The Neighborhood Brand Effect on Housing Prices

Douglas Woodward[1], Marian Manic[2], William R. Hauk[3], Joseph Von Nessen[4], and Paulo Guimarães[5]

[1]woodward@moore.sc.edu, University of South Carolina, Columbia, SC, USA
[2]marian.manic@rfa.sc.gov, South Carolina Revenue and Fiscal Affairs Office, SC, USA
[3]hauk@moore.sc.edu, University of South Carolina, Columbia, SC, USA
[4]Joey.VonNessen@moore.sc.edu, University of South Carolina, Columbia, SC, USA
[5]pguimaraes@fep.up.pt, Banco de Portugal and Universidade do Porto, Portugal

Abstract

We propose that neighborhoods have a measurable effect on housing prices. In theory, searching for houses by neighborhood can be seen as a heuristic process, reducing the time and effort needed to evaluate a plethora of particular local attributes that are associated with the price premium of the house. To test this hypothesis, we estimate the neighborhood's fixed effect that picks up the time-invariant quality of the area within the framework of a hedonic housing price model. Our data base encompasses more than 50,000 housing market transactions in Charleston, South Carolina. Our regression fits the data extremely well, with the neighborhood fixed effect exerting a pronounced effect on regression results and outperforming significantly alternative local areal units. The fixed effect estimates are ultimately rank-ordered to evaluate their heterogeneous effect on the housing price premium. The analysis reveals that neighborhood fixed effects capture effectively the wide range of house price premiums, from high-priced historic districts and ocean-side communities to low-price, poverty-stricken areas damaged by urban redevelopment. In addition, our approach to using spatial fixed effect estimates as a tool to measure the brand effect of regions can be used to assess the value of regional identity in other contexts like industry location and migration.

The second important contribution we make to the literature is a new method of estimating the Moran's I in large data sets. Based on our simplified Moran's I tests, we show that spatial dependence is effectively eliminated when the neighborhood fixed effects are added to the model. This original econometric tool can be applied to a wide range of urban and regional research.

JEL classification codes: R15, R21, R31, D90

Keywords: hedonic housing models, fixed effects, heuristics, spatial autocorrelation, neighborhoods

1 Introduction

Neighborhoods fundamentally define the geography of cities and regions. New York City's borough of Manhattan is famous for Chinatown, Little Italy, SoHo, the Lower East Side, Greenwich Village, Hell's Kitchen, the Upper West Side, and Harlem, among many others. San Francisco is known for the Mission District, North Beach, SoMa, Nob Hill, Haight Asbury, Pacific Heights, and so forth. Beyond the city, neighborhoods express suburban spatial identities, where residential subdivisions are developed and marketed like brands: the Ascot, Forest Acres, or Wild Dunes. On the downside, neighborhood designations may signify crime and social dysfunction, like the South Bronx in New York City and the Tenderloin in San Francisco. Love Canal in western New York State illustrates how an environmental disaster can damage a neighborhood's image.

Many studies related to housing and housing prices use various spatial units in the form of fixed effects to control for heterogeneity across neighborhoods. Interestingly, researchers tend to ignore the magnitude and patterns of these fixed-effects estimates. In turn, we propose that fixed-effects estimates may capture neighborhood identities that not only explain housing price variation but can also be used to measure the brand effects of each neighborhood. Thus, this paper advances the hypothesis that neighborhood choice is central to housing market transactions. Buyers, for example, start with preferences for the structural characteristics of a house and then search for properties within desirable neighborhoods. They have an implicit willingness to pay for houses in particular neighborhoods, with a higher price for houses with a better brand or reputation, all else being the same.

The neighborhood choice overcomes the challenge of evaluating a surfeit of local data. Following behavioral economics, the neighborhood's identity or brand can be viewed as a time- and effort-reducing heuristic for decisions where local area information is incomplete, uncertain, and costly to evaluate (Simon, 1978). Essentially, neighborhoods denote a package of local characteristics, amenities, public goods, and externalities that assist buyers and sellers in sorting and selecting properties and negotiating home prices.

The features that define the neighborhood identity – its quality and relative sophistication – are manifold. They have a similar proximity to central business districts, consumption amenities, and recreational opportunities. Other common attributes include transit access and consistent road surfaces, sidewalks, and walkability. Typically, neighborhoods also express relatively homogeneous socioeconomic characteristics, including ethnic composition and household income levels. They share municipal services (water and sewer, fire, and public safety) and are subject to the same land-use regulations (residential and commercial zoning). Neighborhoods face common environmental hazards like flooding, along with other negative externalities that can influence housing prices like crime and pollution (air, water, and noise).

Beyond common location attributes, neighboring homes will likely have consistent structural features like siding, roofing, and other exterior and interior design characteristics. Some neighborhoods will have homeowner associations and community centers. Many neighbors now interact online through defined groups on Facebook, Nextdoor, and other internet-based platforms. The aggregate of these many aspects of place engender an established reputation and even a brand.

Thus, from the perspective of behavioral economics, we should expect that the prices of individual houses are correlated within neighborhoods. There are economic spillovers influencing home prices that are external to the home but internal to the neighborhood. Accordingly, in hedonic regression models housing prices are subject to geographical dependence, as recognized in prior research (Sedgley, Williams, and Derrick, 2008; Nguyen-Hoang and Yinger, 2011; Wilhelmsson, 2002; Ismail, 2006; Moralı and Yılmaz, 2020). We propose that neighborhood identities represent an appropriate control for spatial autocorrelation.

In hedonic housing price modeling, moreover, we expect the neighborhood to provide greater explanatory power than areal units like school districts, zip codes, or census tracts that are often used in research to account for otherwise unobserved local attributes. Yet governments often do not officially delineate neighborhood boundaries, unlike other areal units.

While some regions record the neighborhood in property tax assessor records, the identities are usually defined by private-sector real estate services and web sites. Real estate agents organize their listings into neighborhoods, recognizing the limitations of market participants' ability to appraise the myriad of local characteristics and make comparisons across the relevant region. In the United States, the neighborhood is commonly defined by the Multiple Listing Service (MLS). Through the MLS, real estate brokers disseminate information about properties that are identified according to branded neighborhood appellations.

Using MLS data, we investigate the influence of neighborhood identity through a model of housing prices in Charleston, South Carolina. The neighborhood enters the model as a fixed effect. This variable picks up the neighborhood's long-term brand identity based on attributes that do not change during the period. In our housing price regression, we use the fixed effect estimate to measure the brand effect of neighborhoods on house price premiums. As expected, we find a highly skewed variation in the housing values across the different neighborhoods, with the highest-priced properties located in neighborhoods near the ocean and the downtown area of Charleston substantially boosting home prices.

Importantly, it is our expectation that any hedonic housing model without neighborhood fixed effects will encounter spatial autocorrelation, given that houses situated in neighborhoods share many similarities. The individual observations are residences that *belong* to a neighborhood, so the error terms in regression analyses are clearly not independent. Anselin and Arribas-Bel (2013) argue that an *a priori* justification for fixed effects as a control for spatial autocorrelation has been missing before in housing research. Thus, our heuristic view of neighborhoods gives a theoretical basis for the empirical strategy using fixed effects to resolve spatially correlated omitted variables.

An accepted spatial autocorrelation measure is Moran's I (Wilhelmsson, 2002). With large data sets, however, estimating Moran's I poses intractable computational problems. Our sample contains 50,174 housing market transactions in the three-county Charleston metropolitan region. With this whole data set, it becomes cumbersome, if not impossible, to calculate and manipulate the spatial weights matrix with distances between every pair of observations, as required by the spatial autocorrelation test. That would involve a 50,174 x 50,174 matrix. Thus, we simplify the spatial weights matrix by assuming that some of the observations share the same location. This is done by rounding off the latitude and longitude coordinates. Essentially, we assume that if two houses are in close proximity then they are at the same location. When we follow this method we have 14,178 unique locations. With a consolidated set of locations, the spatial weights matrix is tractable. We adjust Moran's I formulas and calculate the spatial statistics associated with each of the specifications. This new method is described in the appendix. Using this approach, we find that the neighborhood fixed effect resolves the spatial autocorrelation issue.

Unsurprisingly, adding the neighborhood fixed effect to the structural housing characteristics alone (without any additional local area characteristics) increases the model fit and explains a significant share of the variation in Charleston housing prices. In our analysis, we then examine characteristics that may possibly further explain housing price variation in Charleston. In particular, we test measures of school quality, ethnic composition, as well as distances to major urban and natural amenities. School quality ratings serve as examples of local attributes that may not be picked up in the time-invariant, neighborhood fixed effect. School quality and local demographics are statistically significant, but the neighborhood remains a predominant factor explaining housing price variation across the Charleston region. Moreover, we control for proximity to fundamental amenities in the metropolitan area, including the business center (with all the major financial, administrative, education, healthcare, shopping, cultural, and entertainment institutions) and the airport (as a major transportation hub). We also test the distance to the ocean front (important in Charleston). While these variables are statistically significant, neighborhoods retain their relative effect, which supports our view that these spatial units reflect a heuristic (or brand) that is relevant above and beyond the myriad of attributes in the house buying process.

