2021

The Great Outdoors, Indoors: An Evaluation of Green Spaces and Housing Prices in Pittsburgh, PA

Will McCullough
The College of Wooster, wmccullough21@wooster.edu

The Great Outdoors, Indoors

An Evaluation of Green Spaces and Housing Prices in Pittsburgh, PA

By Will McCullough

Presented in Partial Fulfillment of the
Requirements of Senior Independent Study for the Department of
Economics at the College of Wooster

Supervised by
Dr. Moses Luri
Department of Economics

Spring 2021

Acknowledgments

Immense thanks to Dr. Moses Luri, without whom this IS would be very poorly worded and rife with needlessly complex tables. Additional thanks to the brothers of Xi Chi Psi for supporting me through this process. I am extremely grateful for the advice of Dr. Shelly Judge of the Geology department for allowing me to access GIS software, without which little of my analysis would have been possible. I also want to thank my dog, Bea, for introducing me to the Pittsburgh park system and for lessening the burden of COVID restrictions.

Abstract

Green spaces, or public parks, playgrounds, and athletic fields, are a community hub. Many studies have highlighted the benefits of green space on surrounding neighborhoods. However, comparatively little research has assessed the relationship between park proximity and housing sales price. In Pittsburgh, Pennsylvania, city officials have shown a renewed interest in increasing park access and quality, although no economic literature has examined role of green space in the housing consumption decision in the city. I develop a theory of green space preference in a consumer maximization framework to hypothesize that home prices will increase with proximity to green space. I then use a Spatial Auto-Regressive Moving Average (SARMA) model to estimate the impact of green space proximity on single family home sales prices in Allegheny County, Pennsylvania. My results show that house sales prices will decrease by 2.34% with each additional mile from green space. This indicates a price premium on access to green space, further exacerbating inequalities between high and low socioeconomic status neighborhoods. This research highlights the need for additional research into environmental justice topics in Pittsburgh.

Table of Contents

Table of Figures

Chapter 1: Introduction

Throughout history, as people have moved into larger and larger cities, city planners have emphasized the need to include parks and green spaces into the built environment. Londoners in the 17th century were able to frolic in the Moorfields north of city, allowing an escape from the noise and fumes of city life, while also inspiring colonial Philadelphia to follow suit and add four parks into the layout of the city (Mann, 2016). Allowing people to connect with the natural world and have a public space for recreation has been a central goal of both small-scale urban planners and national governments. In the US, President Woodrow Wilson established the National Park System in 1916, which currently encompasses 443 unique parks across the country (NPS, 2021). With such emphasis on providing parks to the public, it is clear that there is a palpable demand for greenspace.

There is indeed much to enjoy about parks, from the ability to connect with others in a neutral area to being a living classroom for biological and geological phenomenon. The mental and physical health benefits of access to parks are innumerable, including lower risks of childhood asthma and cardiovascular disease (Giles-Corti et al., 2005). With parks providing so many benefits and opportunities, it is likely that they may influence the housing market.

The housing market is one of the largest components of the American economy, consistently accounting for 15-18% of US GDP (NAHB, 2021). Houses are frequently one of the main investments people hold, and have a heavy hand in the broader economy (Stroebel, 2016). This was the case from 2007 to 2009, when the housing prices crashed, sending many peoples' largest investments plummeting and leading to the widespread great recession (Boykin, 2019).

Buying a house is therefore a consequential economic decision, and much research has attempted to measure what individuals' value when it comes time to buy a house. Economists have looked at how people factor in different aspects of houses into their residential decision, from the type of heating system the house uses (Kakejoub et al., 2013), to the tax rate of the community the house is in (Banzhaf & Walsh, 2008). One area that has received less attention is the role of public green space in the individual housing consumption decision.

As the global Coronavirus pandemic has pushed more people to work and learn from home (The Economist, 2020), local public green space has received renewed interest as a space for safe, socially distant gatherings, as well as for exercising and maintaining mental health (Cinderby, 2020). With these many different reasons to enjoy green space, it has become clear that the park down the street may be yielding more value to a homeowner that might have been previously considered.

In addition, as city leaders attempt to pull residents and firms away from other cities in favor of their own, cities have begun investing in increasing and improving their green amenity provisions (Garcia-Lamarca et al., 2019). Many of these new developments have also received attention both for their scope and because of their effect on the surrounding real estate market, as can be seen with the Atlanta Beltline (Immergluck & Balan, 2018), and the New York Highline

(Loughran, 2014). However, comparatively less research has assessed the impact of green space on housing prices in Pittsburgh, Pennsylvania, a former industrial hub that has experienced a notable rebirth as a hotspot for education and medical research (*CBRE Study*, 2017). In an analysis of the cities with the best parks in America by the Trust for Public Land, Pittsburgh ranked 15th in 2020, up from 39th just three years prior (*Park Ranking*, 2020). With so much emphasis on park provision and access from city officials an analysis of the economic effects of green space on the surrounding housing market is necessary (*PGH Parks Rank*, 2020).

This Independent Study thesis is divided into five sections, focused on investigating the hypothesis that housing prices in closer proximity to green space will command a higher price than those farther away. First, I provide background on the rationale for this analysis of Pittsburgh green spaces and housing prices. Next, I develop an economic theory of green space preference in the housing consumption decision using a utility maximization framework with preference heterogeneity. I then discuss the empirical literature, focusing on studies of housing prices that use hedonic price theory to assess the impact of green space on housing price and neighborhoods. This is followed by my own analysis of the effect of green space proximity on housing prices in Allegheny County, Pennsylvania. I conclude the thesis with a discussion of the results of this analysis, including future directions for this research.

Background of Area of Study

I chose Pittsburgh as the case city for this analysis for a number of reasons. As a former industrial hub built into the Pennsylvania hills at the merging of the Monongahela and Allegheny rivers, much of the growth of the metropolitan area has been dictated by geography (*CBRE Study*, 2017). Given the multitude of steep hills, mountains, rivers and ravines, many areas are not suitable to development or are cost prohibitive, leading to landslides and high maintenance costs (Bauder, 2020; Sheehan, 2018). This has led to the growth of distinct neighborhoods with unique identities within the city of Pittsburgh, as well as a burgeoning park system. In addition, public officials have recently begun displaying renewed interest in updating the city's green infrastructure (*PGH Parks Rank,* 2020). This interest has paid off, as demonstrated by Pittsburgh's Park Score ranking, developed by the Trust for Public Land increasing from 39th in the country to 15th over the past 3 years (*Park Ranking*, 2020).

Pittsburgh is located in Allegheny County. While distant exurbs of the Pittsburgh metropolitan area are in neighboring counties, the city itself is completely contained within Allegheny county. Several inner ring suburbs are also located in the county, such as Fox Chapel, Mount Lebanon, and Wilkinsburg. In addition, there are a few post-industrial towns in the area, including Nevel Island, McKeesport, and Clairton. As the second most populous county in Pennsylvania, Allegheny county has a population of 1,216,045. This is a 0.6% decrease in population from the 2010 estimate (*U.S. Census Bureau*, 2021), which is similar to population trends of similar cities located in the Rust Belt region of the US (*World Population Review*, 2021). The county contains an estimated 604,258 housing units, of which a majority, 64.3%, are owned by the resident; The median value of a home is $154,700, just under 75% of the national median home value (*U.S. Census Bureau*, 2021)

However, where other former industrial cities have been on a steady decline due to the loss of manufacturing firms, Pittsburgh has been able to slow the rate of outmigration by reinventing itself as a hub for education and technology in the area once central to the steel industry (Henderson, 2018). Specifically, the city is home to over 29 higher education institutions including the University of Pittsburgh, Carnegie Mellon, and Point Park (*Visit Pittsburgh*, n.d.), in addition to the University of Pittsburgh Medical Center, the largest employer in the county (Hasco, 2018). This influx of highly educated industries and the high paying jobs that accompany them have been beneficial for the local real estate market and the economy in aggregate (*CBRE Study*, 2017), leading to a 24% increase in per capita wages in the city (Henderson, 2018). Broadly, the city is gradually shrinking, and rewarding those that remain.

These gains have not been shared widely. While Pittsburgh continues to grow in global rankings of overall livability (Eberson, 2018; Economist Intelligence Unit, 2019), the city ranks last in livability for black women among US cities (Howell et al., 2019; Mock, 2019). Disparities appear in almost every aspect of life; one in three black women live in poverty and more black children grow up in poverty than in 95% of comparable cities (Dickinson, 2021). This massive disparity has only been exacerbated by recent gentrification trends displacing the few affordable areas of the city. One area where racial inequality could be particularly prevalent is in green space provision, with socioeconomically disadvantaged areas located farther away from parks than wealthier areas. Through examining the proximity effect of green space on housing prices, my analysis aims to demonstrate if this is the case by showing if there is a price premium associated with living close to parks, inadvertently creating a barrier to public good usage.

With this introductory and background information established, I now turn to developing an economic model of green space preference in the housing consumption decision in Chapter 2.

Chapter 2: Theory: Heterogeneity of Green Space Preferences in an Urban Area

Introduction

Cities have been vibrant generators of economic growth but have proven notoriously difficult to measure in an economic framework, so much so that the entire practice of urban economics is devoted to it. An area of research that has received attention in the literature is the urban residential location decision. Urban economists have researched this decision on both the inter-city and the intra-city levels, attempting to determine the factors that pull people to one place over another. In this section, I provide my own theoretical framework to assess the intra-urban location decision in a closed urban area with a heterogeneous green space distribution. I will do this by assessing the predominant theories on the topic and recent advances in these models, ultimately developing a theory and predictions of green spaces' effect on the urban housing decision.

Previous Approaches

The residential housing decision has been modelled in several different ways by economists. At the intra-urban level, these models focus primarily on the microeconomic theory of the utility maximizing individual. Tiebout (1956) examines how residents "vote with their feet" by moving to a municipality, or local governing body, that provided an optimal level of public services respective of the tax rate. In this model, the residential consumption decision is framed as choosing the community that provides the optimal level of social services while balancing other consumption needs. Following Tiebout a decade later, Mills (1967) lays out a model of urban spatial equilibrium that incorporates consumer and producer decisions. Here residents choose their housing location in order to maximize utility while minimizing commute and housing costs. Producers simultaneously aim to minimize their transportation costs. The Mills model theorizes that producers and consumers will satisfy these conditions when all parties are located as close as possible to a central point.

This is further developed by Jan Brueckner (1987) who demonstrates this model's ability to accurately predict the housing density of urban areas in spatial equilibrium, where land use becomes more dense closer to the city center due to the demand from firms and residents to locate in its proximity. More recently, economists have expanded on this model to demonstrate how the housing decision is affected by multiple city centers (Heikkila et al., 1989), and how geographic boundaries influence residential choices (Wu, 2006). It is clear that the common theme among these models is that the urban resident is the figure central to analysis, who can be manipulated in order to examine how different factors will influence the housing decision.

Model Formulation: Definitions

The first step in building my model is defining the variables I will be using in specific terms. Urban planners and economists have used the term green space to refer to many different physical goods, spanning from the size of grass lawns, grassy or tree-lined medians, the amount of tree coverage in an area, to public access parks and forests. While these are all important and could affect the prices of nearby houses, in this study I am focused specifically on the effects of public access green space. Therefore, I define green space in this study as any area of land owned by the municipality that is free to use and open to anyone. This definition of green space allows for examination of public parks, forests and community gardens as public goods, where one person's use of a park does not meaningfully impact others' use of the area. In addition, I assume that the green space designated by the city is fixed, and that all land is either used for green space or for housing development.

It is also important to be specific with the definition of housing consumption. Housing takes many different forms within a city, from the palatial estates of the urban elite to the blocks of single-family homes and apartment buildings. One thing that is unique about the housing market is the wide diversity of living spaces available to a consumer. This array of choices is ultimately beneficial to the consumer because it assumes that each resident will be able to find a house that fits their preferences perfectly. However, this poses barriers to measuring all housing as a unified market. In order to account for the wide diversity of options within the housing market, I define housing as the square footage of residential space being consumed. For example, a large single-family house will simply be defined in the market as a high quantity of residential square feet, compared to a studio apartment which would be a much lower quantity of residential square feet.

Theory of the Consumer

With formal definitions of green space and housing consumption laid out, we begin to develop our theory of the urban housing consumer. Suppose a resident *i* has chosen to live in an urban area and is facing a decision on where to live within the city. Suppose this resident consumes two goods: a housing good, H, and a composite good, C. Further suppose that value of a house to any representative resident is determined by two things: the features (characteristics) of the house *r*, and its proximity to green space *g*. Lastly, suppose that the resident *i*, is a utility maximizer. This second assumption implies that the utility function, and consequently the utility-maximizing housing choice for resident *i*, will both be affected by some degree of substitutability between housing characteristics and proximity to green space. In keeping with these assumptions, I develop a modified Cobb-Douglass function, whose implied assumptions are beneficial to this analysis. This is detailed in Equation 1 for a resident *i* consuming house *j*. I use a modified Cobb-Douglas function in this model because of the algebraic simplicity it provides and the established basis for this functional form in other urban economic theories.

Eq 1

$$U_{ij}(H_{ij}, C_i) = H_{ij}^{\alpha(g,r)} C_i^{\beta}$$

Within the city there are many housing options, *j*, of varying size and quality that the resident must choose from. In addition, each house is built in a unique location, recognized through the houses' distance to a green space, *g*. I assume that apart from distance to green space there is no meaningful difference from one house's location to another. The diversity of choices will also lead to different houses being alluring for different reasons. This is exemplified by α(*g*,*r*), where the weight of preference is determined from an individual's preference for distance to green space, *g*, and the preferences for other characteristics of the house, *r*. This shows that houses at different distances from green space will garner a different level of preference within the utility function. The weight a resident gives to composite consumption is denoted β. This equation shows that different consumers will not all have the same preference sets, and that one circumstance may yield one optimal consumption bundle, while in a different instance a different bundle may lead to optimality for the same individual.

This resident is faced with a budget constraint, as they cannot consume infinite amounts of housing and goods. Buying a home costs money, as does buying any other good. We assume that in order to maximize their level of utility, residents will always consume at a level that exhausts their income. This is described in equation 2:

Eq 2

$$Y_i = p_H H_{ij} + p_c C_i$$

In order to solve for the optimal consumption bundle that will yield maximum utility in relation to this budget constraint, I solve for the objective function:

Eq 3

$$\mathcal{L} = H_{ij}^{\alpha(g,r)} C_i^{\beta} + \lambda(Y_i - p_H H_{ij} + p_c C_i)$$

Deriving in terms of H, C, λ and solving the system of equations, I come to the utility maximizing condition:

Eq 4

$$\frac{p_H}{p_C} = \frac{\alpha(g,r)}{\beta} \times \frac{C_i}{H_{ij}}$$

Which shows that the resident will maximize utility relative to their budget when the opportunity cost of housing is equal to the marginal rate of substitutions. Graphically, this equilibrium condition represents the point of intersection, or tangency, between the resident's budget line and their indifference curve for housing and goods consumption. Any consumption combination where this is not fulfilled will either be unaffordable or not yield as high a level of utility, and the resident would be better served by switching consumption so that this is fulfilled.

Important to this analysis is that the weight of preferences for housing, *a(g,r)*, influences the optimal consumption bundle; when *a* increases, the amount of housing consumed must also increase to maintain the equality. While this is helpful for showing how preferences for housing influence the housing consumption decision, it does not show how this is broken down between preferences for physical housing amenities and for proximity to green space. This requires defining the form of *a(g,r)*. Economic theory demonstrates that proximity to green space and housing amenities are normal goods, so that residents will prefer housing closer to green space than housing farther away. In addition, higher quality housing will be preferred by residents more than lower quality housing. Taking this all into account, consider the following fomulation:

Eq 5

$$a(g,r) = \frac{r}{g}$$

Where the distance to green space, *g*, exhibits a diminishing effect on preference while physical amenities have a constant effect when g is constant. Substituting the right hand of the equation 5 into equation 4, I find equation 6. With this rewritten equation and the budget constraint, I can solve to find the reworked optimal condition in terms of p_h:

Eq 6

$$\mathrm{p}_h^* = \frac{rC_i p_c}{g\beta H_{ij}}$$

Where the price for housing must equal the ratio of the level of house amenity with the cost and quantity of composite consumption over the physical amenity level, housing quantity and composite weight.

Comparative Statics

Taking the above information into account, I now turn to demonstrating the effect of increases in proximity to green space and in quantity of housing consumed on this equilibrium condition using comparative statics. Taking the derivative of Equation 7 with respect to *g*, it is possible to show how a change in the distance to green space effects housing prices, which as seen in equation 8 is negative; however, given that lower values of g correspond to a higher level of green space proximity, this is functionally a positive partial derivative. Furthermore, this shows that houses closer to green space will necessarily increase the equilibrium housing cost.

