Racial Housing Price Differentials and Neighborhood Segregation

Authors: Sébastien Box-Couillard and Peter Christensen*

December 18, 2023

Abstract

We report evidence from the largest study of racial price differentials in the U.S. housing market, constructing a panel of 40 million repeat-sales transactions. We find that price premiums facing Black and Hispanic homebuyers are ubiquitous and systematically higher in neighborhoods with a larger share of non-white residents. We find that non-white buyers purchase at a premium from sellers from a different group. Consistent with predictions from theoretical models (Becker, 1957), we find higher premiums in supply-constrained markets. Leveraging exogenous variation in racial segregation, we find that segregation increases price premiums paid by Black homebuyers.

Keywords: Housing Discrimination, Segregation, Repeat-Sales Method
JEL Classification: R2, R3, J15

*Box-Couillard: University of Illinois, 1301 W. Gregory, Urbana, Illinois 61801 (email: sb38@illinois.edu); Christensen: University of Illinois, 1301 W. Gregory, Urbana, Illinois 61801 (email: pchrist@illinois.edu).

Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.

Introduction

Economic theory predicts that any form of discrimination in the housing market that results in a reduced choice set for minority buyers will lead to a price premium for minority homebuyers (Becker, 1957, Courant, 1978, Masson, 1973, Yinger, 1997). Prior to the Fair Housing Act of 1968, race-restricting covenants and redlining restricted access to housing markets for minorities, particularly African Americans. These policies restricted the supply of housing available to African Americans and were shown to lead to higher housing costs (Cutler et al., 1999, Kain and Quigley, 1975, King and Mieszkowski, 1973). While the Fair Housing Act created a set of institutions that define and enforce anti-discrimination policies, empirical evidence indicates that price premiums persist for minority homebuyers (Bayer et al., 2017, Ihlanfeldt and Mayock, 2009, Myers, 2004). However, the current evidence is limited to a small set of markets and very little work exists on the mechanisms that underlie racial/ethnic price differentials or can explain their persistence in U.S. markets.

To address these gaps, we construct a novel dataset that includes approximately 40M repeat home sales during the two-decade period from 2000-2020. Building on the frontier repeat-sales research design with neighborhood-by-year fixed effects developed by Bayer et al. (2017), we expand the geographic and temporal coverage of the extant evidence on racial housing price differentials from a small number of select markets to include the universe of U.S. metropolitan housing market transaction data.[1] We also extend the repeat-sales approach in a number of ways, including addressing time-varying differences in the likelihood that buyers from different race groups will transact and the impacts of time-varying unobservables such as home renovations. We find that over the past twenty years, the average Black and Hispanic home buyer in the United States paid a 3% premium relative to white homebuyers to purchase an equivalent property. However, we find substantial heterogeneity across the U.S.

The size of the sample in the present study, which includes approximately 20 times

[1]This includes data from 34 "disclosure" states and the District of Columbia, where housing transaction data are publicly disclosed.

more transactions than prior work,[2] enables statistically powered analysis of heterogeneity in price differentials and new empirical tests of the mechanisms that underlie them. We begin by examining Becker (1957) prediction that in supply-constrained markets, discrimination will be less costly for sellers and differential pricing is more likely to arise. We use a measure of market-specific fluctuations in sales inventory to examine variation in the magnitude of price differentials in thick vs. thin housing markets. We find that during low-inventory periods, Black and Hispanic buyers pay respective premiums that are 1.3 and 1.8 percentage points higher than the respective premiums facing each group during a typical period in the same market. During high-inventory periods, Black and Hispanic buyers pay premiums that are 2.3 and 1.5 percentage points lower than the respective premiums observed during a typical period. We observe a similar pattern during "hot" periods, which we measure using a measure of housing search intensity.

We then leverage data on the racial/ethnic identities of buyer-seller pairs to examine heterogeneity in differentials for transactions involving sellers from the buyer's same group (ingroup) versus those from a different racial/ethnic group (outgroup). We find robust evidence of higher price premiums in transactions involving sellers from an outgroup. However, whereas the average "outgroup differential" results in a premium of just 0.2 for white buyers, it ranges from 3.3 percentage points for LatinX buyers to 7.7 percentage points for African American buyers. We also find evidence of "ingroup differentials" that result in discounts for buyers who transact with a buyer from their same group. Ingroup discounts are more pronounced for buyer-seller pairs from a minority group.

We then examine how these patterns vary with neighborhood racial composition. We find that while premiums are ubiquitous for African American and Hispanic buyers, they are systematically larger in neighborhoods with larger shares of own-race residents. Discriminatory constraints, which have also been shown to be stronger in neighborhoods where minority homebuyers are less well-represented (Christensen and Timmins, 2022, Turner et al., 2013), could result in more inelastic demand for housing everywhere else. Homophily preferences could also induce homebuyers from different groups to pay a higher

[2] Bayer et al. (2017) is the next largest study, analyzing a dataset of 2M observations.

prices for the same housing in neighborhoods where their group has greater representation (Aliprantis et al., 2022, Caetano and Maheshri, 2019, Davis et al., 2023). However, within own-race neighborhoods, we find that premiums are stronger for all groups when transactions involve a seller from a racial or ethnic outgroup.

We examine the effect of systemic factors acting as supply constraints by testing for a causal effect of housing market segregation on price differentials. A long-standing literature has examined the dynamics of migration into ethnic enclaves (Altonji and Card, 1991, Card, 2001) and the ways in which contemporaneous sorting behavior among white households influenced the segregation of U.S. cities throughout the 20th Century (Boustan, 2010, Card et al., 2008, Shertzer and Walsh, 2019). Persistent patterns of residential segregation have resulted in systematically lower rates of school performance and wage rates (Boustan, 2012), as well as higher poverty and lower intergenerational mobility for Black residents (Akbar et al., 2019, Ananat, 2011, Chyn et al., 2022). Using an instrumental variables approach based on historical railroad track location (Ananat, 2011, Chyn et al., 2022, Cox et al., 2022), the present study reveals that the patterns of segregation that evolved during reconstruction and the Great Migration also contribute to systemic disparities by exacerbating differential housing prices. A one standard deviation reduction in the level of a city's segregation would eliminate over two-thirds of the price premium paid by Black homebuyers in these cities. We show that segregation acts as a systemic constraint on Black buyers purchasing homes in neighborhoods that are less than 2% Black, neighborhoods that represent over half of the housing stock in our sample. Although segregation increases price differentials everywhere, this is particularly the case in neighborhoods with a relatively larger share of Black households, where segregation increases differentials by 300% more than in low Black share neighborhoods. By constraining the choice set of Black buyers in a majority of neighborhoods, segregation leads to significant price premiums, especially in high Black share neighborhoods.

Data

We combine data from Zillow's ZTRAX database (Zillow, 2020) and the Ethnicolr algorithm (Sood and Laohaprapanon, 2018) to construct a novel dataset covering all housing transactions between 2000-2020 in 34 states and the District of Columbia, representing 80% of the American population.[3] Over the time period of our study, there are 106M transactions in the ZTRAX database in these states. For each transaction, we observe the transaction price, date of sale, year the home was built, location of the home, property characteristics (e.g., number of bedrooms, bathrooms, etc.), buyer name and seller name. Since our research design relies on comparing transactions for homes that sold more than once, we restrict our data to properties that appear at least twice. This yields a total of approximately 58M observations.

We use the Ethnicolr algorithm to match names to a races/ethnicity using first and last names of buyers/sellers.[4] Other popular algorithms tend to use only last names to predict race or ethnicity. This can be problematic in the United States, where the last names of white and African American residents are often less differentiable than first names (Fryer Jr and Levitt, 2004). The Ethnicolr model assigns probabilities that a given name belongs to one of four race/ethnicites corresponding to: non-Hispanic white, non-Hispanic Black, Asian, and Hispanic or Latino.[5] In our base model, the race/ethnicity with the highest probability is assigned to the individual buyer or seller. Of our 58M observations, 12M do not have buyer name information, thus our dataset contains 46M observations for which a race/ethnicity can be assigned. After dropping outliers and observations that are no longer repeat-sales due to a missing buyer name, our final sample is comprised of 39.47M transactions.[6]

[3]Other states are "non-disclosure" states, meaning home sale prices are not public record in these states.
[4]For recent applications of the Ethnicolr algorithm in published research, see: Cheng and Weinberg (2021), Marschke et al. (2018), Millard-Ball et al. (2021)). Ethnicolr is trained on Florida Voter Registration data from 2017 that contains the first and last names of 13M voters in Florida, one of the more diverse states in the US.
[5]The Florida Voter Registration data used to train the Ethnicolr algorithm categorizes anyone from the Asian continent as "Asian", thus the algorithm makes no differentiation between central-Asian, Middle-Eastern, south-Asian or east-Asian origin.
[6]See Section A1.1 for more details on data cleaning and Section A1.3.3 for results based on different outlier definitions.

Methods

Our research design employs a repeat-sales estimation strategy, which compares transactions for the same home with buyers of different predicted races/ethnicities, net of effects of neighborhood price trends, seasonal price trends, and property age. This repeat-sales estimator allows us to address the limitations of previous studies that may have attributed differences in prices paid by minorities to the differing characteristics of a house or neighborhood (Chambers, 1992, Kain and Quigley, 1975, Kiel and Zabel, 1996). Methods that rely purely on higher-level fixed effects, such as census-tract fixed-effects, are likely to be biased, since white buyers are, on average, wealthier than minorities and are likely to buy higher quality housing even within a neighborhood (Bayer et al., 2017).

Our main specification estimates equation 1, where i indexes a transaction, j indexes a property and t indexes time. The dependent variable is the log of the sales price for transaction i and house j at time t. The main independent variables of interest are indicators for the predicted race of a buyer. The indicator takes a value of one if a homebuyer belongs to the Black, Asian and Hispanic groups and zero for the white group (omitted category). As a result, all estimates can be interpreted as the percent difference in purchase price for buyers from each of these groups relative to the white buyer group.

$$ln(p_{ijt}) = \beta_0 + \beta_1 Black_{it} + \beta_2 Asian_{it} + \beta_3 Hisp_{it} + \mu_j + \theta age_{jt} + cty_t + m_t + \epsilon_{ijt} \quad (1)$$

In equation 1, age_{jt} is the age of house j at the time of the transaction, μ_j is a property fixed effect, cty_t is a census-tract by year fixed effect and m_t is a calendar month fixed effect.[7] The property fixed effect controls for unobservable property or neighborhood characteristics, such that price differentials are identified using within-property variation. The age variable controls for the age of the house at the time of the transaction. The census tract by year fixed effect controls for housing price trends in a given neighborhood over time and the calendar month fixed effect accounts for seasonal variation in housing

[7]Note that while some loan information is available in the ZTRAX data, we do not include loan information in our main specification because it is impossible to distinguish homes for which a loan is not reported from homes that are cash sales. Nevertheless, our results are robust to the inclusion of loan controls. For more details see Appendix 1.3.4.

prices.