In essence, this paper's contributions are threefold. First, we recast housing market decisions as a heuristic process based on neighborhood identities. Previously, behavioral economics has not been used to form expectations about housing price determination. Our principal hypothesis is that the neighborhood is a differentiated identity or brand and should have a distinct, quantifiable effect on housing prices. Most previous research using hedonic housing models focuses on estimating the influence of particular local amenities, public goods, or externalities, of which there are many. Yet, the multitude of possible attributes considered in the optimization formulation of housing price determination overshadows the neighborhood itself as a singular determinant. As the adage goes, it is hard to see the forest for the trees.

The paper's second major contribution is demonstrating how fixed effect estimates can identify the relative price premiums demanded across the different neighborhoods. The estimates for each neighborhood can be rank-ordered to identify how the price premiums across the different neighborhoods compare to the average neighborhood. Superstar communities near the ocean and Charleston's renowned downtown area boost housing prices, substantially in some cases. Meanwhile, neighborhoods suffering from large negative externalities (noise, congestion, and pollution) have significant negative effects on house prices. While the relationships between natural or urban amenities and house prices have been studied extensively already, the fact that we can replicate these results based on a single fixed effect that encapsulates all these qualities is novel.

Third, we postulate that all hedonic housing models without fixed effects will necessarily face spatial autocorrelation. We develop a new technique for estimating the Moran's I with large data sets. Using this measure, we find spatial autocorrelation is effectively eliminated once we enter the neighborhood fixed effect in the model specification.

The remainder of this paper is organized as follows. In the next section, we review the literature on heuristic decision-making and the role of neighborhoods in hedonic housing price models. Section 3 explains the database and the model, including the methodology behind the empirical strategy that stresses the influence of the neighborhood fixed effect.

Next, Section 4 presents the results, highlighting our findings about neighborhood implicit prices. The analysis in this section includes an application of our measure of Moran's I (as explained in the appendix) and shows how fixed effects reduce spatial autocorrelation. Section 5 concludes the paper.

2 Neighborhood Heuristics and Housing Models

While the neighborhood choice is central in the housing price decision, it often plays a subordinate role in the empirical housing market literature. Based on neoclassical utility maximization theory, empirical housing research is mostly concerned with the attributes represented by the vector that contains the home's neighborhood characteristics (Li and Brown, 1980). The canonical Rosen (1974) article on hedonic modeling stresses that home buyers "purchase the right to consume a bundle of local public goods" (Kuminoff and Pope, 2014, p. 1227). Each individual property is cast as a differentiated good, with the price determined by the varying characteristics of the structure, size, and location attributes. The location attributes provide utility and their implicit price premiums can be determined through regression analysis.

Consider a straightforward empirical approach to housing price determination using a flexible linear regression model. The model regresses the natural log of the sale price of the house against varying sets of explanatory variables. The reduced form of the hedonic model can be written as:

$$P_i = \alpha\theta_i + \beta X_i + \gamma_i + \varepsilon_i \ ,$$

where P_i represents the natural log of the sale price of house i; θ_i is a vector containing structural attributes of houses; X_i contains all neighborhood characteristics; γ_i is a fixed effect that controls for factors that are unique to each neighborhood; and ε_i is an error term, in part reflecting unobserved neighborhood and housing structural characteristics.

In specifying the model, neoclassical economic theory does not identify the specific neighborhood attributes that are essential or clearly predict the anticipated effect. To be sure, the list of neighborhood characteristics is long and varied: distance to shopping, distance to mass transportation and road arteries, school quality, crime, pollution, proximity to leisure amenities, and even distance to marijuana dispensaries (Basu and Thibodeau, 1998; Figlio and Lucas, 2004; Bogart and Cromwell, 1997; Brasington, 1999; Carlino and Saiz, 2019; Tyndall, 2019; Conklin et al., 2020). Moreover, the preferences for amenities are heterogeneous. For example, it is possible that some home owners value public school quality, which raises property values (Nguyen-Hoang and Yinger, 2011). At the same time, others may not care because they have no school age children or prefer private schools. Further, household preferences for neighborhood ethnic composition vary (Yinger, 2016).

Early hedonic models attempted to encapsulate individual location characteristics into fewer variables by using factor analysis (Kain and Quigley, 1970). Still, even with available data it is not clear what variables should be included (Kuminoff, Parmeter, and Pope, 2010). It is one thing to proclaim that a complete set of spatial characteristics must be included in any well-specified model. It is another to find data that capture the relevant neighborhood attributes, along with environmental and other local characteristics that are necessary to produce accurate estimates of relevant coefficients.

Hence, although researchers depict the housing market decision as involving a scan and assessment of all relevant information, they will inevitably fail to include all neighborhood characteristics. As a practical matter, researchers face considerable data limitations. In turn, empirical models face possible misspecification and omitted variables. Adding characteristics will potentially mitigate the omitted variable problem, but there are still severe constraints, given a lack of accurate measures for many variables. As a result of omitted variable bias, the estimation of the model will likely involve spatial autocorrelation. In any hedonic estimation without adequate controls, the residuals will be spatially correlated (Basu and Thibodeau, 1998).

A possible solution to the problem of spatially correlated omitted variables for cross sectional data is adding local area fixed effects to the hedonic housing price model (Ismail, 2006). In a simulation analysis, Kuminoff et al. (2010) find that spatial fixed effects offer the best approach for dealing with spatial autocorrelation in cross-sectional models (for a critique see Anselin and Arribas-Bel, 2013). Often, contemporary models test various local area fixed effects as controls. Drawing out the implications of the spatial unit on housing prices, however, is not part of the analysis. Instead, most research is concerned with using housing price regression to estimate the implicit price prmiums of specific neighborhood amenities, externalities, or local public goods.

Another complication is that varying neighborhood proxies are used in hedonic models to absorb influences of omitted variables. Carlino and Saiz (2019) study local consumption amenities (focusing on the distance to central recreational districts), using census tracts as the neighborhood proxy. Many other papers use census tracts or blocks as well (Pope, 2008; Leonard and Murdoch, 2009; Bogin and Nguyen-Hoang, 2014). In a recent neighborhood choice model, Anenberg and Kung (2020) rely on U.S. census-designated PUMA districts because they are the most disaggregated spatial units reported in the American Community Survey (ACS). A paper on the influence of aquatic invasive species on housing values defines the neighborhood according to the lake near where the properties are located (Horsch and Lewis, 2009).

Alternatively, in a study related to housing and local amenities, Kuang (2017) identifies 131 "well-defined" District of Columbia neighborhoods (which they argue have a distinct history, culture, and demographic profile). The neighborhoods are listed by the local government. The authors maintain that the neighborhood is the correct spatial unit for housing market studies, yet much of the analysis is based on census tracts because that is the source of data for local amenities. In other studies, the neighborhood fixed effect is based on the MLS definitions used in real estate transactions (Zahirovic-Herbert and Turnbull, 2009). Yet evaluating the impact of the neighborhood itself is not part of the analysis.

The MLS-defined neighborhood would seem most appropriate, given the direct role it plays in the actual housing decision. We posit that the MLS-neighborhood fixed effect best picks up the place-based identity or brand. As we argue in the introduction, this identity saves time and effort in housing market decisions. It serves as a cognitive shortcut for decisions where local area information is too complicated to evaluate at a granular level. Accordingly, we consider γ_i to be more than a control for hedonic housing price models. In our view, the neighborhood fixed effect reflects the actual heuristic buyers and sellers use to streamline housing market decisions.

In behavioral economics and psychology, heuristics represent a form of decision making under bounded rationality. In his Nobel prize address on bounded rationality, Simon (1978) avers that neoclassical assumptions of omniscient rationality are simplistic. In contrast with this "arm-chair" theorizing about human behavior and decision making, he posits that the notion of bounded rationality is based on realistic assumptions about cognitive and computational limitations.

One of the earliest formalizations of heuristics was advanced by Tversky and Kahneman (1986), who analyze judgment under uncertainty. The authors explain that in making estimations and predictions people apply simplifying heuristics by relying on intuitive cognitive processes. The heuristics approach has been applied to understand how consumers use brands to make decisions about products and services when faced with various complex buying decisions. Most of the studies link heuristics with the process of evaluating product or service quality under uncertainty. Erdem, Swait, and Louviere (2002) explain that consumers generally exhibit uncertainty about products when there is imperfect or asymmetric information. In such situations, brand names play a distinct role in learning and evaluating the multitude of product or service attributes.

Allison and Uhl (1964) are among the original researchers to look into the relationship between brand names and consumer preferences. In an experimental study of beer brand identification on consumers' taste preferences, Allison and Uhl find that only when participants are also provided with labels are they able to differentiate among various beer products (product attributes alone may not be enough for proper differentiation). This pioneering study led to a series of other research papers that investigate the relationship between brand names and the effectiveness of product quality evaluation. Jacoby, Olson, and Haddock (1971) carry out an experimental study also involving beer tasting in which the authors test the relative importance of price, brand name, and other product characteristics on consumers' ability to evaluate beer quality. They are also among the first researchers to imply that consumers use a set of heuristics (cues or signals) when evaluating specific quality information. The main finding of this study is that the brand name dominates over price information as a cue used to evaluate quality.