Eq 7

$$\frac{\partial p_h}{\partial g} = \frac{1}{-g^2}$$

In comparison, increasing the equilibrium quantity of housing will require a decrease in the equilibrium cost, which is shown by taking the derivative of equilibrium price with respect to *H*. Shown in figure 9, it is clear that an increase in the quantity of housing demanded will lower the equilibrium housing price. This comfirms that housing is a normal good, on par with the assumptions of the model.

Eq 8

$$\frac{\partial p_h}{\partial H_{ij}} = \frac{1}{-H_{ij}^2}$$

With this in mind, consider the example of two identical houses, with house *A* close to a park and the other, *B*, far away from a park. In this example, *A* will have a lower level of *g* compared to *B* as the distance to green space is less than that of *B.* Now consider two residents, one with a high preference for proximity to green space and one with a lower preference. The resident with higher preferences for living near a green space will be willing to devote a greater share of their finite resources to fulfill those preferences, paying a higher price for *A* compared to *B*. Simultaneously, the other resident who does not have as great a preference for proximity to green space will not be willing to devote as great a share of income to housing in comparison, and therefore will opt for house *B*. All else constant, the price of house *A* will necessarily be higher than house *B,* in order for the equality to be maintained in both houses.

Housing Market

Extending the individual consumer equilibrium model above to identifying equilibrium in the full housing market. To do this, I assume that the housing supply is continuous with respect to green space distance. This implies that there will be a house on every plot of land between the green space and city boundary. In addition, I assume that there are a large number of residents in the market who display continuous green space preferences, so that every resident has different preferences for proximity to green space. Taking both assumptions into account, and assuming that all housing is made at a constant quality level, this shows that every house in the market will fulfill the optimality condition for each resident in the market. Summing for all residents in the urban area with respect to green space preferences demonstrates the relationship between green space and housing across the market, which is shown in equation 10. When quality of housing is constant, the price of housing in the urban area will decrease as the distance from green space increases.

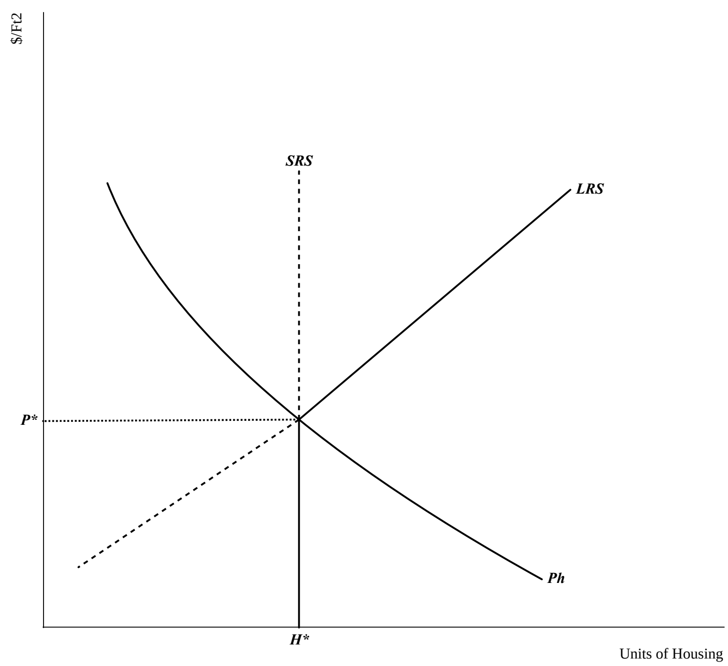
Eq 9

$$P_h = \sum_{g=i} \frac{rC_i p_c}{g\beta H_i}$$

While this is helpful for showing the effect of green space proximity on willingness to pay for a house, it misses out on the decision-making of housing suppliers. To account for this, I add a few assumptions to our model. First, I assume that there is a fixed amount of housing in the market in the short run, because houses cannot be built overnight, but can be increased in the long run.

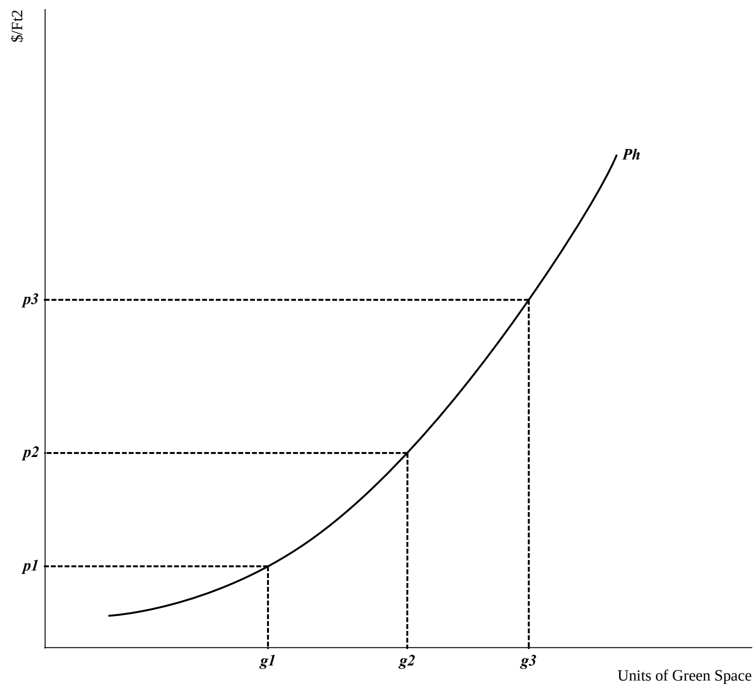
Further, I assume all sellers are profit maximizers working off of the same profit and cost functions. With these assumptions it is clear that in the short run equilibrium will occur where all housing in the urban area is consumed, holding preferences for quality and green space constant. Therefore, housing supply and housing demand will reach equilibrium: every person selling a home will find someone to buy their house for a mutually beneficial price.

Figure 1: Housing Market Demand, Short-Run and Long Run Supply



With this in mind, sellers placing homes on the market near green space will recognize that there are buyers willing to pay a premium for that proximity and will sell the house to the highest bidder, which will coincide with the residential price curve in figure 2. Now supposing that all housing remains fixed while changing the distance to green space as in the residential example, prices will increase with proximity to green space. This allows sellers to recoup the land rent of providing a house near green space. Taken together, differences in consumer preferences will lead to a higher willingness to pay for proximity to green space, which will be mirrored by sellers accepting higher prices for homes closer to green space. This is also consistent with the intuition from my comparative statics analyses.

Figure 2: Market housing price as a function of Green Space



While this is beneficial for theoretically describing the relationship between green space proximity and housing price, it is important to combine both buyer and seller dynamics, which is done with a discussion on Hedonic Modelling.

Hedonic Modelling Theory

The interactions of buyers and sellers in the market can be taken into account with the use of hedonic pricing theory. This is an economic framework which allows for analysis of indirect demand for nonmarket goods. Hedonic pricing was first introduced by environmental economists as a way to measure willingness to pay for nonmarket goods by estimating market goods adjacent to the good in question.

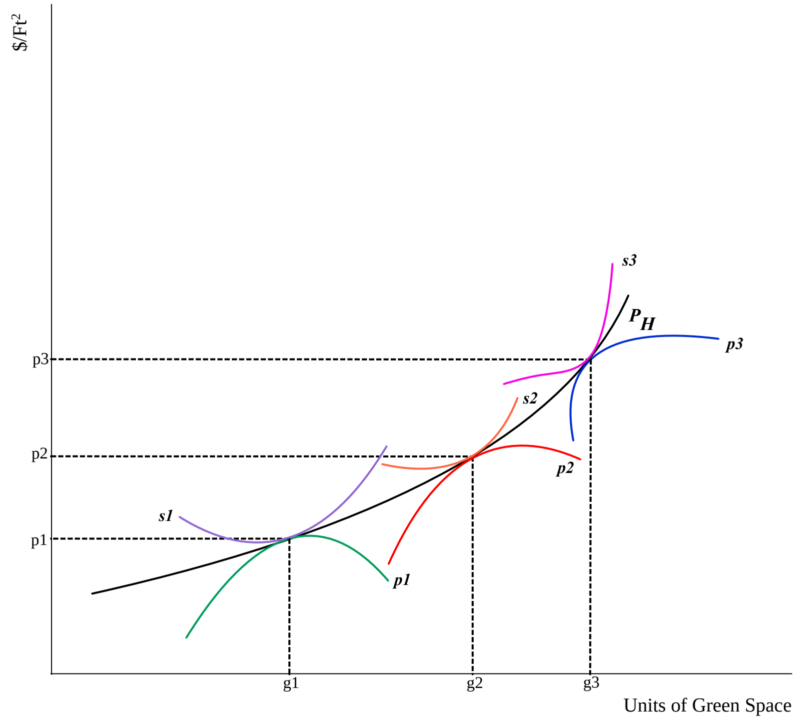
Specifically, original research by Sherwin Rosen (1974) demonstrated that decomposing the market goods into the nonmarket factors that affect the market price allows for analysis of a price schedule for the nonmarket good. In Rosen's explanation of hedonic pricing theory, describes how in a traditional market good, buyers driven by utility maximization and sellers driven by profit maximization, will come to an equilibrium price for a good in an explicit market framework. Hedonic price schedules work similarly, however with no explicit market, prices can be inferred through a more detailed analysis of the consumption and production decisions. By analyzing the indirect utility function of a consumer, it is possible to observe how a change in the level of an indirect attribute can influence the total amount of utility derived from a certain level of income, which ultimately will be reflected in the consumer's willingness to pay for this nonmarket good in a related market. On the producer side, it is possible to analyze how the cost to produce a market good changes with respect to different provisions of a nonmarket good.

In this framework, the price schedule is comprised of a series of simultaneous equilibrium price and quantity combinations for all actors within the market. This allows for examination of how an increase in a nonmarket good is reflected in the equilibrium price of the market good being observed through changes in equilibria in the formal market.

In this case, there is no established market for green space, and therefore no way to directly evaluate willingness to pay for green space. However, the housing market is impacted by a multitude of different indirect markets, including the green space market. Because of this, it is possible to separate out the different components of a house to measure the equilibrium price for everything from the number of windows, number of bedrooms, to the distance to local amenities including green spaces. Isolating these other factors, it is possible to observe the indirect price schedule for green space in terms of housing prices, which in other words is the simultaneous equilibrium of all buyers, *p,* and sellers, *s,* in Figure 3. and sellers in the housing market. For example, when a home buyer is willing to pay more for a house closer to green space they will be willing to pay more while simultaneously a seller will be willing to accept more to sell that house. This price schedule is also consistent with the above analysis of the effect of green space proximity on housing prices.

Using hedonic price modelling to assess valuations of public goods within the housing market has been used extensively in the literature. In Chapter 3, four of the five studies under observation make explicit use of this conceptual framework to underpin their analyses. Specifically, Conway et al. (2010) uses a spatial autocorrelation adaptation of a hedonic pricing model to assess how housing prices change in respect to their distance to green space. They do this by including structural and other environmental variables to isolate the effect of green space. Other studies also use similar hedonic pricing models to determine how green space is valued in the housing market, discussed in detail in Chapter 3.

Figure 3: Hedonic Price Schedule



Conclusion

The theoretical economic construction above outlines a means through which green space can affect housing price. This model performs similarly to other models that have assessed the different factors of the urban housing decision. I hypothesize that when housing quality preferences and the housing quality are held constant across the market, green space preferences will lead to a negative price gradient. This is consistent with the relationship demonstrated in the Muth-Mills bid-rent model, where instead of measuring price in relation to distance from a central business district, it is in relation to a green space (Brueckner, 1987). However, this model comes to this conclusion through different mechanisms than Muth-Mills. Whereas their model is centered around minimizing commuting costs to a CBD and focuses on the budget constraint, this model focuses on the consumer utility function. While the manipulation of preference weight in the utility equation is a relatively novel theory in economics, there is real world basis for the preference heterogeneity, particularly in the housing market, that this model captures. The same reason why someone wants to move into a house may never have occurred to another person. The model is also easily transferrable to empirical analysis. My hypothesis that housing prices will be increase with proximity to green space while holding all else constant can be assessed through hedonic pricing models, as discussed above. The fact that this model already distinguishes between preferences for proximity to green space and other housing quality makes it clear the type of effect an empirical could highlight later on.

Chapter 3: Literature Review

Introduction

In this section, I discuss five articles from the empirical literature that evaluate the relationship between green spaces and housing prices. Each article is presented in five sections: Hypothesis, Conceptual Framework, Empirical Model, Results, and Discussion. This is to provide a clearer understand of how each component of the article is related to the current study.

Article 1

"*Green gentrification or 'just green enough': Do park location, size and function affect whether a place gentrifies or not?"* By: Alessandro Rigolon and Jeremy Németh (2020).

1. Hypothesis:

Equity in access to green space has been a hotly debated issue in environmental justice literature. As more high-profile parks are being built across the country, it is important to take account of what groups are benefitting in this expansion of urban green space. This article argues that green space might not always lead to gentrification when it is built in gentrification eligible areas. Specific park attributes will impact housing prices, including proximity to the city center and the potential to use the park for pedestrian commutes, which could lead to a larger role in the green gentrification relationship. The researchers take these findings and apply them to the theoretical debate that smaller, more evenly distributed parks can have the same positive impact as larger parks but without gentrifying surrounding homes.

2. Conceptual Framework:

The authors' hypothesis is rooted in environmental justice theory, an offshoot of urban economics and sociology. They assume that urban green space is a positive amenity and people will be willing to pay to locate nearby. Specifically, as the quantity of park space, or the size of parks increases, the price to live near it will also increase. This is juxtaposed to the theory that the positive amenity factor of green space comes from the density of the green space and not the raw amount of green area. The authors develop a theory of gentrification in this paper as well, defining gentrification as an influx of new, high income residents into an area made up of previously below average valued homes. For example, a gentrified housing market would be characterized by suddenly high home sales in an area where homes typically would sell for below the city's average price.

3. Empirical model:

The researchers use a "multilevel mixed effects logistic regression" model to estimate the effect of park presence, location, size and function on neighborhood gentrification across 10 cities in America (Rigolon & Németh, 2020). This is similar to how other researchers have modelled multi-city datasets. The authors assess two different time periods, 2000-2008 and 2008-2015, to compare trends before and after the Great Recession.

The regression formula used in this analysis takes into account several controlling variables that have been known to be factors in previous gentrification studies, including variables to account for differences between metropolitan areas, in addition to differences between neighborhoods which are defined here as individual census blocks. These are specified in the following regression equation:

Eq 10

$$\text{Logit}(odds) = B_{00} + B_{10'} * \text{x1}_{ij} + \ldots + B_{N0'} * \text{xN}_{ij} + B_{01'} * \text{X1}_{j} + \ldots + B_{0K'} * \text{XK}_{j} + u_{1j} + \ldots + u_{Nj} + u_{0j}$$

In this equation, B_{00} is the fixed intercept, and $B_{10'}$ – $B_{N0'}$ are the slope coefficients for neighborhood level variables, x1_{ij} – xN_{ij}. The city level variables are denoted by X1_{j} – XK_{j}, with coefficients $B_{01'}$ -- $B_{0K'}$. The common logical thread in this equation is that the neighborhood specific variables are associated with N while city level variables are associated with K. In addition, residual error is represented by several error terms. Each term is linked with one of the neighborhood level variables and an additional term, u_{0j}, to account for error from the city level control variables. The full list of variables is provided in appendix table A1.

Important to the authors' analysis are the variables detailing the presence of a new park, the size of new parks, if the park has pedestrian trails, and if the park is close to downtown. As shown in the regression equation, these predicting variables are used to calculate a prediction of whether a neighborhood has gentrified or not, which is calculated using a combination of dependent variables, including rent and home value.

4. Results

The results indicate that while green spaces increase the likelihood of neighborhoods gentrifying, this is due more to specific park attributes, and was more readily observable in the recent 2008-2015 sample. Overall, 20.3% of the census tracts observed have gentrified over the time period; in the first period 14% of gentrification eligible tracts are within range of a new park, compared to 12.4% in the second period. Specifically, the addition of a new greenway park will increase the likelihood of gentrification by 236%. The distance to downtown will also increase the likelihood a park leading to gentrification; a decrease in distance to the city center increases the likelihood of gentrification by 91%. However, because of this distance to downtown variable, the new park variable also picks up variation in likelihood of gentrification due to distance from the city center, implying a negative sign due to the dis-amenity of living farther from the center city. This is shown in the results, which show the presence of a park reducing the likelihood of gentrification by 44%. The full table of regression results is listed in appendix table A2.

