economics and Rural Sociology
Deacue Fields, Professor of Agricultural Economics and Rural Sociology

Abstract

Chapter 2 attempts to quantify the impacts of green space, by using the hedonic price analysis of the relationship between property values and the green space amenities around the selected single family houses in Delaware County, Ohio. Also, by incorporating spatial-lag term, we can compare the results with and without spatial effect. Eventually, after extending the model by quantile regression, the influence of different green space characteristics on housing price may change across the conditional distribution of housing price. Substantial variation was found between the results with and without spatial effects across quantiles, which indicates that luxury house buyers may value green space differently from middle or low level house buyers.

Chapter 3 employs spatial autoregressive hedonic models within a difference-in-difference and regression discontinuity framework, first investigate the effect of a new high school establishment while controlling for other locational amenities and disamenities in neighborhoods on housing prices in the whole county, and then also investigate those effects on housing price for houses adjacent to the boundary of the new school district in Lee County, North Carolina.

Chapter 4 examines the effect of the legalization of same-sex marriage in Massachusetts. The Massachusetts Supreme Court ruled that a ban on same-sex marriage was unconstitutional, and was thus the first state in the country to legalize gay marriage in 2004. Based on this event, propensity score difference-in-difference and spatial difference-in-difference methods are used to

measure the effect of the law on housing prices in Massachusetts relative to the bordering State of New York.

Acknowledgments

My deepest gratitude goes out to my advisor, Dr. Diane Hite, for her kindness and patience in guiding me through the writing of this thesis. Thanks also to Dr. Valentina Hartarska, Dr. Denis Nadolnyak and Dr. Deacue Fields, who helped me a lot in my study.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

OLS	Ordinary Least Squares
GS2SLS	Generalized Spatial Two State Least Squares
ML	Maximum Likelihood
DID	Difference-In-Differences
PSW	Propensity Score Weighting
2SLS	Two Stage Least Squares
HSQR	Hedonic Spatial Quantile Regression
RD	Regression Discontinuity
IPTW	Inverse Probability of Treatment Weighting

CHAPTER 1

Green space is an important part of the environment around houses. Generally, most research has been focused on the economic impacts of green space on urban planning and environmental pollution cost, but has ignored the impact on single family home values. Limited research was conducted in this area and few studies of green space and housing prices have incorporated spatial econometric techniques. This technique is necessary since housing value also influenced by characteristics of nearby properties. Based on this technique, Chapter 2, Chapter 3 and Chapter 4 use different models to check the effect of different influencing factors on housing price.

Chapter 2 examines the different influences of green space on housing price across the conditional distribution of housing price by using hedonic spatial quantile regression. Regression discontinuity method has recently become a commonly applied framework for empirical work in economics. Chapter 3 introduces the regression discontinuity design by using to shed some light on whether the potential effects on objects just above the cut-off point are very similar to the objects just below the cut-off point and whether the object far from the cut-off point will make these effects more obvious. By using spatial autoregressive hedonic models within a regression discontinuity and difference-in-difference framework, the results indicate that the creation of new high school has a positive influence to property values, but this positive effect only significant near the border of school district, there is no effect when extended to the whole county level. Chapter 4 also uses spatial autocorrelation but with different model and dataset. This chapter examines the effect of the legalization of same-sex marriage in Massachusetts.

Difference-in-difference and propensity score methods are used in this chapter to measure the effect of the same sex marriage law on housing prices. The primary results indicate that the passage of the same-sex marriage law significantly raised property value.

CHAPTER 2

Measuring the Effect of Green Space on Property Value: An Application of the Hedonic Spatial Quantile Regression

1. Introduction

An attractive environment is likely to influence house prices. Houses with an attractive environment will have an added value over similar, less favorably attractive ones. The presence of trees and forests can make the environment a more pleasant place to live, work, and spend leisure time and thus makes substantial improvements in individual well-being, including opportunities for leisure out in the yard or in the neighborhood, reduced heating and cooling costs, privacy, and the lack of a need to construct fences or screens. Moreover, forests can strongly influence the physical/biological environment and mitigate many impacts of development by moderating climate, conserving energy, carbon dioxide and water, improving air quality, water purification, controlling rainfall runoff and flooding, and harboring wildlife thus enhancing the attractiveness of nearby parcels. Furthermore, field tests have shown that properly designed plantings of trees and shrubs significantly reduce noise. In sum, green space provides multiple benefits including recreational opportunities, aesthetic enjoyment, and ecological services.

As Nanette, Jeffery and Laurie (2002) represented, the effect that environmental amenities, such as forested areas and green open space contribute to the value of real estate is often estimated using the hedonic pricing approach, a method that was based on the straightforward premise that the value of a good depends on the stream of benefits derived from that good. Using regression techniques, the hedonic pricing method identifies what portion of the differences in property value can be attributed to environmental amenities, such as green space. The sales value of real estate reflects the benefits that buyers attach to the attributes of that property, including the trees and forest resources found near the property, along the street, and in neighboring parks and greenways.

Although the hedonic model for housing was commonplace and there are a lot of studies that explored the effect of different environmental factors on house prices, few of them focus on the green space effect; even if some of them did, they are quite simple through variables and methods, since they only contain one or a few amenities and rarely contain socio-economic variables, due to a lack of using the large number of census survey data available to us. My approach offers the potential for a richer model: first, beyond the traditional variables to explain residential values, such as housing characteristics of the parcel and distance to amenities, I also create some environment indexes to evaluate its effect on housing price more comprehensively. Second, the idea that location is an important factor in determining the property value is not new, but few people seek the factor effect, which varies with the change in housing price. This paper allows for spatial heterogeneity in estimation by introducing the spatial econometric method and combined with the quantile regression to see the location effect on the different level of housing price, results indicate substantial variation exists across quantiles, suggesting that ordinary least squares (OLS) regression is insufficient on its own.

This paper is organized as follows: section 2 represents some related papers to show the previous work on this topic. Section 3 outlines the basic model specifications. In Section 4, the data is described and a statistical summary is provided. The empirical results and detailed interpretations of the results are presented in Section 5. Finally, section 6 draws some conclusions from the analysis.

2. Literature Review

The study of housing price is a large field on its own, and it is impossible to cover even a small fraction of the research conducted in that field. Thus, this review concentrates mainly on the studies that look at the relationship between the green space and housing prices.

Green spaces provide many environmental and social benefits, which are well documented in the literature (Robinette, 1972; Grey and Deneke, 1978; Laurie, 1979). However, most of the values attached to the green spaces are non-priced environmental benefits. These values include those derived from pleasant landscape, clean air, peace and quiet, and screening, as well as potential recreational activities in wooded green spaces. Other benefits include reduced wind velocity, balanced microclimate, shading, and erosion control. However due to these non-commodity and non-priced nature, and largely intangible benefits, their contribution is usually difficult to assess and quantify, among them, various approaches have been proposed and tested. There are two ways to measure these kinds of amenity values. One is to use a survey-based method, such as travel cost or contingent valuation. The hedonic pricing approach is the other option.

The hedonic method can be traced back to Court (1939) and received considerable application beginning in the 1960s. However, it was not until 1974 that a theoretical model that could serve as a basis for the empirical techniques was developed by Rosen. This model considers a class of differentiated products completely described by a vector of objectively measured characteristics. Observed product prices and the specific amounts of characteristics associated with each good define a set of implicit, or "hedonic", prices.

A hedonic model of price is one that decomposes the price of an item into separate components that determine the price, since every good provides a bundle of characteristics or attributes. This theory was well explained by Brown and Rosen (1982). Rosen's model has been proven to be extremely useful in many years and was cited by the majority of papers in the hedonic field. The most common application of this method is housing price. Hedonic pricing method (HPM) is based on the idea that properties are not homogenous and can differ in respect to a variety of characteristics. This method relies on the fact that house prices are affected by many factors: number of rooms, access to amenities, and so on. As Garrod (1999) represented, the most common application of the HPM is in relation to the public willingness to pay for housing. Each property is assumed to constitute a distinct combination of attributes, which determine the price or buyers' willingness to pay. The price of a housing unit is dependent upon the availability and level of a wide range of attributes, such as structural characteristics, neighborhood characteristics, and amenity characteristics. Among them, one important factor is environment, for example, view or access to a wooded park or watercourse (Palmquist, 1991). Theoretically, HPM can be used in calculating external benefits and costs of forests associated with housing, because the price of a house reflects people's willingness to pay in order to gain easier access to forests and to 'consume' their amenity values. In fact, HPM has been used for estimating the contribution of individual trees to property values (Darling, 1973; Morales, 1980; Morales et al., 1983). Anderson and Cordell (1985) found that tree cover increased property values by 3-5% in Athens, Georgia. A study based on the HPM was carried out in Apeldoorn, a medium-sized town in eastern Netherlands (Fennema et al., 1996). This study analyzed 106 house transactions built around a park; it demonstrated that location within 400 meter of the park attracted a premium of 60% over houses located outside this zone. This result was consistent with the expectation that green has a value-increasing effect on housing price.

There is a long history of using hedonic model to investigate the effects of amenities on sale prices of houses. The most common approach has been to include distance from property to the amenity as an explanatory variable in the model (Milon, Gressel, and Mulkey 1984; Nelson, Genereux, and Generoux 1992; Kiel 1995; Lansford and Jones 1995). In housing price research, a parcel's surroundings also have a major influence on housing value.

When asked, people always said that property values are determined by "location, location, location". A reasonable explanation for this is that spatial econometric techniques should be used in an analysis of housing price, therefore in research area, many hedonic price studies suggested that in a cross-sectional hedonic price analysis, the value of a property in one location may also be affected by the property value in other locations, such as in its neighboring area. Ignoring this spatial effect or spatial dependence may cause hedonic estimation result to be inconsistent or inefficient[1]. Spatial dependence among hedonic regression residuals was initially revealed by Paelinck and Klaassen (1979), who published a small volume entitled Spatial Econometrics, which arguably was the first paper in the field of spatial econometrics and its distinct methodology. Spatial analysis or spatial econometrics in hedonic analysis was introduced by Dubin (1988, 1992) and Can (1990); since then it started to be applied in many more recent studies. Those studies include: Geoghegan et al (1997) employed spatially-explicit indices in that paper, Bockstael and Bell (1997) used a simple spatial error model, He and Winder (1999) demonstrated bi-directional price causality between three adjacent housing markets in Virginia, indicating the existence of spatial effect in housing markets.

[1] See Anselin, 1988 for text-book treatment of spatial econometrics.

Also, there are a number of studies that provide evidence of the existence of spatial effect in hedonic analysis. For example, Legget and Bockstael (2000), Gawande and Jenkin-Smith (2001) estimated a housing price hedonic model using a simple spatial autoregressive model. Bowen, Thrall and Prestegaard (2001) examined housing prices in Cuyahoga County, Ohio. Kim, Phipps and Anselin (2003) measured the benefits of improving air quality on housing prices in Seoul, Korea. Bransington and Hite (2004) discussed the ways to model the influence of different types of omitted variables in the spatial model. There are still many other hedonic studies incorporate the spatial effects, such as Basu and Thibodeau (1998), Dubin, Pace, and Thibodeau (1999), Munneke and Slawson (1999), Gillen, Thibodeau, and Wachter (2001), and Irwin (2002).

The factors mentioned above like amenities surrounding the property and spatial effects are important and can be captured with Geographic Information System (GIS)[2] applications. Din, Hoesli and Bender (2001) argued that GIS have made possible the development of databases that can be used to better measure environmental characteristics. Their environmental parameters refer to the quality of the neighborhood and the quality of the location within a neighborhood. Another benefit of applying GIS in spatial analysis is demonstrated by Clapp and Thrall (1997), he argues that GIS is a powerful tool for supporting research because of its capability of storing and manipulating large data sets on spatial relationships. GIS can quickly assemble large amounts of spatial data, link spatial features to data, and visualize spatial analysis results[3].

In many instances, there may be multiple occurrences of amenities proximate to properties, and GIS can generate variables that distinguish between them. For example, in

[2] A geographic information system (GIS) is a computer system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data.

[3] ArcGIS10 includes a spatial statistics toolbox, with functionality for spatial autocorrelation analysis and spatial regression. Also, GeoDa is good software that can be used in spatial analysis, but it cannot deal with large dataset and the weight it generated cannot be inserted into other software.

examining the influence of wetland amenities on sale prices of residential properties in Portland, Oregon, Mahan, Polasky, and Adams (2000) consider distance to, as well as size and shape of, the nearest wetland area. Similarly, Powe et al. (1997) approximate forest amenities associated with a given property with an index variable that measures the ratio of acreage to squared distance from the home, summed over all woodland areas in the Southampton and New Forest areas of Great Britain. GIS data have also been used by Geoghegan, Waiger, and Bockstael (1997) to construct variables that reflect the extent, diversity and fragmentation of land uses in various buffer sizes around residential properties in the Patuxent Watershed in Maryland. In each of these studies, GIS data have enhanced the ability of the hedonic model to explain variation in sale prices by considering both proximity and extent of environmental attributes.

Until now, statements above show that housing price is affected by many factors at different perspectives, but there is still one issue we need to consider: housing characteristics may have a different effect on housing prices when we analyze it at different points of the distribution of house prices, which is referred to as quantile effects. Quantile regression is based on the minimization of weighted absolute deviations to estimate conditional quantile (percentile) functions as represented by Koenker and Bassett (1978) and Koenker and Hallock (2001). There is a large amount of literature using this model in many different topics: Eide and Showalter (1998), Knight, Bassett and Tam (2000) and Levin (2001) have addressed school quality issues. Poterba and Rueben (1995) and Mueller (2000) studied public-private wage differentials in the United States and Canada. Abadie, Angrist and Imbens (2001) considered estimation of endogenous treatment effects in program evaluation, and Koenker and Billias (2001) explored quantile regression models for unemployment duration data. A paper written by Viscusi and Hamilton (1999) considered public decision making regarding hazardous waste cleanup.

This model was also used in housing value research: Gyourko and Tracy (1999) adopted the quantile regression approach to investigate changes in housing affordability between 1974 and 1997 using the American Housing Survey data. Employing housing transaction data from Chicago in 1993 through 2005, McMillen and Thornes (2006) suggested that quantile regression has advantages over the conventional mean-based approaches to estimating a housing price index. McMillen and Coulson (2007) and McMillen (2008) identified significant variations in values of physical attributes across quantiles after they studied house price appreciation and constructed quantile house price indexes. Since normal quantile regression does not consider spatial autocorrelation that may be present in the data, spatial autocorrelation was incorporated into the quantile regression by adding a spatial lag variable, but adding a spatial lag into OLS regression will cause endogeneity problem (Anselin, 2001). When there are endogenous variables, the estimator of the parameter of interest is generally inconsistent. Researchers have treated some endogeneity problems in quantile regressions. Kemp (1999) and Sakata (2001) studied least absolute error difference (LAED) estimators for estimating a single equation from a simultaneous equation model. Abadie and Imbens (2002) design a quantile treatment effects estimator, which is the solution to a convex programming problem with first-step non-parametric estimation of a nuisance function. MaCurdy and Timmins (2000) propose an estimator for ARMA models adapted to the quantile regression framework. Among them, Kim and Muller (2000) first introduce the Two-Stage Quantile Regression (2SQR)[4]. In 2004, they published another paper in 2SQR with a detailed discussion of the two stages. After that, Zietz et al. (2008) utilize quantile regression, with and without accounting for spatial autocorrelation, to identify the coefficients of a large set of diverse variables across different quantiles.

[4] See Two-Stage Quantile Regression at http://www.nottinghampublications.com/economics/documents/discussion-papers/00-01.pdf

3. Data

The area covered by the data set must be sufficiently wide to ensure a representative spread of variation in the level of any external factors being investigated, and amenities in that area must be fully included in order to cover the location factor that affect the housing price. Therefore, the housing market in Delaware County, Ohio, is chosen as the case study. Delaware County has been a leader in developing a comprehensive land information system (DALIS/Delaware Appraisal Land Information System), and is a source of a variety of spatially explicit data with very detailed characteristics for individual houses. Also, the project provides 2010 Census Geography for Delaware County, which includes amenities and infrastructure in polygon shape files and associated tables in dbf format. At the start of this study, in an investigation of the effects of green space on housing price, variables relating to structural characteristics are designed, including the age of the house, whether it had gas, or heating and the number of bathrooms as has been done by most previous researchers (McLeod, 1984; Des Rosiers et al., 2002). This kind of information is necessary to explain those differences in price attributable to the structural characteristics, as opposed to those which are the result of amenities and socio-economic characteristics.

When making decisions, each house buyer takes the characteristics of neighboring residences into consideration. Thus, socio-economic variables that estimate the quality of neighbors were included in this study. The demographic data are currently available from American Community Survey (ACS), which is a household survey conducted by the U.S. Census Bureau that currently has an annual sample size of about 3.5 million addresses. Socio-economic characteristics were reflected primarily by data from ACS on vacancy ratio, percentage households with medium or high income, percentage of population with different race and percentage of population with different education level. Vacancy rate is included as an indicator to capture prevailing housing market conditions. Table1 lists and defines the explanatory variables that are included in the regression. Table 2 lists the summary statistics of the variables.

Table 1 goes about here

Table 2 goes about here

As we all know, houses in an attractive location attract a premium over houses in a neutral location. Green space, ponds and lakes, smooth traffic and convenience are aspects of an attractive location. Since these factors are valued differently by residents, they will affect house prices differently. Location variables included distance to amenities and disamenities, like distance to nearest medical center, post office, railroad, police office[5] and green space amenities (amenities contain green space) and so on. These distance variables are intended to capture the effect on housing prices of the proximity to various amenities. Negative distance effect was expected for the amenities since shorter distance means more convenience, and opposite effect for the disamenities because of the noise or inconvenience they brought. In fact, the distance to railroad variable is an imperfect measure of traffic effect, for example, Strand and Vagnes (2001) represented that environmental nuisance associated with living close to the railroad. In reality noise and vibrations also depend on topographical properties, e.g. on whether the train line is elevated above the house, on level with it or sunk below it; whether there are objects (such as trees and rocks) that shield the house from noise; and whether there are other houses in between the railroad line and one's own house, and whether the unit has extra protection against noise and vibrations (such as noise-reducing windows). But since this paper is focused on green space, traffic variable is not the primary objects of interest for this research and it is included only for

[5] All the distance variables are in miles.

excluding its effect on housing price. Figure 1-3 are the distribution of these places with the county.

Figure 1 goes about here

Figure 2 goes about here

To capture the effect of green space on the value of a parcel effect, many relevant variables are created, including closest distance to golf course, parks and woodland. The cross product of distance and area (Fanhua K., Haiwei Y., Nobukazu N., 2007) were estimated in the regression. Taken together, the influence of green space to house depends on both the distance and area; this interaction item better describes the characteristic of the surrounding green space environment, which should have an important effect on the value of a parcel. Besides, another variable equaling the cross-product of area of both census block and green space allows the marginal effect of the percentage of green space cover is also created to evaluate the green space effect on housing price. The distributions of woodland around parcels are shown in figure 3.

Figure 3 goes about here

4. Methods

This section first reviews the hedonic, spatial econometric and quantile models, and then introduces the methods to integrate them. Also, it includes specific discussion on my estimation method.

4.1. The Hedonic Housing Price Specification

In general, the hedonic equation for housing relates the sales price of a property to a set of characteristics that determine the property's value. Since this paper deals with owner-occupied housing, three groups of characteristics are included: (1) structural characteristics, (2) amenities characteristics, (3) socio-economic characteristics.

The general functional form of the hedonic price function is:

$$P = f\,(S, A, E) \qquad (2.1)$$

where: P = Log of housing price

S = Structural characteristics of the house

A = Amenities characteristics

E= Socio-economic characteristics

Expansion form is[6]:

$$\ln P_i = \alpha_0 + \alpha_l S_{il} + \alpha_m E_{im} + \alpha_k A_{ik} + \varepsilon_i \qquad (2.2)$$

where ε is assumed to be a normally distributed error term, with E (ε) = 0 and E (εε') = $\sigma^2 I$

P_i is the housing price in nature log form

S_{il} = Structural characteristic *l* of the house i

A_{ik} = Amenities characteristic k of house i

E_{im} = Socio-economic characteristic m of house i

The dependent variable is the natural log of the sale price. A log-linear form allows the marginal effect of each independent variable to vary with the level of the dependent variable, so the marginal effects of independent variables change as the house price changes. Because the predicted hedonic price is the result of the behavior of many different buyers and sellers, the marginal effect of independent variables are not constant for all houses regardless of differences in house price (Taylor, 2003). Therefore, the functional form of the HPM usually was not linear (Freeman, 1993). In addition, this specification was also used by Palmquist (1991) and others to model the determinants of house prices.

4.2. Hedonic Analysis with Spatial Lag and Spatial Error

[6] Detailed information on variables in these categories is in Table 1.

