Department of Economics

Working Paper Series

Agent Intermediation and Racial Price Differences

Adam Nowak

Patrick Smith

Working Paper No. 17-21

Agent Intermediation and Racial Price Differentials

Preliminary Draft - Please Do Not Cite

Adam Nowak *

West Virginia University

Patrick Smith †

San Diego State University

June 15, 2017

Abstract

In most housing transactions the home buyer and seller do not meet, in which case the buyer's race is not revealed and the seller cannot discriminate against them. Despite this fact, previous studies estimate racial price differentials based on the race of the home buyer. In this study we identify the dimensions along which differential treatment occurs in housing markets. We show that home buyers disproportionately hire real estate agents of the same race and that the race of the agent, not the home buyer, is the primary mechanism of discrimination. The results of this study have important implications for fair housing policy.

Key Words: Housing, agent intermediation, racial discrimination, price differentials

*West Virginia University; Email: adam.d.nowak@gmail.com

†San Diego State University; Email: patrick.smith@sdsu.edu

1 Introduction

In most housing transactions the home buyer and seller do not meet or communicate prior to the closing - if at all. Instead, they are represented by real estate agents that act on their behalf. Agent intermediation should, in theory, eliminate racial price differentials in housing markets. If the seller never meets the buyer - discrimination is not feasible.[1] In this study, we identify the dimensions along which differential treatment occurs in housing markets. Specifically, we test whether the race of the selling agent serves as a proxy for the race of the buyer - thereby serving as the mechanism for discrimination.

Previous studies that compare the prices of homes purchased by different racial groups argue that price differentials are evidence of discrimination.[2] The studies do not, however, identify the dimensions along which differential treatment occurs, leaving several important questions unanswered. If the seller and buyer never meet then how is the buyer's race revealed? Does the race of the selling agent representing the buyer serve as the mechanism of discrimination? If so, can a minority buyer avoid discrimination by hiring a non-minority selling agent? These questions have several important policy implications that are directly associated with the U.S. government's fair housing policy.

Fair housing has long been a central focus of governmental policy. The Fair Housing Act (FaHA) of 1968 prohibits discrimination in housing on the basis of race, color, national origin, religion, sex, disability, and the presence of children. In 1988, the Fair Housing Amendments Act (FaHAA) was enacted to help enforce FaHA. Among other things, FaHAA removed the limit on punitive damages, lengthened the statute of limitations, and created a system of administrative law judges to handle claims of discrimination. Although significant progress was made in the late 20th century, racial discrimination is still prevalent in housing markets.

This study makes important contributions to the literature on racial discrimination in housing markets. We provide evidence that home buyers sort not only into neighborhoods

[1] Similar to Yinger [1995] we define discrimination as adverse treatment of an individual based solely on her or his membership in a particular racial or ethnic group.

[2] Section 3 includes a review of the literature on discrimination in housing markets.

based on race, but also in terms of agent representation. Black (white) home buyers are disproportionately represented by black (white) selling agents relative to the underlying real estate agent population. The racial sorting of buyers to selling agents provides a mechanism for racial discrimination at the agent level - where the race of the selling agent proxies for the race of the home buyer.

The primary contribution of this study is its estimation of black-white price differentials based on the race of the selling agent and home buyer. Agent intermediation should, in theory, prevent discrimination in housing markets - especially given the fact that the agents in our study receive training on the topic.[3] Because the home buyer's race is not revealed, racial price differentials should not exist *unless* the racial sorting we document provides a proxy for the buyer's race. We examine whether the race of the selling agent is the mechanism of discrimination using a simple, yet novel test. We compare the price paid for similar housing when a black home buyer is represented by a black selling agent relative to when they are represented by a white selling agent. Using a matched data set we find that black home buyers pay a premium of 3.5 to 5.0 percent when represented by black selling agents. The results suggest that the race of the selling agent serves as one of the dimensions along which differential treatment occurs.

Understanding the mechanism of discrimination in housing markets strengthens and aids in the enforcement of housing discrimination policies that are already in place. We show that agent intermediation does not eliminate racial price differentials because market participants sort based on race. Our results suggest, however, that racial price differentials can be eliminated by (i) making dual agency illegal and (ii) obscuring the race of the selling agent, whose race currently serves as a proxy for the race of the home buyer, from the listing agent.[4]

[3]We focus on real estate agents that are members of the National Association of Realtors (NAR) because they must complete ethics training not only when they become a new member, but once every four years after. Membership in NAR requires Realtors to pledge to conduct business in keeping with the spirit and letter of the Code of Ethics. Article 10 of the Code of Ethics includes a firm statement of support for equal opportunity in housing. This ensures that the agents in our study are aware of and have the knowledge and training required to prevent discrimination.

[4]Eight states already ban dual agency. Although two of the eight states allow designated agency (Maryland and Texas), three allow transaction brokerage (Florida, Kansas and Oklahoma), and three allow both

This could be achieved by building an "anonymous" bidding portal into current Multiple Listing Service (MLS) software.[5]

The remainder of the paper is organized as follows. Section 2 describes the comprehensive data set employed in this study. Section 3 provides an overview of previous studies on racial discrimination in housing markets, highlights their limitations, and then discusses the methodology we implement in this study. Section 4 presents our findings and Section 5 offers our concluding remarks.

2 Data

2.1 Demographics

This study focuses on the metro-Atlanta single-family detached housing market from January 2000 through December 2015.[6] Metro-Atlanta is the ideal location for this study given its size, diversity and changing racial dynamics. According to the 2010 Census, the Atlanta metropolitan statistical area (MSA) was the ninth largest MSA in the United States. From 2000 to 2010 the MSA's population grew approximately 24 percent, eclipsing the 5 million mark for the first time. The MSA's racial demographics also underwent a considerable change. In 2010 the Atlanta MSA was 55.4 percent white - which represents a sharp decline from 2000 when the MSA was 63 percent white. During the same time period the black population in the MSA increased from 28.9 percent in 2000 to 32.4 percent in 2010.

Although the MSA's black population grew - its geographic disbursement changed considerably. The black population in the City of Atlanta, which was the fifth largest black-

(Alaska, Colorado and Texas).