5. Comment

This article is an important contributor to the research on green space and housing prices because of the empirical model it develops to estimate the likelihood that a park would increase housing prices using dummy variables for different park characteristics. This is in addition to the more general neighborhood characteristic variables they use to isolate the effect of green space characteristics, which I will attempt to take into account in a regression analysis. While this article is focused on green gentrification in neighborhoods, the insights gained will serve the closely related study on green spaces' effects on housing prices. By including median home value and median neighborhood rent in their composite dependent variables, this article gives further credence to my study and informs the directionality of my hypothesis. The findings on what specific aspects of parks increase the likelihood of gentrification are helpful in revealing what it is specifically about parks that consumers value in their housing decision.

Article 2

"*A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values"* By: Delores Conway, Christina Q. Li, Jennifer Wolch, Christopher Kahle, Michael Jerrett, 2010

1. Hypothesis:

In this study researchers assess the impact of green space on the housing prices of nearby homes. As a positive amenity, people will value being closer to a park or green space more than being farther away and will be reflected in higher housing prices closer to parks. The hypothesis points to positive willingness to pay for proximity to green space.

2. Conceptual Framework:

Researchers use a hedonic pricing model framework in this analysis, using goods with established markets, in this case the real estate market, to infer a hypothetical market valuation of a nonmarket good, such as green space. This takes into account urban economic spatial theory, which is used to evaluate where people decide to live based on spatial constraints.

3. Empirical Model:

Two empirical models are used: a standard OLS regression and a "linear regression model with a spatial autoregressive disturbance," which produces more accurate results (Conway et al., 2010). Taking home sales data and overlaying this with spatial data on green space and other positive amenities in GIS software, the researchers constructed a dataset comprising both home characteristics and information related to the surrounding neighborhood.

a. Standard Hedonic Pricing Model:

A hedonic pricing model is used to infer the value of green space from the surrounding housing market. This is done by conceptualizing the price of a house as the summation of the prices of the attributes of the home. Regression analysis is perfect for observing hedonic prices in the housing market because of its distinct ability to isolate the effects of several variables simultaneously. As in this hedonic model, physical and environmental attributes can be assessed as contributing to the total price of the house. This is detailed in the following model:

Eq 11

$$\ln\left(Y_{i,t}\right) = \beta_0 + \beta_1 \ln(Area) + \beta_2 \ln(LotSize) + \beta_3(Age) + \beta_4(Age)^2 + \beta_5\,(MedInc) + \beta_6 \ln(RampDist) + \beta_7 \ln(RecDist) + \sum_{t=1}^{T} \delta_t Q_t + \sum_{k=1}^{K} \lambda_k \ln G_k + \varepsilon_{i,t}$$

In this equation, the percentage change in housing price, ln(y), is dependent on the percentage change in house and lot size, the age of the house (nonlinearly), median neighborhood income, distance from the nearest highway exchange, and distance to the nearest outdoor park. In addition to these standard regression variables are summation expressions to account for variation between sale dates, Q_t, as well as a volume measure of green space, G_k, which allows for assessing the value of larger green spaces separately from proximity to them.

b. Spatial Model:

While the standard hedonic pricing approach has empirical validity and has been used in previous analyses, the authors note that it is prone to experiencing "spatial autocorrelation," where observations can be predicted by the values of observations surrounding it. This leads to biased results and a violation of the core error term assumptions, specifically causing the model to behave as though there are omitted variables. Including spatial effects solves the issue.

The proposed fix is to use a regression technique known as a "linear regression model with a spatial autoregressive disturbance," which is established in the following equation:

Eq 12

$$Y = X\beta + \lambda W\varepsilon + v$$

In this form, dependent variables are condensed and represented as a single unit, X, with fixed coefficient β. The spatial element in this model is represented by $\lambda W\varepsilon$ where W is a spatial weight accounting for trends in surrounding observations with coefficient λ. The error term, v, is also corrected in this model to account for spatial trends.

4. Results:

Results are provided for both the standard hedonic model as well as the spatial lag model and are provided in appendix table A3.

The results of the hedonic model indicate that the controlling variables are significant predictors of home price and are important to the model. The explanatory variables relevant to the question of green space effecting housing prices are also significant predictors of housing cost, indicating that a 1% increase in the distance to a park will decrease housing sale price by 0.13. The variables looking at the amount of park area within range of the property at different distances are also significant except for the furthest category, which indicate that park space will increase housing costs more when they are very close with a diminishing effect.

The results of the spatial model indicate that the initial results may not be as robust as initially thought. The coefficient accounting for spatial autocorrelation is significant, indicating that this spatial model is necessary. The green space variables here show that the effect could be overestimated in the hedonic model because the coefficients are slightly greater and have higher p-values, showing a potentially less robust result compared to the spatial model.

5. Comment:

This article is important for this analysis because it looks at the effect of green space on housing prices within a single metropolitan area, which is similar to what my analysis is attempting. This gives a great model to extend in terms of how to create a model and which data points will be necessary to accurately run an empirical model and achieve unbiased results. Of particular help in this paper is the discussion on the effectiveness of applying a spatial lag to the standard hedonic pricing model to more accurately specify the model. The methodology of matching individual housing data to a map of green space and using GIS software to calculate specific amounts of green space within certain vicinities of each house for a cohesive data set is also informative for building a dataset from the available resources.

Article 3

"*Hedonic pricing and different urban green space types and sizes: Insights into the discussion on valuing ecosystem services"* By: Piotr Czembrowski and Jakub Kronenberg, 2015

1. Hypothesis:

The researchers theorize that different types of parks will be valued differently in the housing market. They hypothesize that proximity to larger parks will be valued greater than proximity to smaller parks, and that people will not value living near cemeteries and community gardens. This will all be reflected in housing price variations.

2. Conceptual Framework:

The authors use the hedonic pricing model in this study, which estimates the willingness to pay for nonmarket goods. Embedded in this is consumer price theory, which assumes that people will be willing to pay more for a greater amount of a good which they receive utility from. In this case, green space would be considered a normal good because people enjoy and receive utility from being in a park, which can occur more frequently if they live near a park.

3. Empirical Model:

In developing these concepts, the authors develop an empirical model for estimating the willingness to pay for different levels of green space based on the surrounding housing prices. These models incorporate information on the structural elements, the environmental elements, and the locational elements. In this model, proximity to green space would be considered a locational element, since individuals value living close to positive amenities and away from dis-amenities. Structural elements encompass the physical parts of the house, such as the number of bedrooms, bathrooms, windows, floors and the age of the house. Environmental attributes can be thought of as amenities without a specific location, the most common examples being air, water, and noise pollution. These are represented in the following empirical model:

Eq 13

$$P = \alpha S + \beta E + \gamma L + \varepsilon$$

In this, the price of a house, *P*, is the sum of structural elements, *S*, environmental elements, *E*, and locational elements, *L*, with respective weights α,β, and γ. Each element in the model is a vector of variables, meaning that they are comprised of multiple variables which data is available for. Importantly, the location vector contains the dependent variables measuring distance to different types and sizes of green space. These are broken down into forests, parks, cemeteries, and “allotment” gardens, with small, medium, and large size distinctions for forests and parks. The dependent variables and other controlling variables are listed in appendix table A4.

As mentioned in other studies using regression to analyze housing data, prices are interrelated, with the sale price of one house affecting the sales of surrounding houses, known as spatial autocorrelation. In order to account for this, the authors apply the variables in a “spatial autoregressive model with a spatial autoregressive disturbance” which accounts for correlations between the observations and those in the immediate vicinity. In addition, the authors also incorporate fixed effects to account for changes in home prices due to immeasurable aspects, such as the “character” of the neighborhood the house is in.

This model is run on a sample of apartment sale prices in the Polish city of Lodz, which is overlayed with green space data and distance calculations derived from a Google Maps API. The authors use apartments instead of houses in their analysis because that is the prevailing housing option for the city.

4. Results:

The researchers run two analyses, one without the fixed effects variables and one with fixed effects included. The results indicate there are some disparities between the two analyses, which confirm the importance of including these variables. Only half of the included green space variables are significant, partially confirming the study hypotheses. The theory that proximity to green space will increase house values in all cases was not confirmed, as some green spaces had no measurable effect. However, the green spaces that do have an effect increase housing prices as proximity to them increases, except for cemeteries. For example, the strongest effects came from the largest forest and other large parks, as well as small neighborhood parks. The authors theorize there are two reasons for this: people value the name recognition of living near a well-known park and being near a small open area for easy access to an outdoor venue. The full results of the analysis are listed in appendix table A5.

What is also revealing are the fixed effects variables added in the second model, which are all significant. This indicates that home prices are dependent not only on the houses immediately surrounding them but are also influenced by the broader neighborhood association.

5. Comment:

This article is beneficial to my study because it shows the relationship between green space and housing prices, and the level of specificity required in the data and empirical model necessary to achieve accurate results. The explanation on why the authors chose the spatial autoregressive model was helpful in informing what specificity tests will be necessary to confirm our model is an accurate fit of the data. In addition, their discussion of using the Google Maps API to determine the walking distance from an apartment to specific green spaces was informative for constructing the data set on Pittsburgh. The findings that green space may be valued for the name recognition in addition to the other benefits brought by living in a park will be especially insightful when assessing Pittsburgh's park system, which is centered around 4 major parks well known throughout the city and smaller neighborhood parks for more specific uses. The city of Lodz presents a helpful parallel to Pittsburgh in that the park systems are both similarly configured and the existence of several small neighborhoods with distinct identities within the broader Pittsburgh umbrella.

Article 4

"*Can proximity to urban green spaces be considered a luxury? Classifying a non-tradable good with the use of the hedonic pricing method"* By: Edyta Łaszkiewicz, Piotr Czembrowski, and Jakub Kronenberg, 2019

1. Hypothesis:

The researchers hypothesize that the demand for proximity to green space will disproportionally increase as income increases, and therefore be classified as a luxury good. People with higher incomes will be disproportionally willing to pay for proximity to green space compared to those with lower incomes.

2. Conceptual Framework:

This paper uses hedonic pricing theory to estimate the impact of green space on housing prices. They then look at the income elasticity of the marginal willingness to pay for proximity to green space, a concept based in microeconomic consumption theory.

3. Empirical Model:

There are three main empirical models used in this paper, which build off each other to determine the income elasticity of proximity to green space: a spatial autoregressive model to estimate the hedonic price structure of housing prices, a generalized additive model to estimate heterogeneity in the effect from green space, and a spatial quantile autoregressive model to estimate how the coefficients differ on different subsets of the apartment price distribution. These analyses were completed on a data set of apartment sales in Lodz, Poland.

a. Hedonic Pricing Model:

Hedonic pricing models are used to estimate the value of nontraded goods from the price of goods for which there is a market by accounting for all of the factors that influence the price of that good. This is helpful for valuing greenspaces because there is no defined market for them, but they are frequently located near residential real estate, for which there is a clearly defined market. By looking at the effect of green space on housing prices it is possible to infer the marginal willingness to pay for green space. The authors use a spatial autoregressive model to estimate the hedonic pricing model while accounting for spatial autocorrelation due to the effect of a home sale on the sales of surrounding homes. The model is specified in the following equation:

Eq 14

$$\mathrm{P} = \rho \mathrm{WP} + \mathrm{S}\alpha + \mathrm{E}\beta + \mathrm{L}\gamma + \varepsilon$$

In this model, housing prices, P, are dependent on a spatial weight matrix, W, structural attributes, S, environmental attributes, E, and locational attributes, L, with respective coefficients ρ, α, β, and γ. These attributes are matrices of explanatory variables, grouped by their relationship to housing costs. In this model, the distance to green space is categorized as an environmental attribute. The combined ρWP expression explains how housing prices are dependent on the sale prices of houses immediately surrounding it, incorporating the spatial element into the model.

b. Green Space Classification Model:

One caveat of the hedonic pricing model is that it has difficulty capturing complex nonlinear relationships, which a general additive model is adept at. In the previous model, green space proximity was measured as having a logarithmic relationship with housing prices, although the true shape of the relationship may not closely follow a geometric curve. In doing this the model provides estimates of the impact of specific green spaces, separating out those parks that effect the surrounding housing prices from those that do not. The generalized additive model is specified as follows:

Eq 15

$$\mathrm{P} = \mathrm{S}\alpha + \mathrm{E}_1\beta + f(\mathrm{E}_2) + \mathrm{L}_1\gamma + f(\mathrm{L}_2) + f(\mathrm{Long} \cdot \mathrm{Lat}) + \varepsilon$$

Where P, S, E, and L represent the same attributes as in the first model. This model uses the geographic coordinates of observations to account for spatial trends in the data where the initial model used a spatial weight, which allows for more flexibility in analyzing the nonlinear relationship. Also of note in this model are the unknown functional forms, $f\,()$, that are specified for environmental and locational variables, and the spatial trend term, which represents the unspecified nonlinear relationship this model intends to reveal.

c. Spatial Quantile Autoregressive Model:

The first two models estimate the marginal willingness to pay for proximity to green space on housing prices, while this third model assesses the income elasticity of the marginal willingness to pay for proximity. This is accomplished by assuming the real estate market is highly income elastic, meaning that higher house prices will be reflective of a higher income level. By segmenting the housing market into price deciles, researchers are able to analyze how the effect of green space changes as inferred income increases. The authors use a spatial quantile autoregressive model to do this, which is specified very similarly to the original hedonic model:

Eq 16

$$P(\tau) = \rho(\tau)WP + S\alpha(\tau) + E\beta(\tau) + L\gamma(\tau) + \varepsilon$$

With all the same forms and relationships as before, but with the inclusion of τ, which represents the decile of housing price. By estimating the effect on each subsample it is possible to calculate how a change in income will effect the willingness to pay for proximity to green space without manipulating individual's incomes or observing a change in income over time.

4. Results:

These models taken together provide a nuanced view of how the value of green space is reflected in the housing market, as well as how some homeowners value living closer to green space more than others. The hedonic pricing model demonstrates that people do value living closer to parks and forests, with home values decreasing as they move farther from these green amenities, although the effect of living near a park is more pronounced than living near a forest. The spatial weights measure used was also significant, confirming the need to include a spatial element in the model. The full results are listed in appendix table A6.

The results from the general additive model indicate that not all green spaces have an equal effect on housing prices, and some have no impact at all. Specifically, out of the 107 parks in the study, only 28 demonstrated an amenity effect on surrounding real estate. Only 6 out of the 32 forests showed this relationship.

Having isolated the parks which are valued as environmental amenities, the spatial quantile model shows the effect which proximity to green space has on housing prices for different segments of the housing price distribution.

Specifically, as the price of apartment increases, the effect of green space becomes more pronounced. The effect on price is significant for all price deciles except for the lowest decile, with an increasing effect in each decile, which is detailed graphically in appendix figure A2. This indicates that the marginal willingness to pay for proximity to green space is not equal across the estimated income distribution.

5. Comment:

This article is important to my analysis by more precisely detailing the relationship between green space and housing prices. By showing that green space proximity increases housing prices and going further to show that green space increases housing prices more for higher value houses, the authors provide a basis for further distributional assessment of hedonic pricing modelling. Their method of inferring an income distribution from the distribution of housing prices was an insightful means of analyzing income elasticities from the available empirical data. This helps my study by going into the unequal relationship of the green space effect.

Article 5

"*Green urban development, environmental gentrification, and the Atlanta Beltline"* By: Dan Immergluck and Tharunya Balan, 2017

1. Hypothesis:

The authors hypothesize that the Atlanta Beltway, a major greenspace development initiative underway in Atlanta, Georgia, will increase the value of homes within a half mile radius of completed sections.

2. Conceptual Framework:

The analysis in this paper uses the hedonic pricing framework which is based in microeconomic consumer utility theory. Hedonic pricing looks at the price of a market good as the sum of the nonmarket attributes that go along with the good. When looking at the housing market, this means that house prices are influenced by the amenities surrounding them, including parks and green spaces.