Estimates of $\beta_{1,2,3}$ identify the effect of variation in buyer race/ethnicity on price under the assumption that there is no additional unobserved variation within a property that is correlated with price. We highlight the following concerns with this baseline specification. First, it does not control for changes in the general racial composition of homebuyers over time. For example, Figure A1.2 illustrates that there were relatively more African American and Hispanic buyers in 2006-2007 and 2018-2019 than in 2010-2015 or 2000-2004. These periods of higher prevalence of Black and Hispanic buyers correspond to periods of high housing prices (see Table A1.4). As a result, while housing price differentials estimated using equation 1 identify inter-group housing price differentials, these estimates may conflate differences in the prices facing buyers with the differential probabilities of buying a property in periods of relatively high housing prices. To control for this, we introduce a variant of the standard repeat-sales estimator that includes specifications with race by year time trends (see Appendix 1.3.2).

A second concern is that the accuracy of predictions from the Ethnicolr algorithm may vary in ways that affect the estimated differentials. We test the robustness of our estimates using specifications that restrict the sample to names for which Ethnicolr assigns a high degree of certainty to its race/ethnicity assignment, by bootstrapping a subsample of our results and by using predictions from an alternate race prediction algorithm (see Appendix 1.3.1). We find that estimates are nearly identical.[8]

Finally, if certain groups are more/less likely to purchase a recently renovated or remodeled home, then estimates identified using within-property price variation may be biased (Nowak and Smith, 2018). The ZTRAX data set contains information on the remodeled status of a home as recorded by county assessor offices, allowing us to control for remodeled status of a property in this subset of 3.9 million transactions for the 1.7

[8]We recover similar estimates whether using a binary definition of race based on Ethnicolr predictions, the raw continuous race prediction from Ethnicolr, or the commonly used WRU (Imai and Khanna, 2016) algorithm (see Tables A1.7 and A1.9). Interestingly, we find that estimates of price differentials for African American and Hispanic buyers are significantly larger (5-6.5%) if we restrict the sample to names for which Ethnicolr is more certain of its prediction, implying that Black and Hispanic buyers with particularly distinct names are more likely to pay a premium.

million homes that were at some point remodeled during our study period.[9]

Racial Price Differentials across the U.S.

In this section, we report price differentials across the U.S. by neighborhood racial/ethnic composition and by seller race/ethnicity. In the presence of systemic discriminatory constraints, theory predicts that price differentials are likely to exist across U.S. housing markets, no matter the race of the seller (Becker, 1957, Courant, 1978). However, recent work has shown that minority homebuyers may have strong homophily preferences (Aliprantis et al., 2022, Caetano and Maheshri, 2019, Davis et al., 2023) and face significant constraints in their housing search, especially in higher white share neighborhoods (Christensen and Timmins, 2022, Turner et al., 2013). Both homophily preferences and discriminatory constraints in the housing search process could lead to a search that is constrained to higher own-group share areas. By reducing the choice set of minority homebuyers, homophily preferences and discriminatory constraints are likely to lead to higher premiums in higher own-group share neighborhoods (Courant, 1978). In addition, the existence of explicit discriminatory constraints in the housing market is also likely to be reflected in higher premiums when a buyer transacts with a seller from outside one's own racial or ethnic group (Becker, 1957).

Table 1 reports estimates of the racial housing price differentials for Asian, Hispanic and Non-Hispanic Black homebuyers between 2000-2020. Column 1 reports estimates from a repeat-sales specification without neighborhood-by-year fixed effects. Estimates from this specification imply that Hispanic and Non-Hispanic Black buyers pay average premiums of 4.7 and 3.5%, respectively, compared to white buyers who purchased the same home.

Column 2 reports estimates from a specification that includes tract-by-year fixed ef-

[9]Remodeled status is a time-varying indicator that measures whether a home has been remodeled at the point of a transaction. It is based on the most recent remodel year, taking a value of zero if the property was not remodeled in a prior transaction and a one if a sale takes place following a remodel. The remodeled status of a home may be subject to missing observations of homes that were remodeled and not recorded by assessors. This could introduce sampling error that could affect a robustness analysis. However, consistent estimates of price differentials between the full sample and sub-sample of remodeled properties would likely indicate that recording constraints do not affect the interpretation of results.

fects, which flexibly controls for any difference in the probability that buyers from different groups purchase homes in neighborhoods that are appreciating at different rates. Estimates are somewhat smaller with the inclusion of neighborhood-by-year fixed effects, suggesting that a portion of price differentials estimated using the standard repeat-sales design may be attributed to inter-group differences in the timing of purchases. However, this does not dramatically attenuate our estimates of price differentials. On average, Hispanic and Non-Hispanic Black buyers pay premiums of 3.0 and 3.3%, respectively, compared to white buyers who purchased the same home. Given the $248,155 average house price in our data, these estimates correspond to an average premium of $7,457 for Hispanic buyers and $8,202 for Black buyers. In contrast, Asian buyers receive a discount of 0.7% relative to white buyers.[10] We note that the present estimates suggest premiums that are approximately 90% larger than the prior estimates of premiums facing Black buyers.[11]

In columns 3 and 4, we report racial price differential estimates from the sub-sample of remodeled homes in the Zillow data. Despite other potential differences in these samples, we do not find meaningful differences in price differentials. In addition, adding a remodeled control does not change the main conclusions of the paper. In all the specifications and using both samples, African American and Hispanic buyers pay approximately 3% more for the same home as whites while Asian buyers pay 0.7-0.8% less. These results contrast with Nowak and Smith (2018), who find that racial price differentials are no longer statistically significant when accounting for the effects of home remodels.

In Figure 1, we leverage observations of buyer and seller names in the ZTRAX data to examine price differentials for transactions that are characterized by different home seller and buyer interactions in neighborhoods with varying racial and ethnic composition. While not all observed transactions contain both a buyer and seller name, the sample

[10] These results are in contrast to Bayer et al. (2017), who find that price differential estimates *increase* for Black and Hispanic buyers with the inclusion of neighborhood-by-year fixed effects. This suggests that the effects of neighborhood appreciation differ across study periods and geographies, highlighting the potential importance of samples that cover larger geographies and longer time frames for external validity.

[11] Bayer et al. (2017) estimate average premiums of 1.7% and 1.6% for Hispanic and Black buyers, respectively. For a full comparison to the results of Bayer et al. (2017), see Table A2.1. For a full set of results by State and MSA see Appendix 2.3 and for time-varying results see Appendix 2.2.

of repeat-sales with both a buyer and seller name remains large: 21M transactions. We define low (high) own-share neighborhoods as census tracts where less (more) than 5% of the population is of any given race/ethnicity as of the 2000 census. Ingroup transactions are transactions for which the seller and buyer are of the same group, outgroup transactions are those transactions where the seller and buyer are of different groups.[12]

Three patterns emerge from these results. First, we find that differentials are consistently higher in tracts with higher own-group representation and lower in tracts with minimal own-group representation. Second, ingroup premiums are always lower than outgroup premiums and in almost all cases are negative.[13] Third, transactions in high own-group share neighborhoods with an outgroup seller carry the highest premiums. While Asian, Hispanic and Black buyers obtain discounts of 9.0, 5.1 and 8.6%, respectively, in low own-share neighborhoods when buying from a buyer of the same race/ethnicity, they pay premiums of 0.5, 4.0 and 3.4% in high own-share neighborhoods when transacting with sellers from a difference race/ethnicity.

These results, combined with recent evidence of the continued existence of discrimination in the housing market (Christensen and Timmins, 2022, Turner et al., 2013) and homophily preferences (Aliprantis et al., 2022, Caetano and Maheshri, 2019, Davis et al., 2023), point to an interplay between homophily preferences and discriminatory constraints that leads to a constrained search in higher minority share neighborhoods. In the next section, we further explore the role of supply constraints in exacerbating price differentials.

[12]For a full set of results by neighborhood share, outgroup status and the interaction of the two see Appendix 2.4. Differentials in Figure 1 are calculated using the estimates in column 4 of table A2.2.

[13]These results contrast with the findings in Bayer et al. (2017) who find that ingroup transactions carry the same or a larger premium for Black and Hispanic homebuyers. However, we note that the standard errors on the the estimates in Bayer et al. (2017) for White -> Hispanic and White -> Black premiums imply that their estimates are not statistically different from ours for these same categories. However, estimates of ingroup premiums are significantly different. However, given the relatively low frequency of non-white to non-white transactions (see table A1.6), previous studies may not be sufficiently powered to precisely capture the difference in ingroup vs. outgroup premiums.

Racial Price Differentials in Competitive Markets

Theory predicts that supply constraints are likely to exacerbate price differentials. In markets with less inventory (ie. "thin" markets) or "hot" housing markets, price differentials are more likely to arise since discrimination in these contexts is less costly for sellers (Becker, 1957). We test both of these hypotheses using Realtor.com's number of listings and measure of market hotness.[14] The first measure is the number of listings in each zip code for each month. The second measure assigns a score from 0-100 for each zip code-month combination based on the number of views per listing and days on market. Note that these measures are only publicly available from July 2016 and August 2017 onward.[15]

In Table 2, we interact high/low inventory inventory and high/low market hotness indicators with race/ethnicity dummies.[16] We define high/low thresholds based on the percentile of the distribution of all zip code-month combinations in our study. A high inventory (thick) market is defined as being above the 80th percentile of Realtor.com's distribution of zip code-month listings and a low inventory market (thin) is one that is below the 20th percentile.[17] We define high/low hotness markets (hot/cold markets) in the same way using Realtor's measure of market hotness. In Columns 1 and 3 we show that, despite the much smaller samples, we find price differentials that are consistent with our previous results.

In Column 2 we report that, consistent with theoretical predictions, price differentials for Black and Hispanic buyers are significantly smaller in thick markets and larger in thin markets. In thick markets, Black and Hispanic buyers pay premiums that are 2.3 and 1.5 percentage points lower than in markets within the 20-80th percentiles. In markets with little inventory, Black and Hispanic buyers pay premiums that are 1.3 and 1.8 percentage points higher. In Column 4, we show that Black and Hispanic buyers also pay

[14] https://www.realtor.com/research/data/

[15] The hotness index is available from August 2017 and the number of listings from July 2016. To our knowledge only one published study uses Realtor.com's measure of market hotness (Borges Ferreira Neto and Whetstone, 2022). Use of Realtor.com's listings as an indicator of housing inventory is more widespread. In fact, the Federal Reserve uses this exact measure as an indicator of housing inventory (see: https://fred.stlouisfed.org/series/ACTLISCOUUS).

[16] For results with seller -> buyer race interactions see Appendix 2.5.

[17] For results using alternative definitions of thick/thin or hot/cold markets see Appendix 2.5.

lower premiums in cold markets and higher premiums in hot markets, although the small sample size yields imprecise estimates of these effects.

Taken together, these results provide strong evidence that in markets where discrimination is more costly for sellers the premiums paid by Black and Hispanic buyers are lower and where discrimination is less costly they are higher. They also demonstrate the role of constrained and more competitive searches in exacerbating price differentials. In the following section, we study segregation as a mechanism that causes systematically more concentrated housing searches for Black buyers and leads to higher price differentials.

Price Differentials: The Impact of Housing Segregation

Although we have presented suggestive evidence that is consistent with a mechanism by which supply constraints experienced by minority homebuyers exacerbate price differentials, we have yet to identify a clear mechanism through which these constraints operate. In this section, we test for the causal effect of segregation on racial housing price differentials. We provide evidence that segregation is a mechanism that both increases price differentials and lowers the probability that a Black buyer purchases a home in very white tracts.