Whether or not any pair of houses is neighbors is based on whether or not they are located in neighboring area. Two areas are considered neighbors when they share common borders (contiguity) or when their distance to each other is below a certain level. In order to measure that in the spatial model, we need to use the spatial weight matrix (W). There are two basic types of spatial weights matrices. The first type is contiguity-based; the second type is distance-based. For both types of spatial weights matrices, we must specify two general parameters before their construction. The first is the spatial extent of the influence or the definition of the neighborhood. For a contiguity-based matrix, if two polygons are contiguous, they are considered neighbors. Two basic types of contiguity exist: rook contiguity (e.g., two polygon share a common border) and bishop contiguity (e.g., two polygons share a common vertex). Queen contiguity is a combination of these two. Specifically, a contiguity-based spatial weights matrix (W) is typically specified as

$$w_{ij} = \begin{cases} 1 & \text{if i and j are contiguous} \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

For a distance-based matrix, a critical value of distance must be specified within which two points are thought to be neighbors. The parameter is called the "power" of influence of two neighbors, which indicates that neighbors influence each other's housing price to different degrees, depending on the distance between them. For example, for houses at different locations, the prices/error terms associated with close neighbors are more highly correlated than those of distant neighbors. The relationship in the distance-based spatial weights matrix is typically represented as an inverse function of distance, within the assumed critical value.

$$w_{ij} = \begin{cases} 1/(d_{ij})^{\theta} & if\ i \neq j\ and\ d_{ij} \leq m \\ 0 & if\ i \neq j\ and\ d_{ij} > m\ or\ if\ i = j \end{cases} \tag{2.4}$$

The term d_{ij} is the distance between points i and j, and usually calculated according to their latitude and longitude (or X, Y coordinates). The parameter m is the extent of influence or critical distance value. The choice of its value is an empirical problem that depends on the scale of data and the extent of the perceived neighborhoods. Parameter θ measures the "power" of influence, whose value represents the distance decay effect within neighborhoods. As θ increases, the influence of nearby observations becomes greater than those further away. An alternative distance-based weights matrix uses linear decay. The weight corresponding to points i and j is assumed to be linearly inverse to the distance between them (d_{ij}) and equal zero at a specified distance.

The purpose of including a spatial weights matrix is to correct for potential problems due to spatial correlation and unobserved heterogeneity. Spatial autocorrelation is used to deal with the situation where the price of a house at one location is correlated with the price of neighboring houses. This dependence originates in part from the fact that each house shares with its neighbors influences from location factors that are nearly identical. In practice, parcel level distance variables, or spatial weights matrix approach, are usually used to incorporate spatial effects into hedonic regression models.

There are two kinds of weighting methods. The first one is spatial-lag model, which is weighting the sum of neighboring observations on the dependent variable (y), which is generally accomplished by creating a spatial lag term *Wy* weighted by neighbors' proximities to each observation. The spatial-lag model implicitly assumes that the spatially weighted average of housing prices in a neighborhood affects the price of each house (indirect effects) in addition to the standard explanatory variables of housing and neighborhood characteristics (direct effects). It assumes that the spatially weighted sum of neighborhood housing prices (the spatial lag) enters as an explanatory variable in the specification of housing price formation. The spatial lag model more or less resembled the autoregressive (AR) model in time-series econometrics. However, unlike the AR model, OLS estimation in the presence of spatial dependence will be inconsistent, because of the endogeneity problem. The spatial lag model will be estimated using instrumental variable (IV) estimation.

The second weighting method is the spatial error method, which is done by creating a proximity-weighted error term $W\varepsilon$, where ε is the weighted sum of neighboring errors. Compared to the spatial lag model, the spatial error model does not include indirect effects but is based on the assumption that there is one or more omitted variables in the hedonic price equation and that the omitted variables vary spatially. Due to this spatial pattern in the omitted variables, the error term of the hedonic price equation tends to be spatially autocorrelated.

4.3. Quantile Regression and Spatial Autocorrelation

While ordinary least-squares regression models the relationship between one or more covariates X and the conditional mean of a response variable Y given X = x, quantile regression can be employed to explain the determinants of the dependent variable at any point of the distribution of the dependent variable. It is preferred to the approach that quantile regression uses all data in each quantile, and then minimizes the absolute deviations (LAD model), we can see from this process that quantile regression uses the full sample and avoids the truncation problem that the alternative approach usually encounters. Other important advantages of quantile regression include its superior capability in handling heteroscedasticity, outliers, and unobserved heterogeneity.

In practice, Koenker and Hallock (2001) review this econometric method thoroughly. For hedonic price functions, quantile regression makes it possible to statistically examine the extent to which housing characteristics are valued differently across the distribution of housing prices. The detailed equations for quantile regression are as follows:

For a random variable *Y* with probability distribution function

$$F(y) = \text{Prob}\ (Y \leq y) \tag{2.5}$$

the τth quantile of Y is defined as the inverse function

$$Q\ (\tau) = \inf \{y : F\ (y) \geq \tau\} \tag{2.6}$$

where $0 < \tau < 1$. In particular, the median is $Q\ (1/2)$.

The mechanism to carry out quantile regression is similar to ordinary regression. The difference is, instead of searching for the argmin of sums of squared residuals, quantile regression looks for the argmin of weighted sums of absolute residuals.

As mentioned above, spatial autocorrelation is a special problem must be considered in the housing data. Therefore, in this paper, a spatial lag variable was incorporated into the quantile model. The presence of the spatial lag term in the right-hand side introduces endogeneity in the model, which will make biased and inconsistent estimators. There are two commonly used alternative estimation procedures: instrumental variable (IV) estimation and maximum likelihood estimation. The former is more robust than the latter in the sense that it does not require the error term to be normally distributed. In this study, the IV estimation, or more specifically the two stage least squares (2SLS) was used, this mixed method is called the two-stage quantile regression.

As the name implied, the Two-Stage Quantile Regression (2SQR) includes two steps, but I actually did three steps to finish this model. In the first step, I created spatial lag variable of housing price and all the independent variables. Then, I regressed the spatially lagged independent variables as well as the independent variables themselves against the spatial lag of housing price, and got the predicted value of the spatial lag of housing price. Finally, I ran the quantile regression of the housing price against all the characteristics and the predicted value I got from the last step. The reason why I used the predicted value instead of the true value is that it can eliminate correlation between the spatially lagged endogenous variable and the error term. When combined with hedonic spatial model, 2SQR is also called Hedonic Spatial Quantile Regression (HSQR)[7].

This paper also studies the Generalized Spatial Two-stage Least-Square (GS2SLS)[8] estimation of spatial autoregressive models with autoregressive disturbances where we allow for a spatial lag of the dependent variable. Results from ordinary least square (OLS) regression, quantile regression without spatial effect are also analyzed and compared.

5. Results and Discussion

The results of the variables estimated using OLS, QR, GS2SLS and HSQR are reported in Table 3 and Table 4. The first columns in Table 3 and Table 4 are OLS and GS2SLS separately, and the remaining columns in the two tables are QR and HSQR models; the numbers in the parentheses are the bootstrapped standard errors.

Table 3 goes about here

Table 4 goes about here

---

[7] Kostov, Phillip (2009) A *Spatial Quantile Regression Hedonic* Model of Agricultural Land Prices. *Spatial* Economic Analysis, 4 (1). p. 53.

[8] GS2SLS estimation can be used to incorporate a high degree of flexible spatial weight matrix, and get consistent results.

The OLS and GS2SLS estimates are presented in the first column of Tables 3 and 4[9]. They both are estimated on the entire data set. These regression analyses go beyond simple correlations and allow us to separate the various effects of green space, house quality and location, and socio-economic characteristics, yielding a better picture of the impact of green space on sales price. The statistically significant coefficients for both spatial lag and spatial error terms suggested that both a spatial lag and a spatial error analysis are appropriate ways of accounting for the spatial dependence.

The results from the OLS and GS2SLS estimates show that all of the structural variables are positive and statistically significant except for number of garages and number of family rooms, which means that number of garages and number of family rooms are not very important in determining housing price. Coefficient signs of the structural characteristics variables are as expected.

Variables that represented green space are distance and area interaction term of green space, percentage of forest area in each census block, and nearest distance to a forested amenities area. Besides, there are also some other good indicators for green space; for example, some research use golf courses to represent green space (Bolitzer and Netusil 2000; Lutzenhiser and Netusil 2001). In this paper, I simply use the distance to the nearest golf course since its data are in a point shape file. Both sign and statistical significance are the same in OLS and GS2SLS for all the green space variables. Most importantly, the nearest distance to golf course and parks show a fairly strong negative correlation with selling price, meaning that shorter distance to them around the house are associated with houses that sell for more money. This result suggests that green space amenities have a positive effect on housing price.

---

[9] Coefficients for X_Sphat are all scaled.

The expectations for the effect of percentage of the census block area occupied by green space, effect of distance to woodland and woodland interaction term on housing price are all quite ambiguous, with no a priori expectations for the sign of the coefficients. Although green spaces are the major means to carry out outdoor activities and provide enjoyable green views, and a greater green coverage of area around the house which could result in higher house price, a portion of the value was actually also reduced by location and traffic condition. This is because there was a statistical tendency for areas far from urban areas to have more trees and traffic conditions around forested amenities to be bad. The regression result shows although the coefficient estimate of distance to nearest woodland was negative, it is not statistically significant. This result suggests the negative effects eliminate the positive effects. The coefficient estimate for the green space percentage in the census block is positive in both OLS and GS2SLS but not significant, which indicates there is a probability that this effect is zero.

In addition, the coefficients of neighborhood variables from the census-block group, percentage of Black and Asian people, percentage of household own child, percentage of population not graduate from high school or only have high school diploma and percentage households with medium income are of the predicted sign with statistical significance.

The main difference between QR and HSQR models is they evaluate the characteristics' effect at the different housing price point before and after including the spatial effect. In order to better see the tendency of the coefficients change across quantiles, figures for each variable are created for both QR and HSQR models in Figure 4 and Figure 5. It is the graph version of the results in Table 3 and Table 4.

Figure 4 goes about here

Figure 5 goes about here

The results from both QR and HSQR show most of the structural variables are statistically significant in all quantiles except for number of family rooms and number of garages. Coefficients' signs of the structural variables are all the same across quantiles, but the significance level varies for some quantiles.

Coefficients' signs for the distance variables are as expected. The coefficients for the nearest distance to railroad variable are positive and statistically significant in almost all quantiles, and the magnitude gets larger with the increasing quantile level. This suggests house price increases with increasing distance from a railroad, and the increasing amount grows for more expensive houses. This may be explained by the fact that a railroad is likely to be associated with a noise disamenity or other inconvenience. The same situation also happened to the coefficients of nearest distance to post office variable, they are positive and statistically significant at all quantile level except for 0.1 quantile. The coefficients for the distances to fire district, medical center and police office are not statistically significant and the signs are different in quantiles, which indicates these amenities may not have an effect on housing price. Urban dummy is still a significant variable, the effect increases with the increase of housing price, but not significant at low level quantiles (not significant at 0.1 and 0.2 quantiles).

Besides, the coefficients of socio-economic variables from the census block level, such as percentage of population who travels more than 90 minutes to work, percentage of population only have high school diploma, percentage households own child and percentage of population worked outside place of residence are of the predicted sign with statistical significance in most quantiles.

Visual inspection of the coefficients for the green space variables suggests ambiguous results. Estimations in Table 3 and Table 4, as well as lines in Figure 4 and Figure 5 represent substantial variation of the coefficients' sign and significance level across quantiles, and the coefficients that are not significant overall (in OLS and GS2SLS), are significant at some quantiles. For example, the coefficients of nearest distance to woodland are significant for QR and HSQR at middle level quantiles10, the significance level is a little different between each quantile, but the signs are still negative. This implies that middle level properties with woodland nearby will sell for a higher price, but the woodland effect is not obvious for both low and high level houses. This reveals that both cheapest and luxury houses have no obvious relationship with woodland around, but it has a significant positive influence on the middle level housing price. The underlying economic reason for this result may be tied to the fact that area with woodland is always far away from working places and downtown areas. Wealthy people will buy homes with the features that they place the most value one, which could be a house in a highly urban setting. The households who purchase lowest-priced units would have a premium by enjoying a green environment, but the amount they pay for houses are quite low, they actually are "free-riding", and that's why coefficients are not significant at both highest and lowest quantiles.

The coefficients of the percentage of green space in each census block are not significant in general as mentioned above, and not significant in most quantiles except for 0.8 and 0.9 quantiles. The positive sign reported in Tables 3 and 4 in this quantile range suggests that luxury houses closer to larger green space sell for relatively more amount, and the statistical significance for high quantile coefficients also confirm this effect. Moreover, the increasing magnitudes coefficients reveal that there is a higher positive effect for green space in higher-priced homes. The positive, significant and large magnitude implies the strong preference to green space of richest people. As we all know an obvious interrelation, which is difficult to

[10] QR at 0.3-0.7 quantiles and HSQR at 0.3-0.6 quantiles

disentangle, occurs between social status and attractive location. People who can afford to do so have a tendency to choose attractive, green settings for their homes. As a consequence, certain towns or districts in attractive, green settings have become known as places for the rich. Buyers in these areas are willing to pay more premiums for the attractive environmental setting, such as green space. Consequently, houses in these areas are the highest priced.

The result shows an opposite effect for the nearest distance to green space amenities (nearest distance to golf course and park). For example, the distance to golf course has a high significance level in OLS and GS2SLS analysis (t=-3.77) as previously discussed. But when analyze by QR and HSQR, the significance level suddenly decreases, although the signs of regression coefficients are still negative, which also indicate a positive effect of green space amenities. The coefficients of this variable is statistically significant in 0.1, 0.2, 0.6 and 0.8 quantile and the magnitudes in QR indicate there is more than a 50 percent decrease from 0.1 quantile to 0.2 quantile and even less for 0.6 and 0.8 quantile. The coefficients for rest quantiles are not significant at all with low magnitudes, indicating the negative effects as discussed above almost catch up with the positive green space effects and these negative effects increases a lot compared to the positive ones with the increasing of house price. Taken together, golf course may only have strong effect on low-priced houses.

The nearest distance to parks has a positive effect on housing price in all quantiles, but the coefficients are not statistically significant in the high level percentiles[11]. This reveals that house prices in high quantiles have no obvious relationship with parks around, but it has a significant positive influence on the lowest and some middle level housing price. This may due to the fact that buyers of more expensive houses bear more taxes or fees of provision of park while all home owners nearby equally benefit from it. But usually, the people who buy expensive

[11] Coefficients are not significant after 0.5 quantile.

houses are rich people; they may not care about this amount, that's why coefficients are not significant at high quantiles.

6. Conclusions

Intuitively, we felt that houses in attractive locations will have an added value over similar, less favorably located houses. But, the definition of attractive location is quite complicated; it depends on many factors such as people's income and their preference. By excluding the effect of different characteristics and factors on housing price, this study finds that green space was not always a positive factor that can make a house sell in a higher price; location, traffic, tax and other factors may affect the housing price as green space accessories.

As the result shows that the overall impact of green space was ambiguous. Furthermore, separating houses in different levels makes the relationship more complex. People with different income level may have huge difference on valuation to green space, this partly reflect house price. For example, nearby woodland has increasing positive effect with increasing housing price for middle level houses. And house near forest amenities benefit from theses location factors on low or middle level houses. Green space coverage percentage in census block is not an important factor in determining housing price, except for luxury level houses.

Methodologically, the current study improves on past hedonic modeling efforts by directly incorporating spatial effects into the hedonic model. This model measures both the direct and induced effects of a change in a public good such as green space. It deals with neighborhood effects that cannot be captured by non-spatial techniques, also avoids the econometric problems of biased and inconsistent estimators when spatial dependence is present but ignored. By doing that, this paper demonstrates the importance of incorporating spatial effects in hedonic housing price models when assessing the effect of green space on housing prices.

Except for normal hedonic spatial regression, the hedonic spatial quantile regression, which is the integration of quantile regression and spatial econometric modeling, is also considered in this research. The results indicate this method is helpful, because the variables' effects and the estimated spatial dependence vary substantially across quantiles. Also, as shown in the result section, we can see more detailed relationships between the green space variables and housing price after incorporating quantile regression, this method helps us to see the relationship changes over different price ranges, which is sometimes different from the overall effect. For example, variables that are not statistically significant in the GS2SLS estimation such as the percentage of green space in each census block; are significant in some price ranges as suggested by the hedonic spatial quantile regression.

Table 2.1. Variable Descriptions and Expected Signs

Variable	Description	GIS	Expect Sign
Structural characteristics			
BASE_DUMMY	Availability of Basement (0 or 1)	NO	+
BDROOMS	Number of Bedrooms	NO	+
FAMROOMS	Number of Family-Rooms	NO	+
DINROOMS	Number of Dining-Rooms	NO	+
GARAGE_CAP	Number of Garages	NO	+
FULBATHS	Number of Full Bathrooms	NO	+
FIREPL_STA	Number of Fireplaces	NO	+
Amenities characteristics			
NEARDIST_Fire	Closest Distance to Fire Districts	YES	?
NEARDIST_Meidcal	Closest Distance to Medical Center	YES	-
NEARDIST_Police	Closest Distance to Police Office	YES	-
NEARDIST_Postoffice	Closest Distance to Post Office	YES	-
NEARDIST_Railroad	Closest Distance to Railroad	YES	+
NEARDIST_Hydrology	Closest Distance to Hydrology	YES	+
Urban	Dummy variable for whether house in urban area	YES	+
NEARDIST_Woodland	The nearest distance to woodland	YES	-
NEARDIST_Golf	Closest Distance to Golf Course	YES	-
NEARDIST_Parks	The nearest distance to park	YES	-
Percent_Greenspace	The percentage of green area in each census block	YES	+
Distance*area_Woodland	Closest distance to woodland*Area of the woodland	YES	?
Socio-economic characteristics			
VAC_RATIO	Vacancy ratio = Vacant Housing Unites/Total Housing Unites	YES	-
HH_SIZE	The number of people living in the Respondent's Household	YES	?
WHITE	Percentage of White Population in Census Block	YES	+
BLACK	Percentage of Black Population in Census Block	YES	-
ASIAN	Percentage of Asian Population in Census Block	YES	-
Greater_90_minutes_commute	Percentage of Population Who Travels More Than 90 Minutes to Work	YES	-
Pct_work_outresidence	Percentage of Population Worked Outside Place of Residence	YES	-

Pct_hh_withchild	Percentage of household own child	YES	-
Less_than_HS	Percentage of Population not graduate from high school	YES	-
HSGrad	Percentage of Population only have high school diploma	YES	+
Bachelor	Percent of population with Bachelor degree	YES	+
Graddeg	Percentage of Population with Graduate or Professional School Degree	YES	+
Medium_income	Percentage Households with Medium Income (%)US $ 50,000 - US $ 99,000)	YES	+
High_income	Percentage Households with High Income (%)US $ 100,000 and over	YES	+
Electric_heating	Percentage of House Use Electricity as Heating Fuel	YES	?
X_Sphat	Predicted Value of Spatial Lag of the Housing Price	NO	?
SALEPRICE	Price of the Single-Family House Sale	YES	

Table 2.2. Summary Statistics of the Variables

	Mean	SD	Min	Max
SALEPRICE (in $1000)	275	179	12	2700
BDROOMS	3.5826	0.6871	0.0000	8.0000
FAMROOMS	0.6320	0.5058	0.0000	3.0000
DINROOMS	0.5977	0.4923	0.0000	2.0000
FULBATHS	2.1718	0.7877	0.0000	7.0000
FIREPL_STA	0.8451	0.4766	0.0000	4.0000
BASE_DUMMY	0.9132	0.2816	0.0000	1.0000
GARAGE_CAP	1.3737	1.2228	0.0000	6.0000
NEARDIST_Fire	0.0795	0.0471	0.0059	0.3334
NEARDIST_Meidcal	0.1532	0.1172	0.0067	0.7379
NEARDIST_Police	0.1247	0.0785	0.0045	0.6362
NEARDIST_Postoffice	0.1177	0.0667	0.0025	0.4053
NEARDIST_Railroad	0.0011	0.0013	0.0001	0.0156
NEARDIST_Hydrology	0.0425	0.0272	0.0002	0.1210
Urban	0.3200	0.4666	0.0000	1.0000
NEARDIST_Golf	0.0782	0.0580	0.0039	0.5625
NEARDIST_Parks	0.0057	0.0253	0.0000	0.3489
Percent_Greenspace	0.0415	0.1071	0.0000	0.7827
Distance*area_Woodland	0.0013	0.0262	0.0000	0.9083
vac_ratio	4.8681	6.4983	0.0000	100.0000
white	0.7847	0.2999	0.0000	1.0000
black	0.0298	0.0509	0.0000	1.0000
asian	0.0451	0.0642	0.0000	0.7143
hh_size	2.1041	0.4987	1.0000	4.0000
Pct_hh_withchild	0.3692	0.0704	0.2930	0.5800
Less_than_HS	0.0481	0.0354	0.0020	0.1500
HSGrad	0.2035	0.0824	0.0670	0.4410
Bachelor	0.2810	0.0974	0.1070	0.4860
Graddeg	0.1403	0.0786	0.0070	0.3210
Greater_90_minutes_commute	0.0146	0.0108	0.0000	0.0448
Pct_work_outresidence	0.2508	0.2512	0.0000	0.7060
Medium_income	0.3182	0.0780	0.1864	0.5215
High_income	0.7176	0.2179	0.1169	1.3122
Electric_heating	0.1506	0.0789	0.0430	0.3098
N	2247			