[5]Offers in the bidding portal would only be "anonymous" from the perspective of the listing agent and home seller. In which case the identity, and therefore the race, of both the home buyer (i.e. bidder) and the selling agent who represents the buyer would be obscured. We recognize that there are numerous confounding factors that make the implementation of such a portal extremely difficult. These additional considerations are outside the scope of this project and we leave them for future discussion.

[6]The transaction data set includes Clayton, Cobb, DeKalb, Fulton, and Gwinnett counties.

majority city in 2010, declined by 11.3 percent from 2000 to 2010.[7] Instead of moving in to the city, the black population settled in the suburbs of metro-Atlanta.[8] The changing demographics in metro-Atlanta is often attributed to an influx of white households in to gentrifying neighborhoods within the city, coupled with a movement of black households into the adjacent suburbs [Dewan, 2006].

2.2 Data Overview

The data set employed in this study was constructed using several sources including Georgia Multiple Listing Service (GAMLS), CoreLogic, and Home Mortgage Disclosure Act (HMDA). The GAMLS data includes property characteristics (bedrooms, bathrooms, etc.), transaction details (sales price, time-on-market, etc.), and agent information (sales agent, listing agent, dual agent, etc.) for every house that was listed on the MLS from January 1997 through September 2016. We combine the GAMLS data with proprietary tax assessor data obtained from CoreLogic. The CoreLogic transaction dataset includes every real estate transaction that was recorded from January 2000 through September 2016. The sales transaction file includes information about the buyer and seller, such as their name and address, which we use to identify owner-occupiers and investors. It also includes, among other things, the sales date, purchase price, deed type, lender name, and loan amount.

Next we merge the publicly available loan application registry data gathered under HMDA. The HMDA dataset provides detailed demographic and economic information about the buyer. We merge the datasets based on each transaction's (i) census tract, (ii) transaction year, (iii) lender name, and (iv) loan amount. The HMDA data includes the buyer's race and ethnicity that we use to estimate the housing price differentials.

[7] Detroit (MI), Memphis (TN), Baltimore (MD), and Washington (DC) are the only black-majority cities with a larger black population.

[8] Roughly 90 percent of the City of Atlanta resides in Fulton County and the remaining 10 percent resides in DeKalb County.

2.3 Real Estate Agent Race and Ethnicity

We identify the sex, race and ethnicity of the real estate agents in our study using the Amazon Mechanical Turk (AMT) platform.[9] Anonymous workers on the platform were shown images of real estate agents from www.realtor.com. As such, only agents that are NAR members are included in this study, which ensures that every agent has received training on housing discrimination. The AMT workers were shown an image of the real estate agent and asked to categorize the agent's sex as follows:

$$Sex = \begin{cases} Female \\ Male \\ No\ individual\ in\ photo \\ More\ than\ one\ individual\ in\ photo \end{cases}$$

If the worker selects "No individual in photo" or "More than one individual in photo" then the agent is not included in the study. If the worker selects "Female" or "Male" then separate sets of workers categorize the agent's race and ethnicity as follows:

$$Race = \begin{cases} American\ Indian \\ Asian \\ Black\ or\ African\ American \\ Native\ Hawaiian \\ White \end{cases}$$

$$Ethnicity = \begin{cases} Hispanic \\ Non\text{-}Hispanic \end{cases}$$

The race and ethnicity categories match those provided in the HMDA dataset, allowing for ease of comparison. Each image is categorized by two independent workers. When setting

[9] Additional information on Amazon Mechanical Turk is available here: https://www.mturk.com/

up the categorization project in AMT we restricted the task to only include workers that were rated as "masters".[10] In the case of a disagreement between two workers we personally reviewed and categorized the image. We also randomly validated 1 percent of the remaining sample to ensure the categorizations were accurate.

The approach we employ to identify the real estate agent's race and ethnicity is similar to the approach employed for the HMDA data. Lenders must ask the loan applicant for their sex, race and ethnicity, but cannot require that the loan applicant provide the information. If the loan application is submitted by phone, mail or internet and the applicant does not provide the information, then the information is not required for HMDA reporting purposes (FFIEC [2013]). However, if the application is submitted in person and the applicant does not provide the information, the lender is required to "note the applicant's ethnicity, race, and sex on the basis of visual observation and surname, to the extent possible."[11] Similar to the HMDA collection process, we have the AMT workers classify the real estate agents in this study based on a visual observation of their profile picture from the NAR website.

The agent race and ethnicity classifications are merged with the GAMLS data using a manually created cross reference file. We create the cross reference file to overcome two issues. First, there are several many-to-one and many-to-many relationships based on agent names alone. Second, agent names may change over time due to life events (i.e. marriage, divorce). We overcome these issues by matching agents to their unique ID in GAMLS based on their name *and* transaction set. Using the agents' recent transactions listed on www.realtor.com we match at least one transaction with the GAMLS data based on the property's address, sales price and sale date. We then identify whether the agent represented the buyer or seller, based on the agent's name. If there is a match we associate the sex, race and ethnicity of the agent with their unique GAMLS ID.

[10] Amazon monitors worker performance over time and identifies high performing workers. Workers who demonstrate excellence across a wide range of tasks are awarded the "masters" qualification. "Masters" must continuously pass a statistical monitoring test to maintain the qualification.

[11] Additional information on the HMDA forms and collection process available in 'Appendix B to 12 C.F.R. Part 1003' of the Guide to HMDA Reporting (FFIEC [2013]).

AMT workers were shown 6,164 real estate images. 441 images were categorized as having "No individual in photo" or "More than one individual in photo" and 817 images did not have a recent transaction listed on www.realtor.com. Descriptive statistics for the remaining 4,906 agents are displayed in Table 1. Approximately 70.1 percent (3,441) of the real estate agents are female and 29.9 percent (1,465) are male. White agents represent the largest racial group (79.5 percent) followed by black agents (15.2 percent), Hispanic agents (3.6 percent), and Asian or other (1.7 percent). The NAR reports that 62 percent of their members are female and that the "typical Realtor is a 53 year old white female" (NAR [2016]), so it is not a surprise that white female agents represent a majority (55.2 percent) of the sample.