3. Empirical Model:

The authors use a standard hedonic price model to determine the impact of the Beltline on surrounding home prices. As discussed in other papers using this model, this incorporates structural, environmental, locational, and neighborhood characteristics. These are detailed in the following equation:

Eq 17

$$\ln(\mathrm{P_i}) = \alpha + \beta \mathrm{H_i} + \zeta \mathrm{N_i} + \delta L_i + \theta \mathrm{Q_i} + \kappa \mathrm{T_i} + \phi \mathrm{B_i} + \gamma \mathrm{T_i} * \mathrm{B_i} + \varepsilon_i$$

Where H is a combination of house variables, N is a combination of neighborhood demographics, L is a combination of locational variables, Q represents control variables for the quarter the sale occurred, and T controls for the year of sale, and

B is a dummy variable for whether the house is within a half mile of a section of the Beltline. Of note is the expression, $\gamma T_i * B_i$, which shows that an interaction between distance to the Beltline and year is expected in the data. This is because of the specific situation of the Beltline, which is not intended to be fully completed until 2030, so as more portions reach completion the higher the expected impact on surrounding homes.

The data for this model is made up of housing sales data in the period 2011-2015 from the DeKalb and Fullerton county assessor's offices, as well as demographic data from the American Community Survey. Data on distance from major employment centers and the Beltline were calculated by the authors.

4. Results:

The results broadly indicate that housing saw increases greater than would otherwise be expected closer to the Beltline. This is shown in the time/beltline interaction variables, which are significant and demonstrate that houses near the Beltline are appreciating faster than houses not near the Beltline. However, the size of the effect is different for different sections of the Beltline; for example, houses near the northwest section see an increase of 21.5%, houses near the northeast section increase 17.9%, houses near the southeast section increase 19.2%, and houses near the southwest section increase 26.6%.

5. Comment:

This article is helpful for our analysis by showing an alternative measure of green spaces' effect on housing prices from a hedonic pricing model. Whereas other analyses using a hedonic pricing model with a dummy experimental variable (i.e. within range of green space or not) have only looked at one time period with normalized prices, this study operationalizes the time component to make inferences about how green space effects prices over time. This is particularly important from a policy perspective, as houses increasing in price more than expected could have an outsize impact on the real estate tax burden being levied on residents, leading to an unaffordable neighborhood for some. However, this analysis could potentially be biased because the empirical model made no attempt to account for spatial correlation, which other studies show is frequently present in hedonic analyses of housing data.

Summary of Literature Review

The articles analyzed above provide a compelling basis with which I can empirically test my model. While all approaching the issue of the effect of green space on surrounding housing prices, each article uses a slightly different approach. Taken together, these articles offer a great deal of insight into what relationships to look for and what issues may arise in the data. In three out of the five studies, researchers correct their model to account for spatial autocorrelation, which will likely be present in the Pittsburgh dataset. Whereas three articles looked strictly at the effect of green space on housing prices, the other two extrapolated from this analysis to assess how green space could lead to gentrification, and the policy implications of such an outcome. Four out of the five studies look at housing prices in only a single city, whereas the fifth looks at a multi-city dataset.

The common thread through all five articles was the universal use of hedonic pricing models to estimate the willingness to pay for green space. While used in slightly different ways, the commonality of inferring green space value from the surrounding housing values was useful for seeing the potential of the method. These methods fall into two main groups of analysis: dichotomous, where a house is either within the boundary of the park or outside of the boundary, or continuous, where house prices are shown to be affected by green space as distance decreases. The dichotomous studies, while providing less detailed results about the effect of proximity, did not have biased results. This was not the case in the continuous studies, which all needed to account for serial autocorrelation, and spatial correlation. The researchers all chose different ways of correcting this; Conway et al. (2010) employs a spatial autoregressive disturbance factor in their model, which Czembrowski & Kronenberg (2016) also uses with the addition of fixed effects controls for larger neighborhood level trend. Rigolon & Németh (2020) instead uses a multilevel, mixed effects model to account for trends between adjacent neighborhoods, whereas Łaszkiewicz et al. (2019) uses a generalized additive model in their analysis of green space amenity effects.

The findings from all five studies confirm that green space has a positive effect on housing prices. Each study provides a unique example of how this occurs and adds insights for replication in other cities. The earliest article researched (Conway et al., 2010) provides the confirmation that this empirical framework used on a single city can provide reliable results. Rigolon & Németh (2020) found that green spaces increase the median housing prices of surrounding neighborhoods, but specifically when these neighborhoods are also close to the city center. Czembrowski & Kronenberg (2016) and Łaszkiewicz et al., (2019) taken together show that the positive influence on housing prices is not only observable in single family American households, but also in multi-family apartments in the European housing market, while alluding to the income elasticity of proximity to green space influencing the green space consumption decision. Finally, Immergluck & Balan (2018) demonstrated a method of showing how green space affects housing prices over time in a single city, finding that housing prices increase at a faster rate within walking distance to a new park than otherwise. With a solid understanding of the previous empirical research, it is possible to turn to the study of Pittsburgh green spaces and housing prices.

Chapter 4: Empirical Evaluation: A Case Study of Housing Prices in Allegheny County, Pennsylvania

Introduction

Having assessed the empirical literature on green space and housing prices, I present an analysis of the data. To do so, I define the variables and specify a spatial regression analysis of home prices in Allegheny County, Pennsylvania. I hypothesize that sales prices of houses will be greater nearest to green space and decrease with distance. I establish the empirical model used in this analysis, discuss the data and the variables, and finally present the results with a discussion of the implications of this analysis. I conclude the chapter with a discussion on the limitations and future directions for this research.

Empirical model: Spatial Autoregression

In order to estimate the effect of green space proximity on housing prices, it is necessary to isolate this from other influences on housing price. Typically, the most efficient way to do this statistically is with an Ordinary Least Squares (OLS) regression estimation, shown generally in Equation 18. This type of regression accounts for the amount of variance which each individual explanatory variable has on the dependent variable. OLS is able to do this by assuming the error term is "well-behaved." The assumption of a well-behaved error term implies it fulfills four conditions. The first condition is that the model itself is correctly specified, which ensures the model is not prone to endogeneity, or that there is high correlation between one or more explanatory variables and the error term, and the variables are not independent. The second assumption is that the variance of the error term is equal across observations, or that the model residuals are homoscedastic. Third is that all explanatory variables have a linear relationship with the dependent variable which is done through transforming variables with a nonlinear relationship. Finally, the error term observations must also follow the normal distribution pattern. If these conditions are met, OLS is sufficient for engaging in econometric analysis.

Eq 18

$$Standard\ OLS\ Model: Y = \beta_0 + \beta_n x_n + \varepsilon$$

This is not the case in this study, however. Given the spatial nature of individual housing sales data, there is significant reason to suggest that one or more of these conditions are not completely met and must be accounted for in this statistical analysis. As evidenced from the empirical literature analysis in chapter 3, the prevailing issue is the presence of spatial autocorrelation within the data, which occurs when observations are correlated with those in close spatial proximity to one another or are spatially dependent on each other.

There are a few ways to correct this. Depending on the nature of the spatial dependence, it can fall into two main groups: substantive and nuisance dependency. Substantive spatial dependency refers to expected relationships between observations located near each other, which in the case of this housing price analysis would be that houses located within the same neighborhood would have more in common in both price and structural characteristics than two houses in different neighborhoods. Nuisance dependency is spatial correlation that cannot effectively be placed onto one or more specific variable in the model. For example, nuisance spatial dependency in this analysis could stem from an angry neighbor who threatens unspeakable things on neighbors if they do not properly care for their lawns. This would likely result in lower prices for surrounding homes but is not attributable to a specific variable.

Spatial dependency necessitates a different econometric specification than the typical OLS. Where OLS is a linear estimator of explanatory variables, n, on a dependent variable, as shown in Equation 1, researchers instead account for the autocorrelative nature of spatial dependency in a Spatial Auto-Regressive Moving Average SARMA model which is shown generically in Equation 19. This model differs from OLS with the addition of 2 specific terms, a spatial lag variable, $\rho_{n+1}W\ y,$ included to account for substantive dependency, and a spatial error term, $\lambda W\varepsilon + u$ for nuisance dependency.

Eq 19

$$SARMA\ Model: Y = \beta_0 + \beta_n x_n + \rho_{n+1}W\ y + \lambda W\varepsilon + \ u$$

Spatial dependency terms are computed from a spatial weights matrix which assigns a nonzero value to observations within a specified range of one another and a zero value to observations outside of the range. For this analysis, the spatial weights matrix is calculated using an arc distance of 0.1 Miles, or roughly the length of a standard city block. This was done with the assumption that houses will be most similar to those on the same block. Specifically, they will likely have similar structure and be affected in similar ways by the surrounding environmental amenities such as school districts and business accessibility.

Spatial Diagnostics

With all of this taken into account, it is clear OLS is insufficient for an accurate analysis of the data. However, with many variations in configuring spatial autoregressive models, it is necessary to examine the results of spatial diagnostic tests to determine which configuration is best for this analysis.

Tests on spatial dependency were run on a baseline OLS model using the dataset, discussed in the next section, to determine the specific model corrections necessary for this analysis. Moran's I test of spatial dependency assesses if the assumptions of spatial dependency in the spatial weights matrix are accurate. This was calculated and found to be significant, indicating the presence of spatial autocorrelation. Research has shown that this is a dependable, but nonspecific, test of spatial dependency.

Table 1: Spatial Dependency Tests	Value	p-Value
Moran's I (error)	61.51	0.00*
Lagrange Multiplier (lag)	258.05	0.00*
Robust LM (lag)	112.18	0.00*
Lagrange Multiplier (error)	3779.18	0.00*
Robust LM (error)	3633.31	0.00*
Lagrange Multiplier (SARMA)	3891.36	0.00*

Note: * indicates significance at $\alpha < 0.01$** indicates significance at $\alpha < 0.05$, *** indicates significance at $\alpha < 0.1$.

To determine the spatial autoregressive form most suited to the data, more specific Lagrange Multiplier (LM) are used to test whether including a spatial lag variable, a spatially weighted error term, or both, are most suitable to the data. LM tests for spatial dependency were calculated including robust standard errors and normal standard errors. Typically, these tests indicate if one spatial autocorrelation correction method is preferred over the other. However, in both instances the results of these tests are highly significant. Because of this, the significance of the LM SARMA test is the most informative to our specification decision. Specifically, the LM SARMA test evaluates the effectiveness of a SARMA model to account for spatial dependency through both a spatial error and lag term, and it's significance shows that this is the most suitable empirical model for the data. These diagnostics confirm the necessity of including spatial controls in this analysis. With this in mind, the full empirical specification is:

Eq 20

$$\ln(Price) = \beta_0 + \beta_1 \ln(HouseSize) + \beta_2 \ln(LotSize) + \beta_3 NumBedrooms + \beta_4 NumBaths + \beta_5 HouseAge + \beta_6 HouseAge^2 + \beta_7 Numfloors + \beta_9 GarageSize + \beta_{10} NumFire + \beta_8 \,(GreenDist) + \beta_{n+1} \text{SpatialLag} + \lambda + \varepsilon$$

This specification includes all the explanatory variables that are included in this analysis. The source of these data points and additional explanation on the importance of including these variables are detailed in the following sections.

Data

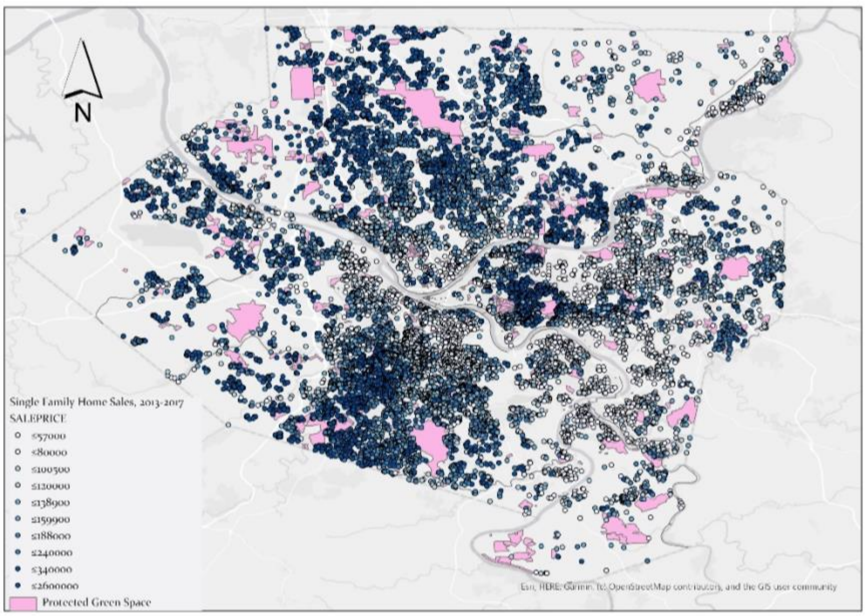
Data comes from the Western Pennsylvania Regional Data Center (WPRDC), which is a free data service provided and maintained by the University of Pittsburgh Center for Urban and Social Research with help from Allegheny County and the City of Pittsburgh (*About • WPRDC*, 2020). Two datasets were obtained from the WPRDC. The first data set consisted of all digitized property records for land parcels in Allegheny County. For the purposes of this analysis, I filtered the sample to include parcels which had a "valid" sale code.

Observations were filtered further to include only parcels with a land use code of "010," indicating the parcel contains only a single-family home. Single family homes are focused on in this study because of both the abundance of this land use code in Allegheny County and because comparing the sale prices of single-family, multi-family, and apartment style housing adds significant variance to our model that the controlling variables cannot account for alone. This is distinction is also consistent with the economic literature on housing studies, which use single-family zoning as the de-facto unit of observation for residential decision studies, due to the relative consistency of single-family homes across the US.

To reduce variance from changes in prices over time, only houses with a sale date in the most recent 5 years available were examined. For this data, 2013 through 2017 were the most recent years available. This time frame was selected for two reasons. First, it aligns with the introduction of Pittsburgh's green space reinvestment program that emphasized green space equity or ensuring that access to green space is not only available to the wealthier areas of the city. By focusing on this period, it is possible to examine whether the "green premium" of purchasing a home near a park was present in the housing market, which could be a factor in green space inequalities. In addition, this period was marked by relatively low inflation and broadly stable prices, further reducing the need for time controls in this analysis.

Data cleaning uncovered 6 observations that needed to be removed from the sample, one for having a sale price of $0, and 5 because they did not have a bedroom. In total, this dataset of valid, single family home sales in Allegheny County between 2013 and 2017 contained 16,590 observations. Sale prices of houses and their location are shown in Figure 4.

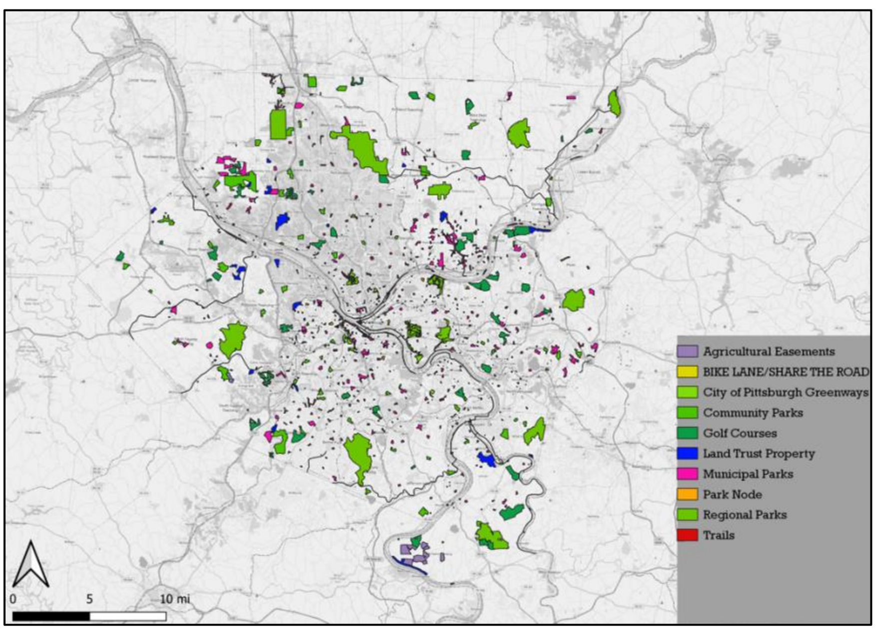
Figure 4: Housing Price by Decile and Green Space, Allegheny County



The second dataset was of green space in Allegheny County, as measured by the Allegheny County Land Trust. This dataset was used to calculate the distance to parks variable as well as to construct maps visualizing house and park distributions across Allegheny County. Apart from the geographic shapes of each park, objects included attributes relating to the status of the area, which was either protected or not protected, and the designation of the land use. Examples of these designations include “municipal parks,” “city park,” or “golf courses.” Golf courses were not removed from the data because they share many of the recreational amenities and scenic amenities of other parks.