Table 2.3. Estimates and Statistical Significance of the Parameters in OLS and QR

	OLS	QR 0.1	QR 0.2	QR 0.3	QR 0.4	QR 0.5	QR 0.6	QR 0.7	QR 0.8	QR 0.9
BDROOMS	0.0790***	0.0559*	0.0829***	0.0983***	0.1162***	0.0977***	0.0915***	0.0876***	0.0667***	0.0571**
	(5.83)	(2.01)	(4.54)	(5.02)	(8.55)	(7.25)	(4.30)	(6.32)	(6.15)	(3.19)
FAMROOMS	0.0253	0.0722	0.0424*	0.0167	0.0155	0.0144	0.0121	0.0071	0.0204	0.0119
	(1.51)	(1.53)	(2.06)	(1.03)	(1.02)	(0.92)	(0.95)	(0.48)	(1.24)	(0.66)
DINROOMS	0.1091***	0.1721***	0.0925***	0.0960***	0.0789***	0.0878***	0.0918***	0.0842***	0.0791***	0.0821***
	(6.02)	(3.50)	(4.06)	(4.36)	(4.68)	(4.08)	(4.98)	(8.98)	(4.87)	(3.33)
FULBATHS	0.2747***	0.2494***	0.2629***	0.2803***	0.2748***	0.2793***	0.2628***	0.2673***	0.2643***	0.2740***
	(22.26)	(11.28)	(14.14)	(13.01)	(18.77)	(19.40)	(21.72)	(14.68)	(22.16)	(14.51)
FIREPL_STA	0.1948***	0.2804***	0.1905***	0.1747***	0.1597***	0.1456***	0.1564***	0.1678***	0.1632***	0.1479***
	(10.59)	(6.32)	(7.54)	(8.29)	(7.14)	(7.22)	(8.65)	(7.07)	(6.56)	(7.73)
BASE_DUMMY	0.3134***	0.5409***	0.3916***	0.3400***	0.3079***	0.2813***	0.2272***	0.2359***	0.2370***	0.2415***
	(10.93)	(3.76)	(7.06)	(6.21)	(6.01)	(6.78)	(6.50)	(6.76)	(8.76)	(5.92)
GARAGE_CAP	-0.0031	-0.0007	-0.0051	-0.0039	-0.0065	-0.0066	-0.0041	-0.0054	-0.0132	-0.0087
	(-0.45)	(-0.04)	(-0.58)	(-0.44)	(-0.67)	(-0.80)	(-0.71)	(-0.67)	(-1.73)	(-1.26)
NEARDIST_Fire	0.1852	0.6243	0.5680	0.1764	-0.0500	-0.1543	-0.1662	-0.2511	-0.2045	0.1964
	(0.88)	(1.35)	(1.27)	(0.58)	(-0.19)	(-0.67)	(-0.65)	(-1.05)	(-0.69)	(0.78)
NEARDIST_Meidcal	-0.0451	0.1950	0.1239	0.0313	0.1339	-0.0596	-0.0855	-0.0461	-0.0292	-0.2274
	(-0.38)	(0.67)	(0.52)	(0.21)	(0.90)	(-0.45)	(-0.65)	(-0.39)	(-0.24)	(-1.24)
NEARDIST_Police	-0.1724	0.0188	-0.4632*	-0.3275*	-0.1711	-0.1701	-0.2264	-0.2112	-0.0793	-0.4389*
	(-1.22)	(0.05)	(-2.15)	(-1.98)	(-1.48)	(-0.95)	(-1.42)	(-1.29)	(-0.46)	(-2.29)
NEARDIST_Postoffice	0.7625***	0.3762	0.6168*	0.8778***	0.9484***	0.9950***	0.9175***	0.8668***	0.8546***	1.0232***
	(4.56)	(1.46)	(2.25)	(4.35)	(5.86)	(5.69)	(5.47)	(6.55)	(4.84)	(4.56)

NEARDIST_Railroad	50.1783***	17.6833	46.4562***	34.8739**	35.4000***	35.6451***	38.4710***	59.8700***	62.4426***	67.0905**
	(7.33)	(0.84)	(5.80)	(3.17)	(4.02)	(3.92)	(5.54)	(3.42)	(10.40)	(3.12)
NEARDIST_Hydrology	0.2237	1.2525	1.1174*	1.3501***	0.5759	0.1521	-0.2485	-0.5509*	-0.8549*	-1.1071**
	(0.68)	(1.78)	(2.07)	(3.70)	(1.60)	(0.40)	(-0.71)	(-2.20)	(-2.17)	(-2.81)
Urban	0.1022***	0.1075	0.0468	0.0565*	0.0849***	0.0904**	0.1013***	0.0980***	0.1148**	0.1260**
	(3.66)	(1.95)	(1.55)	(2.35)	(3.36)	(3.07)	(5.92)	(3.56)	(3.16)	(3.14)
NEARDIST_Woodland	-1.8645	2.5796	-1.3576	-5.1785*	-5.3558*	-6.4169**	-5.9295*	-5.7989*	-1.7171	-1.5386
	(-0.64)	(0.64)	(-0.32)	(-2.33)	(-2.11)	(-2.75)	(-2.32)	(-1.98)	(-0.52)	(-0.46)
NEARDIST_Golf	-0.6523***	-1.3619**	-0.6807*	-0.2428	-0.4174	-0.3932	-0.4725**	-0.4224	-0.4959*	-0.3811
	(-3.74)	(-2.65)	(-2.43)	(-1.01)	(-1.68)	(-1.72)	(-2.69)	(-1.65)	(-2.48)	(-1.81)
NEARDIST_Parks	-1.2065**	-2.0393*	-1.5373	-1.5071	-1.6233**	-1.1820**	-1.0557	-0.1554	-0.6155	-1.1661
	(-2.94)	(-2.04)	(-1.77)	(-1.66)	(-3.01)	(-2.93)	(-1.54)	(-0.29)	(-0.94)	(-1.45)
Distance*area_Woodland	0.3870	1.0884	0.8480	0.6249	0.5205	0.4426*	0.3574	0.2654	0.1827	0.0860
	(1.37)	(1.73)	(1.00)	(0.56)	(1.13)	(2.04)	(0.23)	(0.70)	(1.20)	(0.32)
Percent_Greenspace	0.0629	0.0258	-0.0341	-0.0198	-0.0463	-0.0218	0.1247	0.1596	0.2534**	0.1823**
(in Census Block level)	(0.88)	(0.16)	(-0.30)	(-0.25)	(-0.49)	(-0.20)	(1.16)	(1.77)	(2.78)	(2.70)
vac_ratio	0.0009	-0.0010	-0.0019	-0.0016	-0.0005	0.0005	0.0024	0.0041**	0.0048***	0.0082***
	(0.72)	(-0.39)	(-1.13)	(-1.14)	(-0.30)	(0.31)	(1.66)	(2.67)	(3.93)	(4.04)
hh_size	0.0434	-0.0073	-0.0084	0.0092	0.0340	0.0432	0.0596***	0.0523**	0.0705*	0.1008***
	(1.74)	(-0.14)	(-0.27)	(0.28)	(1.09)	(1.28)	(4.09)	(2.84)	(2.22)	(3.92)
white	-0.0530	0.0786	0.0631	-0.0029	-0.0330	-0.0489	-0.0991**	-0.0884*	-0.0823	-0.1514**
	(-1.63)	(1.04)	(1.28)	(-0.05)	(-1.00)	(-1.10)	(-2.73)	(-2.49)	(-1.49)	(-2.64)
black	-1.0313***	-1.0557*	-0.7232*	-0.5610	-0.4790	-0.3651	-0.3702	-0.3663	-0.3880	-0.5501*
	(-6.68)	(-2.49)	(-2.37)	(-1.40)	(-1.53)	(-1.16)	(-1.56)	(-1.30)	(-1.50)	(-2.27)
asian	0.2790*	0.4458	0.4732***	0.2811**	0.1148	0.0585	0.0695	0.0927	0.0366	0.1037
	(2.02)	(1.33)	(3.38)	(2.92)	(1.13)	(0.54)	(0.63)	(0.81)	(0.46)	(0.51)
Pct_hh_withchild	0.8925***	0.7445	0.8011*	0.5104*	0.7968***	0.7634**	0.8230***	0.8578***	1.1520***	1.1755***

	(4.17)	(1.70)	(2.37)	(2.05)	(3.43)	(2.68)	(5.59)	(3.47)	(3.79)	(3.97)
Less_than_HS	-1.2948**	-2.0074*	-2.2278***	-1.7364***	-0.9642	-1.0522***	-0.7370	-0.7139	-0.2718	0.1053
	(-3.09)	(-2.04)	(-3.63)	(-3.99)	(-1.75)	(-3.65)	(-1.25)	(-1.86)	(-0.47)	(0.23)
HSGrad	-0.6699***	-0.3199	-0.6195*	-0.4701*	-0.6667***	-0.7575***	-0.9499***	-1.0310***	-1.1986***	-1.2074***
	(-3.62)	(-0.85)	(-2.04)	(-2.07)	(-3.44)	(-4.26)	(-5.59)	(-4.45)	(-4.92)	(-4.33)
Bachelor	-0.1143	0.0922	0.0008	-0.0023	0.0032	-0.0606	-0.0379	-0.1868	-0.2118	-0.1543
	(-1.08)	(0.38)	(0.00)	(-0.02)	(0.03)	(-0.48)	(-0.36)	(-1.41)	(-1.65)	(-0.77)
Graddeg	0.1634	0.0835	-0.1138	0.1417	0.1572	0.1944	0.2279	0.3115*	0.3432*	0.2351
	(1.10)	(0.26)	(-0.63)	(1.19)	(0.80)	(1.39)	(1.53)	(2.35)	(2.41)	(1.51)
Greater_90_minutes_commute	2.8353**	1.6327	4.6344***	2.3790*	2.5117*	1.9932	1.8459	3.1933***	3.6438*	2.6161
	(2.58)	(0.76)	(3.80)	(1.98)	(2.14)	(1.72)	(1.73)	(4.13)	(2.25)	(1.55)
Pct_work_outresidence	-0.1444**	-0.2167	-0.3427***	-0.1931***	-0.1400**	-0.1519**	-0.0872	-0.0521	-0.0173	0.0029
	(-2.78)	(-1.57)	(-4.20)	(-3.56)	(-2.76)	(-3.13)	(-1.81)	(-0.85)	(-0.29)	(0.03)
Medium_income	-0.2294	-0.0764	-0.1001	-0.3229*	-0.3812*	-0.4240	-0.4707**	-0.4314*	-0.6371***	-0.7653***
	(-1.29)	(-0.14)	(-0.42)	(-2.09)	(-2.17)	(-1.88)	(-2.91)	(-2.29)	(-3.70)	(-3.90)
High_income	-0.0099	-0.0621	-0.0237	-0.0590	-0.0106	-0.0486	-0.0572	-0.0073	0.0311	-0.0164
	(-0.10)	(-0.22)	(-0.20)	(-0.46)	(-0.08)	(-0.37)	(-0.67)	(-0.07)	(0.20)	(-0.10)
Electric_heating	-0.6299**	-0.9056	-0.8616***	-0.7088	-0.5235	-0.4210*	-0.3247	-0.4534	-0.3166	-0.1516
	(-2.80)	(-1.70)	(-3.59)	(-1.84)	(-1.74)	(-2.30)	(-1.37)	(-1.49)	(-1.05)	(-0.47)
_cons	10.9113***	10.3244***	10.7022***	10.8688***	10.7906***	11.0631***	11.2488***	11.2673***	11.2850***	11.4511***
	(65.47)	(21.98)	(45.53)	(67.95)	(60.04)	(75.67)	(75.65)	(87.87)	(41.14)	(46.79)
N	2247	2247	2247	2247	2247	2247	2247	2247	2247	2247
pseudo R^2										

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.4. Estimates and Statistical Significance of the Parameters in GS2SLS and 2SQR

	GS2SLS	HSQR 0.1	HSQR 0.2	HSQR 0.3	HSQR 0.4	HSQR 0.5	HSQR 0.6	HSQR 0.7	HSQR 0.8	HSQR 0.9
BDROOMS	0.0749***	0.0585*	0.0827***	0.0978***	0.1170***	0.1021***	0.0926***	0.0789***	0.0650***	0.0589***
	(5.64)	(2.07)	(4.03)	(4.93)	(8.40)	(7.05)	(4.09)	(5.37)	(6.79)	(3.48)
FAMROOMS	0.0251	0.0710	0.0425	0.0165	0.0098	0.0142	0.0091	0.0060	0.0196	0.0119
	(1.53)	(1.62)	(1.86)	(0.99)	(0.64)	(0.93)	(0.70)	(0.41)	(1.23)	(0.63)
DINROOMS	0.1105***	0.1724***	0.0926***	0.0950***	0.0865***	0.0890***	0.0947***	0.0841***	0.0780***	0.0889***
	(6.21)	(3.56)	(4.37)	(4.35)	(5.31)	(4.26)	(5.24)	(9.70)	(4.05)	(3.56)
FULBATHS	0.2684***	0.2482***	0.2630***	0.2819***	0.2729***	0.2708***	0.2583***	0.2659***	0.2665***	0.2723***
	(22.00)	(11.85)	(12.70)	(13.12)	(18.34)	(17.25)	(21.04)	(13.70)	(22.79)	(14.89)
FIREPL_STA	0.1898***	0.2794***	0.1908***	0.1749***	0.1638***	0.1466***	0.1556***	0.1641***	0.1685***	0.1493***
	(10.56)	(5.94)	(7.43)	(8.62)	(6.89)	(7.46)	(8.26)	(6.31)	(6.59)	(8.73)
BASE_DUMMY	0.3106***	0.5433***	0.3905***	0.3385***	0.3039***	0.2762***	0.2257***	0.2388***	0.2382***	0.2393***
	(10.97)	(3.89)	(7.06)	(6.12)	(6.19)	(6.93)	(6.66)	(6.57)	(9.43)	(5.76)
GARAGE_CAP	0.0024	-0.0005	-0.0052	-0.0036	-0.0048	-0.0057	-0.0043	-0.0070	-0.0131	-0.0117
	(0.35)	(-0.03)	(-0.57)	(-0.42)	(-0.52)	(-0.73)	(-0.68)	(-0.91)	(-1.77)	(-1.55)
NEARDIST_Fire	0.2154	0.6259	0.5642	0.1946	-0.0923	-0.1937	-0.1266	-0.2471	-0.2132	0.2359
	(1.02)	(1.27)	(1.14)	(0.62)	(-0.35)	(-0.84)	(-0.54)	(-1.14)	(-0.72)	(0.86)
NEARDIST_Meidcal	-0.0939	0.1986	0.1288	0.0297	0.1256	-0.0493	-0.0784	-0.0539	-0.0246	-0.1913
	(-0.78)	(0.66)	(0.48)	(0.19)	(0.86)	(-0.37)	(-0.58)	(-0.44)	(-0.21)	(-1.06)
NEARDIST_Police	-0.1782	-0.0020	-0.4608*	-0.3232	-0.1022	-0.1570	-0.2217	-0.1825	-0.0641	-0.4163*
	(-1.25)	(-0.01)	(-2.15)	(-1.87)	(-0.87)	(-0.85)	(-1.37)	(-1.19)	(-0.39)	(-2.19)
NEARDIST_Postoffice	0.7878***	0.3762	0.6162*	0.8653***	0.9418***	1.0325***	0.8880***	0.9022***	0.8206***	0.9969***
	(4.64)	(1.33)	(2.14)	(4.13)	(5.82)	(6.01)	(5.75)	(6.42)	(4.12)	(4.23)
NEARDIST_Railroad	50.8067***	18.7770	46.5163***	34.9238**	34.0090***	34.0427***	38.5458***	61.8946***	62.6829***	69.2344***

	(7.48)	(0.93)	(5.57)	(3.17)	(3.77)	(3.89)	(5.11)	(3.58)	(10.18)	(3.60)
NEARDIST_Hydrology	0.1733	1.2393	1.1080*	1.3626***	0.6180	0.1498	-0.1957	-0.4987	-0.7979*	-1.1444**
	(0.52)	(1.63)	(2.00)	(3.40)	(1.84)	(0.41)	(-0.56)	(-1.87)	(-2.06)	(-3.05)
Urban	0.1090***	0.1073	0.0475	0.0547*	0.0882***	0.0891**	0.1055***	0.1084***	0.1200**	0.1132**
	(3.85)	(1.73)	(1.53)	(2.26)	(3.56)	(3.17)	(6.38)	(3.43)	(3.25)	(2.70)
NEARDIST_Woodland	-2.6077	2.6580	-1.3461	-5.2306*	-5.5557*	-5.4156*	-6.4673**	-5.4712	-2.3958	-0.5649
	(-0.89)	(0.62)	(-0.33)	(-2.31)	(-2.17)	(-2.30)	(-2.79)	(-1.80)	(-0.78)	(-0.18)
NEARDIST_Golf	-0.6650***	-1.3821**	-0.6855*	-0.2423	-0.3598	-0.3486	-0.4583**	-0.3788	-0.5450**	-0.3888
	(-3.77)	(-2.63)	(-2.36)	(-1.00)	(-1.51)	(-1.53)	(-2.62)	(-1.44)	(-2.85)	(-1.91)
NEARDIST_Parks	-1.2091**	-2.0177*	-1.5284	-1.4938	-1.4963**	-1.2367**	-1.0337	-0.2426	-0.5711	-1.2321
	(-3.00)	(-1.96)	(-1.67)	(-1.64)	(-2.63)	(-3.27)	(-1.61)	(-0.44)	(-0.90)	(-1.54)
Distance*area_Woodland	0.4069	1.0934	0.8508	0.6292	0.3876	0.3833	0.3489	0.2545	0.1626	0.0637
	(1.47)	(1.63)	(0.98)	(0.56)	(0.52)	(1.73)	(0.23)	(0.64)	(1.13)	(0.24)
Percent_Greenspace	0.0527	0.0197	-0.0351	-0.0224	-0.0402	-0.0121	0.0730	0.1552	0.2242*	0.2041**
(in Census Block level)	(0.73)	(0.13)	(-0.32)	(-0.28)	(-0.43)	(-0.12)	(0.67)	(1.73)	(2.44)	(3.09)
vac_ratio	0.0006	-0.0010	-0.0018	-0.0016	-0.0011	0.0008	0.0025	0.0042**	0.0048***	0.0082***
	(0.50)	(-0.34)	(-1.11)	(-1.13)	(-0.61)	(0.56)	(1.60)	(2.61)	(3.59)	(3.86)
hh_size	0.0456	-0.0027	-0.0083	0.0094	0.0300	0.0491	0.0556**	0.0511**	0.0681*	0.1051***
	(1.82)	(-0.05)	(-0.26)	(0.26)	(1.04)	(1.41)	(2.79)	(2.71)	(2.16)	(4.09)
white	-0.0483	0.0730	0.0644	0.0018	-0.0232	-0.0603	-0.0903**	-0.0824*	-0.0797	-0.1582**
	(-1.49)	(1.01)	(1.22)	(0.03)	(-0.68)	(-1.37)	(-2.66)	(-2.09)	(-1.42)	(-2.67)
black	-1.0040***	-1.0454*	-0.7187*	-0.5526	-0.3927	-0.4099	-0.3580	-0.3884	-0.4499	-0.6006**
	(-6.11)	(-2.37)	(-2.16)	(-1.34)	(-1.28)	(-1.28)	(-1.49)	(-1.52)	(-1.69)	(-2.69)
asian	0.3008*	0.4434	0.4739***	0.2929**	0.1381	0.0740	0.0803	0.1087	0.0008	0.1000
	(2.13)	(1.32)	(3.34)	(2.95)	(1.35)	(0.64)	(0.72)	(1.01)	(0.01)	(0.48)
Pct_hh_withchild	0.9183***	0.7697	0.8001*	0.5211*	0.7274**	0.7588**	0.7872***	0.8393**	1.0980***	1.2546***
	(4.26)	(1.60)	(2.29)	(2.00)	(3.28)	(2.79)	(5.48)	(3.25)	(3.59)	(4.62)