2.4 Transaction Data

Prior to merging the agent attributes with the transaction data we remove records that have a sales price less than $30,000 or greater than $3,000,000, have more than six bedrooms or bathrooms, are under construction, are new construction, or are missing data in key fields. Table 2 presents descriptive statistics for the agent and buyer attributes, select property characteristics, and transaction prices. Column 1 presents descriptive statistics for 83,989 transactions in which one of the 4,906 agents from Table 1 represented the home buyer (i.e. they were the selling agent) from 1997 through 2016. Next we merge the MLS data in column 1 with the CoreLogic and HMDA data to incorporate the race of the home buyer. Column 2 displays summary statistics for the resulting 32,614 transactions.[12] Column 3 displays summary statistics for a subset the transactions in column 2 where we know the race of the selling agent, home buyer, and listing agent. Note that a large portion (23 percent) of the transactions in column 3 are dual agent transactions where one real estate agent represents

[12] A large number (51,375) of the MLS transactions in Column 1 are dropped for several reasons. First, we only have CoreLogic data from January 2000 onwards, so the MLS transactions from 1997-1999 are dropped. Second, HMDA data for 2016 data will not be available until September 2017, so the MLS and CoreLogic transactions in 2016 are dropped. Third, the HMDA data does not cover the entire mortgage market, so financed transactions not contained in the HMDA dataset are dropped. Fourth, if a loan application is submitted by phone, mail or on the internet the lender is not required to report the applicants race, so the data is not available and the record is dropped. Finally, cash purchases are dropped because they are not in the HMDA dataset.

both the home buyer and seller.

We recognize that the precision of estimated pricing differentials increases as the number of transactions in which the race of the home buyer increases. Transactions in which both the selling agent and home buyer are either black or white represent approximately 85.7 percent of the transactions in Column 2 of Table 2. For this reason the bulk of our empirical analysis focuses on black-white pricing differentials.

2.5 Selling Agent Selection

When estimating racial price differentials we must consider that home buyers are not randomly assigned to selling agents. If home buyers are more likely to select a selling agent of the same race, then the selling agent's race may be used as a proxy for the home buyer's race during negotiations. Table 3 displays a pairwise comparison of the race of the home buyer and selling agent for the transactions listed in column 2 of Table 2. The descriptive statistics in Table 3 suggest that the race of the selling agent is loosely correlated with the race of the home buyer. In other words, home buyers disproportionately hire selling agents of the same race relative to the underlying real estate agent population.

Asian home buyers were represented by Asian selling agents in 10.1 percent of their transactions even though Asian agents only accounted for 1.7 percent of the agent population. Asian agent's 10.1 percent market share among Asian home buyers is considerably higher than black, Hispanic and white home buyers. Black home buyers were represented by black selling agents in 50.1 percent of their transactions even though black agents accounted for 15.2 percent of the agent population. Black home buyers are the only race that did not employ a white agent the majority of the time - although white agents still represented black home buyers in 47 percent of their transactions.

Hispanic home buyers were represented by Hispanic selling agents in 19.0 percent of their transactions even though Hispanic agents only accounted for 3.6 percent of the agent population. Similar to Asian home buyers, Hispanic home buyers hired white agents the majority of the time (73.6 percent). Similar to the other races, white home buyers were disproportionately represented by selling agents of the same race relative to the agent population. White selling agents represented white home buyers in 94.3 percent of their transactions even though white agents accounted for 79.5 percent of the agent population.

2.6 Neighborhood Sorting

When estimating racial price differentials one must recognize that home buyers are not randomly assigned to neighborhoods. This is particularly important when estimating racial price differentials because if black neighborhoods are systematically different than white neighborhoods and they are not properly controlled for, then price estimates for black home buyers will be biased. To highlight the importance of controlling for neighborhood characteristics, such as racial demographics, we plot kernel density estimates by buyer race using the percent 'Black or African American alone' 2010 census block group level estimates. Panel A of Figure 1 clearly shows that black buyers are more likely to purchase houses in neighborhoods that are majority black - especially in census block groups that are greater than 80 percent black. Whereas, Asian, Hispanic and white buyers are more likely to purchase houses in census block groups that are not majority black.

Panels B and C of Figure 1 plot kernel density estimates by race for selling agents and listing agents, respectively. The plots by race are similar to Panel A. Black selling (listing) agents are more likely to represent home buyers (sellers) in neighborhoods that are majority black. Whereas, non-black selling (listing) agents are more likely to represent home buyers (sellers) in neighborhoods that are not majority black.

3 Racial Price Differentials

When estimating racial price differentials one must compare prices paid by buyers of different races for *identical* housing. The heterogeneous nature of the housing market makes this difficult because no two properties are identical. Previous research has attempted to overcome this obstacle using one of two distinct approaches.

In the first approach, which we will refer to as the hedonic approach going forward, researchers isolate racial pricing differentials using a hedonic model that attempts to control for all characteristics of the house and surrounding neighborhood (see, for example, King and Mieszkowski [1973]; Yinger [1978]; Chambers [1992]; Kiel and Zabel [1996]; Ihlanfeldt and Mayock [2009]). The primary advantage of the hedonic approach is its representativeness, as it includes the entire sample of housing transactions. The primary disadvantage of the first approach is its susceptibility, both at the property and neighborhood level, to omitted variable bias. At the property level, omitted variable bias is an issue if the buyer's race is correlated with unobserved property characteristics. For example, if one race has a higher level of income and wealth, then they will likely purchase higher quality housing that is in better condition relative to other housing in the same neighborhood. If property quality and condition are not properly controlled for, then the price estimates for the wealthier race will be biased upwards. Similarly, if buyers sort into neighborhoods based on their race and those neighborhoods offer different amenities, then the price estimates will be biased if the differing amenities are not properly controlled for (Bayer et al. [2007]). This is an important concern in this study, given the neighborhood selection relationship we document in Section 2.6.

In the second approach, which we will refer to as the repeat sales approach going forward, researchers isolate racial price differentials using a hedonic model that includes house fixed effects (Myers [2004]; Bayer et al. [2012]). The primary advantage of the repeat sales approach is the inclusion of house fixed effects that reduce the model's susceptibility to omitted variable bias. There are, however, several disadvantages to the repeat sales approach. For example, the approach assumes that the quality and condition of the house are constant over time, so it does not entirely address the omitted variable bias discussed in the first approach.[13] The inclusion of house fixed effects also requires that a house sell at least two times, so houses that sold only once are dropped raising serious sample selection concerns. Gatzlaff and Haurin [1997] argue that repeat sales samples are not random, representative samples of the housing stock and are therefore subject to substantial bias.