While explicit exclusionary policies played an important role in the segregation of housing markets across the U.S. in the 20th century, segregation has continued to be a fact of life in America in the 21st century, driven by a complex interplay between sorting behavior, systemic discrimination and public goods provision (Aliprantis et al., 2022, Boustan, 2010, Card et al., 2008, Christensen and Timmins, 2022, Shertzer and Walsh, 2019). To the extent that housing market segregation continues to reduce the supply of housing available to African American homebuyers, it could raise the premium paid by supply-constrained buyers (Becker, 1957, Masson, 1973). Christensen and Timmins (2022) show that African American buyers are constrained in their search and that these constraints are systematically stronger in neighborhoods where they are not represented, and in more segregated markets. Empirical work on institutional discrimination has shown that exclusionary policies restricting the supply of housing available to African

American households resulted in higher housing costs (Cutler et al., 1999, Daniels, 1975, Kain and Quigley, 1975, King and Mieszkowski, 1973). Haugen and Heins (1969) develop a model of the impact of segregation on price differentials. They theorize that the greater the constraints imposed by segregation on the propensity for Black buyers to purchase in predominantly white neighborhoods, the greater price differentials will be in Black neighborhoods. Courant (1978) links search costs and price differentials. He shows that any factor that constrains the search of non-White buyers to certain neighborhoods can lead to price differentials.

We construct dissimilarity indices for each Metropolitan Statistical Area (MSA) in our sample using tract-level Census data from the 2000 census.[18] Following Ananat (2011), Chyn et al. (2022), Cox et al. (2022) and Cutler and Glaeser (1997), we calculate:

$$\text{Seg}_c = \frac{1}{2} \sum_{n \in c} \left| \frac{Black_n}{Black_c} - \frac{White_n}{White_c} \right|, \tag{2}$$

Seg_c measures the level of dissimilarity in each city c, where $Black_n$ is the Black population in census tract n, $Black_c$ is the Black population in city c, $White_c$ and $White_n$ are similarly defined.

We study the effect of segregation on racial housing price differentials by estimating a variant of Equation 1 that adds an interaction term between the indicator for a Black buyer and the city-specific dissimilarity index, Seg_c.[19] This results in the following equation:

$$\begin{aligned} ln(p_{ijct}) = \beta_0 + \beta_1 Black_{itc} + \beta_2 Asian_{itc} + \beta_3 Hispanic_{itc} + \beta_4 Black_{itc} \times Seg_c + \\ \theta age_{jtc} + \mu_j + cty_{tc} + m_t + \epsilon_{ijtc} \end{aligned} \tag{3}$$

However, OLS estimates from Equation 3 are unlikely to yield a causal interpretation of the effect of segregation on housing price differentials due to the fact that omitted variables could have an effect both on housing prices and segregation. For example, time-

[18] Note that using either the 2010 or 2020 census to calculate dissimilarity indices does not materially change our results.

[19] Note that including the uninteracted Seg_c term is not necessary in equation 3 since it is absorbed by the census-tract by year fixed effect.

varying differences in school quality or local government policies could increase home values and simultaneously increase segregation. To overcome this identification hurdle, we adopt a similar methodology to Ananat (2011), Chyn et al. (2022) and Cox et al. (2022), instrumenting the dissimilarity index with the Railroad Division Index (RDI) constructed in Ananat (2011). The RDI is a type of Herfindahl index measuring the amount of division into sub-units of land generated by railroad track placement. Specifically, the RDI for city c is defined as:

$$\mathrm{RDI}_c = 1 - \sum_{r \in c} \left(\frac{\mathrm{area}_r}{\mathrm{area}_c} \right)^2 \tag{4}$$

To construct the RDI, Ananat (2011) divides polygons of land in a city into "railroad neighborhoods", defined as an area that is clearly delineated by the intersection of historical railroad tracks. The RDI is then calculated as one minus the squared sum of the area in "railroad neighborhoods" ($area_r$) divided by the total area in a city ($area_c$). The closer the RDI is to one, the more subdivided by rail lines a city is. The intuition behind the first-stage is that the more divided a city was by its railroads, the easier it was for segregation to arise as this gave rise to smaller physically separated pockets of land in which segregation could more easily be enforced. The exclusion restriction is satisfied if the degree to which a city is divided into parts by railroads (RDI) is only related to current-day housing prices through its effect on segregation. The RDI constructed by Ananat (2011) is limited to 121 non-Southern cities, some of which do not correspond well to current-day MSAs. After matching those cities which clearly map onto current-day MSAs, we are left with a sample of 83 cities and 14.5M observations for this analysis.

Table 3 reports estimates of the effect of segregation on housing price differentials. These estimates indicate that segregation arising at the end of the 19th and throughout the 20th century has a strong positive effect on price premiums paid by Black homebuyers during the period 2000-2020. The estimates in the third Column imply that African Americans in a completely segregated city would pay a premium of 9%.[20] The average dissimilarity index in our data is 0.59, which implies an average price premium for Black

[20] -0.061+(1 $\times$ 0.154)=0.093

homebuyers of approximately 3%.[21] A 1 standard deviation (SD) increase in segregation results in a 2.2 p.p. increase in the price premium paid by a Black buyer.[22] The magnitude of this effect is large given that the average premium paid by a Black buyer is around 3%. These estimates suggest that a one SD reduction in the level of a city's segregation would drastically reduce the price premium for Black homebuyers.[23]

Theory predicts that, by imposing constraints on where Black buyers search for housing, segregation may lead to a more concentrated search by Black homebuyers. This concentrated search is then reflected in higher price differentials caused by a more inelastic demand curve for Black buyers (Courant, 1978, Haugen and Heins, 1969). To test this mechanism, we construct a measure that attempts to capture the likelihood that any Black buyer searches for a home in a given tract and relate it to segregation levels. Since we do not observe searches but do observe where homes are purchased, we proxy for the likelihood of search in a neighborhood with the share of Black buyers in an MSA that purchase a home in a given tract n. We construct:

$$\text{shareblack}_n = \left(\frac{\text{BlackBuyersTract}_n}{\text{BlackBuyersMSA}}\right) \tag{5}$$

We then split our sample into bins of tract-level Black population shares and regress the share of Black buyers in a tract on segregation (instrumented with the RDI) for each bin:

$$shareblack_n = \beta_0 + Seg_c + \epsilon_n \tag{6}$$

Panel A of Figure 2 presents results of these regressions.[24] In tracts that are less than 2% Black, segregation significantly reduces the likelihood that a Black buyer chooses to purchase a home in these tracts. In tracts that are 5-40% Black, segregation increases the

[21] -0.061+ (0.59 × 0.154)=0.030
[22] 1 SD is 0.14 points. 0.14 ×0.154=0.022
[23] Note that we also generate results interacting segregation and outgroup premiums. While outgroup premiums are also exacerbated by segregation, the precision of these estimates is smaller due to the more limited sample. See Figure A2.11 and Table A2.8.
[24] Note that we exclude homes in tracts that are 40-100% Black from our analysis due to the fact that this category is a significant outlier in terms of home price, with median prices almost half of those in the bin with the next lowest prices (see table A1.3). For results using a more complete set of bins see Appendix 2.6.

likelihood that a Black buyer buys a home in this type of tract. Our model implies that at the 25th percentile of segregation, the average likelihood that any Black buyer in an MSA purchases a home in a tract that is less than 2% Black is 0.68% but only 0.27% at the 75th percentile, a 60% reduction. A similar calculation shows that segregation yields a 23% increase in purchase probability in 5-40% Black tracts.

In Panel B, we plot the share of housing units in each MSA that fall within each bin of Black population share. On average, 55% of housing units in the MSAs in our sample are located within census tracts that are less than 2% Black. Our results imply that segregation imposes a significant constraint on the likelihood of purchasing a home in areas that make up more than half of the total housing stock in any given city. In Panel C of Figure 2, we show that although segregation yields higher price premiums for African Americans everywhere, this is particularly the case in census tracts which are 5-40% Black.

In the last three columns of Table 3, we present estimates of the effect of segregation on premiums from 3 different bins of Black population share. The effect of segregation on price differentials is more than three times as large in higher Black share tracts. The premium for a black buyer in a 0-2% Black tract at the 75th percentile of segregation (Philadelphia) would be 2.2% whereas it would be 7.5% in a 5-40% Black tract.[25] In a less segregated city (Seattle, at the 25th percentile), the premium for a Black buyer in a 0-2% Black tract would be 0.8% and 2.5% in a 5-40% Black tract.[26]

These results suggest that segregation causes African American buyers to constrain searches to higher Black share neighborhoods, where they pay much higher premiums for equivalent housing due to a reduced choice set. These findings contribute two new insights to the literature: (1) they further document the harmful effects of segregation on economic inequality and (2) they suggest that policies and other actions that reduce discrimination could have important effects on African American homebuyers and patterns in homeownership.

[25] 0-2%: -0.033+0.078(0.7085)=2%, 2-100%: -0.110+0.261(0.7085)=7.5%.
[26] 0-2%: -0.033+0.078(0.5229)=0.8%, 2-100%: -0.110+0.261(0.5229)=2.5%.

Conclusion

Housing discrimination can lead individuals from minority backgrounds to pay higher prices than white buyers for equivalent housing. This paper provides the most comprehensive evidence to date on racial housing price differentials in the U.S. housing market. We assemble a dataset comprising approximately 40M repeat home sales across most of the contiguous United States between 2000-2020. Our results indicate that on average, African American and Hispanic homebuyers pay approximately 3% more for the same property as white buyers while Asian buyers experience a 0.7% discount.

We examine potential mechanisms underlying price differentials. Our results suggest that price premiums differ based on neighborhood racial and ethnic composition and the race/ethnicity of the seller. Premiums paid by Black and Hispanic buyers are higher when they purchase a home from a seller of a different race/ethnicity and in neighborhoods with higher own-group representation. Consistent with theoretical predictions, we find higher premiums in supply constrained markets, settings in which discrimination may be less costly for home sellers. Using an instrumental variables strategy that leverages historical railroad placement, we find that racial segregation drives significant increases in the premiums paid by African American buyers and limits the set of neighborhoods in which they purchase housing.

Our results have important implications for future work. First, while our study documents the existence of price differentials, it only begins to address the specific causes of these differentials. Examining additional factors that lead to higher or lower differentials in certain areas or during certain time periods requires further study. Second, the effect of these price differentials on minority homebuyers is also of interest. By constraining the choices faced by minority homebuyers, price differentials may prevent individuals from moving to better neighborhoods and cause important welfare losses for current and future generations (Chetty et al., 2011, Chetty and Hendren, 2018, Chetty et al., 2016, Chyn, 2018, Graham, 2018). In addition, premiums may act as a barrier to homeownership for Black and Hispanic individuals, constraining access to one of the largest contributors to intergenerational wealth accumulation. Finally, our results suggest that policies aimed at reducing segregation are likely to both increase access to housing and decrease premiums for African Americans.

References

Akbar, P. A., Li, S., Shertzer, A., and Walsh, R. P. (2019). Racial segregation in housing markets and the erosion of black wealth. Working Paper 25805, National Bureau of Economic Research.