This data came in the form of a complete GeoPackage of all greenspace in Allegheny county, not exclusively public parks. I removed portions of Greenspace listed as unprotected land and those listed as bikeways or biking trails, because unprotected land by definition is not a public park, and contained green features such as hillsides, greenways, and bodies of water. I omitted one bike trail because it covered the banks of all three rivers in the city. I removed a second object listed as “share the road” because it was also not a park. There were several thousand unique geographic shapes, as many larger parks were made up of smaller shapes because of roads or other structures making them discontinuous. These distinctions are shown in Figure 5.

Figure 5: Green Spaces by Type, Allegheny County



Variables

I analyze the effect of green space on housing sales prices for single family homes. In addition, I include several structural variables to account for house size, capacity, and other amenities, which have been shown in previous studies to influence housing sale prices through hedonic pricing theory. This is a method of measuring home prices as the sum of many smaller elements, including both the physical components the make up the structure itself and other environmental aspects of the area around the house. This model includes structural independent variables to account for these influences on housing price.

Specifically, I include house size, measured as the square footage of livable area within the house. House size is important to include because larger houses, or houses with more living space, will command higher prices at sale, given the higher cost of building a bigger house and the added benefits with more personal space. House size is expected to have a positive relationship with sale price.

I also include number of bedrooms in the model. While a house with several bedrooms will likely have a greater living area, the number of bedrooms is important to include additionally because of the specific utility that bedrooms bring to a home. Additional bedrooms can allow more people to occupy a house, or an extra bed can be converted into a room with a specific use such as a home office or a workout area. Number of bedrooms is expected to have a positive sign associated with it.

I also include number of bathrooms into the model, which is calculated by adding the number of full bathrooms and half bathrooms. The number of bathrooms specifically are important to account for in the model because of intricacies of building plumbing, tiling, toilets, and sinks. More bathrooms will inherently lead to a higher sticker price for a home. I include the number of floors of a house in the model because it is more expensive to build structurally sound houses that are taller, leading to a higher total cost for the house.

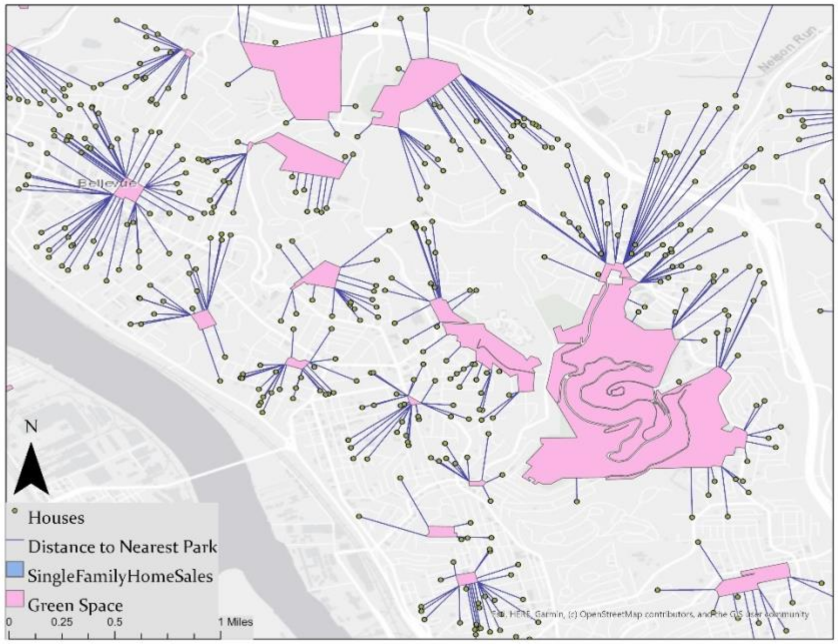
House age is included in the model because older houses will be more expensive to maintain over time compared to a newly built house, and therefore will depress the selling price of a house. However, some old houses may retain their value for nostalgic or historical significance, which further highlights the importance of its inclusion in the model.

I also include the size of a garage in the model, which is calculated as the number of cars the garage is capable of housing. Houses without a garage are indicated by having a zero-car garage. Garage influences house price because of the specific utility a garage brings to house, for storage and protection from the elements of things that require a large area of space. In addition, it is likely cheaper to build houses without a garage, as they require additional materials and construction time.

I also included number of fireplaces into the model because of the utility they contribute to a house and the complexity required in constructing them. Specifically, fireplaces contribute to the aesthetic value of rooms in a house while acting as a piece of furniture that compels people to gather around it, increasing the appeal of a house. With these variables included to control for the many structural components of a house, it is possible to isolate the effect of the variable of interest in the estimation.

I am estimating the effect of proximity to green space on housing price in this model with the calculated direct distance from houses to the park. This is similar to the estimation strategy employed in Conway et al. (2010). The WPRDC dataset of single-family Allegheny county homes did not include explicit geospatial data. Explicit geographic coordinates were calculated by geocoding street addresses using Geocode.io, an online bulk geocoding service. The geocoded data was run through ArcGIS Pro, a geographic information systems software. After combing housing a park data in ArcGIS, I then calculated the minimum distance from a house to the perimeter of park objects using a Near Analysis, producing the distance in miles to the nearest park. An example of this, focused on the area surrounding Riverview Park in Pittsburgh is shown in Figure 6.

Figure 6: Minimum Distance of Houses and Green Space, Brighton Heights Neighborhood, Pittsburgh, PA



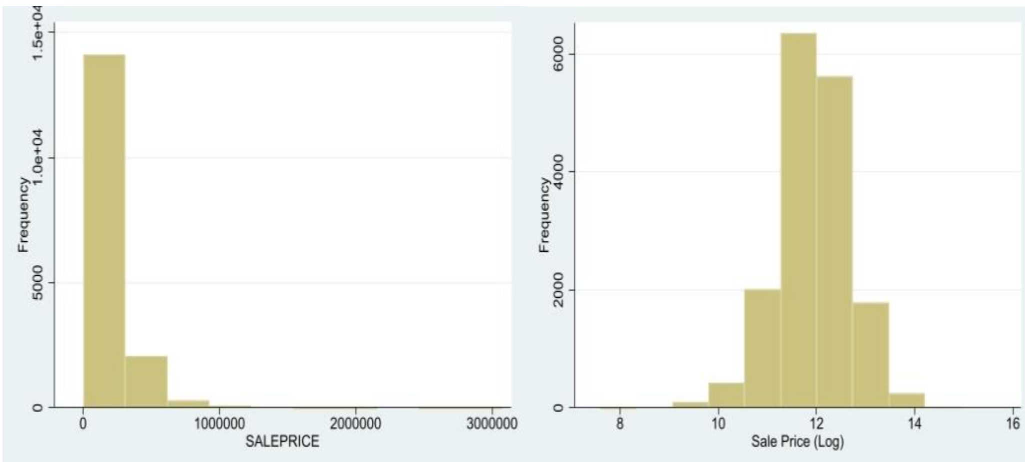
In addition, the spatial autoregressive moving averge model employed in this study includes the calculated spatial lag and spatial error terms, as explained above. With the model specified including the above dependent and spatial variables, I estimate the impact of proximity to green space on housing price. Detailed description of variables and data sources can be found in Table 2.

Table 2: Variables and Data Sources			
Variable	Description	Data Source	Expected Sign
Price	Housing Sales Price	Western Pennsylvania Regional Data Center (WPRDC)	Dependent Variable
HouseArea	House Size, Ft^2	WPRDC	+
LotArea			
NumBedrooms	Number of Bedrooms	WPRDC	+
NumBaths	Number of Bathrooms	WPRDC	+
NumFloors	Number of Floors	WPRDC	+
HouseAge	House Age	WPRDC	-
Garage	Garage Size (# of cars)	WPRDC	+
NumFire	Number of Fireplaces	WPRDC	+
GreenDist	Dist. To Green Space, Mi	WPRDC (Calculated)	-

Descriptive Statistics

Descriptive statistics were calculated using the STATA econometric suite and are listed in Table 2. Mean, standard deviation, minimum and maximum were calculated. Starting with *price*, the average house in Allegheny County sold for $196,865.70 with a standard deviation of $162,945.30, indicating that there is a very wide array of sale prices in this data set. The cheapest house sold for $1998, while the most expensive house cost $3,078,645. The market skews toward the low end of sale prices, as indicated by a standard deviation that is nearly equal to the mean. Given this extremely wide distribution of sale values, it is important to look at proportional, or percentage change effects, in prices. An increase in price of $1,000 due to park proximity indicates very different things for a $2,000 house than for a $2,000,000 house. Examining the natural log of the sale price, instead of the raw data point, shows this as evidenced by Figure X. This variable, *LnSale*, has a mean of 11.93862 and a standard deviation of .7204604. This distribution of values is much less skewed than the raw values, shown in Figure 7, confirming the necessity of examining this value.

Figure 7: Frequency Distributions of Price and LnPrice



Distance to parks ranges from 0 Mi to 0.05 Mi, with a mean of 0.005 Mi and a standard deviation of 0.004 Mi. This indicates that some houses directly border a park, and thus are directly on a park. In addition, all houses are a relatively short distance from a park, which is likely due to the "as the crow flies" method of distance calculation, implying that true walking distance will likely be greater. Because these distances are less than 1, applying the proportional measure through the traditional logarithmic transformation changes the directionality of the relationship. In addition, because several observations have a distance of 0, it is important to assess this variable as it is.

All other variables display reasonable values, indicating that there are no true outliers or errors within the data. To ensure the linear relationship assumption is met, *LotArea* and *HouseArea* are calculated as the logarithm of these values, listed in Table 3 as *LnLot and LnArea* respectively.

Table 3: Descriptive Statistics

Variable	Mean	S.D.	Min	Max
Price	196,865.7	162,945.3	1,998.196	3078645
LnSale	11.93862	0.72046	7.6	14.94
NumFloors	1.597077	0.493436	1	4
Age	67.99813	28.06987	5	221
NumBaths	2.078541	0.992608	0	9
NumBedrooms	3.109162	0.774749	1	10
NumFire	0.450995	0.571683	0	6
Garage	0.854973	0.839421	0	6
LotArea	15,356.78	34800.74	422	1598443
LnLot	9.162159	0.851716	6.045005	14.28454
HouseArea	1,727.613	791.0081	360	10203
LnArea	7.371637	0.393138	5.886104	9.230437
GreenDist	.0050175	.0040539	0	.0488758

Chapter 5: Results

Table 4 contains regression results for three models: OLS, S2SLS (LAG), and S2SLS (SARMA). In each model, I include a reduced form estimation and a full model estimation. The reduced estimationl excludes *Age^2 NumFire, and Garage,* while these are included in the full estimation includes these variables. This is to demonstrate the importance of including both in the model. The distinction is made to show how the results of each model change with the inclusion of additional relevant variables, which helps in confirming the robustness and ultimately the validity of the results

Table 4: Regression Results		OLS		S2SLS (LAG)		S2SLS (SARMA)
Variable	Reduced Model	Full model	Reduced Model	Full Model	Reduced Model	Full Model
LNLotArea	0.1302832*	0.1134114*	0.1349763*	0.1183388*	0.1535432*	0.1447697*
	(.0060679)	(.0061256)	(0.0061117)	(0.0061673)	(0.0064051)	(0.0064102)
LNArea	0.7364534*	0.7191867*	0.7374818*	0.7174655*	0.6671893*	0.6349905*
	(.0171145)	(.0182965)	(0.0170214)	(0.0181884)	(0.0155088)	(0.0162772)
NumFloors	-0.0458779*	-0.0274019*	-0.0469953*	-0.0298104*	-0.0359495*	-0.0296295*
	(.0085867)	(.0088044)	(0.0085427)	(0.0087661)	(0.0075006)	(0.0076530)
Age	-0.0037811*	-0.0032442*	-0.0038265*	-0.0036271*	-0.0054439*	-0.0068101*
	(.0001827)	(.0006812)	(0.0001821)	(0.0006783)	(0.0001892)	(0.0006737)
NumBaths	0.1676161*	0.1518793*	0.16489*	0.1492634*	0.1159399*	0.1072775*
	(.0060705)	(.0061113)	(0.0060337)	(0.0060674)	(0.0050442)	(0.0050384)
NumBedrooms	0.0577963*	0.0529214*	0.0577468*	0.0533392*	0.0371143*	0.036967*
	(.0075699)	(.0075621)	(0.0074981)	(0.0074885)	(0.0061870)	(0.0061643)
Age2		-0.00000426		-0.0000019		0.0000092
		(5.10e-06)		(0.0000051)		(0.0000048)
NumFire		0.1106865*		0.1097374*		0.0733866*
		(.0068108)		(0.0067770)		(0.0056384)
Garage		0.02261*		0.0221874*		0.0165643*
		(.0041974)		(0.0041923)		(0.0036513)
GreenDist	-3.106815*	-2.258265*	-2.8999949*	-2.1333877*	-2.4038193*	-2.3404086*
	(.7366672)	(.7403561)	(0.7339994)	(0.7366525)	(0.8924741)	(0.8973546)
LNSALE_LAG			0.0092641*	0.0089326*	0.0091012*	0.0090906*
			(0.0008201)	(0.0008117)	(0.0009812)	(0.0009746)
LAMDA					0.5482674*	0.5480349*
					(0.0088611)	(0.0089347)
CONST	5.133945*	5.347221*	4.9989031*	5.2454276*	5.606361*	5.9284311*
	(.1098134)	(.1253773)	(0.1100525)	(0.1252759)	(0.1081926)	(0.1195326)
R2 (spatial R2)	0.6131	0.6207	0.6188	0.626	0.6114	0.6175

Note: * indicates significance at α < 0.01, ** indicates significance at α <0.05, *** indicates significance at α <0.1. Robust Standard errors are in parentheses.

Regression Diagnostics

OLS estimation results are shown in Table 4. While previously discussed as being an inaccurate estimator for spatial relationships within the data, the OLS results are important to include for two reasons. First, OLS allows for traditional tests on well-behaved error term assumptions and for determining which specific spatial estimating strategy to include. In addition, they provide a baseline estimation scenario, with which my accurately specified model can be compared with.

Diagnostics were run on the data to assess the error term assumptions of the model. These were calculated in GeoDa, an open-source spatial analysis software made available from the University of Chicago Center for Spatial Data Science and are shown in Table 5. The Bruesh-Pagan test for heteroskedasticity indicated a significant likelihood of heteroscedasticity, or nonconstant error term variance across observations, in the data. In the S2SLS models this is corrected through the use of robust standard errors. The multicollinearity condition number, which indicates the possibility of relationships between one or more of the dependent variables for values greater than 30, shows that there may be further complex relationships within the data. An examination of the pairwise correlations between dependent variables indicate that there is correlation between the measures of house size, specifically living area, number of bedrooms, and number of bathrooms. A variance inflation analysis indicates that no combinations of dependent variables contribute to multicollinearity.

Table 5: Regression Diagnostics

Test	Value	P
Breusch-Pagan Test (Heteroskedasticity)	5106.60	0.00*
Multicollinearity Condition Number	105.697	--
Jarque-Bera (Normality of Errors)	14421.859	0.00*

Note: * indicates significance at α < 0.01, ** indicates significance at α <0.05, *** indicates significance at α <0.1. Multicollinearity Condition Number does not provide a probability statistic.

Spatial Models

Using GeoDa Space, the spatial econometrics subset of GeoDa, I estimate two versions of a Spatial Two Stage Least Squares (S2SLS) model. First, I run a regression including only a spatial lag, followed by the full SARMA model with both a spatial lag and a spatially weighted error term. Building the spatial model incrementally shows how the overall results of the model change with greater specificity and how specific values of variables also change.

Spatial Lag Model

Starting with the S2SLS model with a spatial lag, I calculate a spatial lag on sale price, named *LNSALE_LAG*. This was calculated using the spatial relationships defined in the spatial weights matrix and assesses the impact of sale prices of houses considered nearby in the matrix on the sale price of a house. Results are listed in Table 5. Overall, both estimations have remarkably similar R^2 values compared to the respective runs of the OLS model. The similarities continue throughout these results, with near identical coefficients as the previous model. Once again, the only insignificant controlling variable is Age^2. The addition of *LNSALE_LAG* is significant in both runs, with near identical coefficients. This, in addition to the results from Robust Lagrange Multiplier, indicate that this variable is necessary to include in the model. The coefficient in both estimations shows that a 1% increase in the prices of nearby houses will only increase a houses sale price by 0.93%, or by 0.89% in the fully specified model, which is a relatively small impact, although similar in size to other variables Looking at the effect of distance to green space on prices, it is clear that this variable has an outsize effect, similar to the results from the OLS model.