Less_than_HS	-1.2650**	-1.9586*	-2.2226***	-1.7198***	-1.0086	-0.9892**	-0.7604	-0.7917*	-0.2351	0.1565
	(-3.01)	(-2.19)	(-3.49)	(-3.95)	(-1.92)	(-3.28)	(-1.30)	(-2.06)	(-0.43)	(0.32)
HSGrad	-0.7089***	-0.3405	-0.6192*	-0.4662	-0.6106**	-0.7013***	-0.9344***	-1.0286***	-1.1535***	-1.1926***
	(-3.81)	(-0.91)	(-2.15)	(-1.94)	(-3.22)	(-3.47)	(-5.50)	(-4.78)	(-4.57)	(-4.06)
Bachelor	-0.1321	0.0849	0.0019	-0.0035	0.0055	-0.0384	-0.0170	-0.1656	-0.1998	-0.1729
	(-1.23)	(0.35)	(0.01)	(-0.03)	(0.05)	(-0.33)	(-0.16)	(-1.27)	(-1.64)	(-0.90)
Graddeg	0.1856	0.1183	-0.1140	0.1515	0.2077	0.1945	0.2217	0.3139*	0.3573*	0.2337
	(1.23)	(0.37)	(-0.63)	(1.36)	(1.06)	(1.24)	(1.54)	(2.36)	(2.28)	(1.52)
Greater_90_minutes_commute	2.9253**	1.6212	4.6166***	2.3348	2.1861	1.7236	1.7014	3.1709***	3.3187*	2.7611
	(2.63)	(0.74)	(4.11)	(1.87)	(1.84)	(1.47)	(1.57)	(3.95)	(2.12)	(1.61)
Pct_work_outresidence	-0.1408**	-0.2191	-0.3412***	-0.1990***	-0.1333**	-0.1411**	-0.0796	-0.0372	-0.0122	-0.0175
	(-2.68)	(-1.66)	(-4.42)	(-3.97)	(-2.62)	(-2.80)	(-1.73)	(-0.55)	(-0.21)	(-0.19)
Medium_income	-0.2175	-0.0391	-0.0991	-0.3261*	-0.3845*	-0.4947*	-0.4815**	-0.4277*	-0.6211***	-0.8082***
	(-1.22)	(-0.07)	(-0.41)	(-2.01)	(-2.28)	(-2.10)	(-2.92)	(-2.24)	(-3.46)	(-4.38)
High_income	-0.0120	-0.0494	-0.0206	-0.0582	-0.0307	-0.0773	-0.0574	0.0077	0.0390	0.0192
	(-0.12)	(-0.18)	(-0.17)	(-0.44)	(-0.24)	(-0.59)	(-0.67)	(0.08)	(0.27)	(0.11)
Electric_heating	-0.6035**	-0.9190	-0.8568***	-0.6987	-0.5505	-0.4446*	-0.3100	-0.4064	-0.3179	-0.1365
	(-2.66)	(-1.73)	(-3.81)	(-1.77)	(-1.83)	(-2.41)	(-1.38)	(-1.40)	(-1.15)	(-0.41)
x_sphat		-0.0357	0.0319	-0.0498	0.0190	0.0246	0.0209	0.0223	0.0267	0.0109
		(-0.08)	(0.01)	(-0.27)	(1.21)	(1.15)	(1.03)	(1.25)	(1.74)	(0.71)
_cons	10.9186***	10.2876***	10.6983***	10.8571***	10.8104***	11.0823***	11.2607***	11.2671***	11.2879***	11.3898***
	(65.56)	(21.88)	(47.94)	(60.95)	(67.50)	(74.67)	(74.56)	(91.21)	(44.32)	(46.44)
lambda										
_cons	0.0140*									
	(2.54)									
rho										
_cons	0.6993***									
	(5.45)									
N	2247	2247	2247	2247	2247	2247	2247	2247	2247	2247

pseudo R^2	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
In the parentheses under HSQR estimates are bootstrapping standard errors with setting seed 1001 in Stata
Lambda is the coefficient of the spatial lag term and rho is for error term in GS2SLS

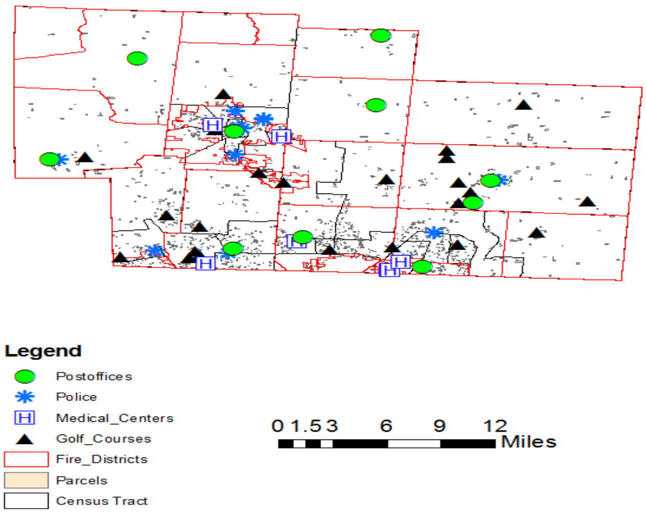


Figure 2.1. Places of Interest and Parcels

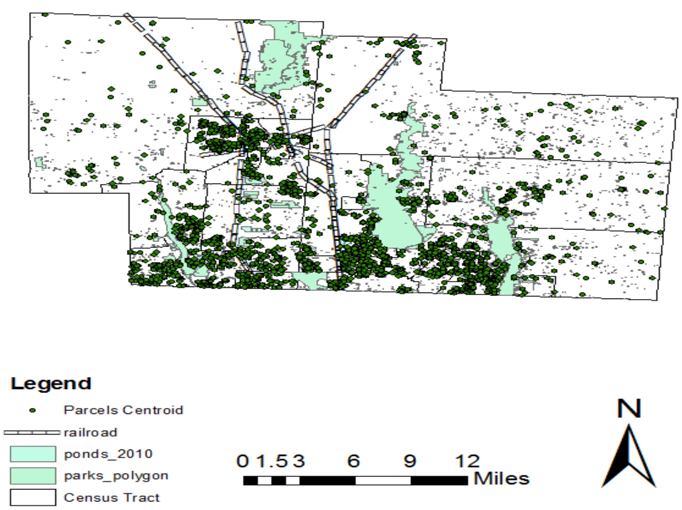


Figure 2.2. Railroad, Ponds and Parks

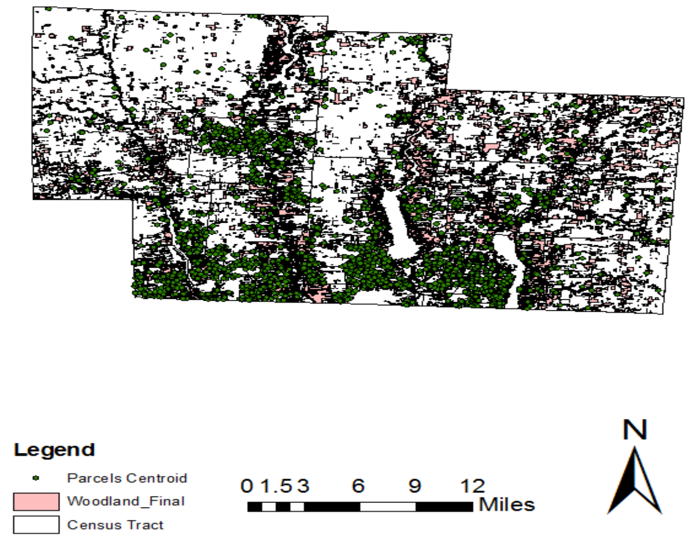


Figure 2.3. Woodland and Parcels

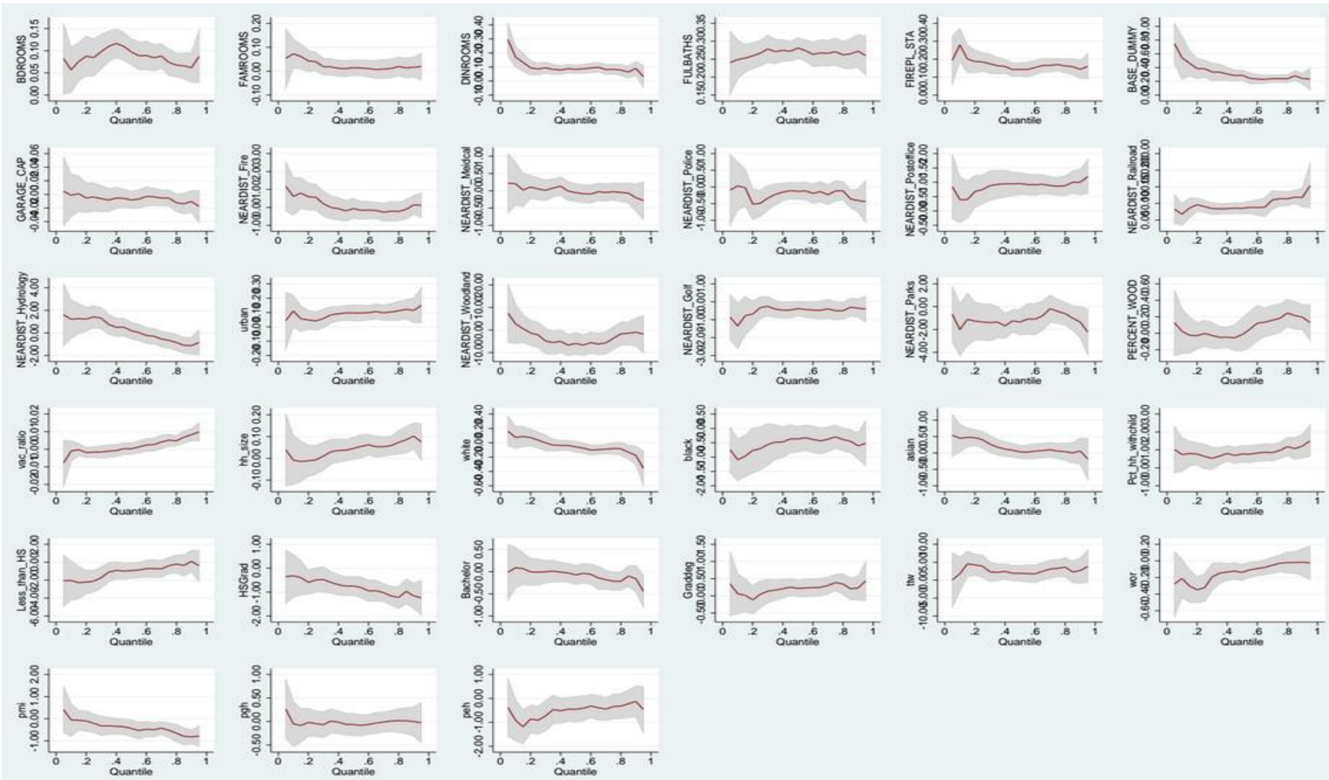


Figure 2.4. Plot of Coefficients in Different Quantile in QR

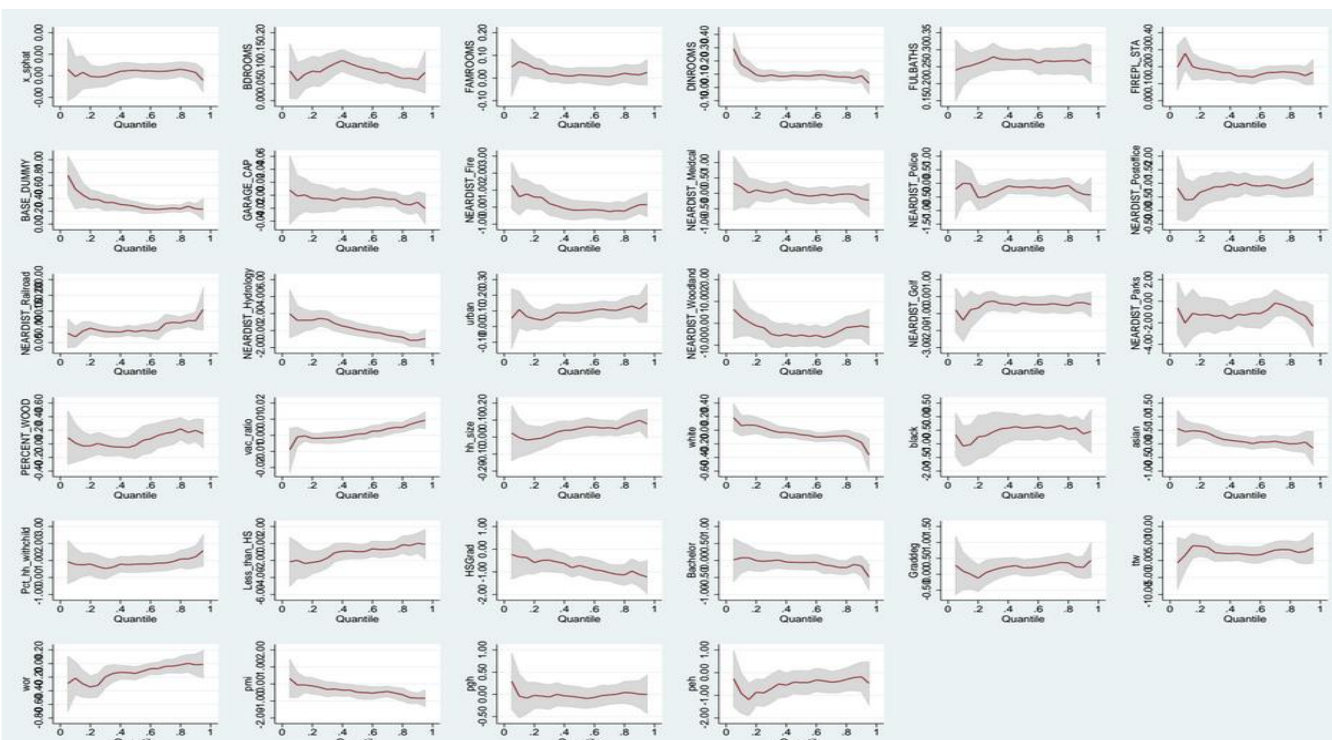


Figure 2.5. Plot of Coefficients in Different Quantile in HSQR

CHAPTER 3

The Impact of a New School District on Property Values

1. Introduction

The connection of land values and value of "place" has a very long history (Hurd, 1903; Alonso, 1964). Location of land largely determines property value. Attributes that impact land value and demand offer consumers a better quality of life related to the quality of the environment, proximity to work, accessibility, quality of education and so forth. This paper is mainly interested in the value placed on each of these commodities separately.

Data on housing costs has historically been used to support the cost of local amenities and determine consumer's willingness to pay for improved services (Sheppard, 1999). The hedonic price theory of Rosen (1974) is used as a model to estimate price differentials of property. According to this theory, goods are valued based on commodity characteristics. Therefore, using this price theory, the price of a single family house is determined by the observable and unobservable characteristics of the house, or product, but also on the demand and local amenities desired by participants in the market. This approach is often used when analyzing consumer preferences in the housing market and determining the value of local amenities like school quality, safe neighborhoods, and environmental factors.

In recent years, analyzing spatial data using geographic information system (GIS) applications have made the hedonic spatial[12] method more popular. Spatial analysis using GIS

[12] Housing price observed at one location depend on housing prices observed at neighboring locations

software can be used to examine the impact of amenities on a house and develop neighborhood characteristics based on distance or convenience. Therefore, as an analytical tool GIS provide more housing price determination variables and improves the hedonic method (Can, 1992) perspective. This method of analyzing geographic data is useful to explain the effect of location and additional neighborhood factors of a study area (Tiefelsdorf and Boots, 1997).

Regression discontinuity (RD) design is used when selecting observations using a program based on a designated threshold or cut-point, only observations just above or just below the cut-point are selected (Jacob, Zhu, Somers and Bloom, 2012). Some recent empirical applications try to solve the selection bias problem by introducing random variations, for example: Lemieux and Milligan (2004), Chen and Van der Klaauw (2004), Martorell (2004), Matsudaira (2004).

This unique design was first introduced by Thistlewaite and Campbell (1960) in an experiment evaluating scholarship programs by identifying causal relationships and excluding arbitrary factors. The advantage of this method is it provides an interpretation of estimates for random threshold; it can also be used even when there is heterogeneity in the treatment effects (Hahn, Todd and Van der Klaauw, 2001).

Some studies in economics, starting in the late 1990's, applied and extended RD methods, including, Chay, McEwan, and Urquiola (2005), McEwan and Shapiro (2007), and Card, Mas and Rothstein (2006). The RD approach has been used to evaluate the impact of unionization (DiNardo and Lee, 2004), anti-discrimination laws (Hahn, Todd, and van der Klaauw, 1999), social assistance programs (Lemieux and Milligan, 2004), limits on unemployment insurance (Black, Galdo, and Smith, 2007), and the effect of financial aid on college enrollment (van der Klaauw, 2002). This approach has been used in primary and secondary education to estimate the impact of class size reduction (Angrist and Lavy, 1999), remedial education (Jacob and Lefgren, 2006), delayed entry to kindergarten (McEwan and Shapiro, 2008), and the impact of the Reading First program on instructional practice and student achievement (Gamse, Bloom, Kemple, and Jacob, 2008)13. Furthermore, Greenstone and Gallagher (2008) uses the RD design to assign Superfund sites based on the Hazardous Ranking Score at the inception of the program.

RD design can also use geography boundary or area as discontinuities, such as school district. Geographic RD design was used to estimate the effect of school quality on house prices (Black 1999). According to Black (1999), the quality of local public school is a key determinant of housing price. His work used the RD approach to compare houses within a very close proximity to each other, but on the opposite sides of school district boundaries in the Boston metropolitan area. Changes in school quality occurred somewhat within district boundaries; however, changes in house characteristics were not noticeable in neighborhoods. This proves his assumption that difference in house price can be attributed to differences in school quality because houses on the opposite sides of the boundaries are actually in the same neighborhood. In this model, the RD method helps distinguish the role of school quality on housing prices separate from neighborhood attributes. To further verify, Black examines the characteristics of houses on opposite sides of attendance district boundaries and finds them to be similar.

Breaking up of school districts can affect the housing price because it provides more choices for households. In particular, the smaller districts may be differentiated horizontally, with schools emphasizing different types of programs and specific classes in math, reading, extra-curricular courses, and so forth (Banzhaf and Bhalla, 2012). Smaller districts can reflect differences and local tastes easier and parental input creates a greater impact. Additionally, new

[13] From Jacob, R. T., Zhu, P., Somers, M. A., and Bloom, H. S. (2012).*A Practical Guide to Regression Discontinuity*. MDRC.

school districts create competition between schools for students which is an incentive to improve performance (Borland and Howsen 1992, Hoxby 2000, Hanushek and Rivkin 2003). In contrast, Bogart and Cromwell (2000) did research on a school district reassignment from a reduction of schools in the Cleveland area. They found a significant housing price decrease in Shaker Heights, where the number of elementary schools reduced from nine to six.

Organization of this study: (i) Section 2 – detail of area of study and data resources; (ii) Section 3 – overview of theory and relevant literature, introduction of empirical model in detail and application to this case; (iii) Section 4 – estimation and preliminary results; and (iv) Section 5 – conclusion.

2. Data Description and Event Background

Lee County, North Carolina had only one high school opened under the name of Sanford Central High School in 1951. As the population of Lee County grow, this high school become one of the largest schools in NC. The county officials tried to renovate the older buildings in this high school because it leads to a high cost of maintenance[14]. Southern Lee High School opened its doors at August 25, 2005; the buildings were begun to build from 2003. It is a brand new public school with newly established buildings, labs and departments such as, CAREERplus, CAREERplus2 and BuildingSkills. They are very important in helping students in their further development after high school, especially in their success at college and careers[15]. Based on these, parents may have positive expectation on the performance of this new high school, and make their choice based on their expectation. The families choose to attend new high school but live in old school district will move to the new one, this action will affect the housing price in both school districts.

[14] http://en.wikipedia.org/wiki/Lee_County_High_School_%28Sanford,_North_Carolina%29
[15] http://www.paxtonpatterson.com/success-stories.aspx

Several GIS and spatial data sources are combined for this study. Geo-coded parcel information was obtained from Lee County GIS Department. The dataset has information about property owner, year built, lot size, land use code, sale year, latitude and longitude coordinates, and sale price. The subsample of the data provides a record of each house transaction, attached and detached, that took place between 2004-2005 and 2006-2007. Each property is uniquely identified in the data which allows the creation of a panel data set.

All the amenities and disamenities data are also from GIS Department of Lee County, which includes parks, traffic count, brown field, railroad, and hydrology and so on. All the socio-economic information are from the year 2010 U.S. Census Bureau data detailed to census block[16], which provides provide a richer set of covariates.

GIS software allows us to merge the individual level single-family data into the census data which is gathered at the block level. GIS software also used for buffering around the school district boundary and the distance to the amenities and disamenities, it is the shortest distance from each single family house to the point of interest. The buffer is a spatial tool that records which objects fall within a specified distance to the geographic boundary. It is used in this research to record houses that are within 1500, 2000, 2500, 3000 meters from the new school district boundary.