Aliprantis, D., Carroll, D. R., and Young, E. R. (2022). What explains neighborhood sorting by income and race? *Journal of Urban Economics*, page 103508.

Altonji, J. and Card, D. (1991). The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives. In *Immigration, Trade, and the Labor Market*. University of Chicago Press.

Ananat, E. O. (2011). The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3(2):34–66.

Bayer, P., Casey, M., Ferreira, F., and McMillan, R. (2017). Racial and ethnic price differentials in the housing market. *Journal of Urban Economics*, 102:91–105.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago press.

Borges Ferreira Neto, A. and Whetstone, K. (2022). The effect of the raiders' relocation to las vegas on residential property values. *Journal of Housing Research*, 31(2):181–195.

Boustan, L. P. (2010). Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration. *The Quarterly Journal of Economics*, 125(1):417–443.

Boustan, L. P. (2012). Racial Residential Segregation in American Cities. In *The Oxford Handbook of Urban Economics and Planning*. Oxford University Press.

Caetano, G. and Maheshri, V. (2019). A unified empirical framework to study segregation. Technical report, Working paper.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics*, 123(1):177–218.

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

Cheng, W. and Weinberg, B. A. (2021). Marginalized and overlooked? minoritized groups and the adoption of new scientific ideas. Technical report, National Bureau of Economic Research.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility 1: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Christensen, P. and Timmins, C. (2022). Sorting or steering: The effects of housing discrimination on neighborhood choice. *Journal of Political Economy*, 130(8):2110–2163.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review*, 108(10):3028–56.

Chyn, E., Haggag, K., and Stuart, B. A. (2022). The effects of racial segregation on intergenerational mobility: Evidence from historical railroad placement.

Courant, P. N. (1978). Racial prejudice in a search model of the urban housing market. *Journal of Urban Economics*, 5(3):329–345.

Cox, R., Cunningham, J. P., Ortega, A., and Whaley, K. (2022). Black lives: The high cost of segregation. In *PAA 2022 Annual Meeting*. PAA.

Cutler, D. M. and Glaeser, E. L. (1997). Are ghettos good or bad? *The Quarterly Journal of Economics*, 112(3):827–872.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (1999). The rise and decline of the american ghetto. *Journal of Political Economy*, 107(3):455–506.

Daniels, C. B. (1975). The influence of racial segregation on housing prices. *Journal of Urban Economics*, 2(2):105–122.

Davis, M. A., Gregory, J., and Hartley, D. A. (2023). Preferences over the racial composition of neighborhoods: Estimates and implications. *Available at SSRN 4495735*.

Elliott, M. N., Morrison, P. A., Fremont, A., McCaffrey, D. F., Pantoja, P., and Lurie, N. (2009). Using the census bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology*, 9:69–83.

Fryer Jr, R. G. and Levitt, S. D. (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 119(3):767–805.

Graham, B. S. (2018). Identifying and estimating neighborhood effects. *Journal of Economic Literature*, 56(2):450–500.

Haugen, R. A. and Heins, A. J. (1969). A market separation theory of rent differentials in metropolitan areas. *The Quarterly Journal of Economics*, 83(4):660–672.

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

Imai, K. and Khanna, K. (2016). Improving ecological inference by predicting individual ethnicity from voter registration records. *Political Analysis*, 24(2):263–272.

Kain, J. F. and Quigley, J. M. (1975). Front matter, housing markets and racial discrimination: A microeconomic analysis. In *Housing Markets and Racial Discrimination: A Microeconomic Analysis*, pages 22–0. NBER.

Kiel, K. A. and Zabel, J. E. (1996). House price differentials in us cities: Household and neighborhood racial effects. *Journal of Housing Economics*, 5(2):143–165.

King, A. T. and Mieszkowski, P. (1973). Racial discrimination, segregation, and the price of housing. *Journal of Political Economy*, 81(3):590–606.

Marschke, G., Nunez, A., Weinberg, B. A., and Yu, H. (2018). Last place? the intersection of ethnicity, gender, and race in biomedical authorship. In *AEA papers and proceedings*, volume 108, pages 222–27.

Masson, R. T. (1973). Costs of search and racial price discrimination. *Economic Inquiry*, 11(2):167–86.

Millard-Ball, A., Desai, G., and Fahrney, J. (2021). Diversifying planning education through course readings. *Journal of Planning Education and Research*, page 0739456X211001936.

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

Nowak, A. and Smith, P. S. (2018). Reexamining racial price differentials in housing markets. *Available at SSRN 3258811*.

Shertzer, A. and Walsh, R. P. (2019). Racial Sorting and the Emergence of Segregation in American Cities. *Review of Economics and Statistics*, 101(3):415–427.

Sood, G. and Laohaprapanon, S. (2018). Predicting race and ethnicity from the sequence of characters in a name. *arXiv preprint arXiv:1805.02109*.

Turner, M. A., Santos, R., Levy, D. K., Wissoker, D., Aranda, C., and Pitingolo, R. (2013). Housing Discrimination Against Racial and Ethnic Minorities 2012: Final Report.

Yinger, J. (1997). Cash in your face: The cost of racial and ethnic discrimination in housing. *Journal of Urban Economics*, 42(3):339–365.

Yun, L., Lautz, J., Evangelou, N., Snowden, B., and Dunn, M. (2021). Snapshot of race and home buying in america. Technical report, National Association of Realtors.

Zillow (2020). Zillow's assessor and real estate database (ztrax).

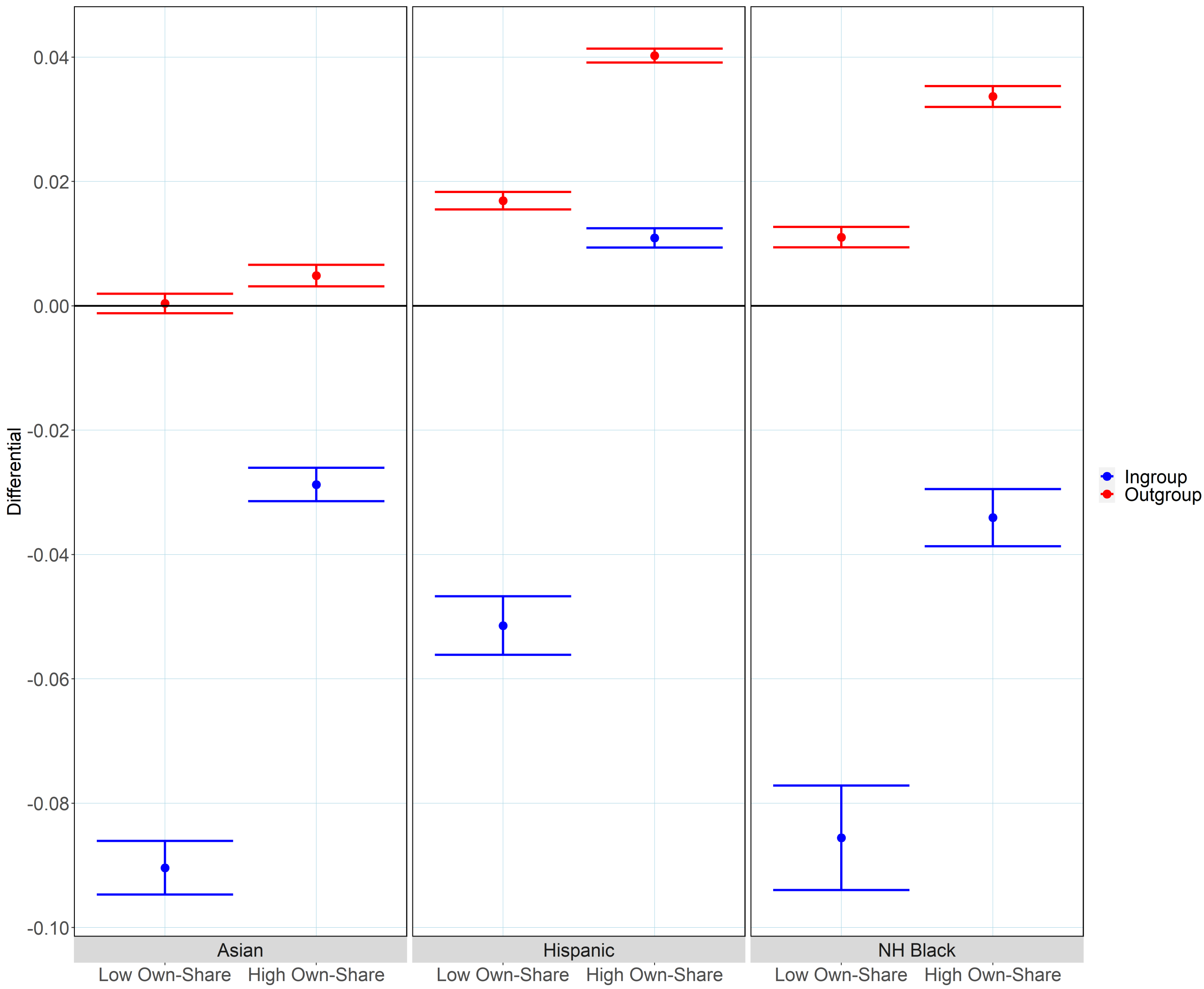
Exhibits

Table 1. Baseline Results: Price Differentials Across the U.S.

	Dependent variable: ln(sales price)			
	Full Sample		*Remodeled Sample*	
	(1)	(2)	(3)	(4)
Asian	-0.009^{***}	-0.007^{***}	-0.007^{***}	-0.008^{***}
	(0.0004)	(0.0003)	(0.001)	(0.001)
Hispanic	0.047^{***}	0.030^{***}	0.031^{***}	0.032^{***}
	(0.0003)	(0.0003)	(0.001)	(0.001)
NH Black	0.035^{***}	0.033^{***}	0.034^{***}	0.034^{***}
	(0.0004)	(0.0004)	(0.001)	(0.001)
Remodeled				0.261^{***}
				(0.001)
Comparison Mean ($)	248,555	248,555	241,066	241,066
Property FE	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes
Tract x Year FE	No	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes
Properties	16,456,528	16,456,528	1,746,192	1,746,192
Observations	39,470,293	39,470,293	3,949,810	3,949,810

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects and calendar month fixed effects. Column 1 presents OLS estimates without tract x year FEs but with year FEs, column 2 adds tract by year fixed effects (our main specification). Column 3 uses the same specification as column 2 but restricts the sample to remodeled properties. In column 4 we add a remodeled control which is equal to one after a remodel occurs. *p<0.1; **p<0.05; ***p<0.01

Figure 1. Ingroup and Outgroup Premiums by Neighborhood Group Share



Notes: Figure shows estimates of housing price differentials and 95% confidence intervals for Asian, Hispanic and NH Black buyers by neighborhood race/ethnicity share and ingroup/outgroup transaction status. Low (high) own-share neighborhoods are defined as census tracts in which less (more) than 5% of the population is of any given race/ethnicity as of the 2000 census. Ingroup transactions are those for which the seller and buyer are of the same group, outgroup transactions are those transactions where the seller and buyer are of different groups. Point estimates and 95% confidence intervals are obtained using the delta method and coefficients from the fully-interacted model in column 4 of Table A2.2.