In a follow-up study, Jacoby, Szybillo, and Busato-Schach (1977) further conceptualize how consumers use heuristics in product quality evaluation. This paper probes the depth of information acquisition required for properly evaluating quality; in other words, the number of items of information that consumers may need for the evaluation decision. Jacoby et al. (1977) also examine the relative importance of brand names in a multi-attribute set that includes price and other information signals. The researchers find that when consumers are faced with a large number of product attributes (more than 10) they tend to focus only on a subset of this information. Crucially, the brand name appears to be the single most important information signal used for quality evaluation. These findings support our hypothesis that the neighborhood identity (or brand) is fundamental in housing choice.

A further explanation of the importance of heuristics in consumer decision-making processes can be found in Dawar and Parker (1994). The authors identify a set of so-called "marketing universals," which are defined as consumer behaviors related to specific products or services that are constant across cultures. They note that due to the time constraints, the need for reducing the perceived risk of the purchase, and the lack of expertise, consumers are highly likely to rely on heuristics when assessing the product or service quality. This study is similar to those done by Jacoby et al. (1971, 1977) as it considers the relative importance of brand name over price, physical appearance, and retailer reputation when evaluating quality. Their findings are robust to those of earlier studies in the sense that brand name cues dominate all other product attributes.

In the housing literature, the construct of branding has been applied in a study based on MLS data from Baton Rouge, Louisiana. Zahirovic-Herbert and Chatterjee (2011) investigate the relationship between the inclusion of the words "country" and "country club" in the names of neighborhoods and the price of houses in such neighborhoods. The paper draws on the concept of "conspicuous consumption," which the authors define as the additional utility that housing market consumers gain by showcasing their greater affluence when purchasing a house located in a specific neighborhood regardless of the house's characteristics. Zahirovic-Herbert and Chatterjee observe that real estate is usually the most expensive product individuals purchase during their lifetime and that the price premiums of these products are large in absolute terms, which may have a significant impact on the ultimate purchasing decision. This study yields interesting results showing that there are significant price premiums for properties located in prestigious neighborhoods. These findings provide further motivation for our emphasis on the brand impact of the neighborhoods on housing prices in general.

In another relevant study, Leonard and Murdoch (2009) employ a hedonic model to estimate the effect of neighborhood quality on housing prices. The authors note that the actual value of a neighborhood's quality is not readily observable and there is little information on the relative effect of neighborhood qualities on housing prices. Leonard and Murdoch further point out that the unobservable neighborhood quality may be approximated through the level of nearby foreclosures. A higher count of foreclosure events nearby may be associated with a lower quality of the neighborhood, and vice versa, which in turn impacts house prices. The study uses three levels for neighborhoods: local, census block, and school district. The latter two levels have distinct boundaries that are externally defined, but the "local" geography of a neighborhood is determined using the "sliding neighborhood" approach (conceptualized in

Dubin, 1992) within distance ranges of 250 to 1,500 feet (using 250 feet increments). The study controls for spatial autocorrelation as well as temporal trends of housing sales using various estimation methods. The main finding is that the variables identifying neighborhood quality (counts of foreclosure events in concentric regions ranging from 250 to 1,500 feet from each housing unit in the data set) have negative and statistically significant coefficients. Moreover, the authors find that the magnitude of the foreclosure effect is highest inside the inner ring of 250 feet.

To our knowledge, the only other papers that relate to our main concern about neighborhood identity are studies of the relationship between the historical neighborhood status and housing prices. Asabere and Huffman (1994), Clark and Herrin (1997), and Diaz, Hansz, Cypher, and Hayunga (2008) employ hedonic modeling to assess the impact of historic status designation on property values. As Zahirovic-Herbert and Chatterjee (2011) note, these authors consistently conclude that the historic designation of the neighborhood exerts a significant impact on price premiums for houses, a finding that we observe in our current study as well.

In sum, hedonic housing models typically follow the neoclassical utility maximization paradigm, in which rational agents assemble complete information regarding location attributes. Behavioral economists allege, however, that agents have bounded rationality, with limited time and computational abilities. Facing cognitive overload, agents rely on heuristics. This behavioral economic hypothesis underlies empirical studies probing the importance of brands in consumer decision making.

Our empirical strategy is to test housing prices against the structural characteristics of the home (size, age, bedrooms, and so forth) and of the location (demographics, school quality, and proximity to amenities), along with neighborhood fixed effects, which estimate the actual heuristic that agents have devised and used in the market. In our model, the fixed effect adjusted by the effect of average housing and amenity characteristics within the neighborhood captures the implicit price premium of the neighborhood as a whole; that is, the time-invariant character of the local area.

3 Data and Empirical Modeling

3.1 Data

To estimate the neighborhood identity and its price premium, we compiled data from the MLS in Charleston, South Carolina from 2003-2007. The MLS database is a marketing tool that seeks to lower transaction costs for buyers and sellers. When a seller hires a real estate agent to assist in the sale of a house, the listing agent will enter the house and parcel attributes into the MLS database, which will include square footage and housing structural characteristics. The realtors who report the MLS data have an incentive to present an accurate appraisal of the house they are selling to prospective buyers and usually inspect the house themselves before listing it. To assess the housing market determinants, we use the structural characteristics available in the MLS data set. Also included is the seller's asking price. Upon the completion of a sale, the selling price of the house, date of sale, and updated square footage (often taken from an appraisal) are entered as well. Crucial to our study, the MLS lists the neighborhood/subdivision designation for the property. The neighborhoods identified in the data base are widely used in practice since the MLS is the main source of information about homes for sale in the United States.

There are 829 neighborhoods/subdivisions defined in our database for the Charleston, South Carolina metropolitan region (Berkeley, Charleston, and Dorchester counties). We use the complete data set available for 2003 through 2007, except for 120 observations where the address could not be verified. All the MLS housing data records were checked for location accuracy using Google Map search algorithms. Several thousand observations had mis-spellings for neighborhood/subdivision names and for the street addresses. These were corrected so we could pinpoint the precise locations. We then determined the geocode (latitude and longitude) for each housing observation, which is necessary in order for us to perform a reliable spatial autocorrelation test on our regression estimation (see the Appendix section for the detailed discussion on our original methodology to compute the Moran's I statistic in such a large data set).

Our housing market data set contains 50,174 observations for single-family home sales in the three counties. We have considerably more housing transactions than Zahirovic-Herbert and Turnbull (2009) and other studies, which nonetheless obtain robust estimates. Besides neighborhood, the MLS variables associated with each observation used in our analysis are:

- Sales price
- Number of bedrooms
- Number of full bathrooms
- Number of half bathrooms
- Acreage
- Heated square feet
- Year built
- Exterior materials used (brick, vinyl, etc.)
- Stories
- Antebellum (historic) home
- Zoned high school and elementary school

3.2 Model Specification

We evaluate the neighborhood identity in a hedonic model where the natural log of the sale price of the house is regressed as a function of the explanatory variables. We build five specifications of the model to determine the incremental impact of various characteristics on the sales price. As the following equations show, we start with the home size and structural characteristics and then add the neighborhood fixed effect in equation (3). Equation (4) follows with another five covariates capturing demographic and education-related factors discussed below. In our final model presented in expression (5), we further include three additional distance covariates. These specifications are as follows:

$$P_i = \beta_1 S_i + \tau_i + \alpha_i + \varepsilon_i \tag{1}$$

$$P_i = \beta_1 S_i + \beta_2 X_i + \tau_i + \alpha_i + \varepsilon_i \tag{2}$$

$$P_i = \beta_1 S_i + \beta_2 X_i + \gamma_i + \tau_i + \alpha_i + \varepsilon_i \tag{3}$$

$$P_i = \beta_1 S_i + \beta_2 X_i + \phi Z_i + \zeta SQ_i + \gamma_i + \tau_i + \alpha_i + \varepsilon_i, \tag{4}$$

$$P_i = \beta_1 S_i + \beta_2 X_i + \phi Z_i + \zeta SQ_i + \theta D_i + \gamma_i + \tau_i + \alpha_i + \varepsilon_i, \tag{5}$$

where, P represents the natural logarithm of the sale price of house i, S represents the size (square footage) of the house, X is a vector containing all other structural housing characteristics, and τ and α are temporal fixed effects (month and year).

The neighborhood fixed-effect is represented by vector γ. This parameter is our main concern, reflecting an actual heuristic encompassing a multitude of local attributes that define a neighborhood's value: sidewalks, natural amenities, water quality, transport access, homeowner associations, neighborhood community centers, and much more. In essence, this fixed effect represents the perceived neighborhood identity and allows us to estimate its implicit housing price premium (we go over the approach to determine the implicit neighborhood price premium in section 4.2). Wooldridge (2016) calls the model in (3) - (5) an unobserved effects model (or a dummy variable regression model). According to Wooldridge, the dummy variable regression yields the exact same coefficient estimates that would be obtained from a regression on time-demeaned data. There are several advantages to estimating the fixed effects this way. Beyond the added benefit of reducing omitted variable bias, another important advantage observed by Wooldridge (2016) on p. 438 is that there may be situations when the estimated intercepts, $\hat{\gamma}$, provide relevant information. This is specifically the case when researchers want to study the distribution of the $\hat{\gamma}$ and determine how these estimates compare to the average or median value in the sample, which is one of the central objectives of our current study.