3. Method

3.1. Regression Discontinuity (RD) Design

There are two main types of regression discontinuity designs (Hahn, et al. 2001): the sharp RD and the fuzzy RD. Sharp design exploits treatment as deterministic on some observable variable, where it takes on a range of values and the point is assumed to be known. Treatment is a random

[16] It provide the most detailed information available so far from the 2010 Census about a community's entire population, including cross-tabulations of age, sex, households, families, relationship to householder, housing units, detailed race and Hispanic or Latino origin groups, and group quarters.

variable and it is not a deterministic function in fuzzy design, but the conditional probability is known to be discontinuous (Hahn, Todd and van der Klaauw, 2011).

With regression discontinuities design, the probability of receiving treatment can be a random function of one or more underlying variables. In this study, it is obvious that treatment, as a geographic boundary, is a fixed threshold. Therefore, there is no chance that any observation in this study will escape from the treatment (Imbens and Lemieux 2008). Therefore in this case, the threshold for the treatment is sharp, and this study employs the method known as a sharp regression discontinuity design. The focus is on observations near the boundary, or a "discontinuity sample" (Angrist and Lavy 1999), varying the size of the sample by adjusting the distance from the boundary for observations.

In order to understand the RD model and why it is effective, certain ideas should be understood. First, choosing a deliberate cutoff criterion is essential. If certain groups accidently avoid the treatment, the outcome is likely to be biased. Second, the pre-post distribution needs to be a polynomial function. Third, the division between groups is determined only by the cutoff value, and other variables must be similar between the groups. Finally, accurate study analysis relies on fixed program variables in all test groups.

3.2. Spatial Model

The spatial autocorrelation effect is also considered in this research. This model is commonly referred to as a spatial-autoregressive model or SAR[17] (Drukker, Egger and Prucha, 2010). The combined spatial-autoregressive model with (spatial) autoregressive residuals is often referred to as SARAR[18] (Anselin and Florax, 1995).

The basic functional form is:

[17] Generalized versions of this model also allow for the dependent variable to depend on a set of exogenous variables and spatial lags, and the disturbances to be generated by a spatial-autoregressive process.
[18] Mixed-regressive-spatial autoregressive model with a spatial autoregressive disturbance (SARAR)

(3.1) $$\mathbf{y} = \boldsymbol{\lambda}\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$$

(3.2) $$\boldsymbol{u} = \boldsymbol{\rho}\mathbf{M}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

where y is an $n \times 1$ vector of observations on the dependent variable, W and M are $n \times n$ spatial-weighting matrices (with zero diagonal elements), Wy and Mu[19] are $n \times 1$ vectors typically referred to as spatial lags, and λ and ρ are the corresponding scalar parameters typically referred to as spatial-autoregressive parameters, X is an $n \times k$ matrix of observations on k right-hand-side exogenous variables (where some of the variables may be spatial lags of exogenous variables), and β is the corresponding $k \times 1$ parameter vector, ε is an $n \times 1$ vector of innovations.

In this study, SARAR model first combined with the difference-in-difference (DID) and then combined with RD model, the basic equation is as follow:

(3.3) $$\ln(P_{it}) = \beta_0 + \rho W_i \ln(P_{it}) + \sum_{k=1}^{K} \beta_k S_{ikt} + \sum_{l=1}^{L} \beta_l A_{ilt} + \sum_{m=1}^{M} \beta_m N_{imt} + \varepsilon_{it}\,,$$

$\varepsilon_{it} = \lambda \ M_i \varepsilon + u_{it}$ $|\lambda| < 1$ and $|\rho| < 1$

in this model, the price of a property,(the price of any house i, sold at time t) P, is modeled as a function of the structural characteristics, S_{ikt}, the kth housing attribute for the ith residence in the period t (e.g. number of rooms, size of the house), the aminities characteristics, A_{ilt}, lth housing attribute for the ith residence in the period t (e.g. distance to river, distance to parks), and the neighborhood characteristics, N_{imt}, mth housing attribute for the ith residence in the period t (e.g. percentage of white population, vacant ratio). Besides, x_i is a dummy variable, it equals 1 if the parcel in the new school district, equals 0 if otherwise.

λ and ρ are the spatial autocorrelation parameter and spatial autoregressive coefficient respectively. ε is a vector of error terms, and u is a vector of independent and identically distributed random error terms. W and M are n*n spatial weighting matrices that are taken to be

[19] In this paper, I use the same W in both the lag term and the error term, in other words, W=M

known and non-stochastic[20]. A normalized spatial contiguity matrix was used to account for the spatial dependence among the nearby properties. In a contiguity matrix, contiguous properties or the neighboring properties are assigned weights of 1, and noncontiguous properties are assigned weights of 0.

The total spatial effects used here are the spatial autoregressive model with autoregressive disturbances (SARAR)[21]. This research uses a generalized moments (GM) estimator for the autoregressive parameter of the disturbance process that was introduced by Kelejian and Prucha (1999) and offers simple computation of large sample sizes. Because of the simple fact that neighboring properties are generally built around the same time and share common location features, property values of a certain area share a spatial dependence. Estimations can be unreliable and misleading if that dependence is overlooked (Anselin and Bera 1998).

4. Estimation

The total sample of the whole county has 1208 parcels outside the new school district, and 1063 within it. Table 1 summarizes the data and presents descriptive statistics. It shows the mean value of housing price for control group before the establishment of new high school is about 108 thousand dollars and 104 thousand dollars after the establishment. Before and after the event, the mean values are 217 and 249 respectively for the treatment group. Parcels inside the new school district have higher average value than outside the district and these average values increase after the event for both of the groups when we evaluate them in the whole county level.

Table 1 goes about here

[20] $W=M$ was imposed in this model

[21] A SARAR model includes a spatial lag, the spatially weighted average of housing price, and also spatial error term. It assumes the missing variables correlated with each other spatially. Each observation has its own spatial weight matrix.

The sub-samples range from 136 parcels for the sample drawn with a 1500 meter buffer to 320 parcels for the sample drawn with a 3000 meter buffer. Table 2 summarizes the data and presents descriptive statistics. The results indicate that the difference of mean value of sale price between treatment and control group is positive in buffer 1500 and 2000, but become negative in buffer 2500 and 3000. Besides, we can also see almost all of the means of covariates are similar across the sub-samples, but not the same. This is true for an obvious reason: because I have selected sub-samples based on the similar distance to the border, factors influencing the housing price may vary at such small areas, but this variation cannot be large.

Figure 1 goes about here

Table 2 goes about here

Table 3 reports my estimates of the effect of the new high school district in Lee county using standard DID and Spatial DID models. The standard DID model indicates a positive but not significant interaction term, which indicates the new school has a positive effect on property prices for houses sold within the catchment after the school was established but this premium is not significant, and there is no obvious increase in housing price after deduction of the general effect shown in the control group. The spatial DID model also represents a positive result for the interaction term, but it is not significant, which indicates that after taking the spatial factor into account, the effect becomes zero. But the t statistic still suggests the presence of error spatial dependence in the data set (The coefficient estimates for Rho (error dependence) =0.07***). The result of spatial DID model confirms that the new school establishment does not have significant effect on housing price when we analyze it in the whole county level.

Table 3 goes about here

Housing price is also expected to rely on the information about the structural characteristics and the location of the property, their proximity to the amenities and disamenities. The results indicate area of the house and proximity to parks increased the property value in Lee County, but disamenities like brownfields and railroads will depress the price. There was no significant premium associated with the spatial effect when we compare the results shown in two columns. When we see the people lived in a neighborhood, it seems where more white people live and the presence of vacant houses in the surrounding area will significantly increase the housing price.

In addition, the result of an assessment of the degree to which the boundary may also demarcate differences in housing prices for the whole county dataset is provided in Table 4.

Table 4 goes about here

This table includes the means of the measured housing prices' exposure to the treatment of access to the new school district boundary ("Within new school boundary" and "Outside new school boundary") for periods before event (2004–2005) and after event (2006–2007) after taking neighborhood and structural characteristics into consideration. The final column shows the relative change in the variables (the "difference in difference") over the two periods. As expected, the real price of properties sold within the school district boundary rose after the deduction of normal increase, which is the effect of other factors except for the new school, but this effect is still not significant.

The outcome of the RD regressions are presented in Table 5, the results are presented in two columns for each of the four sub-samples around the boundary. In general, many of the results are as expected. For example, percentage of black people has no impact in all buffers. A similar situation exists for the vacancy ratio, a proxy for excess supply, has no impact in most buffers. The size of a parcel is positive and highly significant in all buffers except for 2500 meter, both with and without spatial effect, implying that larger parcels usually can be sold at a higher price.

Table 5 goes about here

When estimating the model using the sub-samples, I systematically restrict my sample to parcels that have different distances from the boundary. As the sub-samples are restricted to houses that are close to the boundary, it becomes less likely the differences of housing price on either side of the boundary will come from covariates other than the new school establishment. In comparing the estimates across the sub-samples, some estimates appear significant while others do not. For example, the nearest distance to hydrology and the nearest distance to railroad are not significant throughout. Some other estimates are significant in some sub-samples, but others are not. For example, distance to park is positive and significant across all samples with the exception of the small buffer subsample. Percentage of Asians in census block level has negative effect and remains significant in the long distance buffer sub-sample (2500 and 3000 meter buffers), but drops in significance level with the short distance buffer sub-samples (1500 and 2000 meter buffers). Finally, the age of the structure is negative and significant throughout except for the 2000 meter buffer samples.

The spatial weights matrixes are used here to eliminate spatial dependence. Table 5 also reports the estimation results of both the regression discontinuity design without and with the spatial effect for each sub-sample. In comparing the magnitude of the coefficients across these samples, I find that when comparing the coefficients of regressions without incorporating the spatial weight matrix, the general trend of the magnitudes first decreases slightly as the buffer distance increases, but then increases a little in the largest buffer. In addition, as the results of the four buffer groups showed, when compared with the results of RD and spatial RD within each buffer, the signs of all the coefficients are stable, which means that adding spatial effect does not change the direction of their effects on housing price. Moreover, for most coefficients, the significance level is equal or higher with spatial effect. Estimation results indicate a statistically significant spatial autoregressive coefficient (Lambda) only in a buffer of 2000 meters, suggesting that spatial dependence in sample of housing prices indeed exists in this buffered area sub-sample.

Taken as a whole, the variable we should pay more attention to in the RD model is the dummy variable indicating whether the parcel is within the school district border (Within_schborder). We can see from the result that the sign of the coefficients are positive and significant through short distance buffer sub-samples (1500 and 2000 meter buffers), which indicate the property values of houses within the new school district are higher than those outside it when houses are close to the boundary, and the positive impact was statistically significant. But the magnitudes as showed in RD result experience an increase first, and then a decrease with the increase in buffer size. This is also showed in Figure 2.

Figure 2 goes about here

Figure 2 is the plot of average housing price change with the distance from houses to the school district border. The curve in the left hand side is the housing price of houses within the new school district boundary; the trend of this curve is first decreased with the distance to the boundary, but after arriving at the minimum point, it begins to increase. The curve on the other side is the housing price of houses outside the boundary, the trend is almost the same but more obvious. Therefore, the maximum gap between the housing prices on the opposite sides of the boundary appeared at some point near the minimum point, this point is around 2000 meter buffer. This conclusion is in accordance with RD result in Table 5, where the magnitude of coefficients of within school district dummy increase from 1500 to 2000 buffer, but they decreases both in magnitude and in significance level when we compare them at longer distance buffer.

5. Conclusion and Implications

This paper examines housing price change before and after the establishment of a new high school. More specifically, it explores the use of geographic boundary as discontinuity in RD design; the assignment variable is the distance to a school district boundary and the new school effect on either side of this boundary are compared.

RD and DID methods are used as quasi-experimental evaluation tools because in this event, the new high school effect is implemented along geographic boundaries, rather than by randomly assigning students or citizens, which requires stringent modeling of the assignment process to minimize competing explanations.

The key assumption for the RD model is that sites around the threshold have similar unobservable attributes. Therefore, by focusing on the sales prices of houses in close proximity to the boundary of the new school district, we can more directly compare properties in similar neighborhoods that differ only by the treatment effect of being within the school district.

The results show that once other locational and structural attributes are controlled for, the benefits of access to the new high school have been capitalized into higher property values for houses around the new school district boundary. Since the two schools have developed a rivalry, helping them to make development, and the new school obviously gives householders another choice, there are good reasons for an increase in property value. But this positive effect is less obvious with houses that further away from the boundary. Therefore, we can conclude the new public school has necessarily induced the increase of property value, especially for houses near the boundary.

Besides, spatial RD is a particular type of RD design focuses on discontinuities in geography boundary, more methodological inquiry is needed to judge how well and under what conditions spatial RD yields unbiased estimates. And what buffer size around the boundary is most appropriate to generate high quality results.

Table 3.1. Summary Statistics for the Whole County Dataset

	Control				Treatment			
	Before		After		Before		After	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SALE_PRICE	108121.62	248361.25	104595.04	211017.94	217649.64	280347.24	249132.58	403254.05
Acres	1.18	3.49	1.30	4.37	2.53	7.13	2.86	11.81
Age	22.25	16.57	27.37	14.66	20.77	12.67	23.66	12.91
T	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00
Treatment	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00
Neardist_parks	1.20	1.00	1.28	1.02	1.57	1.23	1.51	1.16
Neardist_CBD	1.91	1.38	2.02	1.40	2.48	1.49	2.32	1.36
urbdum	0.54	0.50	0.50	0.50	0.38	0.49	0.40	0.49
Neardist_Brownfields	1.23	1.11	1.31	1.13	1.93	1.47	1.80	1.32
Neardist_Railroad	0.82	0.64	0.85	0.67	0.96	0.72	0.84	0.67
Neardist_Hydrology	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
WHT_pert	0.74	0.24	0.74	0.25	0.72	0.23	0.71	0.25
BLK_pert	0.15	0.18	0.15	0.20	0.16	0.17	0.18	0.19
ASN_pert	0.01	0.03	0.01	0.02	0.01	0.05	0.02	0.06
VAC_ratio	0.09	0.07	0.09	0.08	0.10	0.13	0.09	0.09
N	681		527		627		436	

Table 3.2. Summary Statistics for the 1500, 2000, 2500, 3000 Meters Sample

	1500				2000				2500				3000			
	Control		Treatment		Control		Treatment		Control		Treatment		Control		Treatment	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SALE_PRICE	114158	81759	211474	710000	104140	68753	232524	614189	165427	514580	97361	69609	156851	466518	100832	73667
LNPRICE	11.39	0.79	11.57	0.85	11.21	1.08	11.84	0.77	11.45	1.08	11.24	0.73	11.41	1.07	11.26	0.76
Acres	0.60	1.22	0.50	0.88	0.46	0.31	0.62	1.35	0.56	1.01	0.47	0.43	0.64	1.33	0.50	0.50
Age	28.08	26.77	34.97	25.49	29.46	26.54	26.85	23.21	26.74	25.65	27.32	24.18	28.80	27.01	27.94	27.13
Within_schborder	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00
Neardist_parks	1.15	1.10	1.50	1.29	1.11	1.07	1.59	1.29	1.19	1.11	1.39	1.21	1.14	1.11	1.43	1.18
Neardist_CBD	1.23	0.78	0.98	0.56	1.30	0.83	1.08	0.52	1.33	0.81	1.04	0.50	1.26	0.80	1.11	0.61
urbdum	0.94	0.24	0.97	0.17	0.92	0.27	0.99	0.11	0.94	0.25	0.96	0.19	0.93	0.25	0.92	0.27
Neardist_Brownfields	0.70	0.59	0.35	0.40	0.71	0.61	0.47	0.39	0.77	0.57	0.43	0.42	0.73	0.57	0.50	0.53
Neardist_Railroad	0.75	0.55	0.46	0.41	0.77	0.56	0.58	0.40	0.82	0.51	0.54	0.43	0.78	0.50	0.59	0.47
Neardist_Hydrology	0.06	0.05	0.10	0.07	0.06	0.05	0.08	0.07	0.05	0.04	0.08	0.07	0.06	0.04	0.08	0.07
WHT_pert	0.62	0.29	0.45	0.22	0.62	0.28	0.50	0.23	0.65	0.28	0.46	0.22	0.64	0.28	0.47	0.24
BLK_pert	0.24	0.27	0.33	0.21	0.24	0.25	0.30	0.19	0.21	0.24	0.32	0.20	0.21	0.24	0.32	0.21
ASN_pert	0.02	0.06	0.02	0.04	0.01	0.05	0.02	0.03	0.01	0.05	0.02	0.04	0.01	0.04	0.02	0.05
VAC_ratio	0.09	0.11	0.08	0.07	0.08	0.10	0.08	0.06	0.08	0.09	0.08	0.06	0.07	0.09	0.08	0.09
N	66		70		117		82		154		113		189		131	

Table 3.3. Difference in Difference Model of the Whole County

	DID without Spatial ln*price*	DID with Spatial ln*price*
Acres	0.01^{***} (3.43)	0.01^{***} (4.23)
Treatment	0.80^{***} (16.44)	0.89^{***} (17.90)
T	-0.05 (-1.02)	-0.05 (-1.10)
Interaction	0.06 (0.85)	0.03 (0.51)
Neardist_parks	0.75^{***} (12.34)	0.86^{***} (11.76)
Neardist_CBD	0.29^{***} (5.37)	0.24^{***} (3.77)
urbdum	0.18^{**} (3.20)	0.14^{*} (2.16)
Neardist_Brownfields	-1.05^{***} (-13.95)	-1.09^{***} (-12.60)
Neardist_Railroad	-0.12^{***} (-3.64)	-0.07 (-1.60)
Neardist_Hydrology	-2.07^{***}	-2.20^{***}

	(-5.34)	(-5.43)
WHT_pert	0.85***	0.72***
	(6.18)	(5.14)
BLK_pert	-0.03	-0.06
	(-0.15)	(-0.36)
ASN_pert	0.49	0.45
	(1.16)	(1.06)
VAC_ratio	0.82***	0.76***
	(4.62)	(4.24)
_cons	10.55***	10.64***
	(71.78)	(68.12)
lambda		
_cons		-0.03
		(-0.40)
rho		
_cons		7.66***
		(10.71)
N	2271	2271
pseudo R^2		

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.4. Evaluating the Treatment Effect of the Whole County Dataset

	Before Event		After Event		
	within new high school boundary	**Outside new high school boundary**	**within new high school boundary**	**Outside new high school boundary**	**Difference in Difference**
SALE_PRICE(in $100,000)	1.1	2.2	1.0	2.5	0.3
(t statistics)					(1.44)

"difference in difference" refers to the change in the value of the variable over two periods (prior to the creation of the high school and after the creation of the high school).
Means and t Statistics are estimated by linear regression
Inference: *** p<0.01; ** p<0.05; * p<0.1

Table 3.5. Results of Spatial RD for Four Sizes of Buffers

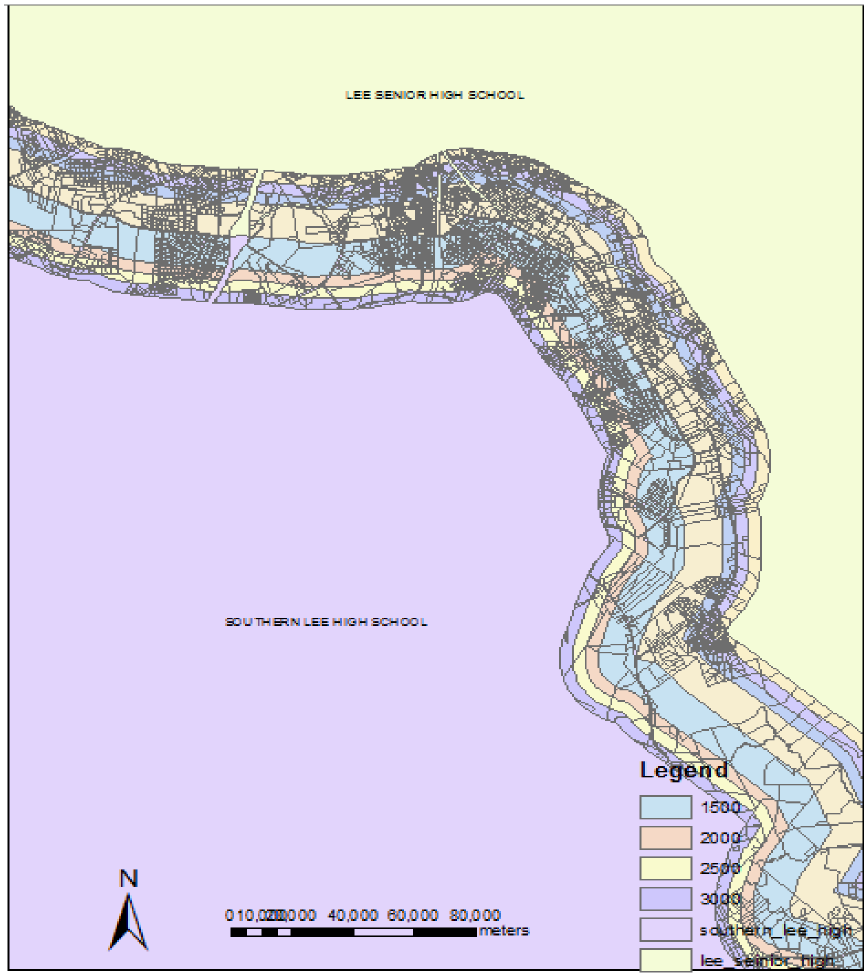
	1500 RD	**1500 RD+SARAR**	**2000 RD**	**2000 RD+SARAR**	**2500 RD**	**2500 RD+SARAR**	**3000 RD**	**3000 RD+SARAR**
Acres	0.18**	0.20**	0.30***	0.42***	0.04	0.04	0.20***	0.20***
	(2.70)	(2.73)	(3.76)	(4.62)	(0.64)	(0.67)	(3.73)	(3.80)
Age	-0.02*	-0.02*	-0.01	-0.01	-0.02**	-0.02**	-0.02***	-0.02***
	(-2.33)	(-2.48)	(-1.49)	(-0.74)	(-2.62)	(-3.17)	(-3.92)	(-4.27)
Age2	0.98	0.98	0.77	0.31	0.52	0.80	1.15*	1.14*
	(1.22)	(1.30)	(0.92)	(0.38)	(0.69)	(1.18)	(2.52)	(2.57)
Within_schborder	0.36*	0.36*	0.59***	0.58***	-0.19	-0.20	-0.08	-0.09