Another challenge for the repeat sales approach is the level at which the repeat sales indices must be estimated so that they accurately capture the price dynamics of a neighborhood over time. Bayer et al. [2012] include census tract-by-time fixed effects in their models. The authors argue that the inclusion of the tract-by-time fixed effects ensures that the race parameters are identified by comparing transactions within the same neighborhood during the same time period. Myers [2004] notes, however, that census tracts contain approximately 1,527 houses each, so they are relatively large and represent several neighborhoods. Census block groups are a more appropriate neighborhood measure based on their smaller size. However, their use in the repeat sales approach is precluded because of their smaller size.

3.1 Notation

There are a total of $i = 1, ..., N$ black buyers. The N_T black buyers who use a white agent are the *treatment* group, and the N_C black buyers who use a black agent are the *candidate* group. For any individual i, the treatment indicator $d_i = 1$ if individual i is in the treatment group and $d_i = 0$ otherwise. Without loss of generality, data is ordered so that the first N_T individuals are in the treatment group. We use $\mathcal{S}_T$ when referring to the $i = 1, .., N_T$ units in the treatment group and $\mathcal{S}_C$ when referring to the $i = N_T + 1, ..., N$ individuals in the control group. For each unit in $\mathcal{S}_T$ and $\mathcal{S}_C$, the $K \times 1$ vector of location and housing attributes is represented by, $x_i \in \mathbb{R}^K$. Likewise, log transaction prices are given by $p_i \in \mathbb{R}^1$. We use $z_i \in \mathbb{R}^Q$ to represent the $Q < K$ attributes specific to the property excluding location attributes.

Individuals in $\mathcal{S}_C$ are matched to individuals in the treatment group based on the distance

[13] Bayer et al. [2012] try to address this concern using tax assessor records that identify houses with major renovations. However, the tax assessor records only identify major renovations that require a permit.

$d(x_i, x_j) = x_i' W x_j$ where W is a positive definite weighting matrix. The set of $1 \leq M$ individuals in $\mathcal{S}_C$ matched to individual i is denoted by $m(i) \subseteq \mathcal{S}_C$. The *matched* set, $\mathcal{S}_M$, is the union $\bigcup_i m(i) \subseteq \mathcal{S}_C$. For any $i \in \mathcal{S}_T$, we use $\bar{x}_{m(i)} = M^{-1} \sum_{j \in m(i)} x_j$ to denote the average value of x_j across the individuals in $m(i)$ and $\Delta x_i = x_i - \bar{x}_{m(i)}$ to denote the difference between x_i and $\bar{x}_{m(i)}$. Similar average and difference notation is used for p_i and z_i.

3.2 Methodology

Regardless of the approach employed, previous research finds that racial pricing differentials exist in housing markets - although the magnitude and direction of the price differential varies across approaches and studies. Recent studies using the hedonic approach generally find that black buyers pay a discount relative to white buyers (Chambers [1992]; Kiel and Zabel [1996]; Ihlanfeldt and Mayock [2009]). In contrast, recent studies that use the repeat sales approach find that black buyers pay a premium relative to white buyers (Myers [2004]; Bayer et al. [2012]). The difference in findings is often attributed to an omitted variable bias where the omitted variables (condition, quality, neighborhood amenities, etc.) are correlated with the race of the home buyer in the hedonic approach.

In this study, we use the hedonic approach. However, we create a matched data set to address the omitted variable bias identified in previous studies. Therefore, we are interested in comparing the expected transaction price for black buyers with and without the treatment given by:

$$\tau | x_i \equiv \mathbb{E}[p_i | x_i, d_i = 1] - \mathbb{E}[y_i | x_i, d_i = 0] \tag{1}$$

In 1, $\mathbb{E}[p_i | x_i, d_i = d]$ is the expected log transaction price conditional on house attributes x_i and $d_i = d$. Thus, $\tau | x_i$ is the difference in the transaction price when the treatment is applied.

In a controlled experimental setting, we would compare the log transaction prices for the same black buyer purchasing the same house using both a black real estate agent and a white real estate agent. For obvious reasons, this is not feasible. Instead, we pursue a practical approach used in many observation studies. We match transaction $i \in \mathcal{S}_T$ to transactions $m(i) \subseteq \mathcal{S}_C$ using observables x_i. After standardizing the observables to have mean 0 and unit variance, we form $m(i)$ by solving

$$m(i) = \min_{j \in \mathcal{S}_C} x_i' W x_j \tag{2}$$

In 2, W is a positive-definite weighting matrix, and $m(i)$ are the M transactions in $\mathcal{S}_C$ that are as close as possible to X_i as measured by the distance $d(X_i, X_j) = x_i' W x_j$. The researcher can choose any $0 \prec W$. In the interest of transparency, we explore $W = I$ and $W = \hat{\Sigma}^{-1}$. The set $\mathcal{S}_M = \bigcup_i m(i) \subseteq \mathcal{S}_C$ is then the set of matched observations.

Using $\mathcal{S}_M$, we then compare the log transaction price p_i to the average transaction price for the transactions in $m(i)$ given by $\bar{p}_{m(i)}$. For unit i, the individual level treatment effect is given by $\hat{\tau}_i = \Delta p_i$. The average treatment effect for the treated is then calculated as the average $t\hat{a}u_i$

$$\hat{\tau} = N_T^{-1} \sum_{i \in \mathcal{S}_T} \hat{\tau}_i \tag{3}$$

Alternatively, $\hat{\tau}$ is the difference in the mean log transaction price in the treatment group and the mean log transaction price across the matched transactions.