In addition to the role of capturing the 'brand' effect on housing prices, we also use neighborhood fixed effects to control for spatial autocorrelation. Autocorrelation stems from the fact that when individual observations are clustered together spatially, the error terms stop being independent. In regional studies, the Moran's I is a popular measure to test for the presence of spatial autocorrelation. However, estimating Moran's I in large data sets becomes computationally problematic: a data set of 50,174 housing market transactions implies computing a 50,174 x 50,174 spatial weights matrix with distances between every pair of observations.

However, a solution to this complication could be to assume that the houses that are close to each other share the same location. This is achieved by rounding off the latitude and longitude coordinates for the location of each house. Applying this technique, we reduce the number of unique locations to 14,178, which allows us to compute the spatial weights matrix in a tractable way. We provide the full details of how we adjust the Moran's I formulas and calculate the spatial statistics in the Appendix section.

While the neighborhood parameter picks up the time-invariant value of the location, there are no doubt other factors in the Charleston metropolitan area that could alter the desirability of certain areas. In particular, it is possible that home prices are affected by the demographic makeup and school quality reputation of the home's location as well as its proximity to various urban and natural amenities.

Vector Z (included in specification 4) captures local demographic characteristics: total population size and the share of the population that is Black and Hispanic in the census tract. Hedonic housing prices research suggests that the local ethnic composition can matter, although preferences diverge substantially (Yinger, 2016). A negative effect on housing prices may be expected because of prejudice and discriminatory practices restricting residential loans in some neighborhoods because of the racial or national origin makeup of the community. Notorious redlining practices began in the 1930s, with the Home Owners' Loan Corporation grading neighborhoods into categories according to the racial composition. Predominately Black neighborhoods were marked red, which denoted a high-risk for mortgage lenders. The long-lasting effect of discriminatory lending practices that limit home ownership may be picked up by the neighborhood fixed effect. Nevertheless, given the strong effects that have been shown in previous literature, we include the ethnicity variables in our model.

Local school quality should also be evaluated, given the prominent role it plays in many housing purchase decisions and the large literature showing its influence (Nguyen-Hoang and Yinger, 2011). Vector SQ (also included in specification 4) denotes school quality measures. Many attributes of the location are difficult to assess and evaluate, but school quality is generally well known. We measure school quality with No Child Left Behind (NCLB) Adequate Yearly Progress (AYP) ratings, *ayp*, and with average SAT scores, *satscore*. No Child Left Behind was a major effort to evaluate school quality during the early 2000s. School ratings were widely reported in the press and, as a result, may have affected home values. The South Carolina Department of Education has data indicating whether each school successfully achieved AYP in each year since 2003 (when No Child Left Behind ratings were first available). We also test the local school quality with average SAT scores (in mathematics and reading) for the high school in which the house was zoned. Each house sale is matched with SAT score data from South Carolina Department of Education data.

One extra final group of variables, identified by vector D, we control for in our analysis (included in specification 5) is related to the proximity of the house (in miles) to various urban and natural amenities. We identify three types of distances: 1) distance to Charleston's business center (*buscenter_dist*); 2) distance to Charleston International Airport (*airport_dist*); and 3) average distance to the beach (*avgbeach_dist*) measured as the average of the distance to the two most popular beaches in Charleston, Folly Beach and Isle of Palms Beach. We calculate distances using the great circle method based on the longitude and latitude coordinates for each house and the respective coordinates of the three amenities.

Table 1 provides basic descriptive statistics. The mean values show that the average

Table 1: Descriptive Statistics

	Mean	Minimum	Maximum
Sales price (*sold_price*)	288,374.42	50,000.00	6,500,000.00
Size of the house in heated square feet (*tot_heat_sqft*)	1981.02	381.00	20,000.00
Number of bedrooms (*bedrooms*)	3.42	1.00	6.00
Number of full baths (*baths_full*)	2.13	1.00	6.00
Number of half baths (*baths_half*)	0.49	0.00	4.00
House built before 1861 (*antebellum*)	0.01	0.00	1.00
House built between 1960 and 1980 (*built19601980*)	0.14	0.00	1.00
House built between 1980 and 2000 (*built19802000*)	0.31	0.00	1.00
House built in 2001 or later (*built2001plus*)	0.46	0.00	1.00
House used brick on exterior (*brickdummy*)	0.31	0.00	1.00
House has 2 stories (*story2*)	0.44	0.00	1.00
House has 3 stories (*story3*)	0.02	0.00	1.00
House has 1.5 stories (*storyhalf*)	0.02	0.00	1.00
Census tract 2000 population (*pop2000*)	6,101.76	350.00	13,097.00
% of census tract pop. that is black in year 2000 (*blackpercent*)	23.73	0.33	97.49
% of census tract pop. that is hispanic in year 2000 (*hispanicpercent*)	2.29	0.24	15.70
Average score on SAT reading/math (*satscore*)	489.14	362.00	530.75
AYP pass/fail for zoned school (*ayp*)	0.16	0.00	1.00
Distance to the business center in miles (*buscenter_dist*)	13.56	0.16	50.42
Distance to the airport in miles (*airport_dist*)	10.04	0.69	41.90
Average distance to the beach in miles (*avgbeach_dist*)	19.24	6.38	56.80

house in the data set sells for $288,374, has 3.4 bedrooms, 2.1 full bathrooms, and 1,981 heated square feet. In addition, approximately 91 percent were built after 1960, 31 percent contain some exterior brick, and 48 percent have more than one story.

4 Results

4.1 Regression Estimates

Our empirical strategy is to determine the extent to which neighborhood identities explain housing price variation and to evaluate the implicit price premium of these identities based on fixed effect estimates. With the hedonic regression model outlined in the previous section, we measure the proportion of Charleston region's price variation that can be attributed to the neighborhood. We then take a closer look at the fixed effect results and contrast

MLS-neighborhoods with alternative fixed effects (census tracts, census blocks, zip codes, and school districts). Finally, we assess how the estimated neighborhood price premiums vary across the region and compare them to the average value (as explained in section 4.2). This analysis reveals the implicit neighborhood price premiums across the Charleston metropolitan area.

Table 2 displays the results for our five specifications. The first specification is restricted to a single explanatory variable, the square footage of the house, as well as the month and year fixed effects. In column (1) of this table, we observe that square footage alone along with temporal fixed effects explain 59.3 percent of the variation in housing prices. The second specification then adds housing structural characteristics. In column (2), we can observe that this addition causes the model's adjusted R^2 to rise to 66.6 percent. Although the reported coefficients for all of the structural characteristics are statistically significant, some maintain unexpected signs since no location-based factors have yet been introduced.

The third specification introduces the neighborhood fixed effect. Adding this fixed effect significantly alters the results, as shown in column (3) of Table 2. The adjusted R^2 increases to 92.5 from 66.6 percent. In addition, in contrast to the first two specifications, many of the coefficients change once the neighborhood is introduced. Notably, the estimates for the housing structural characteristics now have the expected signs. Housing square footage and the number of bathrooms both matter, as does the Antebellum (pre-Civil War) historic status of homes built before 1861 for which Charleston real estate is famous. Note that *bedrooms* has a negative and statistically significant coefficient. One might expect an additional bedroom to have a positive impact on housing prices, but adding a bedroom to a house while holding everything else constant (including the square footage) may decrease the house sale price. The reason is that for each additional room, all other rooms must become smaller to compensate. Thus, a homeowner might prefer two larger bedrooms to three smaller bedrooms, which would make the negative coefficient on the *bedrooms* variable reasonable.

Charleston's housing price variation appears to be largely determined by the fundamen-