While this model appears to perform slightly better than OLS, regression diagnostics indicate that it is not properly specified without a spatially weighted error term. In addition, the persistence of the negative sign on number of floors also gives credence to the assertion that this specification contains a biased error term. Estimating the model with both a spatial lag on sale prices as well as a spatial weight in the error term will more accurately address this issue.

Spatial Lag and Error (SARMA)

Results for this third estimation of the spatial hedonic pricing model are found in Table 5. The S2SLS SARMA model contains both the spatial lag on *LnPrice*, *LNSALE_LAG*, and the spatial error term, λ, which are both included in the results. Overall, results are similar to the previous two models. The spatial R^2 values of 0.6114 and 0.6175 are still nearly identical to the other two estimations. In addition, nearly all of the relationships are within 0.2 of the values of the previous models. The unexpected negative sign is still present on *Floors* as well. This model also only has one insignificant variable, *Age*^{2}. *Age* has a small effect, with a year of age decreasing price by 0.0068%.

Looking specifically at the effects of the structural variables indicate the model is accurately specified. Of interest are the effects of *NumBaths* and *NumBedrooms* on *LnPrice*. Specifically, an additional bathroom will increase the sale price of a house by 0.107%, while an additional bedroom will increase the price by 0.037%. These values show that a higher number of bedrooms will increase the sale price of a house, while an additional bathroom will increase the price by a greater amount. In addition, the effects of *LnArea* and *LnLot* on the sale price show that a 1% increase of these variables will lead to an increase of 0.635% and 0.145%, respectively. This indicates that out of these two variables focused on the size of space being purchased, the size of the house outweighs the amount of land.

Turning to the relationship between *GreenDist* and *LnPrice,* the results show that it is still negative and significant. Interesting is that the difference between the reduced and full model with this specification is much lower compared to the other specifications; an additional mile of distance to a park decreases prices by 2.40% and 2.34%, a difference of 0.06%. This indicates that *GreenDist* is robust in this model. This is consistent with the expectations of my hypothesis and shows that green space proximity is preferred in the housing market.

Spatial error λ is significant, and increases house sale price by 0.548%, which shows that unidentifiable spatial trends influence the model. In addition, the lagged independent variable is also significant, indicating that it is also important to include in the model. Specifically, a 1% increase in the price of surrounding houses will increase the price of a house by 0.009% in the fully specified model. An additional bedroom will increase the price of the house by 0.037%, while an additional bathroom shows an increase of 0.11%, which is consistent with the assumption that bathrooms contribute more to the price of a house than a bedroom. Additionally, a percentage increase of living area translates to a 0.635% increase in sale price. Additional fireplaces and larger garages also increase the price of a house by 0.073% and 0.017% respectively, indicating that preferences for fireplaces are greater than that of larger garages.

Chapter 6: Discussion

To assess the hypothesis that houses near green space will be preferred over houses farther from green space all else equal, I estimated a spatial autoregressive price model on single family homes sold in Allegheny County from 2013-2017. Overall, the assumptions of the model were confirmed in the data, with physical characteristics generally being associated with higher housing prices, in addition to increases in park distance with lower housing prices. By connecting these results with the economic theory developed in Chapter 2, fascinating trends emerge.

The empirical model estimated the percentage change of explanatory variables on house sales price. While this was helpful for accurately assessing the impact of various house attributes statistically, it also allows for drawing conclusions about the elasticities of these attributes as unique goods. Specifically, many of the structural variables in the model indicate that they are inelastic goods, which is consistent with the notion that housing as a whole is an inelastic good. The amount, or size, of a house which a person or family demands is unlikely to change with respect to the price of a house. This extends to the components of the house as well, as evidenced from the inelastic nature of living area, number of bedrooms, and, to a lesser extent, number of bathrooms. When buying a home, consumers are unlikely to purchase a home that has fewer bedrooms than would be necessary to adequately house themselves and complete the necessary aspects of living. The global COVID-19 pandemic has made this point acute, with living spaces converted into work, exercise and study spaces as social distancing restrictions require people to do as much as possible from home.

The spatial effects in the empirical model confirm the necessity of including them in a housing price study as well as further validating this method of analysis for housing data. By including both a spatial lag and spatial corrected errors I was able to correct the standard OLS approach to account for significant autocorrelation issues. This is similar to the approaches used in the empirical literature discussed in Chapter 2. Specifically, Conway et al. (2010) uses a spatial lag autoregressive model, noting however that their LM specificity tests indicated that a lag and a spatial error term are both efficient for correcting spatial autocorrelation. In addition, spatial autoregression was used by Laszkiewicz et al. (2019) to show the effect of differing types of green space on nearby housing prices, and also found this to a be an efficient corrector of spatial autoregression in the data through the inclusion of a “spatial parameter” calculated similarly to λ in my results. Overall, the spatial configuration of my model is consistent with the results found in other hedonic pricing spatial autoregressions.

My main findings on the distance to green space are also in line with the findings from these studies. My model estimates distance to green space in miles, which therefore leads to a large effect size at this unit of measurement. The range of distances in the dataset shows that this large effect is not an outlier, because no house is greater than twentieth of a mile from a green space making an increase in distance of one mile a purely hypothetical case. This large of a change is therefore expected to have a large effect. Given a standard deviation of 0.004 miles, the implications of distance to green space become more beneficial when measured at a $1/100^{th}$ of a mile. It is possible to estimate this effect by dividing the coefficient result by 100, assuming a distance has a linear effect on prices, so that the first 0.01 mile does not change the price by more than the last 0.01 mile. At this unit of observation, distance to green space increases by 0.01 miles, the price of a house will decrease by 0.0234%, much more in line with the effect sizes of the structural variables.

My analysis of distance to green space as a continuous variable is similar to the methods of Laszkiewicz et al. (2019) and Czembroski and Kronenberg (2016). These specific model results show that there is widespread demand for proximity to green space in the housing decision, not specific to certain regional markets. In addition, looking at the results of other studies assessing the effect of green space quantity, it is clear that my results of green space factoring positively into housing sales values is not an outlier in the economic literature. Specifically, Immergluck and Balan's (2018) finding that housing prices rose between 18% and 26.6% more in neighborhoods surrounding a new park development shows than those not near the development show an even greater effect than my analysis, indicating that newer parks may be more greatly desired than parks overall. While these models broke down parks into different subsets by type, the size of the effect and the directionality are very similar to the results of this model, indicating that preference to live near a park shows up in housing data from Poland to Pittsburgh.

Looking at the results of the model in relation to my consumer maximization theory framework, it is clear that this analysis backs up the assumptions of this economic theory. My model showed that consumers with a high penchant for living near a green space would increase housing prices for houses near parks for the entire market of consumers, while other consumers would place a higher preference on other housing characteristics and would therefore choose a house farther away from green space. When extrapolated to the entire market, this indicates that housing prices will decrease with distance from green space, while greater levels of housing goods will increase prices. This is exactly what is seen in the data on the Pittsburgh housing market.

These results have important implications for the distribution of public goods in Pittsburgh. As discussed at the beginning of Chapter 4, Pittsburgh is a highly segregated and unequal city, with extreme racial disparities among residents (Dickinson, 2021). My results show that proximity to green space leads to a price premium on a house, which in relation to distribution indicates that people unable to pay this premium will miss out on access to this vital public good. This places green space in the same realm as other public goods with access limited by ability to pay, such as schools. In addition, this pay-to-play aspect of access to green space could help explain some of the racial health inequalities present in Pittsburgh. A wide array of environmental justice literature has demonstrated a link between park proximity and health outcomes due to easier access to exercise and cleaner, cooler air from the presence of trees and other fauna (Estabrooks et al., 2003; Su et al., 2011). Houses and neighborhoods farther away from green space will more readily trap heat in the abundance of asphalt and concrete, known as the heat island effect, leading to higher levels of cardiovascular diseases in these neighborhoods (Popovich & Flavelle, 2019). This is in addition to the carbon and pollution sequestering abilities of green space which aid in removing air pollution from surrounding areas, lowering rates of childhood asthma and other respiratory diseases (Roy et al., 2012). When looked at in relation to the price premium on park proximity, this reveals that negative health outcomes could likely be more prevalent in lower socioeconomic neighborhoods of Pittsburgh.

Taking this all into account, a price premium on houses near green space explains more than just the willingness to pay for access to nature. While a public good provided for the benefit of all, this analysis shows that those with the ability to pay are able to benefit more than those who cannot. In a time when inequality across the country has reached historic proportions in nearly every aspect of life from income inequality to racial and gender inequality (Farrell, 2015), this analysis shows how inequality penetrates markets of public goods that are explicitly designed to not favor one group over another.

Overall, analyzing the results of the data in relation to the economic theory, important trends emerge for the Pittsburgh housing market and beyond. Housing prices are influenced by the physical pieces of a house, but these do not fully explain variations in the prices, as consumers will have predetermined needs for house size and capacity that are inelastic. Housing price variation therefore is better explained by locational attributes, for instance the proximity to parks and green spaces. This has important implications for green space and ultimately health inequality in the Pittsburgh metropolitan area, as well as the increasingly unequal landscape of public good provisions.

Limitations

Overall, this spatial econometric analysis was successful in estimating the impact of structural housing and distance to green space on the Pittsburgh housing market. The results are consistent with previous empirical studies, while also confirming the assumptions of the economic model presented in Chapter 2. However, this study does have limitations.

Given the many steps and programs necessary to compute geographic coordinates, calculate distances from a second spatial file system, and run spatial econometrics on the resulting dataset, it is possible some data was lost in translation or miscalculated despite many efforts to minimize such error. In addition, it is possible that the systematic geocoding of addresses to coordinates performed with Geocodio may have resulted in misplaced addresses, although a random sampling of 15 results indicated this was unlikely to be a widespread issue. While the calculation of green space distance using a Near Analysis is the most sophisticated method of this type of calculation, it does not take topography into account, potentially biasing areas with significant elevation change. In addition, since the spatial elements of the dataset were self-generated using ArcGIS Pro, they were not encoded using the same syntax as spatial datasets from an official source such as the US Census Bureau, which limited their compatibility with STATA. Because of this, the open-source GeoDa Space was used to estimate the spatial models. Research notes that this software is as accurate as STATA, however it is not able to compute a post-estimation marginal effects analysis, which is possible with more comprehensive econometric software. A marginal effects analysis would have separated the overall coefficient effect into direct and indirect effects. This limits interpretation of the results, and potentially overlooks effects that would have stood out in this type of analysis.

The results of the model are likely biased by a number of factors. Diagnostics showed a significant likelihood of heteroskedasticity in the data, which was corrected for with robust standard errors. However, this correction does not completely resolve the issue. In addition, the diagnostic tests indicated that the errors in the specified model are non-normal, which would invalidate the assumptions of a well-behaved error term. Together, these error term biases emphasize the importance of examining this statistical analysis with a grain of salt.

The model may be missing additional spatial controlling variables, due to the unavailability of spatial data for other potential spatial factors in the housing decision. Specifically, other analyses include distances to schools, highways, and differentiate between different types of green space, all of which may influence the effect of green space proximity on housing prices. In addition, it is possible that the unavailability of this data is a contributing factor to endogeneity issues suggested by non-normal errors. With these caveats brought to light, I turn to discussing ways this research can be extended in the future.

Future Directions

This economic analysis of the relationship between housing prices and green space proximity can be extended and applied in several directions. The availability of sale dates in the dataset could be employed in a time series analysis to show how park quality changes over time influence the prices of surrounding houses. In addition, the findings from this analysis could be used to further explore park access inequality in the

Pittsburgh metropolitan area, and if plans to improve the quality of the park system are inadvertently contributing to gentrification rates, a phenomenon shown in other American cities (Checker, 2011). In addition, the dataset included both the most recent sale date and the second most recent sale date, which could be used to estimate the impact of green space proximity on house “flipping” rates, where developers purchase a run-down house with the sole intention of renovating and selling it for a substantial profit. In addition, future urban planning research could explore ways to reduce to the price gradient on green space proximity.

Chapter 7: Conclusion

This Independent Study examined the effect of green space proximity on housing prices in Allegheny County, Pennsylvania. By using a novel consumer utility maximization framework with heterogeneous preferences, this analysis is grounded in microeconomic theory. This model, which formalized the mechanisms through which different consumers would value a house for the different attributes of the house, was extended into an empirical spatial autoregression equation based on hedonic pricing theory. This method of analysis is consistent with and informed by the previous literature on the relationship between environmental attributes and the real estate market.

The results of the empirical analysis were significant and indicate a negative price gradient as distance from a green space increases. This result, that surrounding green space results in higher housing prices, is similar to the results found in empirical studies on other metropolitan areas. While the model appeared to be an effective estimator of the data, it was not without its flaws; multicollinearity, non-normal errors, and heteroskedasticity were all shown to have a highly likely presence in the data, potentially skewing the results.

This research holds important implications for environmental justice causes, and inequalities in access to public goods generally. When public goods are more accessible in more expensive neighborhoods, priority access can be auctioned off to the highest bidder, which is seen with green space in the housing market. In conclusion, the economic value of green space, thought of in many ways, from the sum of ecosystem services provided to the amount of tourism it brings in, can also be seen in higher prices in the housing market.

Appendix A

Table A1: Variables and Data Sources (Rigolon & Németh, 2020)

Dependent variables to define gentrification (binary outcome variable)				
Variable	Description	Data source	Type	Level
Income	Median household income	ACS, LTDB	DV	1, 2
Percent bachelor	Percentage of people aged 25 and above with at least a bachelor's degree	ACS, LTDB	DV	1, 2
Rent	Median gross rent	ACS, LTDB	DV	1, 2
Home value	Median home value for owner-occupied units	ACS, LTDB	DV	1, 2
Predictor variables				
Variable	Description	Data source	Type	Level
Percent Black	Percentage of non-Hispanic Black residents	ACS, LTDB	CV	1
Percent Latino	Percentage of Latino or Hispanic residents	ACS, LTDB	CV	1
Income	Median household income	ACS, LTDB	CV	1, 2
Rent	Median gross rent	ACS, LTDB	CV	1
Percent vacant housing units	Percentage of vacant housing units	ACS, LTDB	CV	1, 2
Population density	Number of residents per acre	ACS, LTDB	CV	1, 2
Percent multifamily	Percentage of multifamily housing units	ACS, LTDB	CV	1
Percent older housing units	Percentage of housing units older than 30 years	ACS, LTDB	CV	1
Variable	Description	Data source	Type	Level
Distance from downtown	Distance from each city's downtown	City data	IV, CV	1
Access to rail transit*	Presence of a rail transit station within half a mile	City data	CV	1
Income change in previous decade	Change in median household income in the previous decade	ACS, LTDB	CV	1
Percent HUD units	Percentage of housing units subsidised by HUD	HUD	CV	1
New park*	Presence of a new park within half a mile	City data	IV	1
Size of new parks	Size of new parks within half a mile	City data, TPL	IV	1
New greenway park*	Presence of a new greenway park with walking/cycling trails within half a mile	City data, TPL	IV	1
New park close to downtown*	Presence of a new park located close to downtown (less than median distance to downtown of gentrification-eligible tracts for each city)	City data, TPL	IV	1
ParkScore	ParkScore index describing the quality of urban park systems	TPL	CV	2

Notes: * denotes a dummy variable. DV: dependent variable. IV: independent variable. CV: control variable/covariate. ACS: American Community Survey. LTDB: Longitudinal Tract Database. HUD: US Department of Housing and Urban Development. TPL: The Trust for Public Land. Level 1: Tract. Level 2: City. All data was collected at the beginning of the two study periods (2000 and 2006–2010 ACS).