In our analysis, we match on variables specific to the house: latitude, longitude, square footage of living area, date of sale, percent white by census block group, percent black by census block group, and indicators for real estate owned (reo) sales, foreclosures, and short sales. Because we use continuous variables, we are not able to exactly match x_i to M units in $\mathcal{S}_C$. Because we do not match exactly, there is a bias in the estimate of $\hat{\tau}$. In order to correct for this bias, we then adjust $\hat{\tau}$ using a linear approximation of the true hedonic pricing model and a subset of variables $Z_i \subset X_i$

$$\hat{\tau}_{adj} = N_T^{-1} \sum_{i \in \mathcal{S}_T} \hat{\tau}_i + N_T^{-1} \sum_{i \in \mathcal{S}_T} \hat{\beta} \Delta z_i \quad (4)$$

Here, Δz_i is the difference between z_i and the average of the z in $m(i)$. The vector $\hat{\beta}$ are implicit prices that adjust for difference in z_i and $\bar{z}_{m(i)}$. Following Imbens and Rubin [2015], we estimate $\hat{\beta}$ by estimating $p_i = \alpha + \beta x_i + \tau d_i + u_i$ using least squares for $i = 1, ..., N$.

We use square footage of living area, and the reo, shortsale, and foreclosure indicators to adjust the price. The remaining variables in $x_i \setminus z_i$ indicate date of sale and references to location. Properties nearer the city center are more expensive than properties further from the city center. Furthermore, dynamic prices in Atlanta do not appear to be linear, Figure 2. In the presence of these non-linearities, we do not adjust prices based on location.

4 Results

Figure 3 displays the location of the 4,468 transactions with a black buyer and a random sample of 4,468 transactions without a black buyer. It is clear that black buyers purchase homes south of the city, while non-black buyers purchase homes directly north of the city. Were we to test for price effects using the entire sample of data, we would unavoidably oversample properties north of the city where few blacks purchase properties. Instead, the matching procedure finds properties purchased by the $N_T = 2,164$ black buyers in $\mathcal{S}_C$ close to properties purchased by the $N_T = 2,304$ black buyers in $\mathcal{S}_T$.

In order to illustrate the appropriateness of our matching estimator, we compare the empirical cumulative distribution function (ecdf) of the variables in X for observations in $\mathcal{S}_T, \mathcal{S}_C$, and the combined sample of all black buyers. The ecdf's are presented in Figure 4. Figure 4 indicates that the treatment group is comparable to the entire sample of transactions by black buyers as the ecdf for days, square footage of living space, and longitude are comparable. However, the ecdf's for latitude, percent black, and percent white are significantly different across the treatment group and the full set of purchases by black buyers. However, after matching, the treatment and control groups are well-balanced as the ecdfs for all variables in X_i are nearly identical.

Table 4 presents the $\hat{\tau}$ for various M and W both with and without bias adjustments. Across the various models, the average treatment effect ranges from 3.5 percent to just under 5.0 percent. The results suggest that the race of the agent representing the home buyer affects the transaction price. Black home buyers that are represented by black real estate agents pay a premium relative to black home buyers that are represented by white real estate agents.

4.1 Study Limitations

Although the dataset utilized in this study is rich in terms of observable property and neighborhood characteristics it does have several limitations. The collection and classification of the real estate agents' race was performed once, so the racial profile of the agent network is static. In reality, the agent network is dynamic so its racial profile likely changed over time - especially in the late-2000s when the economy went into a recession.[14] Because the agent network in this study represents a snapshot in time we recognize that there may be a survivorship bias in terms of the agents included in this study. However, given the time consuming task and difficulty/cost of identifying the race of agents who are no longer active in the network we leave this inquiry for future research.

We also recognize that membership in the National Association of Realtors (NAR) is optional and that we were forced to drop 1,258 of the 6,164 agent records listed on the NAR website because they were missing a photo or did not have a matching transaction in the MLS dataset. If membership in NAR or posting a photo on www.realtor.com is correlated with agent race, then the racial profile of the agent network we present may be skewed. To partially address this concern we compared the number of agents we classified (4,906) to

[14] Membership in NAR dropped from a high of 1,357,732 in 2006 to 999,824 in 2012 according to the NARs historical data on membership (NAR [2017]).

the number of unique agent IDs in the MLS data set that had at least one transaction in 2016. The results show that the agent network we examine represents a large proportion (approximately 68.7 percent) of the agents that had at least one transaction in 2016. In addition to the representativeness of the NAR website, its use also ensures that the agents included in our study received ongoing training on housing discrimination and fair housing policy.

5 Conclusion

The primary contribution of this study is the identification of the dimensions along which differential treatment occurs in housing markets. We show that home buyers are disproportionately represented by selling agents of the same race relative to the underlying real estate agent population. Because of this the selling agent's race may serve as a proxy for the home buyer's race during negotiations. To test this conjecture we compare transaction prices for black home buyers when they are represented by black selling agents relative to when they are represented by white selling agents. Using a matched data set we find that black home buyers pay a 3.5 to 5.0 percent premium when they are represented by black selling agents. The results suggest that black home buyers looking to "avoid" discrimination can hire a white selling agent. Such action would, however, result in a form of indirect discrimination that adversely affects black selling agents.

This finding provides evidence of racial discrimination in housing markets and more importantly identifies the dimension along which differential treatment occurs. Agent intermediation should eliminate racial price differentials in housing markets - especially given the fact that every agent in this study received training on the topic. Still, we find that racial discrimination remains prevalent in housing markets. The results have important implications for fair housing policy and suggest that "masking" the race of both the home buyer *and* selling agent is necessary to prevent racial discrimination.

References

John Yinger. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. Russell Sage Foundation, December 1995.

Shaila Dewan. Gentrification Changing Face of New Atlanta. *The New York Times*, March 2006. URL https://www.nytimes.com/2006/03/11/us/gentrification-changing-face-of-new-atlanta.html.

FFIEC. A Guide to HMDA Reporting - Getting it Right!, 2013.

NAR. National association of realtors member profile, 2016. URL https://www.nar.realtor/field-guides/field-guide-to-quick-real-estate-statistics.

A. Thomas King and Peter Mieszkowski. Racial Discrimination, Segregation, and the Price of Housing. *Journal of Political Economy*, 81(3):590–606, 1973.

John Yinger. The Black-White Price Differential in Housing: Some Further Evidence, 1978.

Daniel N. Chambers. The racial housing price differential and racially transitional neighborhoods. *Journal of Urban Economics*, 32(2):214–232, 1992.