Table 2: Regression Results

Variable	(1)	(2)	(3)	(4)	(5)
tot_heat_sqft	0.000622***	0.000507***	0.000330***	0.000330***	0.000330***
	(0.000007)	(0.00002)	(0.00001)	(0.00001)	(0.00001)
bedrooms		-0.135***	-0.0254***	-0.0252***	-0.0249***
		(0.00541)	(0.00360)	(0.00359)	(0.00358)
baths_full		0.279***	0.0818***	0.0807***	0.0805***
		(0.00912)	(0.00554)	(0.00552)	(0.00550)
baths_half		0.0831***	0.0301***	0.0300***	0.0304***
		(0.00740)	(0.00455)	(0.00453)	(0.00453)
antebellum		0.318***	0.0964***	0.0969***	0.0928***
		(0.0337)	(0.0258)	(0.0257)	(0.0255)
built19601980		-0.223***	0.00805	0.00435	0.00492
		(0.00992)	(0.00815)	(0.00807)	(0.00792)
built19802000		-0.222***	0.0807***	0.0774***	0.0780***
		(0.00916)	(0.00877)	(0.00868)	(0.00850)
built2001plus		-0.329***	0.148***	0.147***	0.153***
		(0.00921)	(0.00891)	(0.00884)	(0.00868)
brickdummy		-0.112***	0.0238***	0.0234***	0.0254***
		(0.00433)	(0.00246)	(0.00245)	(0.00243)
story2		0.0280***	-0.00551*	-0.00580*	-0.00625**
		(0.00585)	(0.00305)	(0.00304)	(0.00302)
story3		0.137***	0.00443	0.00278	0.00277
		(0.0173)	(0.00960)	(0.00958)	(0.00954)
storyhalf		0.102***	0.00301	0.00333	0.00221
		(0.0120)	(0.00593)	(0.00587)	(0.00585)
pop2000				-0.000001	-0.000001
				(0.000001)	(0.000001)
blackpercent				-0.000395*	-0.000303
				(0.000219)	(0.000205)
hispanicpercent				-0.0195***	-0.0118***
				(0.00250)	(0.00222)
satscore				0.00110***	0.00121***
				(0.000135)	(0.00012)
ayp				0.0349***	0.0342***
				(0.00284)	(0.00283)
buscenter_dist					-0.0276***
					(0.0063)
airport_dist					0.0232***
					(0.00205)
avgbeach_dist					0.00765
					(0.00532)
Temporal FE	Yes	Yes	Yes	Yes	Yes
Neighborhood FE	No	No	Yes	Yes	Yes
Observations	50,174	50,174	50,174	50,174	50,174
Adj R2	0.593	0.666	0.925	0.926	0.927
Morans' I	0.236	0.179	0.00740	0.00724	0.00713
Moran statistic	3,028.7	2,341.6	95.05	93.16	91.87

Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

tals of age, size, structure, and neighborhood. These findings confirm our central hypothesis that the neighborhood quality has a pronounced and quantifiable impact on housing prices.

The fourth and fifth specifications shown in columns (4) and (5) of Table 2 reveal that the results do not change substantially when we add more covariates to the fixed effect model. Note that the adjusted R^2 is approximately the same as in column (3) after introducing variables that capture demographic and school quality factors as well as distances to various amenities. Specifically, column (4) displays five new covariates, most of which are statistically significant. The coefficients capturing race are all negative, implying that regions with a greater Black or Hispanic percentage are correlated with lower housing prices. Nevertheless, the absolute magnitude of these effects is relatively small. For example, these results suggest that a percentage point increase in the percentage of the population that is Black – all else being equal – is associated with a decrease of housing prices of about 0.04 percent (although the coefficient for the percentage of population that is Hispanic appears to indicate a stronger effect of about 2 percent).

The coefficients capturing school quality are both positive and statistically significant. The specific estimate capturing the effect of changes in average standardized test scores, *satscore*, is generally consistent with the effects of SAT test score results uncovered in previous research (Sedgley et al., 2008). The second measure, *ayp*, shows that home buyers were willing to pay approximately 3.5 percent more for homes zoned for a high school that had achieved the AYP rating. The AYP rating expressed new information on local educational quality that was made available to the public from 2003-07. No Child Left Behind generated considerable media attention across the United States. As such, *ayp* appears to have had an effect on housing prices in the first years after NCLB was implemented. It should be noted, however, that *ayp* is a measure of marginal annual progress, not average school quality. Moreover, the improvement is assessed relative to the school itself, not to other schools. Because it is simply reported to the public on a pass-fail basis, it can actually be a confusing measure of school quality.

This result about AYP supports the findings of Bogin and Nguyen-Hoang (2014) in an investigation of poor neighborhoods in the Charlotte, North Carolina metropolitan area during the early and mid-2000s. Their results suggest that the property value impact of a "failing" AYP designation is initially distorting by leading to a six percent price decline. However, the housing price effect of a failing AYP designation declines over time and gradually returns to a normal trend. As for the AYP rating that we tested, once home buyers became aware of the controversial nature of NCLB later in the decade and the fact that it was an imperfect measure of school quality, the program's influence over home prices disappeared. Our findings suggest that there can be transitory influences on home prices that have some degree of explanatory power beyond neighborhood identity.

One caveat about these estimates is in order. Billings (2015) shows that without complete data that includes home renovations, hedonic models may yield biased estimates for housing structural characteristics and neighborhood amenities. Yet the paper found that the omitted renovation bias is small and a "second-order concern" in research design for hedonic housing price regressions (Billings, 2015, p. 689).

Finally, in column (5) of Table 2 we also include the estimates for the effect of proximity to three different types of amenities. Interestingly, even after controlling for these distances, our main results remain mostly unchanged. All the other coefficients maintain their signs and estimated values in the previous specifications. Of the three distances, only the distance to the business center and the distance to the airport are statistically significant.

The signs of the estimated distance coefficients also seem to follow normal economic expectations. Thus, an increase in the distance to the business center is associated with a decrease in the value of houses (houses further away from major urban amenities like education and healthcare institutions, banks, restaurants and entertainment, high-end shopping centers, administrative institutions, etc., have lower values), whereas an increase in the distance to the airport is associated with an increase in values (houses that are further away from the noise, pollution, and traffic congestion of a major international airport have higher values).

An important reason for including distance controls in our model is to test the robustness of the fixed effect estimates, which we discuss in more detail in the section below. One may argue that the neighborhoods are simply picking the proximity of houses to various amenities and by controlling for such distances the fixed effects would lose their importance in the model (in other words, the distribution of the fixed effects should not show substantial deviations from the mean values). Interestingly, we find that despite controlling for these distances both the distribution of neighborhood implicit price premiums over the average neighborhood as well the magnitude of these premiums does not change significantly. This further supports our main hypothesis of the major role that neighborhoods, as a proxy for various time invariant characteristics of the place where the house is located, hold in the house buying process.

We also provide Moran's I spatial autocorrelation test results for each specification. As shown in Table 2, what stands out is that the presence of spatial autocorrelation is essentially removed once we introduce the neighborhood fixed effect (see columns 3-5). Observe that Moran's I is positive in the first two specifications, indicating spatial dependence, and drops close to zero in the specifications with spatial fixed effects, which implies that the neighborhood controls for spatially correlated omitted variables.

Importantly, in order to assess the relative strength of neighborhood identity as a fixed effect, we re-ran the specifications shown in columns (3) – (5) of Table 2 using four alternative areal units as fixed effects in lieu of the neighborhood. A summary of the results is given in Table 3. Our original specification using the MLS-designated neighborhood fixed effect yields the tightest fit for the model with an adjusted R^2 of 93 percent for specifications (3) – (5). This is then followed by census tract (90 percent), zip code (87 – 88 percent), elementary school (86 – 88 percent), and census block (70 – 84 percent). In our data set, the MLS-defined neighborhoods, counting a total of 829 subdivisions, represent the most granular spatial unit for the Charleston region. On the other hand, census tracts in the same region

Table 3: *Adjusted R2* and *Moran's I* Values for Different Fixed Effects

		Fixed Effect Level				
		Census Block	Zip Code	Elem. School	Census Tract	Neighborhood
	Spec. (3)	0.70	0.87	0.86	0.90	0.93
Adj R2	Spec. (4)	0.72	0.87	0.86	0.90	0.93
	Spec. (5)	0.84	0.88	0.88	0.90	0.93
	Spec. (3)	0.130	0.0314	0.0459	0.0127	0.00740
Moran's I	Spec. (4)	0.123	0.0305	0.0379	0.0124	0.00724
	Spec. (5)	0.051	0.0295	0.0302	0.0123	0.00713

Adj R2 for specifications (1) and (2) are 0.59 and 0.67.
Moran's I for specifications (1) and (2) are 0.236 and 0.179.

have fewer spatial units (150) than census blocks (523), but more than elementary school zones (90). Census tracts are considered to be relatively uniform in terms of demographic and economic conditions and are in a sense as government-defined neighborhoods (the boundaries are determined by data user groups). Although many households buying a house may be constrained by elementary school choice in the absence of open enrollment policies, the key finding remains that MLS neighborhoods explain more of the housing price variation than any of the alternatives. After all, the housing market is organized around these identities, which in some cases act like consumer brands (exemplified by names like Wild Dunes and South of Broad in Charleston).

4.2 Neighborhood Implicit Price Premium Estimates

Our next step is to use the fixed effects that we estimated in the previous section to create an implicit price premium of each neighborhood. The estimated neighborhood fixed effects capture the average value of a home sale in the neighborhood, and the other estimated coefficients in the regressions capture the effect of deviations from the neighborhood mean in measurable housing characteristics on housing prices. In this section, we calculate a neighborhood price premium for each neighborhood by subtracting out the effect of average housing and amenity characteristics from each neighborhood. That is, if we use our hedonic pricing model of:

$$P_i = \alpha\theta_i + \beta X_i + \gamma_i + \varepsilon_i \tag{6}$$

where γ_i is the neighborhood fixed effect, we run the following regression:

$$\hat{\gamma}_i = \hat{\beta}^{BE}\bar{X}_i + \bar{\varepsilon}_i \tag{7}$$

where $\hat{\gamma}_i$ is the estimated neighborhood effect from equation (6), $\bar{X}_i$ is the within-neighborhood average of the measured neighborhood characteristics, and $\hat{\beta}^{BE}$ is the "between estimator" estimating the effect of these averaged regressors on the average neighborhood price. The variable that we focus on here is the residual term $\bar{\varepsilon}_i$ from regression (7). This residual represents the price premium (or discount) that houses in a neighborhood have relative to the Charleston average after controlling for the average measurable determinants of housing prices in that neighborhood. It can be thought of as the neighborhood's "brand" value.