	(2.20)	(2.32)	(4.12)	(4.33)	(-1.50)	(-1.72)	(-0.74)	(-0.89)
Neardist_parks	-0.06	-0.06	0.05	0.03	0.14**	0.15**	0.11*	0.12**
	(-0.95)	(-0.99)	(0.84)	(0.57)	(2.87)	(3.07)	(2.46)	(2.64)
Neardist_CBD	-0.39	-0.40	0.02	0.03	0.02	0.04	0.03	0.05
	(-1.82)	(-1.95)	(0.11)	(0.16)	(0.13)	(0.26)	(0.18)	(0.36)
urbdum	-0.17	-0.16	0.79*	0.37	0.58*	0.69*	0.80**	0.94***
	(-0.40)	(-0.40)	(2.12)	(0.98)	(2.01)	(2.09)	(3.05)	(3.48)
Neardist_Brownfields	0.27	0.26	0.73	1.05*	0.59	0.51	0.07	0.07
	(0.46)	(0.46)	(1.46)	(2.21)	(1.49)	(1.31)	(0.20)	(0.22)
Neardist_Railroad	0.23	0.26	-0.26	-0.53	-0.70	-0.65	-0.03	-0.10
	(0.40)	(0.47)	(-0.47)	(-1.02)	(-1.67)	(-1.72)	(-0.08)	(-0.33)
Neardist_Hydrology	-0.50	-0.32	0.61	0.47	0.54	0.56	0.33	0.57
	(-0.36)	(-0.23)	(0.46)	(0.38)	(0.46)	(0.54)	(0.30)	(0.55)
WHT_pert	-0.41	-0.40	-0.39	-0.33	-0.36	-0.34	0.07	0.03
	(-0.69)	(-0.71)	(-0.68)	(-0.61)	(-0.79)	(-0.81)	(0.21)	(0.10)
BLK_pert	-0.47	-0.49	-0.37	-0.10	-0.88	-0.94*	-0.27	-0.40
	(-0.82)	(-0.90)	(-0.65)	(-0.19)	(-1.81)	(-2.08)	(-0.70)	(-1.06)
ASN_pert	0.28	0.32	-1.86	-0.50	-1.92	-2.73*	-2.51*	-3.00*
	(0.19)	(0.23)	(-1.23)	(-0.33)	(-1.43)	(-2.13)	(-2.05)	(-2.47)
VAC_ratio	0.67	0.65	-1.33	-1.11	-1.29	-1.31	-0.34	-0.51
	(0.79)	(0.82)	(-1.68)	(-1.51)	(-1.79)	(-1.93)	(-0.59)	(-0.93)
_cons	12.26***	12.25***	10.58***	10.93***	11.67***	11.58***	10.76***	10.67***

	(16.66)	(17.68)	(14.49)	(15.81)	(19.97)	(20.12)	(22.16)	(22.59)
lambda								
_cons		-0.77		-5.32^{**}		0.19		0.22
		(-0.57)		(-2.62)		(0.69)		(1.56)
rho								
_cons		2.63		-6.68		-8.16		-9.33
		(0.18)		(-0.52)		(-1.54)		(-0.91)
N	136	136	199	199	267	267	320	320
pseudo R^2								

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Coefficients of Lambda and rho are scaled

Figure 3.1. New School District and 1500, 2000, 2500, 3000 Buffer Around the Boundary



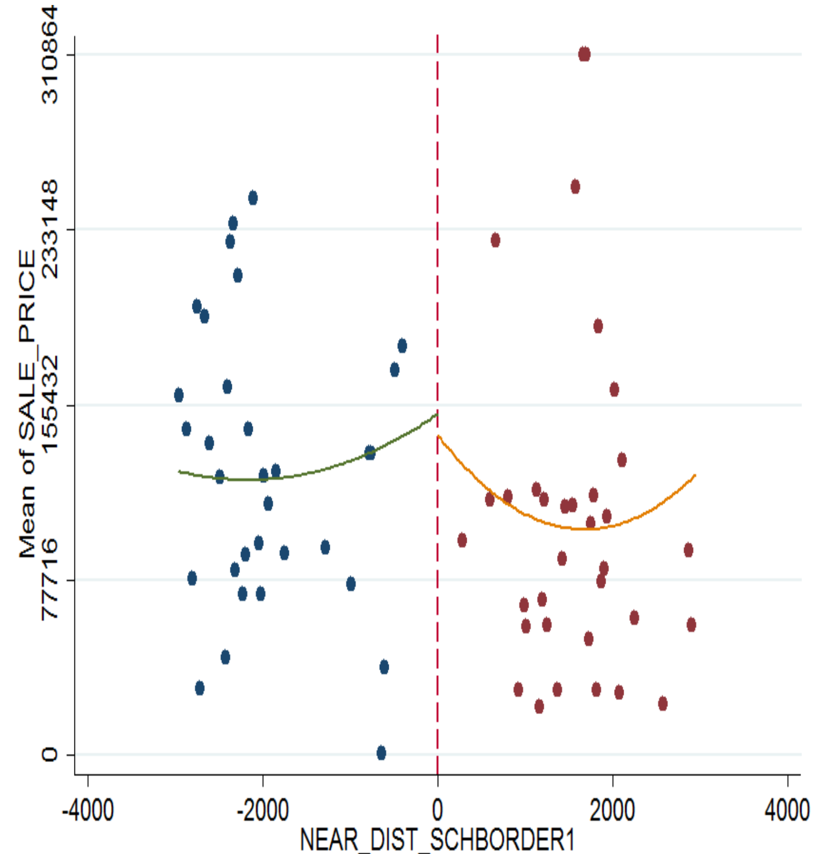


Figure 3.2. Graph for Regression Discontinuity Design

CHAPTER 4

Is There a Housing Price Premium for Legalized Same-Sex Marriages? A Difference-in-Differences Analysis Using Propensity Score Weighting

1. Introduction

There are 8.8 million self-identified gay and lesbian individuals in the United States, approximately 4% of the adult population (Gates and Ost, 2004). It has been thought that the community preferences of the gay and lesbian population increase housing values by identifying like-minded neighbors and quality neighborhoods. However, it can also be argued that income, wages, and social and political views have a direct correlation to housing values as well. For example, Stephen Murray (1996) argues that the reason San Francisco has been by gays for settlement is because “as a place in which being gay is celebrated, accepted, or at least tolerated as no big deal” – whichever was enough to be “better than where I am now”. Nonetheless, a number of studies have shown that gay and lesbian serve as a creative class, and the places they reside have a higher housing price than other similar places (Castells, 1983; Ley, 1994; Zukin, 1995; Smith, 1996). Additionally, social and political views of a community greatly impact an individuals’ decision on which urban area or “place” they choose to live. Besides, Noland (2005) finds that tolerant attitudes toward gays and lesbians are associated with positive attitudes toward global economic activity and international financial outcomes, which in my view will also cause increases in housing prices.

Jacobs (2013) pointed out that married couples have the additional advantage of being able to own real property as tenants by the entirety[22]. Thus, there is high probability that same sex couples who get married in Massachusetts also buy houses there. Same sex couples who do not get married can incur financial consequences from owning real property as joint tenants. Joint tenancy does not offer protection from claims against either tenant, nor does it ensure 100 % sole owner ship upon death of either partner.

Divorce is another problem to consider, as noted by Blake Ellis (2013). If the same sex couple resides in a state that does not recognize the marriage they may be denied a divorce and have to dissolve the marriage out of court instead of filing for a divorce. Divorce of a same-sex marriage can be complex and costly depending on domestic partnership state laws**.**

The theoretical foundation of hedonic modeling laid out by Rosen (1974) is based on the insight that a product price can be separated into distinctive price characteristics of the good. Hedonic models have been extensively used to estimate the value of individual property characteristics to the total value of a property. Hedonic analysis methodology is widely used to estimate the economic value of localized amenities that affect individuals' quality of life such as infrastructure, climate, and pollution.

A quasi-experimental approach to handling effect of treatment by subtracting change of treatment group from control group in the event period is difference-in-differences (DID) analysis (Greenstone & Gayer 2007). Ashenfelter and Card (1985) introduced an application of this method. In a basic DID estimation, two time periods (before and after event) and two groups (treatment and control group) are used. Neither group is exposed to the event during the first time period, but only one of the groups is exposed at the second time period. The inclusion of

[22] A type of concurrent estate in real property held by a Husband and Wife whereby each owns the undivided whole of the property, coupled with the Right of Survivorship, so that upon the death of one, the survivor is entitled to the decedent's share.

unexposed groups in this model removes any bias stemming from the difference of property value caused by other factors (Imbens & Wooldridge, 2009). Previous research on property value using hedonic difference-in-differences analysis include: Galster, Tatian, and Pettit (2004), Hallstrom & Smith (2005), and Tu (2005).

Besides, DID designs have also been used to address a lot of important policy issues, like the effects of minimum wages on employment (Card and Krueger, 1994), the effects of training and other activities on employees (Ashenfelter and Card, 1985, Heckman and Robb, 1986, Heckman and Hotz, 1989, Blundell, Meghir, Costa Dias, and van Reenen, 2004), and the effect of immigration on the local labor market (e.g. Card, 1990).

It is important to discuss the influence of neighborhoods and the socioeconomic relationship of neighborhoods on property values. Advances in computer software now permit the use of spatial econometrics to control for previously unobserved spatial characteristics and quasi-experimental techniques have become more prevalent (Kim et al. 2003). Furthermore, Kuminoff et al. (2010) discusses the increased use of fixed effects or spatial econometrics, by searching on more than 60 published papers, they find that more than half of the hedonic studies apply either spatial fixed effects or spatial econometric models to address omitted spatial variables. Similarly, Veie and Panduro (2013) looked at 21 hedonic studies published since 2010 and found approximately half of these studies used spatial econometrics to control for spatial correlation while the other half used fixed effects[23] and differences-in-differences. Therefore, this research used both the DID model with fixed effects and the spatial DID model to not only estimate the before-and-after effects of the marriage law, but also to simultaneously account for

[23] Fixed effects are generally based on available spatial entities such as provinces, census block, could also base on time.

both bias and inefficiency problems associated with spatially-correlated unobserved neighborhood effects.

Propensity score was first used by Rosenbaum and Rubin (1983) to make treated and control groups as similar as possible by estimating the probability of observations in the treatment group.

The propensity score is estimated from the logistic or probit model, and can be used in a variety of ways such as matching and weighting. Matching is a technique used to match subjects in two groups with the closest propensity score. Another approach is weighting, which is implemented by re-weighting the subjects in comparison groups using propensity scores to make them representative of the population of interest (Hirano and Imbens 2001). Benefit of weighting is the estimator remains consistent if either the first-stage propensity model or the second-stage outcome model is correctly specified, but not necessarily both (Lunceford & Davidian 2004).

Robins and Rotnitzky (1995) suggested using propensity score weighting to make model adjustments. Hirano and Imbens (2001) adopted this approach, but also made some improvements by using more flexible outcome regression functions.

This research also applies the propensity score weighting method and combines with DID model in determining the impact of same sex marriage law on housing price. To the best of my knowledge, few literatures have assessed similar questions about gay marriage effect on housing price. And no literature has used a quasi-experimental hedonic price function to examine this problem.

The paper is organized as follows: (i) Section 2 presents the application and the data used for estimation; (ii) Sections 3 and 4 present modeling approaches and estimation results; and (iii) concluding thoughts are offered in section 5.

2. Data

Several data sources are used to construct the dataset. Data for local property sales, characteristics and locations of the control group—New York State (NY) data were obtained from the Rensselaer County Geographic Information System and MassGIS datalayer for MA. Both datasets represent all home sales during an 8-year period between 2000 and 2008. Observations were restricted to single-family homes zoned for residential use and owned by individuals. There are totally 7099 observations, 3078 of them are in the control group and 4021 are in the treated group, which means that about 56.64% of the transacted parcels are in MA.

The study area encompasses multiple towns near the boundary of NY and MA. Individual property sales data contain information on housing characteristics such as number of rooms, number of stories and age of house, in addition to last sale date and sale price.

Most of the neighborhood characteristics were calculated from the New York State's GIS Clearinghouse, MassGIS datalayer. All of them are calculated in linear miles. Social Characteristics are collected from the 2000 and 2010 United Stated Decennial Census and represents statistics in the Census Block Group within which a house is located. Property sale prices were adjusted to 2004 in constant dollars, using the housing price index for NY and MA from the FHFA. [24]

Table 1 goes about here

Table 1 provides descriptive statistics for key characteristics included in the final empirical models of each state before and after the treatment. We can see from the table parcels in MA had a higher average price, more rooms and lower age than in NY. The parcels in MA were also on average closer to hydrology, police office, hospital, school, library, road and park,

[24] Federal Housing Finance Agency was created on July 30, 2008, when the President signed into law the Housing and Economic Recovery Act of 2008.

but further away from restaurants. In terms of neighborhood, the average percent of white people in census block group level is higher in MA than in NY, but percentage of other race, (e.g. Black, Asian) are lower. The average values also indicate that parcels in MA had a lower vacancy rate and less average family size.

3. Methods

3.1.The Basic Difference-in-Differences (DID) Model

The effect of a same sex marriage law cannot be measured simply by comparing housing prices in Massachusetts before and after legalizing gay marriage. This approach does not take into account other factors that may have changed over the same time period and could have affected housing prices: economic growth and population. In addition, this strategy fails to control different state characteristics such as tax differences, population density, and school quality.

Therefore, I based my strategy on a hedonic difference-in-differences analysis specified with fixed neighborhood and time effects to identify the effect of same sex marriage law on property value. This DID analysis is based on eight years of property transactions that include observations on properties before and after the introduction of same-sex marriage laws. Using DID analysis better identifies of the effects because fixed effects control all observable and unobservable neighborhood amenities that affect property values. Specifically, DID method allowed isolation of individual effects from same sex marriage law and effects of other existing variables. Under the DID approach, parcels in Massachusetts were set as the treatment group (Treatment dummy = 1), while the parcels in New York State were set as the control group (Treatment dummy = 0), because the control group in the DID approach is comprised of properties that are in place where gay marriage is banned.

The basic Hedonic DID model employed to properties is:

$\text{Log (Price)} = \beta_0 + \beta_1\text{Treatment dummy} + \beta_2\text{Time dummy} + \beta_{\text{Diff-n-Diff}}\,(\text{Treatment dummy}) * (\text{Time dummy}) + \beta_3\text{ Structure characteristics} + \beta_4\text{ Homeowner characteristics} + \beta_5\text{ Neighborhood characteristics} + \theta + \delta + u$ (4.1)

The focus is the coefficient of the interaction term ($\beta_{\text{Diff-n-Diff}}$), representing the treatment effect. The treatment effect is the difference between the treatment group and the control group after the same sex marriage law passed, adjusting for basic differences between the two states before the law. DID assumes that the gap between the treatment group and the control group in the follow-up period would have been unchanged if the law not been passed by using a group fixed effect and a "time period fixed effect. In other words, to get an unbiased treatment effect estimate it is assumed that housing price determinants were the same in Massachusetts and New York. This assumption requires both the treatment group and the control group to change over time at the same pace. If this assumption is incorrect, it would be difficult to estimate the effect of same sex marriage law in this experiment.

In the dataset of this research, repeat sales of the same home in both time periods are not observed, because only the last sales record of each parcel have been found. However, controlling for housing characteristics allowed different homes to be compared over time. In order to minimize the risk, I selected not only the neighboring state, but also parcels near the border of the two states, as sample groups.

Neighboring parcels are more likely to experience the same or similar economic and social changes. Besides, Census tract fixed effects (θ) were included to control for possible omitted variables such as crime rate or other unobserved characteristics in the community that are constant over time. Year fixed effects (δ) were included to capture yearly shocks that affect all the properties.

3.2.The Spatial DID Model

Spatial effects (spatial lag and spatial error method) are incorporated into the DID model to account for spatial autocorrelation. The spatial lag method represents the idea that the spatially weighted average of housing prices in a neighborhood affects the price of each individual house. In contrast, the spatial error method assumes there is one or more omitted variable in the hedonic price equation, and these omitted variables correlate spatially. Ignoring the spatial effect will cause an estimate to be both inefficient and inconsistent. Therefore, in this research, the Spatial Autoregressive Model with Spatial Autoregressive Disturbances and exogenous regressors, which is frequently referred to as a SARAR model (Anselin & Florax, 1995), were used.

The Spatial Difference-in-Differences model is:

$$\ln(P_{it}) = \beta_0 + \rho M_i \ln(P_{it}) + \sum_{k=1}^{K} \beta_k S_{ikt} + \sum_{l=1}^{L} \beta_l A_{ilt} + \sum_{m=1}^{M} \beta_m N_{imt} + \gamma_1 T_i + \gamma_2 t_i + \gamma_3 t_i T_i + \varepsilon_{it},$$

$$\varepsilon_{it} = \lambda\ M_i \varepsilon + u_{it} \qquad |\lambda| < 1 \text{ and } |\rho| < 1 \tag{4.2}$$

in this model, the price of a property,(the price of any house i, sold at time t) P, is modeled as a function of the structural characteristics, S_{ikt}, the *k*th housing attribute for the *i*th residence in the period *t* (e.g. number of rooms, size of the house), the aminities characteristics, A_{ilt}, *l*th housing attribute for the *i*th residence in the period *t* (e.g. distance to river, distance to parks), and the neighborhood characteristics, N_{imt}, *m*th housing attribute for the *i*th residence in the period *t* (e.g. percentage of white population, vacant ratio). λ and ρ are the spatial autocorrelation parameter and spatial autoregressive coefficient respectively. ε is a vector of error terms, and *u* is a vector of independent and identically distributed random error terms.

T_i is a dummy variable taking the value 1 if the individual is in the treatment group and 0 if they are in the control group. t_i is a dummy variable taking the value 1 in the post-treatment period and 0 in the pre-treatment period.

3.3.Propensity Score Weighted Regressions and Doubly Robust Estimation

As mentioned above, the treatment group and the control group need to be as similar as possible for a better comparison. But in observational studies, assignment of subjects to the treatment and control groups is not random. Thus, the estimate of the treatment effect may be biased by the existence of confounding factors. Four different propensity score[25] methods were introduced from previous studies for removing the effects of confounding when estimating the effects of treatment on outcomes: propensity score matching, stratification on the propensity score, inverse probability of treatment weighting (IPTW) using the propensity score, and covariate adjustment using the propensity score (Austin & Mamdani, 2006; Rosenbaum, 1987a; Rosenbaum & Rubin, 1983a).

As proposed by Rosenbaum and Rubin (1983), IPTW uses weights based on propensity score to make baseline covariates' distribution not correlated with the incident thus reduce the bias in the estimation of treatment effects with observational data (Morgan & Todd, 2008; Austin, 2013).

Each parcel was assigned a propensity score weight based on its own propensity score. For example, in IPTW method, the IPTW is defined as $W/ps + (1-W)/(1-ps)$[26]. For parcels in the treatment state ($W = 1$), the weight is equal to the inverse of propensity score ($1/ps$). For parcels in the control state ($W = 0$), the weight was equal to the inverse of 1-propensity score, $1/(1-ps)$. In general, for each parcel, IPTW is equal to the inverse of probability of receiving the treatment that the parcel received. The weighted treatment group and the weighted control group should have similar baseline characteristics.

---

[25] A propensity score is the probability of being in the treatment group as a function of all observable covariates. In other words, it is the probability of treatment selection conditional on observed baseline covariates. It can balance the conditional distribution of covariates given a certain propensity score for the treatment and control group. This method has become increasingly popular in medical trials and in the evaluation of economic policy interventions.