Katherine A. Kiel and Jeffrey E. Zabel. House Price Differentials in U.S. Cities: Household and Neighborhood Racial Effects. *Journal of Housing Economics*, 5(2):143–165, 1996.

Keith Ihlanfeldt and Tom Mayock. Price discrimination in the housing market. *Journal of Urban Economics*, 66(2):125–140, 2009.

Patrick Bayer, Fernando Ferreira, and Robert McMillan. A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, 115(4):588–638, 2007.

Caitlin Knowles Myers. Discrimination and neighborhood effects: understanding racial differentials in US housing prices. *Journal of Urban Economics*, 56(2):279–302, 2004.

Patrick Bayer, Marcus D. Casey, Fernando Ferreira, and Robert McMillan. Estimating Racial Price Differentials in the Housing Market. Working Paper 18069, National Bureau of Economic Research, 2012.

Dean H. Gatzlaff and Donald R. Haurin. Sample Selection Bias and Repeat-Sales Index Estimates. *The Journal of Real Estate Finance and Economics*, 14(1-2):33–50, 1997.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

NAR. National association of realtors historic membership report, 2017. URL https://www.nar.realtor/membership/historic-report.

Table 1: Agent Race and Ethnicity by Sex

	(1) Female	(2) Male	(3) Total
White	2,709	1,192	3,901
	(55.2%)	(24.3%)	(79.5%)
Black	546	202	748
	(11.1%)	(4.1%)	(15.2%)
Hispanic	129	46	175
	(2.6%)	(0.9%)	(3.6%)
Asian and Other	57	25	82
	(1.2%)	(0.5%)	(1.7%)
Total	3,441	1,465	4,906
	(70.1%)	(29.9%)	(100.0%)

Note: This table tabulates the race/ethnicity by sex of the 4,906 real estate agents included in this study. The real estate agents ethnicity takes precedence in the classification process. Agents whose race is white and ethnicity is Hispanic are listed as Hispanic in this table. The same goes for agents whose race was classified as black or Asian and other. The race and sex of the real estate agents are sorted in descending order.

Table 2: Descriptive statistics for transaction data

	(1) Sell Agent Only	(2) Sell Agent and Buyer	(3) Agents and Buyer
Sales Price	267,057	290,580	314,960
Lot Size (sqft)	2,444	2,528	2,593
Living Area (sqft)	19,788	20,439	22,420
Bedrooms	3.78	3.85	3.89
Bathrooms	2.49	2.55	2.61
Age	28.40	27.76	28.83
Dual Agent	0.13	0.11	0.23
In-house	0.05	0.05	0.05
Agent Owned	0.01	0.01	0.01
Agent Related	0.01	0.01	0.00
SA: Asian or Other	0.02	0.02	0.01
SA: Black	0.11	0.10	0.07
SA: Hispanic	0.04	0.03	0.03
SA: White	0.83	0.86	0.89
Buyer: Asian or Other		0.06	0.07
Buyer: Black		0.14	0.11
Buyer: Hispanic		0.05	0.05
Buyer: White		0.75	0.77
LA: Asian or Other			0.01
LA: Black			0.04
LA: Hispanic			0.03
LA: White			0.92
Total	83,989	32,614	15,433

Note: Column 1 provides descriptive statistics for every transaction in which one of the 4,906 real estate agents from Table 1 represented the home buyer (i.e. was the selling agent). Column 2 represents a subset of column 1 in which both the race of the selling agent and buyer is known. The subset is created by merging the Multiple Listing Service (MLS) data from column 1 with CoreLogic and HMDA data. Column 3 represents a subset of column 2 in which we know the race of the selling agent, buyer, and listing agent.

Table 3: Selling agent and buyer race

	Buyer's Race				
	(1)	(2)	(3)	(4)	(5)
	Asian and Other	Black	Hispanic	White	Total
Asian and Other	210	52	16	212	490
	(10.1%)	(1.1%)	(1.0%)	(0.9%)	(1.5%)
Black	138	2304	98	572	3,112
	(6.6%)	(50.1%)	(6.3%)	(2.3%)	(9.5%)
Hispanic	133	81	294	613	1,121
	(6.4%)	(1.8%)	(19.0%)	(2.5%)	(3.4%)
White	1,596	2,164	1,136	22,995	27,891
	(76.8%)	(47.0%)	(73.6%)	(94.3%)	(85.5%)
Total	2,077	4,601	1,544	24,392	32,614
	(6.4%)	(14.1%)	(4.7%)	(74.8%)	

Note: This table displays the total number of transactions by selling agent and buyer race. The number of transactions by buyer race is tabulated vertically and the number of transactions by selling agent race is tabulated horizontally. Additional descriptive statistics for the transactions in this table are displayed in column 2 of Table 2. The selling agent and buyers race are both sorted alphabetically in this table.