Some notable results stand out from this analysis. First, the unconditional standard deviation in the log of home prices across the entire sample is 0.6164. Roughly about half of this variation, 0.2961, can be accounted for by fixed effects in the neighborhood fixed effects regression. Then, roughly a little more than half of this variation, 0.1697, is accounted for in the standard deviation of the residuals from regression (7). The message here is that heuristics such as assigning a "value" to a neighborhood can explain a very large fraction of the variation in housing prices, even after controlling for such common determinants of house prices such as size, age, etc. Thus, we have good evidence for the importance of neighborhood brand heuristics.

Secondly, we can use these residuals to identify both "superstar" neighborhoods where housing prices attract a high premium (again, even after controlling for common determinants of home values) as well as more blighted neighborhoods.

Table 4: Top 10 Neighborhoods by Fixed Effect Residuals

Neighborhood	Fixed Effect Residual
Beachside	1.1900
Morris Square	0.7552
Goat Island	0.6858
Radcliffe Place	0.6521
Folly Beach	0.6521
Wadmalaw Island	0.6002
Isle of Palms	0.5982
Plantation Isle	0.5955
Sullivans Island	0.5892
Isle de Nemours	0.5819

For example, Table 4 shows the top 10 neighborhoods with the largest price premium. The Charleston metropolitan area's Beachside neighborhood exerts the highest residual value at 1.1900. In relative terms, a home in this MLS-defined neighborhood is worth 119 percent more than the average Charleston neighborhood value. Beachside homes command particularly high prices even after controlling for proximity to the ocean. Thus, the implicit price reflects more than just proximity to the ocean. Indeed, real estate agents advertise a neighborhood as having a unique package of attributes, with "*all the amenities you would want*," including private beach access, a family friendly atmosphere, and a relatively central location in the Charleston metropolitan area.

Other exclusive oceanfront communities such as Sullivan's Island and the Isle of Palms rank highly as well, with implicit prices 58.92 and 59.82 percent above the average, respectively. These neighborhoods offer amenities like proximity to pristine Atlantic Ocean beaches that are fixed spatially and not possible to reproduce elsewhere. Real estate advertisements for Folly Beach (65.21 percent above the average) tout its "ocean wildlife, great surfing spots, eclectic and a laid back lifestyle." At the same time, some Charleston downtown neighborhoods have high brand values and are also found in Table 4. These top neighborhoods enjoy a well-known reputation in the Charleston market. The ambiance and quality of life is a combination of many attributes that are hard to quantify individually. All these superstar neighborhood brands command extremely high values.

However, because in our regression we control for the house's *antebellum* characteristic (an indicator variable equal to 1 if the house is built before 1861), the values of neighborhoods that make up Charleston's historic district fall closer to the average value. In other words, the antebellum characteristic captures the main unique and time invariant characteristic of these neighborhoods, leaving little behind to the plethora of other urban and natural amenities of these neighborhoods. Nevertheless, because Charleston's historic neighborhoods are characterized on the positive side by close proximity to some of the nation's best restaurants, exquisite waterfront parks, colleges, museums, high-end shopping, romantic cobblestone streets, and other beautiful, Instagram-worthy amenities (like those analyzed in Carlino and Saiz 2019) and on the downside by high tourist congestion and risk of flooding, some of these historic neighborhoods may retain a considerable price premium or discount beyond the premium conveyed by their historical character.

For example, the South of Broad neighborhood has a premium of 21.87 percent (still considerably higher than the average, which points to the unique amenities described above adding value beyond the historical aspect of the neighborhood), whereas the value of the French Quarter neighborhood is slightly lower than the average with a discount of -0.38 percent. Above all, Charleston's historical area contains incomparable 18th and 19th-century architecture, including the rare U.S. examples of Georgian, Federal, Classical Revival, Greek Revival, Italianate, and Gothic Revival styles, and this cluster of historic homes is known to increase property values (Asabere and Huffman, 1994; Clark and Herrin, 1997; Diaz et al., 2008; Zahirovic-Herbert and Chatterjee, 2011).

We see the less-fortunate neighborhoods of the Charleston area in Table 5, which captures the neighborhoods with the 10 lowest fixed effects residuals. Close to the central business district we find Rosemont, which our analysis shows has the lowest value for any of the 829 neighborhoods in the Charleston area. The neighborhood value is approximately 77 percent below the average.

Table 5: Bottom 10 Neighborhoods by Fixed Effect Residuals

Neighborhood	Fixed Effect Residual
Rosemont	-0.7734
Bulls Bay Overlook	-0.5939
Cherokee Place	-0.5188
Dorchester Terrace	-0.4794
Olde Park	-0.4403
Summerville Country Estates	-0.4370
Nafair	-0.4167
Eastside	-0.4004
Daniel Island Park	-0.3966
Waylyn	-0.3919

Finding Rosemont at the bottom of the neighborhood implicit prices revealed by our analysis, we were motivated to find out more about the area's legacy. It turns out that this neighborhood, along with other North Charleston communities near Interstate 26, have an extremely tortured history of environmental degradation and misguided urban and land-use policy (Wilson et al., 2012). In one of many examples of disastrous U.S. urban redevelopment in the 1960s, Interstate 26 was built through Rosemont, dividing the neighborhood, and forever devaluing the long-standing Black residential area. After emancipation from slavery in the 1880s, Black families settled in this area. In what had been a tight-knit community for many decades, homes were removed to make way for the infrastructure project. The busy highway separated neighbors from each other. It became a permanent source of pollution, congestion, noise, and other negative externalities. In the early 2000s, having suffered for years, the neighborhood organized against another destructive infrastructure project–a new port access road along Interstate 26. With other parties, the Rosemont Neighborhood Association filed an environmental justice complaint with the U. S. Department of Justice. They won a small settlement and were promised a sound wall.

Even so, homes in this predominantly low-income, African-American neighborhood have suffered from a permanent wealth loss following the construction of Interstate 26. Rosemont is situated close to the high-price neighborhoods in the Charleston historic district, but in quality of life, it is a long way away. Many of the remaining low-valued neighborhoods listed in Table 5 follow the spine of Interstate 26 to the north, away from the central business district. It could be argued that proximity to highway access would be desirable for commuters to the central city. Yet some of these communities are impacted by nearby toxic release inventory facilities (Wilson et al., 2012).

As we saw with some historical neighborhoods mentioned above, controlling for their specific historic character distills most of the value captured in the fixed effect. However, identifying the myriad of possible characteristics to control for is impossible and, as the behavioral studies described above show, consumers usually focus on the main heuristic, which is the brand identity. As a result, it is important to consider the robustness of the neighborhood brand values. In Figure 1, we show the graphical distribution of the these values as we build up the models and control for an increasing number of factors. In panel (a), we show the distribution of the implicit neighborhood price premiums based on model (3) in Table 2, which only accounts for house structural characteristics. In panel (b), we enhance the model from panel (a) with the addition of demographic characteristics (corresponding partially to model (4) in Table 2). In panel (c), we further add to the model from panel (b) the school quality characteristics (which represents model (4) in Table 2). Finally, in panel (d), we represent the distribution of neighborhoods' implicit price premiums based on the full specification represented in model (5) of Table 2, in which we control for all of the above-mentioned characteristics as well as the distances to the three types of amenities (business center, airport, and the beach).

Notice that even after controlling for distances to various urban and natural amenities, the distribution of fixed effects does not get affected (i.e., the neighborhoods do, in fact, generally maintain their relative rankings as we add more control variables). The only change we notice after building up our specification is that the neighborhoods' price premiums shrink (similar to what we observed in the case of historic district neighborhoods discussed above). For example, the Beachside neighborhood, an oceanfront community and the top ranked