[26] W is treatment dummy ($W = 1$ treatment; $W = 0$ control), and ps is the propensity score.

Doubly robust (DR) estimation is another propensity score weighting method. Different from IPTW, the calculation of DR weighting is more complicated, it uses both the propensity score and regression model. But this weighting method can provide some benefits, if the propensity score model (the probit model in this research) is correctly specified, but regression model is not; or regression model (DID model in this research) is correct but propensity score model is not, the result will all consistent, this is why the weighting method is called "double robust". (Van der Laan and Robins 2003, Bang and Robins 2005, Li, Zaslavsky and Landrum 2013).

4. Result

From the descriptive statistics in Table 1, we can see that the mean of property values for the treated group is larger than control group. Over the same period, we can also see from the table that there are no clear changes in the magnitude of most characteristics. It is because the two groups are adjacent to the state border, as discussed above this makes their environment more similar. We can see location of the properties in Figure 1.

Table 1 goes about here

Figure 1 goes about here

The change in the magnitude of property value between the two groups is partly due to the same sex marriage law and partly a result of wider changes of other factors, such as distance to different amenities and structural characteristics. Because of this, it is necessary to isolate the impacts of the marriage law from other types of variation. This is done by controlling for both observable and unobservable factors that are expected to contribute to changes in property value from different sources. More specifically, this is done either through the use of covariates in the DID models, or by estimating a propensity score for participation in the event which is then used to weight the observations for estimating the impacts of the same sex marriage law on housing price.

The difference-in-differences (DID) estimator of housing price is shown in Table 2. The last column of the table is the DID estimator, which indicates a positive price premium of about $19460 for houses in MA. The results in this table indicate that the same sex marriage law has positive effect on housing price.

Table 2 goes about here

The probability of a particular house in the state which received the treatment as a function of that house's characteristics is estimated by the probit model. Propensity score is predicted based on this process. For identification purposes, one variable (Urban_Dummy) was added only in the probit model for the estimation of propensity score. As Lunceford and Davidian (2004), Bang and Robins (2005) indicated that the estimator is consistent when an important confounder is omitted entirely from one of the two models. In addition, Heckman et al (1997) stress that the data for treated and control groups should came from the same survey, which holds for the dataset used in this study. The results of the propensity score estimation are shown in Table 3, and the distribution of the scores for treated and control groups are shown in Figure 2.

Table 3 goes about here

Figure 2 goes about here

The density of propensity scores in Figure 2 shows that houses in the MA have higher propensity scores than those in NY. An impact of this is that in the matching process, MA houses with higher propensity scores will be compared with a relatively small number of control observations. However, there are at least some control observations with high propensity scores, and related to this, there is a large area of common support on which to estimate the impacts of the marriage law.

Table 4 reports the estimates of the effect of the same sex law on housing prices in MA. For this analysis, the coefficient on the treatment (T) and time (t) interaction variable is the key parameter, which is the same as a dummy variable equal to one for those properties that are in MA during the time period after the law began. The treatment variable controls for all properties that are in MA. My analysis will test whether the parameter of the interaction term is statistically different than zero, thus signifying a potential property value impact from the marriage law.

Additionally, the DID method is a flexible form of causal inference because it can be combined with some other procedures, such as Quantile Regression (Meyer et al., 1995) and Propensity Score (Heckman et al., 1997, 1998), DID combined with quantile regression was showed in Table 2. Both propensity score and DID are popular in their own research area, especially the propensity score method, which is mostly used in medical area. Not many studies combine these two methods. In this paper, I combined DID with propensity score approach and identified a comparison with the standard regression of DID, DID with DR and Spatial DID models in Table 4. Both census block group fixed effects and year fixed effects are included to control for possible omitted variables and yearly shocks. These two fixed effects are included in all the DID models except for Spatial DID model, since the spatial error part[27] have already account for such possible omitted variables.

To the best of my knowledge, these four methods have not been used together in a single paper before.

Table 4 goes about here

[27] Spatial error autocorrelation arises if error terms are correlated across observations, i.e., the error of an observation affects the errors of its neighbors, it is used to account for the spatially autocorrelated omitted variables.

We can see first from the result that the interaction terms for all the four models are positive and significant. This estimate results show that there is a significant premium of almost 28% with normal DID, 14% and 16% when using the DID with IPTW, DID with DR models respectively. This indicates the marriage law has a significant large positive effect on housing price in MA; the positive law effect was capitalized in property prices after the marriage law passed, and we get more modest results after using the propensity score weighting.

As mentioned above, in the spatial model, a characteristic change of one parcel affects not only its own price, but also the prices of the neighboring parcels, which may further influence some units far away, that is, the outcomes are determined simultaneously. Therefore, the magnitudes of coefficients in the spatial model (in this paper, it is the spatial DID model) does not represent the marginal effects that measure how changes in the independent variables affect the dependent variable. But the significance level and sign of the coefficients still make sense.

When making further investigation by comparing the t statistic of the interaction term coefficients in the DID with IPTW and DR models, we can see that DID with DR method ($t = 3.41$) has higher significant level than DID with IPTW ($t = 2.46$). In general, the resulting estimation of the IPTW, DR estimations is different in different situations: If the propensity score model (Probit model in this research) is correctly specified, the double-robust estimator will have a smaller variance than the IPTW estimator. If the outcome regression (DID model) is correctly specified, the double-robust estimator will have a larger variance[28] (Emsley, Lunt, Pickles, & Dunn, 2008).

There is still another possibility, when both the propensity score and outcome regressions are misspecified, DR estimators provide biased and inefficient estimates (Kang and Schafer,

[28] But it is offering protection against the misspecification of this model.

2007, Porter et al., 2011, Basu et al., 2011). But in this condition all alternative methods, such as DID or spatial DID regression, would also have this bias. Unfortunately, in the real study, we can never know whether or not the model we have constructed is misspecified. Thus, correct specification of the regression model is an unverifiable assumption.

The change of housing price is also expected to rely on the information about the other factors of the property, their characteristics, their location, and their neighbors. The results indicate proximity to hydrology, library and road decreased the property prices. There was no significant premium associated with distance to hospital and school. The increase of stories and number of rooms would increase the property prices. It seems residents would pay less for old houses and this price decrease was significant.

5. Conclusion

Using normal DID, spatial DID and propensity score weighting-based DID method, the aim of this study was to identify whether there is a housing price premium in MA as a result of passage of a law providing for same-sex marriages. These techniques, such as fixed effects, spatial autocorrelation, IPTW and DR, assure the comparability of the two groups[29] (treated and control) with respect to relevant characteristics. Furthermore, those approaches implemented in this research also help overcome shortcomings of regression-based approaches, especially when it comes to the comparisons of effects among multiple methods.

Normal DID with fixed effects to control for time effects, observed and unobserved neighborhood effects, while exploiting changes in the property value of MA during the time period of the data set. The result show a positive and significant interaction term, but the magnitude of the coefficient indicates a high percentage of increase (almost 28%) in housing price after deduction of the general effect showed in control group.

[29] They could make the two groups as similar as possible.

At the same time, the two-step procedure that uses IPTW and DR with DID to allow us to identify the different factors causing differences in housing price. The results also indicate that for properties under almost the same condition, houses in MA that passed the same sex marriage law have a premium in terms of housing price, but the magnitude of coefficients of interactions indicate modest premiums (14% and 16%) as compared to normal DID. Another identification strategy is based on a spatial DID specification, which also generate a statistically significant result.

The findings of this paper reveal that properties in MA experienced increases in values from the DID, DID with IPTW, DID with DR and spatial DID methods respectively, after passage of the same sex marriage law. These results are significant across linear and non-linear specifications. Therefore, a unique phenomenon in economic area that same sex marriage law can raise property value is improved by using a quasi-experimental methodology. The results have policy implications for other states to contemplate as they consider similar laws.

Table 4.1. Descriptive Statistics of control and treatment group

	Mean	SD	Mean	SD
	Control		Treated	
LS_PRICE (in $1000 of 2004, 1st quarter)	161	458	195	438
T	0.7014	0.4577	0.7593	0.4276
W	0.0000	0.0000	1.0000	0.0000
Interaction	0.0000	0.0000	0.7593	0.4276
Stories	1.4737	0.4383	1.4312	0.5195
Num_Rooms	4.5420	2.3722	5.5747	2.4188
Age	55.0408	43.0141	50.9167	32.4155
NEARDIST_Hydrology	0.0169	0.0248	0.0110	0.0210
NEARDIST_Police	0.0700	0.0494	0.0441	0.0450
NEARDIST_Hospital	0.2424	0.2161	0.1644	0.2112
NEARDIST_School	0.0710	0.0633	0.0371	0.0509
NEARDIST_Library	0.1239	0.1143	0.0424	0.0581
NEARDIST_Food	0.4899	0.5136	0.8825	0.4754
NEARDIST_Road	0.0004	0.0018	0.0002	0.0011
NEARDIST_Park	0.0675	0.0773	0.0164	0.0304
Pct_white	0.9645	0.0479	0.9486	0.0515
Pct_black	0.0126	0.0271	0.0191	0.0283
Pct_asian	0.0060	0.0136	0.0098	0.0126
Pct_multirace	0.0115	0.0180	0.0143	0.0148
Vac_ratio	0.3198	1.6672	0.2420	1.0176
Ave_family_size	3.2093	0.8628	3.1932	1.0214
N	3078		4021	

Table 4.2. Difference in Differences Estimator of Housing Price

	DIFFERENCE-IN-DIFFERENCES ESTIMATION						
	Before (2000-2003)			After (2005-2008)			
Outcome Variable	**Control**	**Treated**	**Diff(BL)**	**Control**	**Treated**	**Diff(FU)**	**DIFF-IN-DIFF**
PRICE (in 1000$)	296.113	309.437	13.324	138.801	171.585	32.784	19.46
Std. Error	145.823	144.91	17.719	30.821	32.253	10.979	19.007
z	2.03	296.2	0.75	291.01	152.73	15.1	1.02
P>z	0.042	0.033	0.452	0	0	0.003***	0.306

* Means and Standard Errors are estimated by linear regression
**Inference: *** p<0.01; ** p<0.05; * p<0.1

Table 4.3. Estimation of Propensity Score

Variable	Coef.	Std.Err.	t
Stories	-0.08976	0.063019	-1.42
Num_Rooms	-0.02998	0.015926	-1.88
Age	0.00157	0.000848	1.85
NEARDIST_Hydrology	-10.6974	1.611376	-6.64
NEARDIST_Police	20.30455	1.369796	14.82
NEARDIST_Hospital	6.269115	0.351184	17.85
NEARDIST_School	12.52498	1.243854	10.07
NEARDIST_Library	-26.2042	1.241683	-21.1
NEARDIST_Food	2.808639	0.193434	14.52
NEARDIST_Road	5.304629	20.80794	0.25
NEARDIST_Park	-3.99611	1.259849	-3.17
Urban Dummy	0.24469	0.067383	3.63
Pct_white	-6.69008	3.286585	-2.04
Pct_black	-5.97028	3.90081	-1.53
Pct_asian	-7.21296	4.117208	-1.75
Pct_multirace	-7.51319	4.553032	-1.65
Vac_ratio	-0.00771	0.020128	-0.38
Ave_family_size	-0.16968	0.028408	-5.97
_cons	4.655237	3.294699	1.41
Log likelihood	-3729.961		
Pseudo R2	0.2322		
Number of observations	7099		

Table 4.4. DID Estimation Results with Fixed Effects, IPTW, DR and Spatial effects

	DID lnP	DID with IPTW* lnP	DID with DR* lnP	Spatial DID lnP
T	0.037	0.091	0.161	0.020
	(0.27)	(0.56)	(1.30)	(0.12)
W	-0.323***	-0.232***	-0.200***	1.055**
	(-7.18)	(-4.27)	(-4.74)	(3.01)
Interaction	0.278***	0.143*	0.159***	0.193**
	(5.77)	(2.46)	(3.41)	(2.83)
Stories	0.208***	0.227***	0.142***	0.068**
	(10.96)	(7.28)	(8.01)	(2.79)
Num_Rooms	0.001	0.010	0.017***	0.067***
	(0.28)	(1.49)	(3.86)	(12.52)
Age	-0.002***	-0.003***	-0.003***	-0.002***
	(-8.15)	(-7.89)	(-11.07)	(-6.27)
NEARDIST_Hydrology	-0.251	0.443	-0.879	-3.763***
	(-0.54)	(0.79)	(-1.88)	(-5.89)
NEARDIST_Police	0.768**	0.490	0.711	2.854***
	(2.61)	(1.16)	(1.50)	(6.87)
NEARDIST_Hospital	0.164	0.260*	1.100***	0.942***
	(1.92)	(2.16)	(8.00)	(7.33)
NEARDIST_School	0.404	0.753	2.238***	0.550

	(1.37)	(1.76)	(5.28)	(1.36)
NEARDIST_Library	-0.990***	-0.752*	-2.008***	-1.473***
	(-4.59)	(-2.32)	(-4.50)	(-5.14)
NEARDIST_Food	0.827***	0.872***	1.200***	-0.576
	(17.13)	(11.67)	(18.18)	(-1.96)
NEARDIST_Road	-35.835***	-35.357***	-29.264***	-2.076
	(-6.16)	(-3.53)	(-5.74)	(-0.30)
NEARDIST_Park	1.047***	0.882*	0.848*	0.025
	(3.78)	(2.10)	(2.13)	(0.07)
Pct_white	0.705	1.373	-0.138	-1.662
	(0.84)	(1.36)	(-0.16)	(-1.18)
Pct_black	-0.462	0.138	-1.511	-2.392
	(-0.46)	(0.12)	(-1.48)	(-1.53)
Pct_asian	5.484***	7.281***	7.773***	-4.715*
	(5.04)	(5.35)	(7.09)	(-2.54)
Pct_multirace	-0.474	-0.483	-1.881	-0.599
	(-0.42)	(-0.36)	(-1.47)	(-0.35)
Vac_ratio	-0.004	-0.004	-0.007	-0.013
	(-0.68)	(-0.54)	(-1.04)	(-1.17)
Ave_family_size	-0.014	-0.014	-0.022**	0.042***
	(-1.58)	(-1.29)	(-2.84)	(3.52)
_cons	10.596***	9.731***	10.860***	12.797***

	(12.60)	(9.52)	(12.47)	(9.06)
lambda				
_cons				0.010
				(0.53)
rho				
_cons				0.233
				(1.01)
N	7099	7099	7099	4716
pseudo R^2				

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

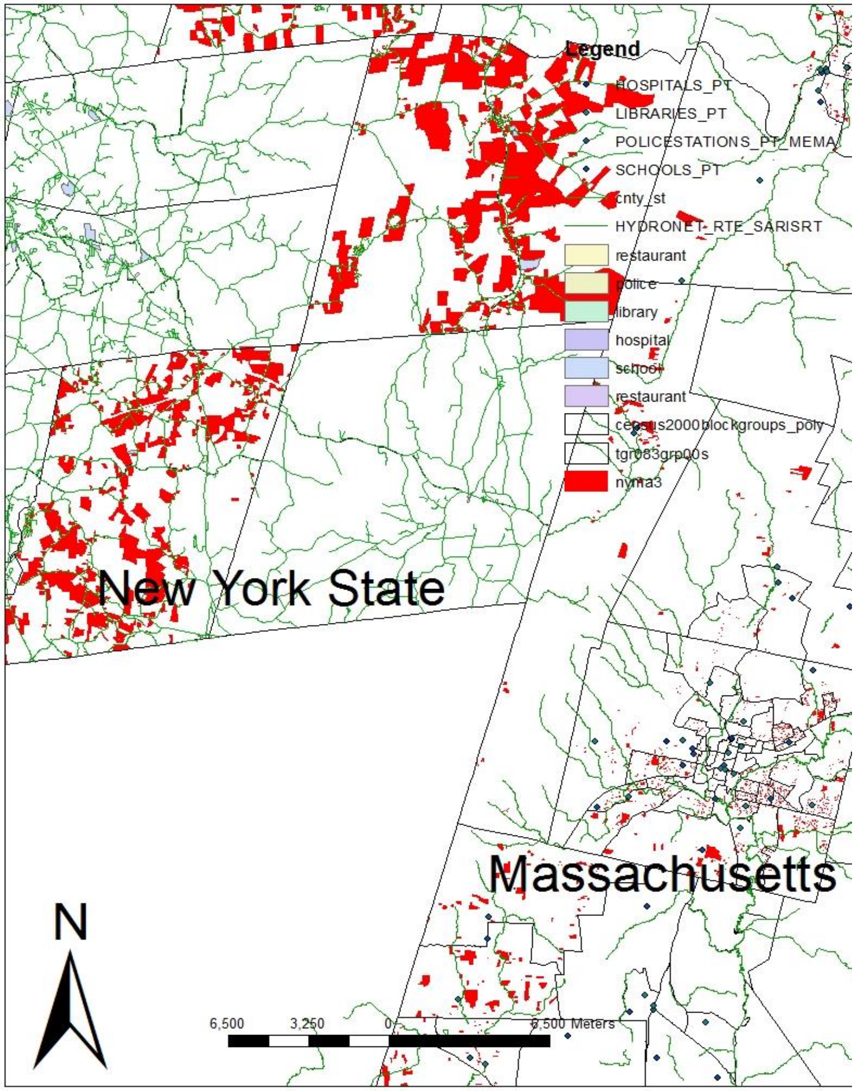


Figure 4.1. Places of Interest and Parcels

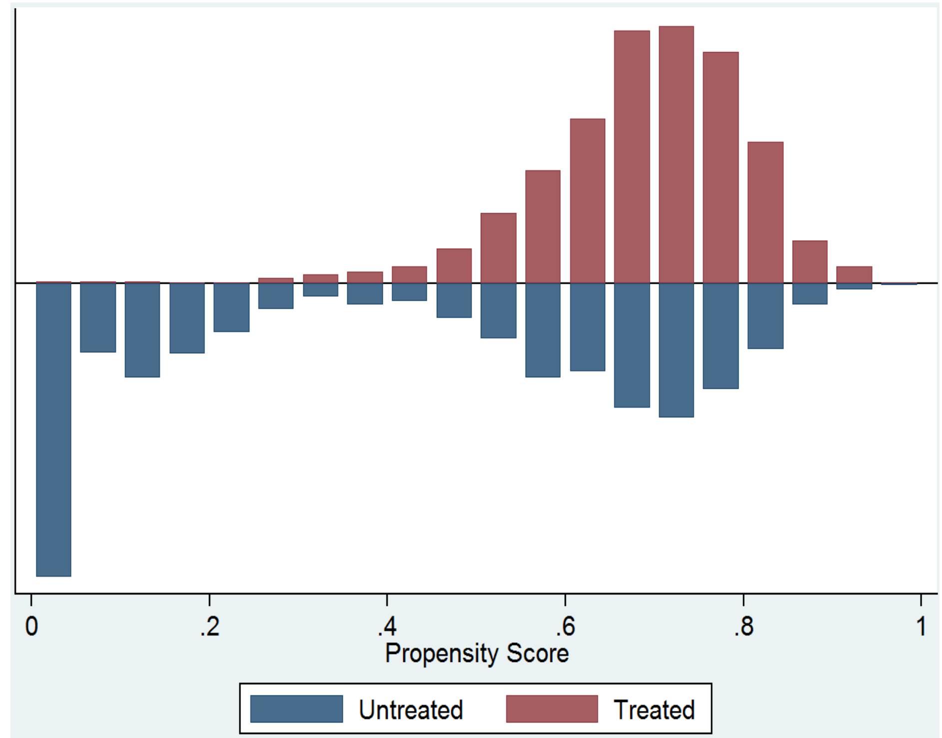


Figure 4.2. Propensity Score Distribution

REFERENCES

Abadie, A., J. Angrist, and G. Imbens. 2002. “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings.” Econometrica 70:91–117.

Abadie, A., J. Angrist and G. Imbens. 2001. “Instrumental Variables Estimation of Quantile Treatment Effects.” Econometrica. Forthcoming.

A.T. Court. 1939. “Hedonic Price Indexes with Automotive Examples, in the Dynamics of Automobile Demand,” General Motors, New York. pp. 98-119.

Anderson, L.; Cordell, H. 1985. “Residential Property Values Improved by Landscaping with Trees.” Southern Journal of Applied Forestry 9: 163-166.

Austin P.C., Mamdani M.M. A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. Statistics in Medicine. 2006;25:2084–2106.

Anselin, L., & Florax, R.J.G.M. (1995). Small sample properties of tests for spatial dependence in regression models: Some further results . In Anselin, L., & Florax, R. J. G. M. (Eds.), New directions in spatial econometrics (pp. 21-74).

Ashenfelter, O. and D. Card (1985), “Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs on Earnings,” The Review of Economics and Statistics 67 (1985), 648- 66.

Anselin, L. and Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, A. and Giles, D. E., editors, Handbook of Applied Economic Statistics, pages 237-289. Marcel Dekker, New York.