Table 2. Racial Price Differentials in Competitive Markets

	Dependent variable: ln(sales price)			
	Inventory		*Hotness*	
	(1)	(2)	(3)	(4)
Asian	−0.016***	−0.016***	−0.014***	−0.013**
	(0.002)	(0.003)	(0.004)	(0.005)
Hispanic	0.031***	0.032***	0.038***	0.043***
	(0.002)	(0.002)	(0.003)	(0.004)
NH Black	0.047***	0.051***	0.049***	0.052***
	(0.003)	(0.003)	(0.004)	(0.005)
High		0.007**		0.006
		(0.003)		(0.004)
Low		0.0001		−0.008*
		(0.003)		(0.004)
NH Black*High		−0.023***		0.016
		(0.006)		(0.011)
NH Black*Low		0.013*		−0.021**
		(0.007)		(0.010)
Hispanic*High		−0.015***		0.011
		(0.004)		(0.007)
Hispanic*Low		0.018***		−0.036***
		(0.005)		(0.007)
Asian*High		−0.001		0.015
		(0.006)		(0.010)
Asian*Low		0.003		−0.015*
		(0.006)		(0.008)
Comparison Mean ($)	262,752	262,752	259,990	259,990
Property FE	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes
Properties	675,387	675,387	291,003	291,003
Observations	1,348,606	1,348,606	586,184	586,184

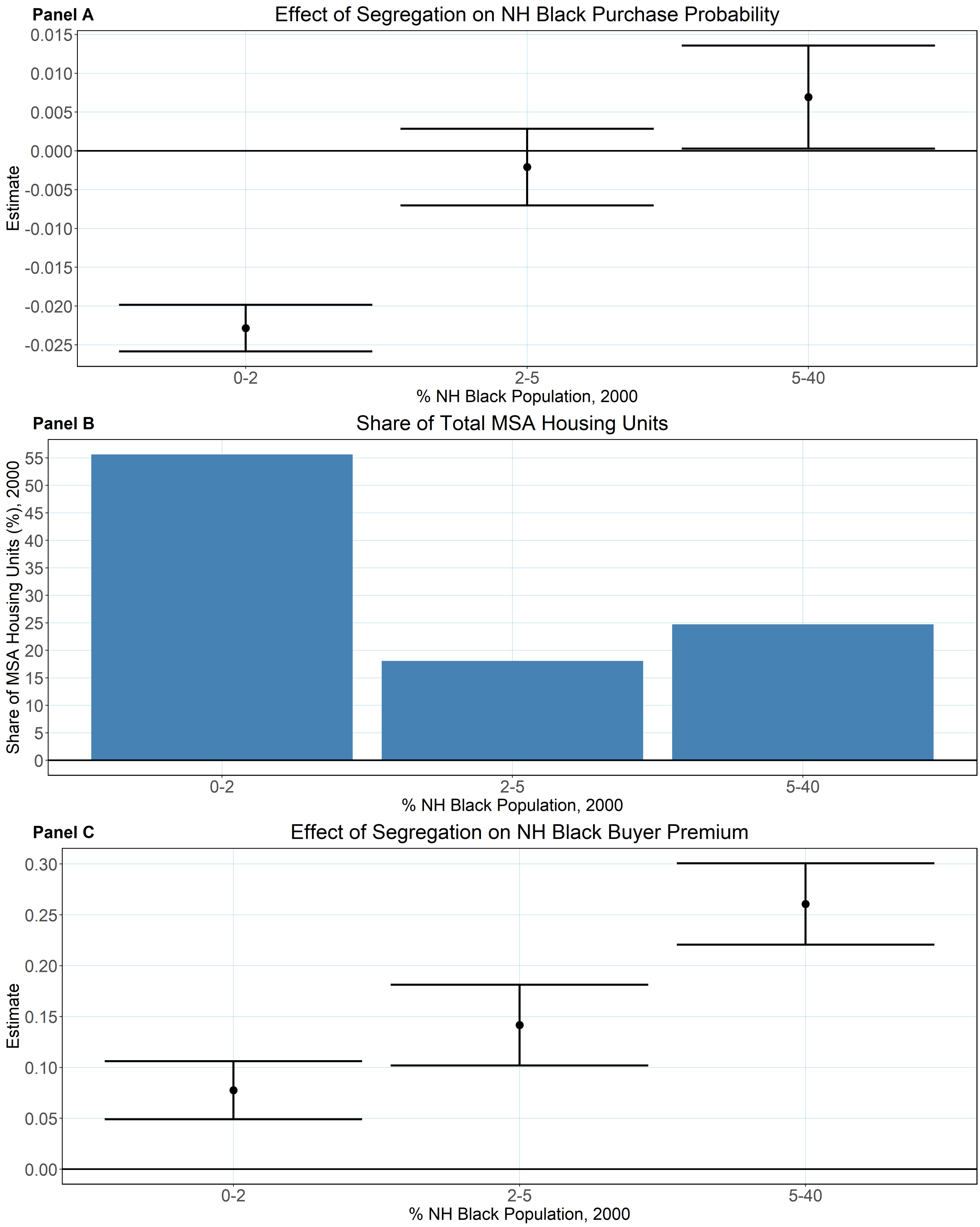
Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. Columns 1 and 3 present baseline estimates using our main specification for the limited samples of properties that sold twice between July 2016 and December 2019 (Column 1) or August 2017 and December 2019 (Column 2), for which Realtor.com data is available. In column 2, we add thick/thin market dummies and race*thick/thin dummy interactions using Realtor.com's measure of listings by ZIP code. In column 4, we add hot/cold market dummies and race*hot/cold dummy interactions using Realtor.com's measure of market hotness. *p<0.1; **p<0.05; ***p<0.01

Table 3. Segregation and Price Differentials

	Dependent variable: ln(sales price)					
			IV			
	Baseline	OLS	Full Sample	0-2% Black Share	2-5% Black Share	5-40% Black Share
Asian	−0.005***	−0.005***	−0.005***	0.003***	−0.004***	−0.013***
	(0.0005)	(0.0005)	(0.0005)	(0.001)	(0.001)	(0.001)
Hispanic	0.034***	0.034***	0.034***	0.025***	0.031***	0.044***
	(0.0004)	(0.0004)	(0.0004)	(0.001)	(0.001)	(0.001)
NH Black	0.031***	−0.020***	−0.061***	−0.033***	−0.060***	−0.110***
	(0.001)	(0.003)	(0.005)	(0.008)	(0.012)	(0.013)
NH Black*Seg		0.085***	0.154***	0.078***	0.142***	0.261***
		(0.005)	(0.009)	(0.015)	(0.020)	(0.020)
Comparison Mean ($)	266,652	266,652	266,652	277,638	261,399	217,100
Comparison Mean (Dissimilarity)	NA	0.59	0.59	0.58	0.58	0.60
Property FE	Yes	Yes	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Census Tract x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
1st Stage F-Stat	NA	NA	303.3	278.9	28.4	49.0
Properties	6,488,019	6,488,019	6,488,019	3,213,856	1,517,664	1,495,846
Observations	14,471,577	14,471,577	14,471,577	7,097,371	3,488,592	3,379,159

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects for the sample of transactions in cities for which the RDI is available. In column 1, we present baseline estimates of the race differentials with the same specification as in column 2 of Table 1 but for this more limited sample. In the second column, we interact the NH Black indicator with the dissimilarity index. In the third column, we instrument for the dissimilarity index with the RDI. In the last three columns we split the sample based on the Black population share in a tract. *p<0.1; **p<0.05; ***p<0.01

Figure 2. Segregation as a Constraint



Notes: Panel A shows point estimates and 95% confidence intervals from a regression of the probability that a Black buyer in an MSA purchases in any given tract on dissimilarity (instrumented by the RDI), by bin of % Black population. Panel B shows the average share of total housing units in an MSA in each % Black population bin. Panel C shows the effect of segregation on price differentials for Black buyers as estimated in table 3 with the sample split into bin of % Black population.

Appendix: Racial Housing Price Differentials in the U.S. Housing Market

1 Data and Methods

1.1 Data

1.1.1 ZTRAX housing data

Zillow's ZTRAX database contains more than 400 million public property transaction records across the entire United States, including sales, deed transfers, mortgages, foreclosures, auctions and property tax delinquencies (Zillow, 2020). There are over one-hundred variables available describing each transaction, such as: the transaction price, date of sale, year the home was built, location of the home, property characteristics (e.g. number of bedrooms, bathrooms, etc.), buyer name and seller name. In addition, each property is assigned a unique identifier which allows us to identify the transaction history of a particular home over time.

The ZTRAX database is available for most states starting in the mid 1990s up to 2021, however samples tend to be small pre-2000. Due to reporting lags, samples are also smaller post-2020. Therefore, we restrict our sample to the period 2000-2020. Fourteen states are "non-disclosure" states,[27] meaning home sale prices are not public record in these states. Although transactions are observed in each of these states, they are excluded from our sample due to the absence of reliable information on sales prices. All of these states are relatively small real-estate markets except for Texas, which is the second largest market in the United States. Nevertheless, excluding these states still leaves us with data for over 80% of the population of the United States.

We further refine the ZTRAX data in the following manner:

1. Transaction types: There are many types of transactions in the data, but many of them do not represent a home sale, we keep only those observations that are coded as deed transfers in the data (DataClassStndCode=="D" or "H").

2. Duplicate transactions: We remove any duplicate transactions in the data, as identified by duplicate transaction IDs.

3. Intra-family transfers: We drop observations flagged as "Intra-family transfers" by Zillow.

4. Repeat-sales: Since our identification strategy hinges on observing multiple transactions for the same property, we drop observations where a property is only sold once.

[27] These states are: Alaska, Idaho, Kansas, Louisiana, Maine, Mississippi, Missouri, Montana, New Mexico, North Dakota, South Dakota, Texas, Utah and Wyoming.

5. Sales price and outliers: We further restrict the sample to include only transactions with an observed sales price greater than 0. For the remaining transactions, we remove properties with a median sales price above the 99th percentile and below the 1st percentile, in line with previous work (Bayer et al., 2017). For state and MSA-level estimates, we similarly restrict the sample but with percentiles defined at the state and MSA level. Similarly, we also drop observations above the 99th percentile and below the 1st percentile of the ratio of the observed sales price to the median sales price for any given property. This alows us to discard observations where a home has experienced an exceptional increase or decrease in value. Again, for state and MSA-level estimates, we restrict the sample with percentiles defined at the state and MSA levels.

Table A1.1 presents descriptive statistics of the raw repeat-sales data without removing outliers and observations for which the buyer name variable is not populated. Our dataset contains over 58M home sales for approximately 21M homes having been sold at least twice between the years 2000-2020. For 21% of these sales, no buyer name is provided which means our effective sample is 46M observations before removing any outliers. Of the observations for which we have a buyer name, 79% are associated to a non-hispanic white buyer, 12% to a Hispanic buyer, 5% to an Asian buyer and 4% to a non-hispanic Black buyer. These percentages are in line with figures from the National Association of Realtors (NAR) which put these respective shares at 81%, 11%, 6% and 7% in 2020 (Yun et al., 2021).[28]

Home prices vary across races/ethnicities with Asian buyers paying the most on average and African Americans paying the least. The gap between prices paid by Asian buyers and buyers from other race/ethnicities is large but may be explained by the fact that Asian homebuyers buy houses with more bedrooms and bathrooms as well as newer homes. Non-hispanic White, Hispanic and African American buyers buy houses with broadly similar characteristics, although both Hispanic and African American buyers tend to purchase older homes. Note that these discrepancies in housing characteristics are all controlled for by our repeat-sales design.