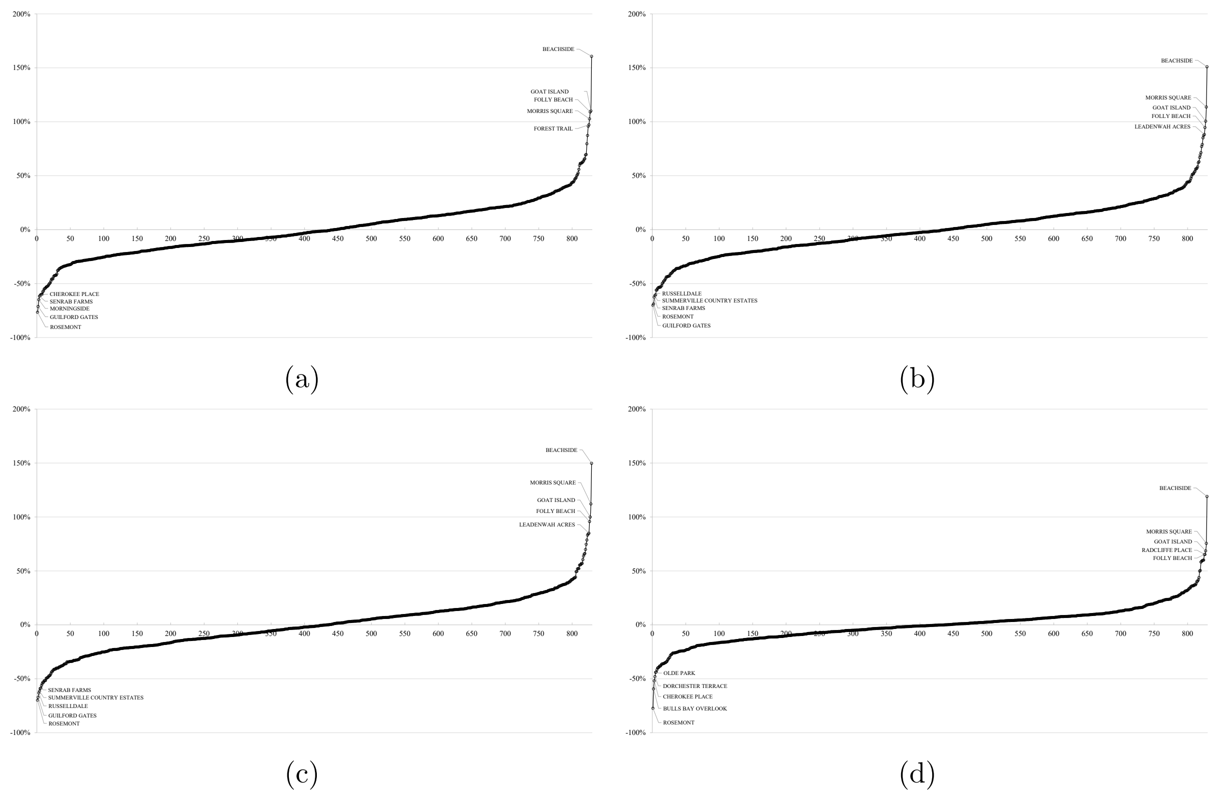

Figure 1: Distribution of neighborhood brand effects across model specifications.

neighborhood in Charleston, based on its fixed effect residual, decreases from a premium of 161 percent in panel a to 119 percent in panel d. One may assume that proximity to the ocean is one of the main price drivers for top beach front communities, however our analysis confirms that even after controlling for distance to the beach there are other intrinsic unobserved characteristics driving the strong price premiums. Controlling for distances does not change significantly the pronounced clustering of top- and bottom-ranked neighborhoods. Clearly, one may find additional nuanced factors to control for and further refine the fixed effect of each neighborhood; however, this is not the objective of this study. Instead, these results support our hypothesis that fixed effects in hedonic housing models may proxy the heuristics that individuals use during their house hunting process and should, as a result, be valuable metrics used in house value estimations.

In sum, our analysis brings to light the large disparities in the implicit prices embedded in neighborhood identities. Neighborhood values appreciate sharply at the upper end and diminish at the lower end. This pattern is shaped by the utility of many long-run location attributes and amenities, which have been studied extensively in previous housing market research. Hedonic models are designed to measure implicit prices of many attributes that are capitalized in housing values. No doubt hedonic housing models provide useful information on the willingness to pay for these particulars. In contrast, our approach examines the long-run hedonic value of the neighborhood as a whole.

5 Conclusion

This paper postulates that neighborhood designations act as heuristics used in housing market decisions. Previously, the heuristic concept has been applied to understand many types of decisions, including consumer choices and the role of brand recognition. Behavioral economists and psychologists argue that decisions are often made with mental shortcuts, allowing for "fast-thinking" (Kahneman, 2003). The standard hedonic housing market theory assumes "slow" thinking; that is, complex cognitive and computational abilities, with agents optimizing across all available information. Following behavioral theory, neighborhoods organize a plethora of local attributes into a recognizable and useful identity. In the real world, the housing market is clearly organized around choices among neighborhoods. Yet to date there has been no research that measures the effect of neighborhood brands on house values.

Applied to housing markets, the heuristic theory puts the neighborhood at the forefront of empirical analysis. The results presented in this paper validate that housing prices are significantly determined by the neighborhood where the property is situated. Based on a large and reliable data set, our hedonic regression analysis points to the promise of our novel approach. Future research will be needed to corroborate our findings, drawing on the increasing availability of housing data from real estate web sites and other sources.

For subsequent research, our paper suggests that housing price models can better integrate neighborhood fixed effects. Residences are tethered to neighborhoods and, *a priori*, we should expect spatial dependence. Researchers can benefit from our finding that neighborhoods control for spatial autocorrelation. Moreover, our paper offers a new spatial autocorrelation measure for large data sets. Our Moran's I technique is potentially valuable for many applications in urban and regional research.

To be sure, there are many specific local factors that can alter housing prices independent of the long-run neighborhood effect. Influences like school quality will no doubt continue to be investigated, but we contend that they should be studied in the context of neighborhood identities. Implicit price premium estimates can help assess the wealth loss or gain from land use and urban/regional policies.

Beyond the housing literature, our approach to analyzing fixed effects estimates has wider implications for urban and regional research. The implicit premium of regions and urban areas can be assessed in other contexts such as industry location and migration choice. Like neighborhoods, cities and regions such as the Silicon Valley, Shanghai, and Paris (along with many others) project an image or brand that can be hard to define by particular attributes alone. When people or firms choose locations, they buy into a place-based identity. We propose that the value of this identity can be quantified using the methods underlying our research in this paper.

References

Allison, Ralph I. and Kenneth P. Uhl. 1964. “Influence of Beer Brand Identification on Taste Perception,” Journal of Marketing Research, 1(3), 36–39.

Anenberg, Elliot and Edward Kung. 2020. “Can more housing supply solve the affordability crisis? Evidence from a neighborhood choice model,” Regional Science and Urban Economics, 80, 103,363.

Anselin, Luc. 1988. Spatial Econometrics: Methods and Models, volume 4, Springer, Dordrecht.

Anselin, Luc and Daniel Arribas-Bel. 2013. “Spatial fixed effects and spatial dependence in a single cross-section,” Papers in Regional Science, 92(1), 3–17.

Asabere, Paul K. and Forrest E. Huffman. 1994. “Historic designation and residential market values,” The Appraisal Journal, 62(3), 396–401.

Basu, Sabyasachi and Thomas G. Thibodeau. 1998. “Analysis of Spatial Autocorrelation in House Prices,” The Journal of Real Estate Finance and Economics, 17(1), 61–85.

Billings, Stephen B. 2015. “Hedonic Amenity Valuation and Housing Renovations,” Real Estate Economics, 43(3), 652–682.

Bogart, William T. and Brian A. Cromwell. 1997. “How much more is a good school district worth?” National Tax Journal, 50(2), 215–232.

Bogin, Alexander and Phuong Nguyen-Hoang. 2014. “Property left behind: An unintended consequence of a no child left behind “failing” school designation,” Journal of Regional Science, 54(5), 788–805.

Brasington, David. 1999. “Which Measures of School Quality Does the Housing Market Value?” Journal of Real Estate Research, 18(3), 395–413.

Carlino, Gerald A and Albert Saiz. 2019. “Beautiful city: Leisure amenities and urban growth,” Journal of Regional Science, 59(3), 369–408.

Clark, David E. and William E. Herrin. 1997. “Historical Preservation Districts and Home Sale Prices: Evidence from the Sacramento Housing Market,” Review of Regional Studies, 27(1), 29–48.

Cliff, Andrew and Keith Ord. 1972. “Testing for Spatial Autocorrelation Among Regression Residuals,” Geographical Analysis, 4(3), 267–284.

Conklin, James, Moussa Diop, and Herman Li. 2020. “Contact high: The external effects of retail marijuana establishments on house prices,” Real Estate Economics, 48(1), 135–173.

Dawar, Niraj and Philip Parker. 1994. “Marketing Universals: Consumers’ Use of Brand Name, Price, Physical Appearance, and Retailer Reputation as Signals of Product Quality,” Journal of Marketing, 58(2), 81–95.

Diaz, Julian, Andrew Hansz, Matthew Cypher, and Darren Hayunga. 2008. "Conservation Status and Residential Transaction Prices: Initial Evidence from Dallas, Texas," Journal of Real Estate Research, 30(2), 225–248.

Dubin, Robin A. 1992. "Spatial autocorrelation and neighborhood quality," Regional science and urban economics, 22(3), 433–452.

Erdem, Tülin, Joffre Swait, and Jordan Louviere. 2002. "The impact of brand credibility on consumer price sensitivity," International Journal of Research in Marketing, 19(1), 1–19.

Figlio, David N. and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market," American Economic Review, 94(3), 591–604.

Horsch, Eric J and David J Lewis. 2009. "The effects of aquatic invasive species on property values: evidence from a quasi-experiment," Land Economics, 85(3), 391–409.

Ismail, Suriatini. 2006. "Spatial autocorrelation and real estate studies: A literature review," Malaysian Journal of Real Estate, 1(1), 1–13.

Jacoby, Jacob, Jerry C. Olson, and Rafael A. Haddock. 1971. "Price, brand name, and product composition characteristics as determinants of perceived quality," Journal of Applied Psychology, 55(6), 570.