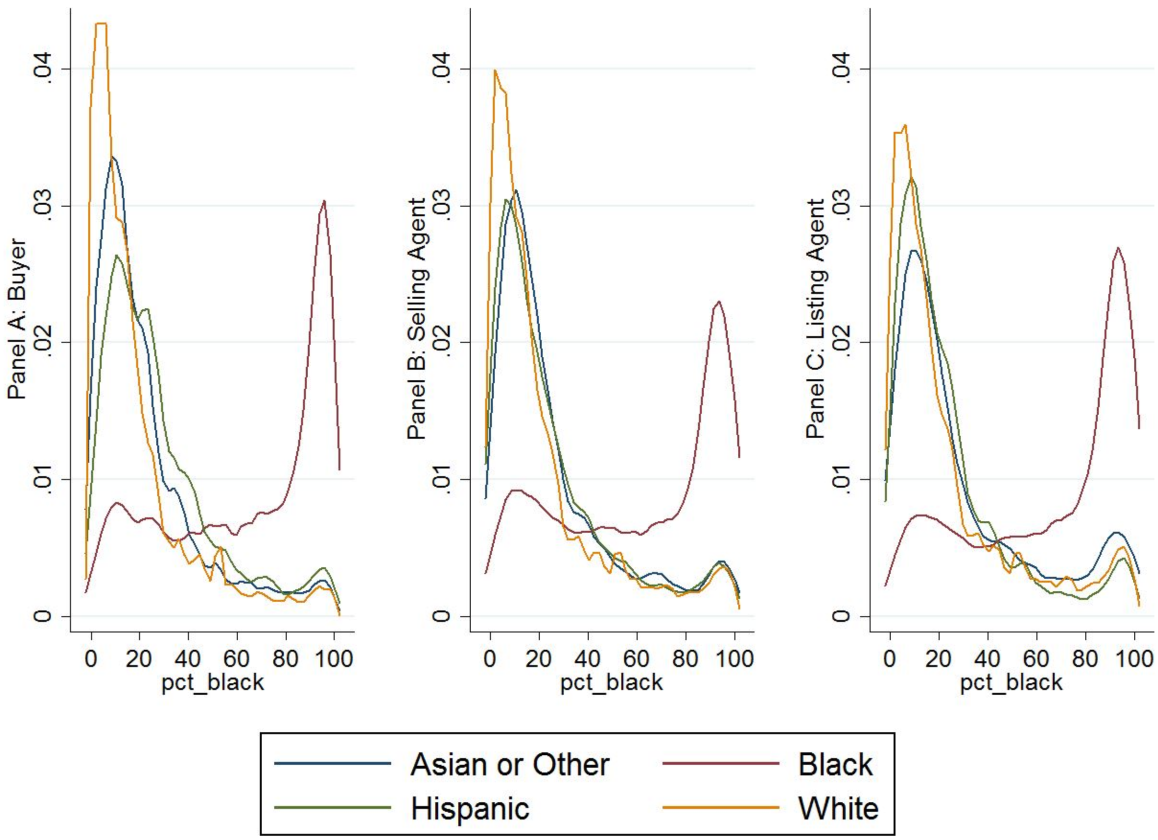
Table 4: Average Treatment for the Treated

	Model 1	Model 2	Model 3	Model 4	Model 5
	0.0448***	0.0434***	0.0499***	0.0357***	0.0418***
	(0.0107)	(0.01)	(0.0099)	(0.0096)	(0.0098)
N	4,468	4,468	4,468	4,468	4,468
M	1	2	3	2	2
Bias Adjust				✓	✓
W	I	I	I	$\hat{\Sigma}^{-1}$	$\hat{\Sigma}^{-1}$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

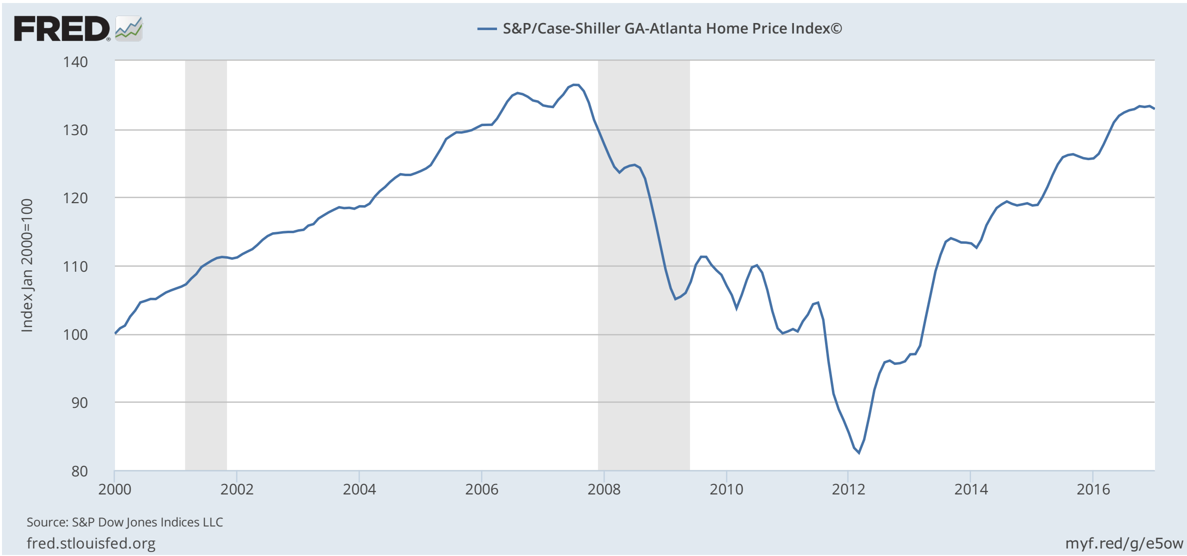
Note: This table presents the average treatment for the treated for various matching estimators. M is the number of matches in the control group, a bias adjustment can be made using a linear approximation of the hedonic pricing model, and W is the weighting matrix used in the distance calculation.

Figure 1: Kernel Density by Race



Note: This figure plots kernel density estimates by race and neighborhood composition for three distinct market participants. Panel A plots the transaction density by home buyer race relative to the neighborhoods percent black. Panel B and C plot similar densities for selling agent and listing agent race, respectively. Neighborhood percent black is measured at the 2010 census block group level.

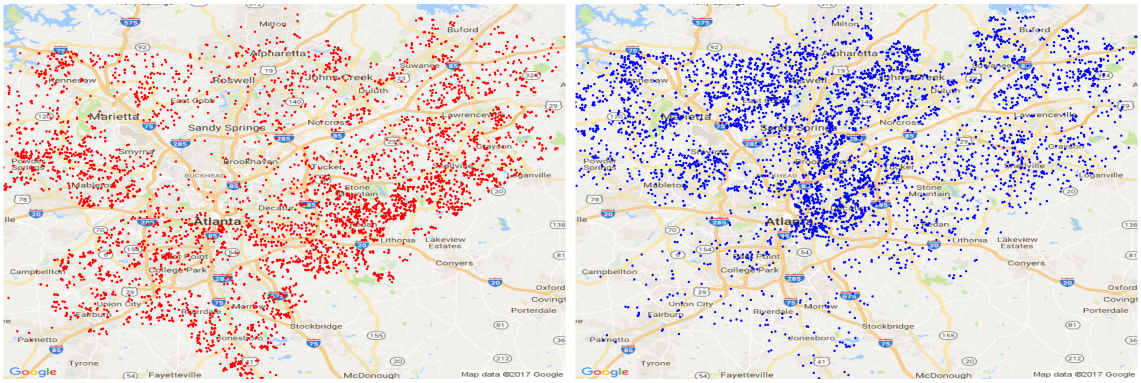
Figure 2: Case-Shiller Price Index for Atlanta, GA



Note: This figure is the Case-Shiller house price index for Atlanta, GA.