The last two rows in table A1.1 present the mean and median level of certainty of the race/ethnicity predictions produced by the Ethnicolr algorithm. While Ethnicolr produces very confident estimates for most race/ethnicities, it is much less confident about its predictions for African Americans.

Table A1.2 presents descriptive statistics for the entire sample of ZTRAX home sales for the years 2000-2020, including homes that sold only once. The characteristics of the properties sold when including homes sold only once are very similar to those for repeat-sales only. In addition, the distribution of sales by race is almost identical.

In A1.1 and table A1.4, we show price distributions by race and year. Asian buyers consistently buy costlier homes. In addition, we can see that homes were at their most expensive in 2006 and after 2018 regardless of race.

[28]The NAR data allows for respondents to select multiple races meaning percentages add up to more than 100%.

Table A1.1. Descriptive Statistics: Zillow ZTRAX Repeat-Sales Dataset

	All	Non-Hispanic White	Asian	Hispanic	Non-Hispanic Black	NA
# of Observations	58,067,183	36,497,197	2,449,868	5,342,062	1,776,804	12,001,252
# of Properties	20,988,949	18,499,507	2,040,709	4,009,592	1,625,422	8,816,255
Fraction of Total Obs.	1	0.63	0.04	0.09	0.03	0.21
Fraction of Identified Obs.	1	0.79	0.05	0.12	0.04	*NA*
Mean Sales Price ($)	257,617	265,069	382,060	246,989	220,524	219,773
Median Sales Price ($)	178,900	190,000	288,000	197,000	164,000	119,900
Mean # of Bedrooms	2.88	2.87	2.94	2.90	2.87	2.88
Median # of Bedrooms	3	3	3	3	3	3
Mean # of Full Bathrooms	1.83	1.83	1.98	1.81	1.79	1.81
Median # Full Bathrooms	2	2	2	2	2	2
Mean House Age	31.65	30.94	27.61	32.63	32.31	34.10
Median House Age	23	22	19	27	24	27
Mean Ethnicolr Certainty	0.86	0.87	0.83	0.81	0.69	*NA*
Median Ethnicolr Certainty	0.90	0.92	0.88	0.87	0.66	*NA*

Notes: This table presents descriptive statistics for the repeat-sales dataset used in the analysis, before removing outliers. Column 1 presents statistics for the entire dataset, columns 2-5 present statistics for observations where a buyer name is listed and hence a race/ethnicity is predicted, column 6 shows statistics for observations without a buyer name. The number of properties per race/ethnicity refers to the number of unique properties with at least one buyer of a given race/ethnicity. For example, there are 18.5M properties that were sold to a white individual at least once (i.e. there are only 2.5M properties which were never sold to a white individual).

Table A1.2. Descriptive Statistics: Zillow ZTRAX Full Sample

	All	Non-Hispanic White	Asian	Hispanic	Non-Hispanic Black	NA
# of Observations	78,928,546	51,825,982	3,530,952	7,171,693	2,513,054	13,886,865
# of Properties	41,850,312	33,828,292	3,121,793	5,839,223	2,361,672	10,701,868
Fraction of Total Obs.	1	0.66	0.04	0.09	0.03	0.18
Fraction of Identified Obs.	1	0.80	0.05	0.11	0.04	*NA*
Mean Sales Price ($)	266,885	268,697	410,140	249,945	224,099	240,191
Median Sales Price ($)	182,500	190,000	310,000	199,000	165,000	123,382
Mean # of Bedrooms	2.87	2.86	2.95	2.90	2.88	2.86
Median # of Bedrooms	3	3	3	3	3	3
Mean # of Full Bathrooms	1.81	1.81	1.99	1.79	1.78	1.80
Median # Full Bathrooms	2	2	2	2	2	2
Mean House Age	32.29	31.79	27.39	33.21	32.21	34.92
Median House Age	25	24	19	28	25	28
Mean Ethnicolr Certainty	0.86	0.87	0.83	0.81	0.69	*NA*
Median Ethnicolr Certainty	0.91	0.92	0.89	0.87	0.66	*NA*

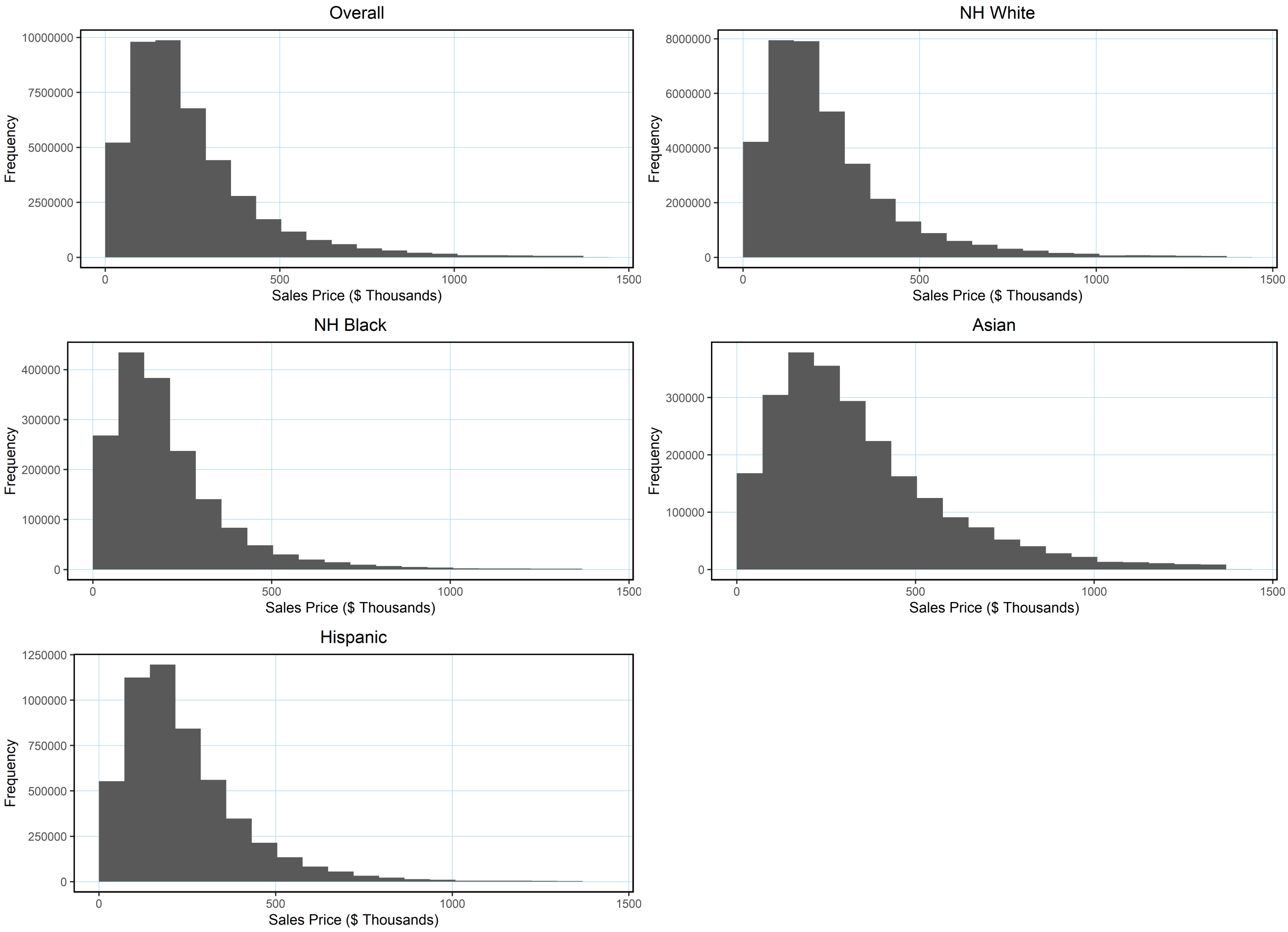
Notes: This table presents descriptive statistics for all transactions in the Zillow Ztrax dataset between 2000-2020, before removing outliers. Column 1 presents statistics for the entire dataset, columns 2-5 present statistics for observations where a buyer name is listed and hence a race/ethnicity is predicted, column 6 shows statistics for observations without a buyer name. The number of properties per race/ethnicity refers to the number of unique properties with at least one buyer of a given race/ethnicity.

Table A1.3. Descriptive Statistics by Tract Black Share

	0-0.5%	0.5-1%	1-2%	2-3%	3-5%	5-7%	7-10%	10-20%	20-40%	40-100%
Home Price: 25%ile	128,000	147,600	152,000	139,000	136,000	125,000	117,500	108,000	84,500	37,000
Home Price: Median	200,000	230,000	238,500	220,000	218,000	195,243	195,000	175,000	160,000	89,500
Home Price: 75%ile	321,000	370,000	380,000	354,000	348,000	310,000	319,950	285,000	280,000	189,900
% Asian	3	5	7	8	9	8	10	9	9	6
% Black	2	2	2	2	3	3	3	4	6	14
% Hispanic	6	7	10	12	16	20	18	24	22	13
% White	89	85	80	77	73	69	69	63	63	67
Transactions	2,032,804	2,313,517	2,959,864	1,703,011	1,839,447	966,952	732,693	1,169,082	568,044	474,404

Notes: This table presents descriptive statistics of the ZTRAX transactions data in the bins used for the segregation analysis in exhibit 4. *p<0.1; **p<0.05; ***p<0.01

Figure A1.1. Home Sales Price Distributions by Race/Ethnicity



Notes: This figure presents histograms of the sales price for each race/ethnicity in our sample.

Table A1.4. Median Home Sales Prices ($) by Year and Race/Ethnicity

Year	NH White	Asian	Hispanic	NH Black	Overall
2000	140,000	210,000	125,000	112,500	131,635
2001	149,000	217,000	139,000	122,000	140,000
2002	160,000	240,000	155,000	133,000	152,000
2003	175,000	269,500	180,000	146,500	167,250
2004	197,810	304,900	225,000	165,000	190,000
2005	210,676	330,000	270,000	181,770	210,000
2006	215,900	337,200	280,000	189,900	217,000
2007	210,000	347,000	249,900	178,000	206,500
2008	187,000	288,500	190,000	155,000	180,000
2009	162,000	235,000	141,400	136,000	152,000
2010	162,000	232,000	140,000	135,000	149,591
2011	156,500	205,000	131,000	128,500	142,000
2012	165,000	225,000	140,000	128,500	145,388
2013	185,500	275,000	160,840	152,000	165,000
2014	195,000	295,000	173,000	160,000	173,000
2015	209,900	315,000	190,000	174,800	188,600
2016	218,000	327,000	205,000	185,000	200,000
2017	228,000	349,900	219,000	197,500	215,000
2018	237,500	355,000	229,000	205,000	225,000
2019	250,000	365,000	245,000	219,858	240,000

Notes: This table reports median housing prices in our sample for each year and each race/ethnicity.