Jacoby, Jacob, George J. Szybillo, and Jacqueline Busato-Schach. 1977. "Information Acquisition Behavior in Brand Choice Situations," Journal of Consumer Research, 3(4), 209–216.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics," American Economic Review, 93(5), 1449–1475.

Kain, John F. and John M. Quigley. 1970. "Measuring the Value of Housing Quality," Journal of the American Statistical Association, 65(330), 532–548.

Kuang, Chun. 2017. "Does quality matter in local consumption amenities? An empirical investigation with Yelp," Journal of Urban Economics, 100, 1–18.

Kuminoff, Nicolai V., Christopher F. Parmeter, and Jaren C. Pope. 2010. "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities?" Journal of Environmental Economics and Management, 60(3), 145–160.

Kuminoff, Nicolai V. and Jaren C. Pope. 2014. "Do "capitalization effects" for public goods reveal the public's willingness to pay?" International Economic Review, 55(4), 1227–1250.

Leonard, Tammy and James C Murdoch. 2009. "The neighborhood effects of foreclosure," Journal of Geographical Systems, 11, 317–332.

Li, Mingche M. and H. James Brown. 1980. "Micro-Neighborhood Externalities and Hedonic Housing Prices," Land Economics, 56(2), 125–141.

Moralı, Orçun and Neslihan Yılmaz. 2020. "An Analysis of Spatial Dependence in Real Estate Prices," The Journal of Real Estate Finance and Economics, pp. 1–23.

Nguyen-Hoang, Phuong and John Yinger. 2011. "The capitalization of school quality into house values: A review," Journal of Housing Economics, 20(1), 30–48.

Pope, Jaren C. 2008. "Do seller disclosures affect property values? Buyer information and the hedonic model," Land Economics, 84(4), 551–572.

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

Sedgley, Norman H., Nancy A. Williams, and Frederick W. Derrick. 2008. "The effect of educational test scores on house prices in a model with spatial dependence," Journal of Housing Economics, 17(2), 191–200.

Simon, Herbert A. 1978. "Rational decision-making in business organizations. Nobel Memorial Lecture," Economic Sciences, 1978, 343–371.

Tversky, Amos and Daniel Kahneman. 1986. "Judgment under uncertainty: Heuristics and biases," Judgment and decision making: An interdisciplinary reader, pp. 38–55.

Tyndall, Justin. 2019. "Getting High and Low Prices: Marijuana Dispensaries and Home Values," Real Estate Economics.

Wilhelmsson, Mats. 2002. "Spatial models in real estate economics," Housing, theory and society, 19(2), 92–101.

Wilson, Sacoby M, Herb Fraser-Rahim, Edith Williams, Hongmei Zhang, LaShanta Rice, Erik Svendsen, and Winston Abara. 2012. "Assessment of the distribution of toxic release inventory facilities in metropolitan Charleston: An environmental justice case study," American Journal of Public Health, 102(10), 1974–1980.

Wooldridge, Jeffrey M. 2016. Introductory Econometrics: A Modern Approach, Cengage Learning, sixth edition.

Yinger, John. 2016. "Hedonic estimates of neighborhood ethnic preferences," Public Finance Review, 44(1), 22–51.

Zahirovic-Herbert, Velma and Swarn Chatterjee. 2011. "What is the Value of a Name? Conspicuous Consumption and House Prices," Journal of Real Estate Research, 33(1), 105–125.

Zahirovic-Herbert, Velma and Geoffrey K. Turnbull. 2009. "Public School Reform, Expectations, and Capitalization: What Signals Quality to Homebuyers?" Southern Economic Journal, 75(4), 1094–1113.

Appendix: A note on the calculation of Moran's I statistic with repeated locations

To check for spatial dependence of the residuals we implement the Moran's I statistic. Following Anselin (1988, page 101) this statistic is given by

$$I = (n/s)\left[\mathbf{e}'\mathbf{W}\mathbf{e}\right]/\mathbf{e}'\mathbf{e} \tag{8}$$

where n is the total number of observations, $\mathbf{W}$ is the non-standardized matrix of distances, $\mathbf{e}$ is the vector of residuals from the regression and $s = \mathbf{i}'\mathbf{W}\mathbf{i}$ is the sum of all weights ($\mathbf{i}$ is a vector of ones). Letting $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ be the regression projection matrix we can write the expected value of I as

$$E(I) = (n/s)\, tr(\mathbf{MW})/(n-k) \tag{9}$$

and the asymptotic variance of I as

$$V(I) = (n/s)^2\left[tr(\mathbf{MWMW}') + tr(\mathbf{MW})^2 + (tr(\mathbf{MW}))^2\right]/d - E(I)^2 \tag{10}$$

where

$$d = (n-k)(n-k+2)\ .$$

The asymptotic distribution of Moran's I follows a normal distribution (Cliff and Ord, 1972) meaning that we can test for spatial dependence on the residuals using

$$Z = [I - E(I)]/V(I)^{1/2}\ .$$

Since we have the exact locations (the latitude and longitude) of each house we can use the actual physical distances to construct the $\mathbf{W}$ matrix. More specifically, the off-diagonal elements of $\mathbf{W}$ equal the inverse of the distance between locations while the diagonal elements equal zero. But with more than 50,000 observations we found that calculation of the $\mathbf{W}$ matrix was impractical given its dimension ($n \times n$). To deal with this problem we rounded the latitude and longitude measured in decimal degrees to three decimal places. By doing this we forced observations that were very close to each other to share the same location (have the same latitude and longitude coordinates) and were thus able to reduce the number of unique locations to 14,189. Next, we modified Moran's I formula to accommodate the fact that we had repeated locations. Let ϖ represent a matrix with all unique locations. This matrix has dimension $G \times G$ where G is the number of unique locations (14,189 values in our case). Let $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_G]$ where each $\mathbf{g}_j$ represents a column vector that takes the value 1 if the observation is located at unique location j and zero otherwise. This means that we can write $\mathbf{W}$ as

$$\mathbf{W} = \mathbf{G}\varpi\mathbf{G}' \tag{11}$$

where $\mathbf{G}$ is $n \times G$ and ϖ is $G \times G$. Replacing $\mathbf{W}$ into the definition of I we obtain

$$\begin{aligned} I &= (n/s)\left[\mathbf{e}'\mathbf{G}\varpi\mathbf{G}'\mathbf{e}\right]/\mathbf{e}'\mathbf{e} \\ & (n/s)\left[\mathbf{e}^{*\prime}\varpi\mathbf{e}^{*}\right]/\mathbf{e}'\mathbf{e} \end{aligned}$$

where $\mathbf{e}^*$ is a vector with dimension G whose individual elements are the group summations of the regression residuals. Likewise, s can be expressed as $s = \mathbf{i}'\mathbf{G}\varpi\mathbf{G}'\mathbf{i} = \mathbf{i}^{*\prime}\varpi\mathbf{i}^*$ where $\mathbf{i}^*$ is a vector with dimension G whose elements are a count of the number of observations in each group. Note that with this approach we reduce the dimension of the matrices needed to calculate Moran's I statistic. To calculate the expected value we note that

$$\begin{aligned} tr(\mathbf{MW}) &= tr[(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{W}] \\ &= tr(\mathbf{W}) - tr[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}] \\ &= -tr[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}] \end{aligned}$$

and replacing $\mathbf{W}$,

$$tr(\mathbf{MW}) = -tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}\varpi\mathbf{G}'\mathbf{X}) = -\,tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{*\prime}\varpi\mathbf{X}^{*}) \tag{12}$$

where $\mathbf{X}^{*}$ is a $k \times G$ matrix containing the group sums of the $\mathbf{X}$ variables in the regression model. To compute the variance we note that $\mathbf{W}$ is symmetric and thus (10) simplifies to

$$V(I) = (n/s)^2 \left[2tr(\mathbf{MWMW}) + (tr(\mathbf{MW}))^2\right]/d - E(I)^2\ . \tag{13}$$

This means that we need to calculate

$$tr(\mathbf{MWMW}) = tr(\mathbf{G}\varpi\mathbf{G}'\mathbf{M}\mathbf{G}\varpi\mathbf{G}'\mathbf{M}) = tr(\varpi\mathbf{G}'\mathbf{M}\mathbf{G}\varpi\mathbf{G}'\mathbf{M}\mathbf{G}).$$

Since

$$\mathbf{G}'\mathbf{M}\mathbf{G} = \mathbf{G}'\mathbf{G} - \mathbf{G}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{G} = \mathbf{G}'\mathbf{G} - \mathbf{X}^{*}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{*\prime}$$

we do not need to deal with matrices of dimension n. Note that $\mathbf{G}'\mathbf{G}$ is a $G \times G$ diagonal matrix with the number of elements of each group in the main diagonal.

In the example discussed above we tackled the problem of dealing with a large $\mathbf{W}$ matrix (more than 50,000 lines in our application). However, some of the regressions that we run have an additional complication, namely the fact that the $\mathbf{X}'\mathbf{X}$ matrix also has a large dimension due to the introduction of a fixed effect with many categories. However, this is not really a problem because we can apply the above formulas to the within-transformed regression. The results will be the same as if we had used the full $\mathbf{X}$ matrix.