Angrist, J.D., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. Quarterly Journal of Economics 114, 533–575.

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

Bell, K. P. and Bockstael. N. E. 2000. "Applying the Generalized-Moments Estimation Approach to Spatial Problems involving Micro-level Data." Review of Economics and Statistics 82: 72-82.

Bockstael, N.E. and Bell, K.P. 1997. "Land Use Patterns and Water Quality: The Effect of Differential Land Management Controls. Conflict and Cooperation in Trans-Boundary Water Resources. Just, R.E., Netanyahu, S., eds. Boston, USA: Kluwer Academic Publishers: 169– 191.

Bowen, W. M., Brian Mikelbank, and D. M. Prestegaard. 2001. "Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications." Growth and Change 32: 466–490.

Bransington D. M. and D. Hite. 2005. "Demand for Environmental Quality: A Spatial Hedonic Analysis." Regional Science and Urban Economics 35: 57– 82.

Brown, J. and H. Rosen. 1982. "On the Estimation of Structural Hedonic Price Models" Econonmetrica 50: 765-768.

Bolitzer, B., and N. Netusil. 2000. “The Impact of Open Space on Property Value in Portland, Oregon.” J. Environ. Manage. 59:185-93.

Blake Ellis (2013), The high cost of same-sex divorce. CNN Money. May 13, 2013

Blundell, R., C. Meghir, M. Costa Dias, and J. van Reenen (2004): "Evaluating the Employment Impact of a Mandatory Job Search Program", Journal of the European Economic Association, 2, 569-606.

BASU, A., POLSKY, D. & MANNING, W. 2011. Estimating treatment effects on healthcare costs under exogeneity: is there a ‘magic bullet’? Health Services and Outcomes Research Methodology, 11, 1-26.

Bang, H., and J. M. Robins. 2005. Doubly robust estimation in missing data and causal inference models. Biometrics 61: 962–973.

Bogart, W. T., & Cromwell, B. A. (2000). How much is a neighborhood school worth? Journal of Urban Economics, 47(2), 280-305.

Black, Sandra E. 1999. “Do better schools matter? Parental valuation of elementary education.” Quarterly Journal of Economics, 114 (1999), pp. 577–599

Can, A (1992), "Residential quality assessment: alternative approaches using GIS", The Annals of Regional Science, Vol. 26 pp.97-110.

Chen, S., and W. Van der Klaauw. “The Effectof Disability Insurance on Labor Supplyof Older Individuals in the 1990s”, unpublished manuscript, 2004.

Card, David. 1990. “The Impact of the Mariel Boatlift on the Miami Labor Market,” Industrial and Labor Relations Review, 43: 245-257.

Castells M(1983) The City and the Grassroots: A Cross-Cultural Theory of Urban Social Movements, Cultural identity, sexual liberation and urban structure: The gay community in San Francisco. In (Edward Arnold, London), pp 138–170.

Can. A. 1990. “The Measurement of Neighborhood Dynamics in Urban House Prices,” Economic Geography 66: 254–272.

Can, A (1992), "Residential quality assessment: alternative approaches using GIS", The Annals of Regional Science, Vol. 26 pp.97-110.

Clapp, J. M., M. Rodriguez, and G. Thrall. 1997. “How GIS Can Put Urban Economic Analysis on the Map”. Journal of Housing Economics 6: 368-386.

Card, D., and A. B. Krueger (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," The American Economic Review, 84, 772-793.

Din Garrod, G. and G.W. Kenneth. 1999. “Economic Valuation of the Environment: Methods and Case Studies.” Massachusetts: Edward Elgar Publishing Inc.

Dubin, R. A. 1988. “Estimation of Regression Coefficients In the Presence of Spatially Auto-Correlated Error Terms.” Review of Economics and Statistics 70: 466-474.

Dubin, R. A. 1992. “Spatial autocorrelation and neighborhood quality.” Regional Science and Urban Economics, 22: 433-452.

Dubin, R., R.K. Pace, and T. Thibodeau. 1999. “Spatial Autoregression Techniques for Real Estate Data,” Journal of Real Estate Literature 7: 79-95.

Darling, A.H. 1973. “Measuring Benefits Generated by Urban Water Parks.” Land Econ XLlX (I): 22-34.

Din, A., Hoesli, M. and Bender, A., 2001. "Environmental Variables and Real Estate Prices," Ecole des Hautes Etudes Commerciales, Universite de Geneve-.

Drukker, D. M., Egger, P., Prucha, I. R. (2010) “On Two-step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors”. Working paper.

Eide, E. and Mark S. 1998. “The Effect of School Quality on Student Performance: A Quantile Regression Approach.” Economics Letters 58: 345-50.

Emsley, R., Lunt, M., Pickles, A., & Dunn, G. (2008). “Implementing double-robust estimators of causal effects.” Stata Journal, 8, 334–535.

Fennema, A.T., Veeneklaas, F.R., Vreke, J. 1996. “Meerwaarde woningen door nabijheid van groen (Surplus Value of Dwellings in the Vicinity of Green Areas).” Stedebouw en Ruimtelijke Ordening 3: 33-35.

Fanhua K., Haiwei Y., Nobukazu N. (2007) “Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China”. Landscape and Urban Planning 79: 240–252

Freeman, A. M. “Measuring Environmental and Resource Values”. Washington, DC: Resources for the Future Press, 1993.

Gawande, K., and Jenkins-Smith, H. 2001. “Nuclear Waste Transport and Residential Property Values: Estimating The Effects Of Perceived Risks.” Journal of Environmental Economics and Management 42: 207-233.

Geoghegan, J, L. A. Wainger, and N. E. Bockstael. 1997. “Spatial Landscape Indices in a Hedonic Framework: an Ecological Economic Analysis Using GIS”. Ecological Economics 23: 251-264.

Gillen, K., Thomas T., and S. Wachter. 2001. “Anisotropic Autocorrelation in House Prices.” Journal of Real Estate Finance and Economics 23: 5-30.

Grey. C. and Dcneke. F. 1978. Urban Forestry. John Wiley, Chichester 279.

Gates, Gary J., Jason Ost. 2004. The Gay and Lesbian Atlas. Washington, D.C.: Urban Institute Press.

Greenstone, M., and T. Gayer. 2007. "Quasi-experimental and Experimental Approaches to Environmental Economics." Available at SSRN: http://ssrn.com/abstract=1001330.

Galster, G., P. Tatian, and K. Pettit. 2004. "Supportive Housing and Neighborhood Property Value Externalities." Land Economics 80(1): 33-54.

Gyourko, J. and Tracy, J. 1999. "A Look at Real Housing Prices and Incomes: Some Implications for Housing Affordability and Quality." Federal Reserve Bank of New York Economic Policy Review, September, pp. 63–77.

He, L.T. and Winder. R.C. 1999. "Price Causality between Adjacent Housing Markets within a Metropolitan Area: A Case Study." Journal of Real Estate Portfolio Management 5 (1): 47–58.

Heckman, J.J., Ichimura, H., Todd, P., 1998. "Matching as an Econometric Evaluation Estimator." The Review of Economic Studies 65, 261–294.

Heckman, J.J., Ichimura, H., Todd, P.E., 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." The Review of Economic Studies 64, 605–654.

Harvey S. Jacobs 2013. "Same-sex marriages create new questions in property laws." The Washington Post. January 4, 2013

Hallstrom, D.G., and K. Smith. 2005. "Market Responses to Hurricanes." Journal of Environmental Economics and Management 50: 541-561.

Hirano K, Imbens G. “Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization.” Health Services & Outcomes Research Methodology 2001; 2:259–278.

Hirano K, Imbens G, Ridder G. “Efficient estimation of average treatment effects using the estimated propensity score.” Econometrica 2003; 71:1161–1189.

Heckman, J. J., and R. Robb (1986): “Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes,” in H. Wainer (ed.), Drawing Inferences from Self-Selected Samples, 63-113.

Heckman, J. J., and V. J. Hotz (1989): “Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training.” Journal of the American Statistical Association, 84, 862-880.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” Econometrica, 69(1): 201–09.

Hahn, J., P. Todd, W. van de Klaauw. 2001. “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” Econometrica 69 (1), 201-209.

H. Spencer Banzhaf and Garima Bhalla. 2012. “Do Households Prefer Small School Districts? A Natural Experiment.” Southern Economic Journal: January 2012, Vol. 78, No. 3, pp. 819-841.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." Quarterly Journal of Economics, 115(4): 1239-85.

Hanushek, E. A., & Rivkin, S. G. (2003). “Does public school competition affect teacher quality?” InC. Hoxby (Ed.), The economics of school choice (pp. 23-48). Chicago: University of Chicago Press.

Imbens, G. W., and T. Lemieux. 2008. “Regression DiscontinuityDesigns: A Guide to Practice.” Journal of Econometrics 142 (2): 615-635.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." Journal of Economic Literature, 47(1): 5-86.

Irwin, Elena G. 2002. “The Effects of Open Space on Residential Property Values.” Land Economics 78: 465-480.

Jacob, R. T., Zhu, P., Somers, M. A., and Bloom, H. S. (2012).A Practical Guide to Regression Discontinuity. MDRC.

Kelejian H.H. and Prucha I.R. 1998. “A Generalized Spatial Two Stage Least Squares Procedure For Estimating A Spatial Autoregressive Model With Autoregressive Disturbances.” Journal of Real Estate Finance and Economics 17: 99-121.

Kemp, G. 1999. “Least Absolute Error Difference Estimation of a Single Equation From A Simultaneous Equations System.” Mimeo, University of Essex.

Kestens, Y., Thériault, M. and Des Rosiers. F. 2004. “The Impact of Surrounding Land Use and Vegetation On Single-Family House Prices.” Environment and Planning B: Planning and Design 31: 539-567.

Kim, C. W. and Phipps, T. T. and Anselin, L. 2003. “Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach.” Journal of Environmental Economics and Management 45(1): 24-39.

Knight, Keith, Gilbert Bassett and Mo-Yin S. Tam. 2000. “Comparing Quantile Estimators for the Linear Model.” Preprint.

Kuminoff N, Parmeter C, Pope J (2010) “Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?”. J Environ Econ Manag 60:145–160

Koenker, Roger and Yannis Bilias. 2001. “Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments.” Empirical Economics 26: 199-220.

Koenker, R. and G. Bassett. 1978. “Regression Quantiles.” Econometrica 46: 33-50.

Koenker, R. and K. F. Hallock. 2001. “Quantile Regression.” Journal of Economic Perspectives 15: 143-56.

Kiel, Katherine A.1995. "Measuring the Impact of the Discovery and Cleaning of Identified Hazardous Waste Sites on House Values." Land Economics 71:428-35.

K. Hirano and G. Imbens. 2001. "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," Health Services and Outcomes Research Methodology 2: 259-278.

KANG, J. D. Y. & SCHAFER, J. L. 2007. “Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data.” Statistical Science, 22, 523-539.

Kathrine Lausted Veie & Toke Emil Panduro, 2013. "An alternative to the standard spatial econometric approaches in hedonic house price models," IFRO Working Paper 2013/18, University of Copenhagen, Department of Food and Resource Economics.

Leggett, C. G., and Bockstael, N. E. 2000. “Evidence of the Effects of Water Quality on Residential Land Prices.” Journal of Environmental Economics and Management 39: 121-144.

Li F, Zaslavsky AM, Landrum MB. “Propensity score weighting with multilevel data.” Stat Med 2013;32:3373-87.

Levin, Jesse. 2001. “For Whom the Reductions Count: A Quantile Regression Analysis of Class Size on Scholastic Achievement.” Empirical Economics 26: 221-46.

LeSage, J. P., and R. K. Pace. 2009. Introduction to Spatial Econometrics. Boca Raton, FL: CRC Press Taylor & Francis Group.

Lutzenhiser, M., and N. Netusil. 2001. “The Effect of Open Space on a Home’s Sale Price.” Contemporary Econ. Pol. 19:291-98.

Laurie, I. 1979. “The Natural Environment in the Design and Development of Urban Green” John Wiley, Chicbester. 428.

Lansford, Notie H, and Lonnie L. Jones. 1995."Marginal Price of Lake Recreation and Aesthetics: A Hedonic Approach." Journal of Agricultural and Applied Economics 27: 212-23.

Lunceford JK, Davidian M. “Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study.” Stat Med 2004;23:2937–60.

Ley D (1994) Gentrification and the politics of the new middle class. Environment and Planning D: Society and Space 12:53–74.

Ludwig, Jens, and Douglas L. Miller. 2007. “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design.” Quarterly Journal of Economics, 122(1): 159–208.

Lemieux, K. Milligan. “Incentive Effects of Social Assistance: A Regression Discontinuity Approach.” NBER Working Paper 10541, 2004.

Lee, D. “Randomized Experiments from Non-random Selection in U.S House Elections.” Journal of Econometrics(Forthcoming). Manuscript, University of California at Berkeley, 2005.

MaCurdy, T. and C. Timmins 2000. “Bounding the Influence of Attrition on Intertemporal Wage Variation in the NLSY.” Mimeo Stanford University.

McMillen, D. P. and Thornes, P. 2006. “Housing Renovation and the Quantile Repeated-Sales Price Index.” Real Estate Economics 34: 567–584.

McMillen, D.P., 2008. “Changes in the Distribution of House Prices over Time: Structural Characteristics, Neighborhood, or Coefficients.” Journal of Urban Economics 64: 573-589.

McMillen, D.P., and Coulson N.E. 2007. “The Dynamics of Intra-urban Quantile House Price Indexes.” Urban Studies 44: 1517-1537.

Mueller, R. 2000. “Public- and Private-Sector Wage Differentials in Canada Revisited.” Industrial Relations 39: 375-400.

Munneke, Henry J. and V.Carlos Slawson. 1999. “A Housing Price Model with Endogenous Externality Location: A Study of Mobile Home Parks.” Journal of Real Estate Finance and Economics 19: 113-131.

Morales, D.J. 1980. “The Contribution of Trees to Residential Property Value.” J. Arboricol 7: 109- 12

Morales, D.J., Micha, F.R. and Weber. R.L. 1983. “Two Methods of Valuating Trees on Residential Sites.” J. Arboricol 9: 21-24.

Milon. J. Walter, Jonathan Gressel, and David Muikey. 1984. "Hedonic Amenity Valuation and Functional Form Specification." Land Economics 60: 378-87.

McLeod, P.B. 1984. “The demand for local amenity: a hedonic price analysis.” Environ. Plann. 16: 389–400.

Murray, Stephen O (1996), American gay, University of Chicago Press, ISBN 978-0-226-55193-7

Meyer, B.D., Viscusi, W.K., Durbin, D.L., 1995. “Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment.” The American Economic Review 85, 322–340.

Morgan S.L., Todd J.L. “A diagnostic routine for the detection of consequential heterogeneity of causal effects.” Sociological Methodology. 2008; 38:231–281.

Martorell, F. “Do Graduation Exams Matter. A Regression-Discontinuity Analysis of the Impact of Failing the Exit Exam on High School and Post-High School Outcomes.” Manuscript, UC Berkeley, 2004.

Matsudaira, J. “Sinking or Swimming. Evaluation the Impact of English Immersion vs. Bilingual Education on Student Achievement.” Manuscript, University of Michigan, 2004.

Noland M (2005) Popular attitudes, globalization and risk. International Finance 8:199–229.

Nelson, Arthur C, John Genereux, and Michelle G.1992. “Price Effects of Landfills on House Values.” Land Economics 68: 359-65.

Nelson, Jeffrey D. and Laurie F. 2002. “Hedonic Analysis of the Enhancement Value of Georgia's Trees.” Working Paper.

Poterba, James and Kim Rueben. 1995. “The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods.” NBER Working Paper No. 4734.

Powell, J. 1983. “The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators.” Econometrica 51: 1569–75.

Paterson, R., and K. Boyle. 2002. “Out of Sight, Out of Mind? Using GIS to Incorporate Visibility in Hedonic Property Value Models.” Land Economics 78:417–25.

Paelinck, J.H.P. and Klaassen, L.H. 1979. Spatial Econometrics, Saxon House, Farnborough

Powe, Neil A., Guy Garrod, Chris Brunsdon, and Ken Willis. 1997. “Using a Geographic Information System to Estimate a Hedonic Price Model of the Benefits of Woodland Access.” Forestry 70: 139-49.

Palmquist. R. 1991. “Measuring the Demand for Environmental Quality.” Elsevier, North-Holland. pp 77-120.

PORTER, K. E., GRUBER, S., LAAN, M. J. V. D. & SEKHON, J. S. 2011. “The Relative Performance of Targeted Maximum Likelihood Estimators.” The International Journal of Biostatistics, 7.

Porter, J. “Estimation on the Regression Discontinuity Model.” Unpublished manuscript, 2003.

P. C. Austin. (2013) The performance of different propensity score methods for estimating marginal hazard ratios. Stat Med. 2013 Jul 20;32(16):2837-49.

Rosen, S.1974. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.” Journal of Political Economy 82: 34-55.

Robinette, C. 1972. “Plants, People and Environmental Quality.” Department of the Interior. National Fark Service, Washington DC, 40

Rubin D. “Using multivariate matched sampling and regression adjustment to control bias in observational studies.” Journal of the American Statistical Association 1979; 74:318– 324.

Robins J, Rotnitzky A. “Semiparametric efficiency in multivariate regression models with missing data.” Journal of the American Statistical Association 1995; 90 (429):122–129.

Robins J, Rotnitzky A, Zhao L. “Analysis of semiparametric regression models for repeated outcomes in the presence of missing data.” Journal of the American Statistical Association 1995; 90 (429):106–121.

Rosenbaum P.R. “Model-based direct adjustment.” The Journal of the American Statistician. 1987a;82: 387–394.

Rosenbaum P.R., Rubin D.B. “Constructing a control group using multivariate matched sampling methods that incorporate the propensity score.” The American Statistician. 1985;39:33–38.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." Biometrika 70: 41-55.

Robins, J. M. 2000. “Robust estimation in sequentially ignorable missing data and causal inference models.” In Proceedings of the American Statistical Association Section on Bayesian Statistical Science 1999, 6–10. Alexandria, VA: American Statistical Association.

Stephen Gibbons & Stephen Machin, 2008. "Valuing school quality, better transport, and lower crime: evidence from house prices." Oxford Review of Economic Policy, Oxford University Press, vol. 24(1), pages 99-119, spring.

Strand, J., Vagnes, M., 2001. “The relationship between property values and railroad proximity: a study based on hedonic prices and real estate brokers appraisals.” Transportation 28, 137–156.

Sheppard, Stephen, 1999. "Hedonic analysis of housing markets," Handbook of Regional and Urban Economics, in: P. C. Cheshire & E. S. Mills (ed.), Handbook of Regional and Urban Economics, edition 1, volume 3, chapter 41, pages 1595-1635 Elsevier.

Smith N (1996) The New Urban Frontier: Gentrification and the Revanchist City (Routledge, London).

Sascha O. Becker and Andrea Ichino. 2002. “Estimation of average treatment effects based on propensity scores.” The Stata Journal (2002) 2,Number 4, pp. 358-377.

Sun, Yixiao, “Adaptive Estimation of the Regression Discontinuity Model.” working paper, University of California, San Diego, 2005.

Thayer, and Don Waldman.1988. "The Robustness of Hedonic Price Estimation: Urban Air Quality." Land Economics 64: 220-33.

Taylor, L. “The Hedonic Method.” In A Primer on Non-Market Valuation, eds., P. Champ, T. Brown, and K. Boyle, pp. 331-394. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.

Tu, C. 2005. “How Does a New Sports Stadium Affect Housing Values? The Case of FedEx Field.” Land Economics 81(3): 379-395.

Tiefelsdorf M, Griffith D A, Boots B, 1999, "A Variance-Stabilizing Coding Scheme for Spatial Link Matrices" Environment and Planning A 31(1) 165 – 180

Tiefelsdorf, M, Boots, B (1997), "A note on the extremities of local Moran's its and their impact on global Moran's I", Geographical Analysis, Vol. 29 No.3, pp.248-57.

Trochim, W. Research Design for Program Evaluation: The Regression-Discontinuity Approach. Sage Publications, Beverly Hills, 1984.

1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." Journal of Educational Psychology, 51(6): 309–17.

Van der Laan, M. J., and J. M. Robins. 2003. Unified Methods for Censored Longitudinal Data and Causality. New York: Springer.

Viscusi, W. Kip and James Hamilton. 1999. "Are Risk Regulators Rational Evidence from Hazardous Waste Cleanup Decisions." American Economic Review 89: 1010-27.

Wang-Sheng Lee (2013), "Propensity score matching and variations on the balancing test." Empirical Economics, Vol. 44, 47-80.

Zukin S (1995) "The Cultures of Cities" (Blackwell Publishers, Cambridge, MA and Oxford).

Zietz, J., Zietz, E.N. and Sirmans, G.S. 2008. "Determinant of Housing Prices: A Quantile Regression Approach." Journal of Real Estate Finance and Economics 37: 317–333.