1.1.2 Race and Ethnicity Prediction

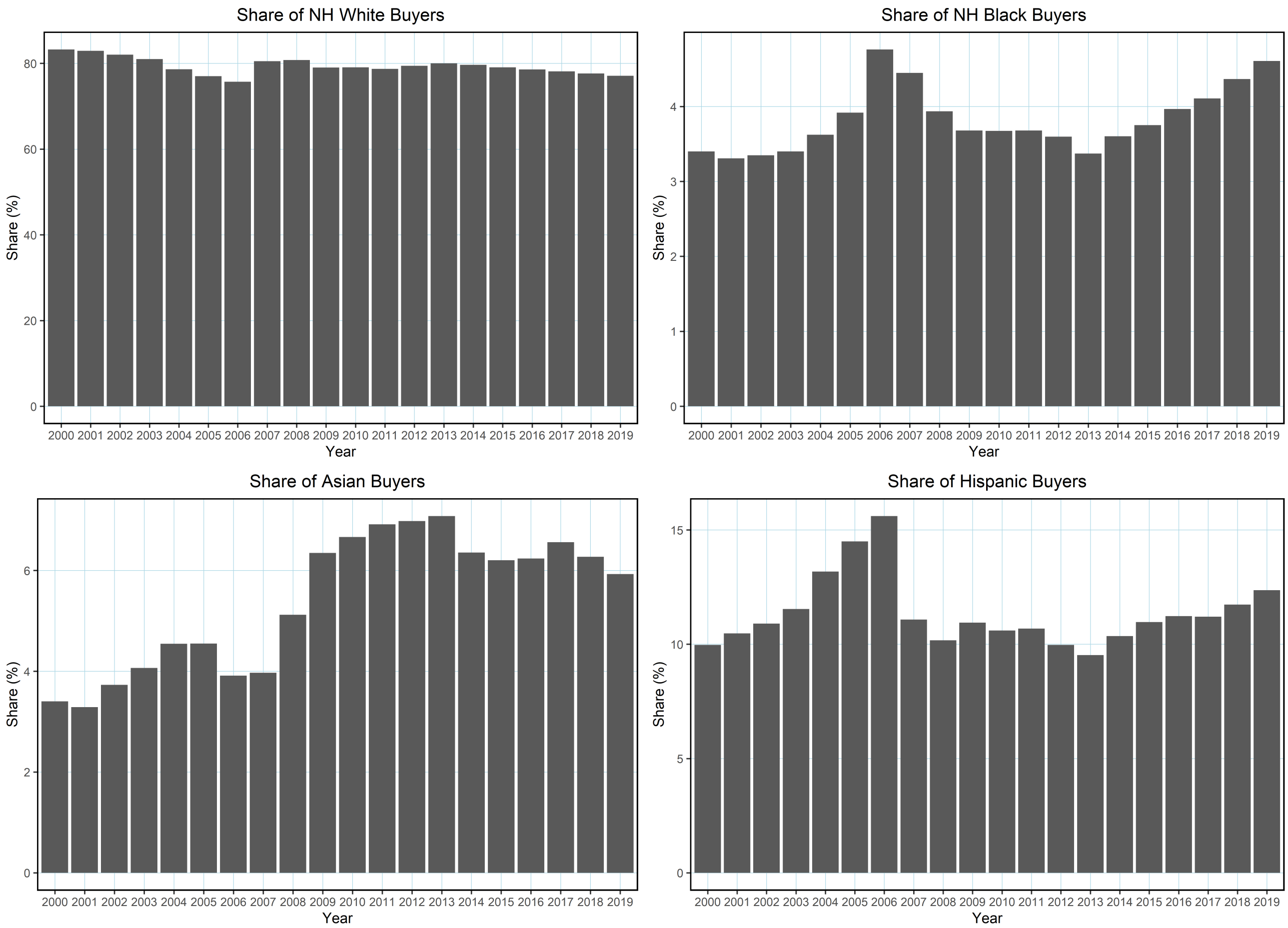
Although the Zillow ZTRAX data contains the names of both property buyers and sellers, it does not contain any information on the race or ethnicity of homebuyers and sellers. We must therefore impute race and ethnicity information based on the names in the Zillow data. To do so, we use Ethnicolr, a Python package that predicts race and ethnicity (Sood and Laohaprapanon, 2018). The main advantage of the Ethnicolr algorithm is that it takes into account both first and last names. Other algorithms tend to use only last names data to predict race or ethnicity. This is particularly problematic in the United States where whites and African Americans often have the similar last names but different first names. Ethnicolr is trained on Florida Voter Registration data from 2017, which contains the first and last names of 13M voters in Florida, one of the more diverse states in the US. Ethnicolr uses a Long Short Term Memory (LSTM) model to model the relationship between the sequence of characters in a name and race and ethnicity (Sood and Laohaprapanon, 2018). This model assigns probabilities that a given name belongs to four race/ethnicites: white, non-Hispanic Black, Asian and Hispanic. The race/ethnicity with the highest probability is the one that is assigned as being the race of individual. This package has been used before in published research (Cheng and Weinberg, 2021, Marschke et al., 2018, Millard-Ball et al., 2021). Since names in Zillow include middle names, we split the names into first and last names and ignore middle names.

Figure A1.2 presents distributions of race and ethnicity by year. Non-Hispanic white buyers represent around 80% of buyers. The share of white buyers decreases between 2000-2006 before rebounding during the financial crisis of 2007-2008 and slowly declines between 2008-2020. Both the share of non-Hispanic Black and the share of Hispanic buyers follow the opposite trajectory, with shares peaking right before the financial crisis and again towards 2019. Asian buyers see a large jump in their share of housing purchases after 2008. We employ various robustness checks to insure that these differential patterns in purchase timing are not driving our price differential estimates (see Table A1.10 column 3).

In tables A1.5 and A1.6, we present the total number of transactions by race/ethnicity and year as well as the number of transactions between different combinations of race/ethnicity. Transactions between white buyers represent the overwhelming majority of transactions in our sample. Nevertheless, transactions involving other races as buyers still represent over 5M observations.

In figure A1.3, we present histograms of the certainty level of Ethnicolr race predictions. For all races/ethnicities except the non-hispanic Black category, the mean and median certainty is above 90%. For names Ethnicolr predicts as belonging to African Americans, Ethnicolr is much less certain about its prediction. See section A1.3.1 for robustness checks limiting the sample to names for which there is a high degree of certainty and an alternate algorithm.

Figure A1.2. Percentage of Buyers of each Race/Ethnicity by Year



Notes: This figure presents the percentage of buyers of each race or ethnicity by year in our sample. Transactions with missing buyer names are not included in the calculation of these percentages.

Table A1.5. Transactions by Race/Ethnicity and Year

Year	NH White	Asian	Hispanic	NH Black	Total
2000	1,649,583	67,426	197,508	67,404	1,981,921
2001	1,708,143	67,746	215,680	68,106	2,059,675
2002	1,833,872	83,402	243,840	74,883	2,235,997
2003	2,030,608	101,949	289,421	85,260	2,507,238
2004	2,277,307	131,623	381,675	104,876	2,895,481
2005	2,742,498	162,002	516,218	139,484	3,560,202
2006	2,321,727	119,941	478,545	145,919	3,066,132
2007	1,871,248	92,314	257,478	103,380	2,324,420
2008	1,546,702	98,052	194,755	75,343	1,914,852
2009	1,558,765	125,176	215,815	72,564	1,972,320
2010	1,476,240	124,456	197,854	68,594	1,867,144
2011	1,472,767	129,370	199,815	68,808	1,870,760
2012	1,549,113	136,053	194,295	70,119	1,949,580
2013	1,711,354	151,402	203,667	72,119	2,138,542
2014	1,636,532	130,540	212,731	73,967	2,053,770
2015	1,740,177	136,518	241,403	82,529	2,200,627
2016	1,813,282	143,953	259,142	91,526	2,307,903
2017	1,888,411	158,567	270,926	99,257	2,417,161
2018	1,860,185	150,307	281,221	104,581	2,396,294
2019	1,808,683	139,071	290,073	108,085	2,345,912

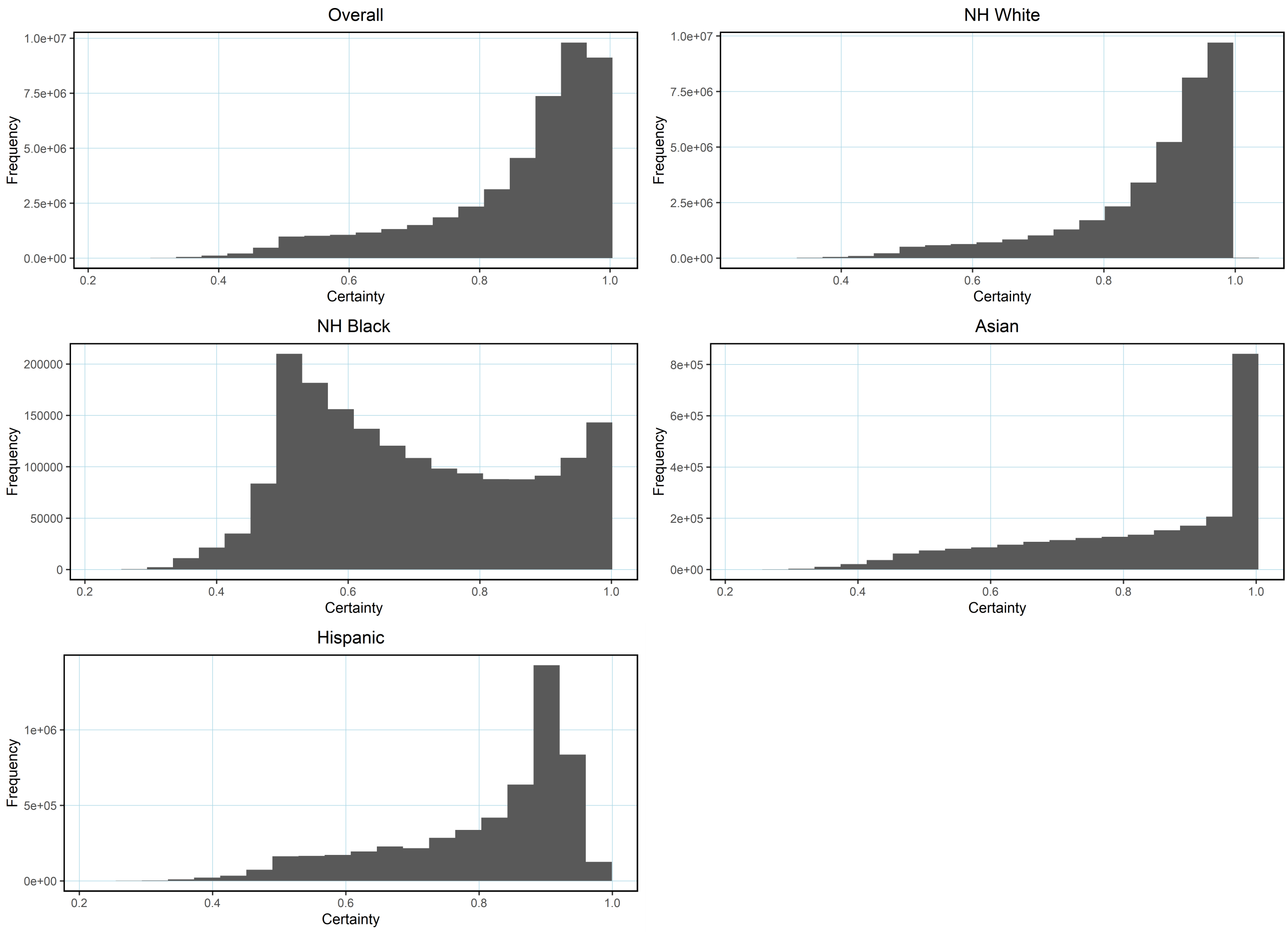
Notes: This table presents counts of the number of transactions for buyers of each race/ethnicity in each year of our sample.