Table A2: Odds ratios of the likelihood of gentrification (Rigolon & Németh, 2020)

	2000–2008 (n = 2836)		2008–2016 (n = 2779)	
	Model 1 (Level 1)	Model 2 (Levels 1, 2)	Model 1 (Level 1)	Model 2 (Levels 1, 2)
Fixed effects				
Intercept	**0.104*****	0.419	**0.270***	7.135^
Percent Black	**0.990*****	**0.989*****	**0.994****	**0.993****
Percent Latino	**0.988*****	**0.987*****	**0.990*****	**0.989*****
Income	**1.032****	**1.032****	**1.023***	**1.022***
Rent	**1.002****	**1.002****	1.000	1.000
Percent vacant housing units	**1.052*****	**1.053*****	1.004	1.006
Population density	**0.996***	**0.996***	0.998	0.998
Percent multifamily	0.999	0.998	1.000	0.999
Percent older housing units	1.002	1.002	1.002	1.000
Distance from downtown	**0.856*****	**0.852*****	**0.884*****	**0.885****
Access to rail transit	1.061	1.059	**1.330***	**1.324***
Income change in previous decade	1.007	1.008	**0.970****	**0.971****
Percent HUD units	1.002	1.003	0.992^	0.991^
New park	0.720	0.749	0.542^	0.567^
	Model 1 (Level 1)	Model 2 (Levels 1, 2)	Model 1 (Level 1)	Model 2 (Levels 1, 2)
Size of new parks	1.004	1.003	1.003	1.002
New greenway park	0.313	0.299	**3.222***	**3.367****
New park close to downtown	1.411	1.392	2.045^	1.916^
City – Income		0.989		0.983
City – Percent vacant housing units		0.992		**0.921*****
City – Population density		1.020		**1.064*****
City – Park Score		0.983		**0.970*****
Random effects				
Level 1 intercept	0.000	0.000	0.010	0.000
Level 2 intercept	0.112	0.198	0.118	0.000
Akaike Information Criterion	14,491	14,551	13,654	13,627

Notes: ^$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table A3: Estimated coefficients from models with greenspace effects and time indicators (Conway et al., 2010)

Predictor	Model 1: Standard hedonic model	Model 2: Spatial lag model
Constant	0.839	−0.118
	(0.213)	(0.874)
Log Building Area	0.571	0.556
	(0.000)	(0.000)
Log Lot Size	0.115	0.124
	(0.031)	(0.015)
Log Distance to Ramp	0.226	0.206
	(0.000)	(0.000)
Log Distance to Rec. Area	−0.128	−0.120
	(0.000)	(0.000)
Log Median Income	0.290	0.260
	(0.000)	(0.000)
Age of House	0.027	0.029
	(0.038)	(0.021)
Age Squared	−0.00021	−0.0002
	(0.0027)	(0.014)
Quarter 2 1999	0.158	0.144
	(0.001)	(0.002)
Quarter 3 1999	0.101	0.091
	(0.029)	(0.040)
Quarter 4 1999	0.177	0.162
	(0.001)	(0.001)
Quarter 1 2000	0.175	0.170
	(0.000)	(0.000)
Quarter 2 2000	0.217	0.204
	(0.000)	(0.000)
Log Green 200 to 300	0.076	0.070
	(0.039)	(0.048)
Log Green 300 to 400	0.068	0.070
	(0.130)	(0.102)
Log Green 400 to 500	0.004	0.006
	(0.908)	(0.867)
Adjusted R^2	0.830	0.833
rho		0.110
		(0.017)

Table A4: Variables used in the model (Czembrowski & Kronenberg, 2016a)

Name	Description	Expected sign	Mean*
PRICE_PER_SQ_M	Price per square meter (explained variable)	n/a	3935.61
LIV_AREA	Living area in square meters	–	52.65
NUM_ROOM	Number of rooms (not included due to the near multicollinearity)	–	2.94
AGE_1850_1880	Building erected between 1850 and 1880 (dummy variable)	–	0.038
AGE_1881_1918	Building erected between 1881 and 1918 (dummy variable)	–	0.038
AGE_1919_1939	Building erected between 1919 and 1939 (dummy variable)	–	0.036
AGE_1945_1970	Building erected between 1945 and 1970 (dummy variable)	–	0.31
AGE_1971_1989	Building erected between 1971 and 1989 (dummy variable)	–	0.177
AGE_1990_2010	Building erected between 1990 and 2010 (dummy variable)	–	0.152
AGE_2011_2014	Building erected between 2011 and 2014 (dummy variable)	+	0.246
STORY1 – . . . – STORY14	The story the apartment is located on (dummy variable)	?	n/a**
DIST_FOREST_LARGE, DIST_LAGIEWNIKI, DIST_FOREST_MEDIUM, DIST_FOREST_SMALL, DIST_PARK_LARGE, DIST_PARK_MEDIUM, DIST_PARK_SMALL	Walking distance to the nearest entrance to large forests except for Lagiewniki, the Lagiewniki forest, medium and small forests, and large, medium and small parks	–	4822 7116 5135 6452 2464 1069 1572
DIST_CEMETERY, DIST_ALLOTMENT	Walking distance to the nearest entrance to a cemetery or allotment garden	?	2196 1063
PERCENT_GREEN	Percentage of greenery in a 500 m radius	+	8.33
GORNA	Location in the Gorna district (dummy variable)	–	0.21
BALUTY	Location in the Baluty district (dummy variable)	–	0.27
WIDZEW	Location in the Widzew district (dummy variable)	–	0.19
POLESIE	Location in the Polesie district (dummy variable)	–	0.19
SRODMIESCIE	Location in the Srodmiescie district (city center) (dummy variable)	+	0.13
SOLARIS	Apartment located in the newly built "Solaris" building	+	0.006
DIST_PRE-KINDERGARDEN, DIST_KINDERGARDEN, DIST_ELEMENTARY_SCHOOL, DIST_MIDDLE_SCHOOL, DIST_HIGH_SCHOOL, DIST_UNIVERSITY	Walking distance to the nearest educational facilities	–	1167 588 723 917 852 2373
DIST_PLAYING_FIELD, DIST_SWIMMING_POOL	Walking distance to the nearest publically accessible sports facility (most swimming pools form part of larger sports complexes)	–	1086 2142
EU_DIST_CENTER	Euclidean distance to the city center (not included due to the near multicollinearity)	–	3704
EU_DIST_INDUSTRIAL	Euclidean distance to the nearest industrial area	–	1998
DIST_TRANSPORT_HUB	Walking distance to the nearest transport hub	–	2099
DIST_SHOPPING_CENTER	Walking distance to the nearest shopping center	–	2961
QUARTER	The number of the quarter the transaction took place in (Jan–Mar 2011 – "12", Apr–Jun 2011 – "11" etc.)	+	n/a

* In the case of dummy variables, mean values can be interpreted as the share of given category in the whole sample.

** Story 1—0.16, 2—0.21, 3—0.2, 4—0.18, 5—0.13, 6—0.04, 7—0.02, 8—0.02, 9—0.01, 10—0.01, 11—0.01, 12—0.003, 13—0.0003, 14—0.0005.

Table A5: Ordinary least squares regression results (Czembrowski & Kronenberg, 2016a)

		Standard model		Fixed effects model	
		Coefficient	Sig	Coefficient	Sig
1	Lagged dependent variable	0.22	***	0.21	***
2	Const	3656.90	***	4364.00	***
3	AGE_1850_1880	−245.42	***	−234.68	***
4	AGE_1881_1918	−1518.80	***	−1588.40	***
5	AGE_1919_1939	−1354.20	***	−1390.00	***
6	AGE_1945_1970	−924.30	***	−941.57	***
7	AGE_1971_1989	−775.27	***	−821.06	***
8	AGE_1990_2010			−76.58	**
9	STORY1	−149.92	***	−160.59	***
10	STORY2			−27.10	*
11	STORY6	140.26	***	145.68	***
12	STORY7	77.92	**	78.04	**
13	STORY8	90.20	**	98.93	**
14	ln(DIST_LAGIEWNIKI)	−60.12	**	−110.24	***
15	ln(DIST_FOREST_SMALL)	−67.75	***	−107.06	***
16	ln(DIST_PARK_LARGE)	−90.14	***	−57.33	***
17	ln(DIST_CEMETERY)	95.99	***	91.44	***
18	PERCENT_GREEN	2.85	**	3.95	***
19	SOLARIS	2301.50	***	2141.30	***
20	GORNA			−343.54	***
21	BALUTY			−371.74	***
22	WIDZEW			−270.64	***
23	POLESIE			−222.18	***
24	ln(DIST_PRE-KINDERGARDEN)	−30.78	**		
25	ln(DIST_MIDDLE_SCHOOL)	58.58	***	42.77	***
26	ln(DIST_UNIVERSITY)	−104.33	***		
27	ln(DIST_PLAYING_FIELD)	35.78	**		
28	ln(DIST_TRANSPORT_HUB)	38.91	**		
29	ln(DIST_SHOPPING_CENTER)	96.45	***	75.47	***
30	QUARTER	50.21	***	51.12	***
	Lambda	0.3919		0.37127	
	Residual variance	332,530		332,540	
	N	9346		9346	

Table A6: Results of baseline hedonic pricing model (Łaszkiewicz et al., 2019)

Variable	Coeff.	*P* value	VIF
Const.	4969.996***	0.000	
LN LIV AREA	−187.595***	0.000	1.155
AGE 1850–1880	−639.079***	0.000	1.395
AGE 1881–1918	−1639.048***	0.000	1.320
AGE 1919–1939	−1504.720***	0.000	1.197
AGE 1945–1970	−1092.528***	0.000	1.753
AGE 1971–1989	−972.390***	0.000	1.555
AGE 1990–2010	−150.308***	0.000	1.258
STORY 2	152.226***	0.000	1.016
STORY 3	169.538***	0.000	
STORY 4	161.173***	0.000	
STORY 5	139.947***	0.000	
STORY 6	280.469***	0.000	
STORY 7	253.827***	0.000	
STORY 8	229.873***	0.000	
STORY 9	138.609***	0.000	
STORY 10	202.940***	0.000	
STORY 11 +	141.868***	0.000	
QUARTER	6.477***	0.000	1.051
LN DIST PRE-KINDERGARDEN	−28.766**	0.016	1.789
LN DIST ELEMENTARY SCHOOL	26.177**	0.032	1.720
LN DIST MIDDLE SCHOOL	37.437***	0.001	1.690
LN DIST HIGH SCHOOL	42.335***	0.000	1.732
LN DIST UNIVERSITY	−9.378	0.578	3.167
LN DIST PLAYING FIELD	−82.145***	0.000	1.827
LN DIST SWIMMING POOL	−70.563***	0.000	1.972
LN DIST TRANSPORT HUB	23.122	0.131	2.666
LN DIST SHOPPING CENTER	−53.247**	0.038	2.988
NOISE 61–65 dB	8.562	0.522	1.082
NOISE 66–70 dB	−80.885***	0.000	
NOISE > 71 dB	−225.991***	0.000	
PM 2.5 or 10	−111.677***	0.000	1.816
PM 2.5 and 10	−201.229***	0.000	
LN DIST CEMETERY	136.627***	0.000	1.839
LN DIST ALLOTMENT	15.996	0.189	1.476
PERCENT GREEN	0.692	0.290	2.373
LN DIST PARK	−54.179***	0.000	1.582
LN DIST FOREST	−45.351**	0.022	2.549
Fixed effects	Yes		1.189
ρ	0.298***	0.000	
R^2	57%		
σ_ε	722		

Significance levels: 0.01 ‘***’ 0.05 ‘**’ 0.1 ‘*’.

Appendix B: Statistical Outputs

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          105.697

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE            PROB
Jarque-Bera                       2       14421.859          0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE            PROB
Breusch-Pagan test                7        5106.597          0.0000
Koenker-Bassett test              7        1660.248          0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE            PROB
Moran's I (error)              0.4810        61.511          0.0000
Lagrange Multiplier (lag)         1         258.048          0.0000
Robust LM (lag)                   1         112.176          0.0000
Lagrange Multiplier (error)       1        3779.182          0.0000
Robust LM (error)                 1        3633.310          0.0000
Lagrange Multiplier (SARMA)       2        3891.358          0.0000

================================ END OF REPORT =====================================
```

```
summarize SALEDATE STORIES AGE TOTALBATHS BEDROOMS FIREPLACES BSMTGARAGE NEAR_DIST LOTAREA HOUSEAREA SALEPRICE

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    SALEDATE |     16,590     20232.4    526.5444      19360      21059
     STORIES |     16,590    1.597077    .4934356          1          4
         AGE |     16,590    67.99813    28.06987          5        221
  TOTALBATHS |     16,590    2.078541    .9926082          0          9
    BEDROOMS |     16,590    3.109162    .7747487          1         10
-------------+---------------------------------------------------------
  FIREPLACES |     16,590    .4509946    .5716839          0          6
  BSMTGARAGE |     16,590    .8549729    .8394209          0          6
   NEAR_DIST |     16,590    .0050175    .0040539          0   .0488758
     LOTAREA |     16,590    15356.78    34800.74        422    1598443
   HOUSEAREA |     16,590    1727.613    791.0081        360      10203
-------------+---------------------------------------------------------
   SALEPRICE |     16,590    196865.7    162945.3   1998.196    3078645

summarize LNSALE LNLOT  LNAREA

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      LNSALE |     16,590    11.93862    .7204604        7.6      14.94
       LNLOT |     16,590    9.162159    .8517159   6.045005   14.28454
      LNAREA |     16,590    7.371637    .3931382   5.886104   9.230437

pwcorr LNLOT AGE STORIES TOTALBATHS BEDROOMS LNAREA FIREPLACES BSMTGARAGE

             |    LNLOT      AGE  STORIES TOTALB~S BEDROOMS   LNAREA FIREPL~S
-------------+---------------------------------------------------------------
       LNLOT |   1.0000
         AGE |  -0.4924   1.0000
     STORIES |  -0.1655   0.1279   1.0000
  TOTALBATHS |   0.4496  -0.4962   0.2334   1.0000
    BEDROOMS |   0.3026  -0.3015   0.3612   0.5974   1.0000
      LNAREA |   0.3813  -0.2968   0.4782   0.7190   0.6800   1.0000
  FIREPLACES |   0.3115  -0.1171   0.0465   0.3466   0.2486   0.3280   1.0000
  BSMTGARAGE |   0.2801  -0.4846  -0.1943   0.2790   0.1830   0.1255   0.1823

             | BSMTGA~E
-------------+---------
  BSMTGARAGE |   1.0000
```

```
. estat vif

    Variable |       VIF       1/VIF  
-------------+----------------------
      LNAREA |      3.30    0.303370
  TOTALBATHS |      2.71    0.368913
    BEDROOMS |      2.00    0.500190
         AGE |      1.90    0.527390
       LNLOT |      1.74    0.573667
     STORIES |      1.71    0.585143
  BSMTGARAGE |      1.39    0.720538
  FIREPLACES |      1.24    0.807141
   NEAR_DIST |      1.11    0.899558
-------------+----------------------
    Mean VIF |      1.90
```

```
 regress LNSALE LNLOT LNAREA STORIES AGE TOTALBATHS BEDROOMS NEAR_DIST, vce(robust)

Linear regression                               Number of obs     =     16,590
                                                F(7, 16582)       =    4048.08
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6131
                                                Root MSE          =     .44823

------------------------------------------------------------------------------
             |               Robust
      LNSALE |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       LNLOT |   .1302832   .0060679    21.47   0.000     .1183894     .142177
      LNAREA |   .7364534   .0171145    43.03   0.000      .702907    .7699997
     STORIES |  -.0458779   .0085867    -5.34   0.000    -.0627086   -.0290471
         AGE |  -.0037811   .0001827   -20.70   0.000    -.0041392   -.003423
  TOTALBATHS |   .1676161   .0060705    27.61   0.000     .1557173    .1795149
    BEDROOMS |   .0577963   .0075699     7.64   0.000     .0429585     .072634
   NEAR_DIST |  -3.106815   .7366672    -4.22   0.000    -4.550762   -1.662869
       _cons |   5.133945   .1098134    46.75   0.000     4.918699    5.349191
------------------------------------------------------------------------------
```

```
. regress LNSALE LNLOT LNAREA STORIES AGE AGE2 TOTALBATHS BEDROOMS FIREPLACES BSMTGARAGE NEAR_DIST, vce(rob
> ust)

Linear regression                               Number of obs     =     16,590
                                                F(10, 16579)      =    2952.62
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6207
                                                Root MSE          =     .44384

------------------------------------------------------------------------------
             |               Robust
      LNSALE |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       LNLOT |   .1134114   .0061256    18.51   0.000     .1014046    .1254183
      LNAREA |   .7191867   .0182965    39.31   0.000     .6833237    .7550497
     STORIES |  -.0274019   .0088044    -3.11   0.002    -.0446594   -.0101443
         AGE |  -.0032442   .0006812    -4.76   0.000    -.0045795    -.001909
        AGE2 |  -4.26e-06   5.10e-06    -0.84   0.404    -.0000143    5.74e-06
  TOTALBATHS |   .1518793   .0061113    24.85   0.000     .1399004    .1638582
    BEDROOMS |   .0529214   .0075621     7.00   0.000     .0380989    .0677439
  FIREPLACES |   .1106865   .0068108    16.25   0.000     .0973366    .1240365
  BSMTGARAGE |     .02261   .0041974     5.39   0.000     .0143826    .0308374
   NEAR_DIST |  -2.258265   .7403561    -3.05   0.002    -3.709442   -.8070872
       _cons |   5.347221   .1253773    42.65   0.000     5.101468    5.592974
------------------------------------------------------------------------------
```