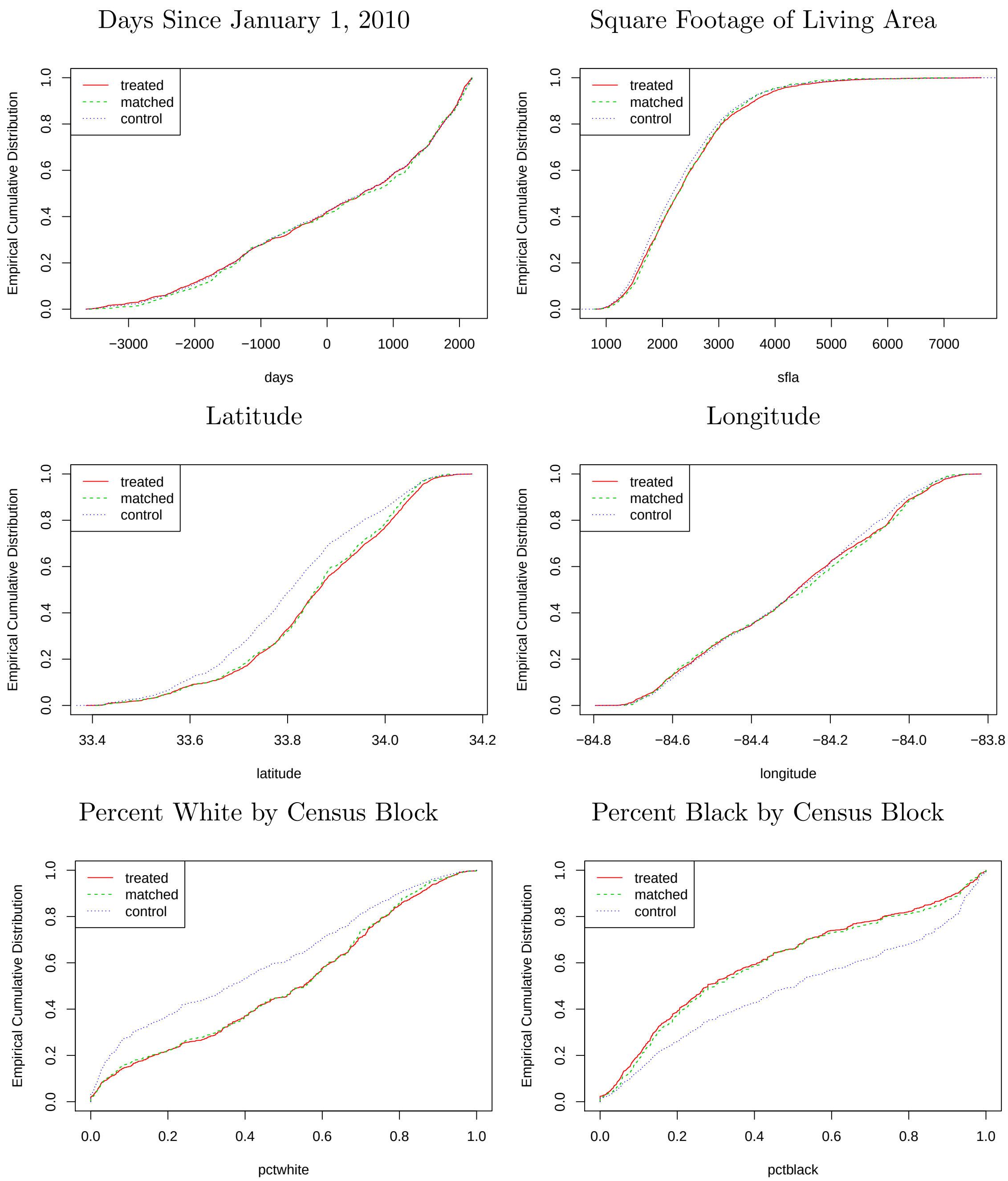
Figure 3: Transactions in Atlanta, Georgia

Black Buyer Transactions | All Transactions



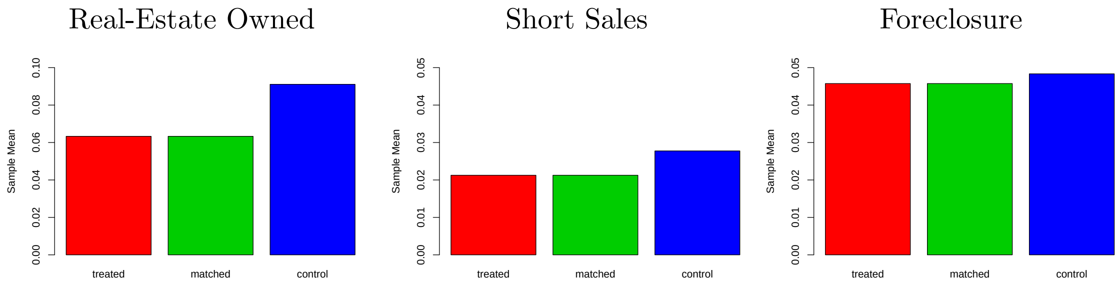
Note: This figure plots the locations of all 4,468 transactions with an identified black buyer and the locations of a random sample of 4,468 transactions from the entire data set.

Figure 4: Assessing Balance



Note: This figure plots the empirical cumulative distribution function for the continuous variables associated with black buyers who used a white agent, *treatment*, black buyers who used a black agent in the matched sample, *matched*, and all black buyers, *control*. The control group is matched using $M = 1$ and an identify weighting matrix.

Figure 5: Assessing Balance



Note: This figure plots the sample average for the binary variables associated with black buyers who used a white agent, *treatment*, black buyers who used a black agent in the matched sample, *matched*, and all black buyers, *control*. The control group is matched using $M = 1$ and an identify weighting matrix.