```
REGRESSION LAG ONLY
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :SingleFamilyHomeSales.dbf
Weights matrix      :SingleFamilyHomeSales.shp: distance: Threshold, 0.1
Dependent Variable  :      LNSALE              Number of Observations:       16590
Mean dependent var  :     11.9386              Number of Variables   :           9
S.D. dependent var  :      0.7205              Degrees of Freedom    :       16581
Pseudo R-squared    :      0.6188
Spatial Pseudo R-squared:  0.6159

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       4.9989031       0.1100525      45.4228771       0.0000000
                 AGE      -0.0038265       0.0001821     -21.0099658       0.0000000
            BEDROOMS       0.0577468       0.0074981       7.7015299       0.0000000
              LNAREA       0.7374818       0.0170214      43.3267236       0.0000000
               LNLOT       0.1349763       0.0061117      22.0848215       0.0000000
           NEAR_DIST      -2.8999949       0.7339994      -3.9509498       0.0000778
             STORIES      -0.0469953       0.0085427      -5.5012338       0.0000000
          TOTALBATHS       0.1648900       0.0060337      27.3282007       0.0000000
            W_LNSALE       0.0092641       0.0008201      11.2957105       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_LNSALE
Instruments: W_AGE, W_BEDROOMS, W_LNAREA, W_LNLOT, W_NEAR_DIST, W_STORIES,
             W_TOTALBATHS

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1        3333.870          0.0000
================================ END OF REPORT =====================================

REGRESSION LAG ONLY
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :SingleFamilyHomeSales.dbf
Weights matrix      :SingleFamilyHomeSales.shp: distance: Threshold, 0.1
Dependent Variable  :      LNSALE              Number of Observations:       16590
Mean dependent var  :     11.9386              Number of Variables   :          12
S.D. dependent var  :      0.7205              Degrees of Freedom    :       16578
Pseudo R-squared    :      0.6260
Spatial Pseudo R-squared:  0.6233

White Standard Errors
------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       5.2454276       0.1252759      41.8709952       0.0000000
                 AGE      -0.0036271       0.0006783      -5.3473323       0.0000001
                AGE2      -0.0000019       0.0000051      -0.3750290       0.7076389
            BEDROOMS       0.0533392       0.0074885       7.1228617       0.0000000
          BSMTGARAGE       0.0221874       0.0041923       5.2923778       0.0000001
          FIREPLACES       0.1097374       0.0067770      16.1927029       0.0000000
              LNAREA       0.7174655       0.0181884      39.4464206       0.0000000
               LNLOT       0.1183388       0.0061673      19.1880300       0.0000000
           NEAR_DIST      -2.1333877       0.7366525      -2.8960572       0.0037788
             STORIES      -0.0298104       0.0087661      -3.4006388       0.0006723
          TOTALBATHS       0.1492634       0.0060674      24.6008772       0.0000000
            W_LNSALE       0.0089326       0.0008117      11.0052314       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_LNSALE
Instruments: W_AGE, W_AGE2, W_BEDROOMS, W_BSMTGARAGE, W_FIREPLACES,
             W_LNAREA, W_LNLOT, W_NEAR_DIST, W_STORIES, W_TOTALBATHS

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Anselin-Kelejian Test             1        3182.170          0.0000
================================ END OF REPORT =====================================
```

```
REGRESSION SPATIAL AND LAG (SARMA)
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
-------------------------------------------------------------------
Data set            :SingleFamilyHomeSales.dbf
Weights matrix      :SingleFamilyHomeSales.shp: distance: Threshold, 0.1
Dependent Variable  :      LNSALE               Number of Observations:       16590
Mean dependent var  :     11.9386               Number of Variables   :          12
S.D. dependent var  :      0.7205               Degrees of Freedom    :       16578
Pseudo R-squared    :      0.6175
Spatial Pseudo R-squared:  0.6141
N. of iterations    :           8               Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       5.9284311       0.1195326      49.5967778       0.0000000
                 AGE      -0.0068101       0.0006737     -10.1089651       0.0000000
                AGE2       0.0000092       0.0000048       1.9323917       0.0533112
            BEDROOMS       0.0369670       0.0061643       5.9969590       0.0000000
          BSMTGARAGE       0.0165643       0.0036513       4.5365946       0.0000057
          FIREPLACES       0.0733866       0.0056384      13.0155005       0.0000000
              LNAREA       0.6349905       0.0162772      39.0110972       0.0000000
               LNLOT       0.1447697       0.0064102      22.5841378       0.0000000
           NEAR_DIST      -2.3404086       0.8973546      -2.6081200       0.0091041
             STORIES      -0.0296295       0.0076530      -3.8716345       0.0001081
          TOTALBATHS       0.1072775       0.0050384      21.2917675       0.0000000
            W_LNSALE       0.0090906       0.0009746       9.3280211       0.0000000
              lambda       0.5480349       0.0089347      61.3375666       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_LNSALE
Instruments: W_AGE, W_AGE2, W_BEDROOMS, W_BSMTGARAGE, W_FIREPLACES,
             W_LNAREA, W_LNLOT, W_NEAR_DIST, W_STORIES, W_TOTALBATHS
================================ END OF REPORT =====================================
```

```
REGRESSION SPATIAL AND LAG (SARMA)
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
-------------------------------------------------------------------
Data set            :SingleFamilyHomeSales.dbf
Weights matrix      :SingleFamilyHomeSales.shp: distance: Threshold, 0.1
Dependent Variable  :      LNSALE               Number of Observations:       16590
Mean dependent var  :     11.9386               Number of Variables   :           9
S.D. dependent var  :      0.7205               Degrees of Freedom    :       16581
Pseudo R-squared    :      0.6114
Spatial Pseudo R-squared:  0.6080
N. of iterations    :           8               Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       5.6063610       0.1081926      51.8183173       0.0000000
                 AGE      -0.0054439       0.0001892     -28.7778191       0.0000000
            BEDROOMS       0.0371143       0.0061870       5.9987187       0.0000000
              LNAREA       0.6671893       0.0155088      43.0199678       0.0000000
               LNLOT       0.1535432       0.0064051      23.9719923       0.0000000
           NEAR_DIST      -2.4038193       0.8924741      -2.6934330       0.0070720
             STORIES      -0.0359495       0.0075006      -4.7928884       0.0000016
          TOTALBATHS       0.1159399       0.0050442      22.9849600       0.0000000
            W_LNSALE       0.0091012       0.0009812       9.2752455       0.0000000
              lambda       0.5482674       0.0088611      61.8735744       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_LNSALE
Instruments: W_AGE, W_BEDROOMS, W_LNAREA, W_LNLOT, W_NEAR_DIST, W_STORIES,
             W_TOTALBATHS
================================ END OF REPORT =====================================
```

References

About • WPRDC. (2020). WPRDC. http://www.wprdc.org/about/

Banzhaf, H. S., & Walsh, R. P. (2008). Do People Vote with Their Feet? An Empirical Test of Tiebout's Mechanism. *The American Economic Review*, *98*(3), 843–863. JSTOR.

Bauder, B. (2020, January 22). *Pittsburgh spending $2.6 million on landslide repairs | TribLIVE.com*. https://triblive.com/local/pittsburgh-allegheny/pittsburgh-spending-2-6-million-on-landslide-repairs/

Boykin, R. (2019, October 25). *The Great Recession's Impact on the Housing Market*. Investopedia. https://www.investopedia.com/investing/great-recessions-impact-housing-market/

Brueckner, J. (1987). *THE STRUCTURE OF URBAN EQUILIBRIA: A UNIFIED TREATMENT OF THE MUTH-MILLS MODEL**.

CBRE study: Pittsburgh housing buoyed by eds and meds. (2017, September 13). Pittsburgh Business Times. https://www.bizjournals.com/pittsburgh/news/2017/04/13/pittsburgh-housing-buoyed-by-eds-and-meds-study.html

Checker, M. (2011). Wiped Out by the "Greenwave": Environmental Gentrification and the Paradoxical Politics of Urban Sustainability. *City & Society*, *23*(2), 210–229. https://doi.org/10.1111/j.1548-744X.2011.01063.x

Cinderby, S. (2020). *Lockdown highlights the value of green space in cities* (Perspectives). Stockholm Environment Institute. https://www.sei.org/perspectives/covid19-value-of-green-space-in-cities/

Colleges & Universities in Pittsburgh | About Pittsburgh. (n.d.). Visit Pittsburgh. Retrieved February 22, 2021, from https://www.visitpittsburgh.com/about-us/relocation/education/colleges-universities/

Conway, D., Li, C. Q., Wolch, J., Kahle, C., & Jerrett, M. (2010). A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values. *Journal of Real Estate Finance and Economics*, *41*(2), 150–169.

Czembrowski, P., & Kronenberg, J. (2016a). Hedonic pricing and different urban green space types and sizes: Insights into the discussion on valuing ecosystem services. *Landscape and Urban Planning*, *146*, 11–19. https://doi.org/10.1016/j.landurbplan.2015.10.005

Czembrowski, P., & Kronenberg, J. (2016b). Hedonic pricing and different urban green space types and sizes: Insights into the discussion on valuing ecosystem services. *Landscape and Urban Planning*, *146*, 11–19. https://doi.org/10.1016/j.landurbplan.2015.10.005

Dickinson, J. (2021, February 22). *Commentary: Pittsburgh is America's apartheid city*. PublicSource | News for a Better Pittsburgh. https://www.publicsource.org/commentary-jerry-dickinson-pittsburgh-is-americas-apartheid-city/

Eberson, S. (2018, August 20). *Pittsburgh is the No. 2 most livable city in America*. Pittsburgh Post-Gazette. https://www.post-gazette.com/local/neighborhood/2018/08/20/Pittsburgh-No-2-most-livable-city-America-32-global-liveability-index-Economist/stories/201808200090

Economist Intelligence Unit. (2019, October 22). *The Global Liveability Index 2019*. Economist Intelligence Unit. https://www.eiu.com/n/the-global-liveability-index-2019/

Estabrooks, P. A., Lee, R. E., & Gyurcsik, N. C. (2003). Resources for physical activity participation: Does availability and accessibility differ by neighborhood socioeconomic status? *Annals of Behavioral Medicine*, *25*(2), 100–104. https://doi.org/10.1207/S15324796ABM2502_05

Farrell, P. B. (2015, April 20). *Opinion: The inequality bubble is accelerating, worse than '29, even 1789*. MarketWatch. https://www.marketwatch.com/story/the-inequality-bubble-is-accelerating-worse-than-29-even-1789-2015-04-14

Garcia-Lamarca, M., Anguelovski, I., Cole, H., Connolly, J. J., Argüelles, L., Baró, F., Loveless, S., Pérez del Pulgar Frowein, C., & Shokry, G. (2019). Urban green boosterism and city affordability: For whom is the 'branded' green city? *Urban Studies*, 0042098019885330. https://doi.org/10.1177/0042098019885330

Giles-Corti, B., Broomhall, M. H., Knuiman, M., Collins, C., Douglas, K., Ng, K., Lange, A., & Donovan, R. J. (2005). Increasing walking: How important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine*, *28*(2, Supplement 2), 169–176. https://doi.org/10.1016/j.amepre.2004.10.018

Hasco, L. (2018, April 20). *Here are Pa.'s 20 largest employers, and the biggest employers in each county*. Pennlive. https://www.pennlive.com/news/erry-2018/04/eb7d0c4952761/pas_top_20_employers_and_the_t.html

Heikkila, E., Gordon, P., Kim, J. I., Peiser, R. B., Richardson, H. W., & Dale-johnson, D. (1989). *What happened to the CBD-distance gradient?. Land Values in a polycentric city, Environment and Planning A*.

Henderson, T. (2018, December 12). *The Mystery of Pittsburgh: How Some Shrinking Cities Are Thriving in the New Economy*. https://pew.org/2La3FID

Howell, J., Goodkind, S., Jacobs, L., Branson, D., & Miller, E. (2019). *Pittsburgh's Inequality Across Gender and Race* (p. 96). City of Pittsburgh, Gender and Equity Commission.

Immergluck, D., & Balan, T. (2018). Sustainable for whom? Green urban development, environmental gentrification, and the Atlanta Beltline. *Urban Geography*, *39*(4), 546–562. https://doi.org/10.1080/02723638.2017.1360041

Kakejoub, A. R., Eisalou, A. A., Mosammam, H. M., & Rahmati, A. (2013). An Investigation into Effective Factors Deciding Urban Housing Prices via Hedonic Pricing Model: The Case of District 2 of the City of Sanandaj. *Journal of Urban Economics and Management*, *1*(3), 33–43.

Łaszkiewicz, E., Czembrowski, P., & Kronenberg, J. (2019). Can proximity to urban green spaces be considered a luxury? Classifying a non-tradable good with the use of hedonic pricing method. *Ecological Economics*, *161*, 237–247. https://doi.org/10.1016/j.ecolecon.2019.03.025

Loughran, K. (2014). Parks for Profit: The High Line, Growth Machines, and the Uneven Development of Urban Public Spaces. *City & Community*, *13*(1), 49–68. https://doi.org/10.1111/cico.12050

Mann, E. (2016, March 22). *Story of cities #7: Philadelphia grid marks birth of America's urban dream*. The Guardian. http://www.theguardian.com/cities/2016/mar/22/story-of-cities-7-philadelphia-grid-pennsylvania-william-penn-america-urban-dream

Mills, E. S. (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *The American Economic Review*, *57*(2), 197–210.

Mock, B. (2019, September 20). *How Pittsburgh Fails Black Women in 6 Charts—Bloomberg*. https://www.bloomberg.com/news/articles/2019-09-20/how-pittsburgh-fails-black-women-in-6-charts

NAHB. (2021). *Housing's Contribution to Gross Domestic Product—NAHB*. https://www.nahb.org/news-and-economics/housing-economics/housings-economic-impact/housings-contribution-to-gross-domestic-product

NPS. (2021). *National Park Service History: National Park System Timeline*. https://www.nps.gov/parkhistory/hisnps/NPSHistory/timeline_annotated.htm

Pittsburgh parks ranked 15th in nation | Pittsburgh Post-Gazette. (2020, May 20). https://www.post-gazette.com/local/city/2020/05/20/Pittsburgh-parks-ranked-15th-US-Trust-for-Public-Land-tax-city-council/stories/202005200139

Popovich, N., & Flavelle, C. (2019, August 9). Summer in the City Is Hot, but Some Neighborhoods Suffer More. *The New York Times*. https://www.nytimes.com/interactive/2019/08/09/climate/city-heat-islands.html

Rigolon, A., & Németh, J. (2020). Green gentrification or 'just green enough': Do park location, size and function affect whether a place gentrifies or not? *Urban Studies*, *57*(2), 402–420. https://doi.org/10.1177/0042098019849380

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55. JSTOR.

Roy, S., Byrne, J., & Pickering, C. (2012). A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban Forestry & Urban Greening*, *11*(4), 351–363. https://doi.org/10.1016/j.ufug.2012.06.006

Rust Belt States 2021. (2021). https://worldpopulationreview.com/state-rankings/rust-belt-states

Sheehan, A. (2018, January 30). Homes Hang From Cliff, Owners Blame Nearby Luxury Development. *KDKA*.

https://pittsburgh.cbslocal.com/2018/01/30/kdka-investigates-squirrel-hill-homes-falling-off-cliff/

Stroebel, J. (2016). *What Can Housing Markets Teach Us about Economics?* (No. 4). NBER. https://www.nber.org/reporter/2016number4/stroebel.html

Su, J. G., Jerrett, M., de Nazelle, A., & Wolch, J. (2011). Does exposure to air pollution in urban parks have socioeconomic, racial or ethnic gradients? *Environmental Research*, *111*(3), 319–328. https://doi.org/10.1016/j.envres.2011.01.002

The Economist. (2020, September 12). Covid-19 has forced a radical shift in working habits. *The Economist*. https://www.economist.com/briefing/2020/09/12/covid-19-has-forced-a-radical-shift-in-working-habits

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, *64*(5), 416–424. JSTOR.

U.S. Census Bureau QuickFacts: Allegheny County, Pennsylvania. (2021). https://www.census.gov/quickfacts/alleghenycountypennsylvania

Which city is best for parks? (2020). The Trust for Public Land. https://www.tpl.org/parkscore

Wu, J. (2006). Environmental amenities, urban sprawl, and community characteristics. *Journal of Environmental Economics and Management*, *52*(2), 527–547. https://doi.org/10.1016/j.jeem.2006.03.003