Table A1.6. Transactions by Seller and Buyer Race/Ethnicity Combination

Seller->Buyer	Observations	Percent
NH White ->NH White	16,303,420	73.34
Different ->NH White	1,752,525	7.88
Different ->Hispanic	1,651,629	7.43
Different ->Asian	853,638	3.84
Hispanic ->Hispanic	724,141	3.26
Different ->NH Black	650,890	2.93
Asian ->Asian	240,993	1.08
NH Black ->NH Black	51,533	0.23

Notes: This table presents counts and percentages of the number of transactions for each seller and buyer race/ethnicity combination in our sample.

Figure A1.3. Histogram of Ethnicolr Estimated Certainty of Race/Ethnicity Predictions



Notes: This figure presents histograms of the estimated certainty of the race/ethnicity prediction made by the Ethnicolr algorithm for each race/ethnicity in our sample.

1.2 Methods

1.2.1 Repeat-Sales and time-varying indicators

Our main set of results estimates equation 1 over our entire sample and the subset of all housing transactions in each state and MSA. To estimate price differentials over time, we introduce a specification in which we interact two-year indicator variables with our race indicators as shown in equation 7.

$$\begin{aligned} ln(p_{ijt}) = \beta_0 + \beta_1 Black_{it} + \beta_2 Asian_{it} + \beta_3 Hispanic_{it} + \\ \sum_{t=2002-2003}^{t=2018-2019} \beta_0^t white_{it} + \sum_{t=2002-2003}^{t=2018-2019} \beta_1^t Black_{it} + \sum_{t=2002-2003}^{t=2018-2020} \beta_2^t Asian_{it} + \\ \sum_{t=2002-2003}^{t=2018-2019} \beta_3^t Hispanic_{it} + \theta age_{jnt} + \mu_j + m_t + \epsilon_{ijt} \end{aligned} \tag{7}$$

With the above specification, we are able to estimate the price differential in each two-year window and compare the differentials over time. The omitted category here is white buyers in 2000-2001, thus all estimates are relative to transactions for white homebuyers in 2000 and 2001. Price differentials in 2000-2001 are simply the intercepts for each race/ethnicity. For all the other two-year periods, the differential is equal to the intercept for each race plus the coefficient on the interaction between each race and a given two-year period minus the coefficient for whites in that same time period. For example, the price differential for Black buyers in 2006-2008 would be equal to $\beta_1 + \beta_1^{t=2006-2008} - \beta_0^{t=2006-2008}$. Figure A2.1, plots these estimates, while figures A2.2-A2.4 show the evolution of housing prices for whites and each other race/ethnicity compared to the omitted category. The difference between the two lines in these figures is what is plotted in figure A2.1. Averaging the estimates for each two year period from this specification also allows us to control for the effect of the changing composition of homebuyers over time (see Column 3 of A1.10).

1.3 Robustness

1.3.1 Race/ethnicity imputation

Proper attribution of names to races/ethnicities is crucial to our study design. In the presence of a premium for African American and Hispanic buyers, improperly attributing a name from either of these groups to a white buyer means our estimates would under-estimate the premium paid by members of these groups. Improperly attributing a white name to a minority group would likewise cause us to under-estimate the premium paid by members of historically disadvantaged groups. Similarly, for Asian buyers who pay a discounted price, estimates would be attenuated by an inaccurate attribution of names. The Ethnicolr algorithm assigns probabilities that any given first/last name combination belongs to one of four races/ethnicities: non-Hispanic white, non-Hispanic Black, Asian and Hispanic. In our main estimates, we assign the race/ethnicity of the individual as the one with the highest Ethnicolr predicted probability.

We verify the degree to which our estimates are sensitive to Ethnicolr predictions in a number of ways. First, in the second and third columns of table A1.7, we replace the race/ethnicity indicators by the race/ethnicity probabilities assigned by Ethnicolr. In column 2, estimates represent the premium paid for a buyer that is predicted to be of race/ethnicity x with a 100% probability. In column 3, estimates represent the premium paid for a buyer that has the mean probability of being of race/ethnicity x. In both cases, estimates are larger than when simply imputing race as a binary variable. In the last two columns of table A1.7, we test the robustness of our estimates to the inclusion of observations for which Ethnicolr is relatively uncertain about its prediction. To do so, we keep only observations which are above the 50th and 75th percentiles of certainty for each race. Estimates of the premiums paid by Black and Hispanic buyers tend to increase significantly when only observations for which Ethnicolr is relatively certain of its name imputation are included in the sample. This implies that our results are a conservative estimate of the premium paid by African Americans and Hispanics and that those with names that are more easily identifiable as a member of one of these minorities may be even more disadvantaged.

Another approach for verifying the robustness of our results to the uncertainty in race predictions is to bootstrap our standard errors. However, this procedure is very computationally heavy and would take too long to implement for our entire sample. Therefore, we randomly select 500,000 properties and obtain standard errors from our main estimation and by bootstrap. These results are presented in table A1.8. The signs and magnitudes of the coefficients are similar to those with the full sample. Standard errors are larger with the smaller sample but comparing columns 2 and 3, we can see that standard errors are almost identical whether we bootstrap them not.

While we believe the Ethnicolr algorithm is best suited to our analysis due to its ability to take first names into account, there are other name attribution algorithms that have been used in the literature. Most notably, the WRU algorithm (Imai and Khanna, 2016) that attributes race/ethincity based on last name and geographical location using census data. While 91% of our observations have the same race/ethnicity attribution using WRU and Ethnicolr, WRU identifies 40% less non-Hispanic Black individuals. This is likely due to the fact that African American and White Americans often have similar last names (Elliott et al., 2009). In table A1.9 we can see that while estimates for Asian and Hispanic buyers are very similar no matter the algorithm, the estimate for Non-Hispanic Black buyers is 1 percentage point lower. This is likely due to the fact that WRU mistakes some Black buyers for White buyers and vice versa. Indeed, if we omit from the sample those buyers that are deemed to be Black in the Ethnicolr sample but white in WRU or white in the Ethnicolr sample but Black in the WRU sample (third column), the coefficient on Black buyers is larger than in our baseline estimates. This makes sense given our earlier results indicating that names for which we are more certain of the race see higher premiums.

Table A1.7. Robustness to Ethnicolr Predictions

	Dependent variable: ln(sales price)				
	Main	Continuous Race (100%)	Continuous Race (Mean)	> Median	> 75th Pctile
Asian	−0.007***	−0.006***	−0.005***	−0.002***	0.003
	(0.0003)	(0.0004)	(0.0003)	(0.001)	(0.002)
Hispanic	0.030***	0.046***	0.037***	0.046***	0.055***
	(0.0003)	(0.0003)	(0.0003)	(0.001)	(0.001)
Non-Hispanic Black	0.033***	0.056***	0.039***	0.053***	0.065***
	(0.0004)	(0.0005)	(0.0003)	(0.001)	(0.002)
Comparison Mean ($)	248,555	248,555	248,555	247,342	247,645
Mean Ethnicolr Certainty: Asian	0.83	0.83	0.83	0.97	0.99
Mean Ethnicolr Certainty: Hispanic	0.81	0.81	0.81	0.91	0.93
Mean Ethnicolr Certainty: NH Black	0.69	0.69	0.69	0.83	0.93
Mean Ethnicolr Certainty: NH White	0.87	0.87	0.87	0.96	0.98
Property FE	Yes	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes	Yes
Properties	16,456,528	16,456,528	16,456,528	6,079,118	1,954,874
Observations	39,470,293	39,470,293	39,470,293	13,084,856	3,990,435

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. In the first column, we present estimates of the baseline model, in which we include all transactions for which we have a name. In the second and third columns, we replace race/ethnicity dummies with the Ethnicolr estimated probability each buyer is of a given race/ethincity and show coefficients assuming 100% ethnicolr certainty and the mean level of certainty for each race/ethnicity. In the fourth column, we restrict the sample to observations with race/ethnicity predictions that fall above the 50th percentile of "certainty" for each race. In the fifth column, we restrict the sample to observations with race/ethnicity predictions that fall above the 75th percentile of "certainty" for each race. Note that the samples in the last two columns are much smaller due to the repeat-sales design requiring that we drop any property where dropping a "less certain" prediction results in only one observation for that property. *p<0.1; **p<0.05; ***p<0.01

Table A1.8. Bootstrap

	Dependent variable: ln(sales price)		
	Main	Subsample	Subsample: Bootstrap
Asian	−0.007***	−0.003	−0.003
	(0.0003)	(0.003)	(0.005)
Hispanic	0.030***	0.028***	0.028***
	(0.0003)	(0.002)	(0.003)
Non-Hispanic Black	0.033***	0.017***	0.017***
	(0.0004)	(0.003)	(0.005)
Comparison Mean ($)	248,555	248,773	248,773
Property FE	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes
Properties	16,456,528	495,294	494,498
Observations	39,470,293	1,388,869	1,388,869

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects and calendar month fixed effects. Column 1 presents our main estimates for the entire sample. In column 2, we present the same estimates for a subsample of approximately 500,000 randomly selected properties. Column 3 uses this same subsample but estimates standard errors by bootstrap. *p<0.1; **p<0.05; ***p<0.01

Table A1.9. Ethnicolr vs. WRU

	Dependent variable: ln(sales price)		
	Main	WRU	WRU: Omitting Different Black/White Imputation
Asian	−0.007***	−0.003***	−0.001***
	(0.0003)	(0.0003)	(0.0003)
Hispanic	0.030***	0.033***	0.035***
	(0.0003)	(0.0003)	(0.0003)
Non-Hispanic Black	0.033***	0.023***	0.042***
	(0.0004)	(0.0005)	(0.001)
Comparison Mean ($)	248,555	248,555	250,021
Property FE	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes
Properties	16,456,528	16,456,528	16,394,685
Observations	39,470,293	39,470,293	39,230,446

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. In the first column, we present estimates of the baseline model estimated using Ethnicolr race/ethnicity predicitons. In the second and third columns, we replace race/ethnicity predictions from Ethnicolr with those from WRU. In the third column, we omit from the sample those transactions that are imputed as White buyers using WRU but Black buyers using Ethnicolr and vice versa. *p<0.1; **p<0.05; ***p<0.01

1.3.2 Composition of Buyers over Time

Table A1.10 reports our main estimates in Columns 1 and 2 (as in 1) but adds results from a regression with race-by-year time trends in Column 3. This specification controls for changes in the racial and ethnic composition of homebuyers over time. Since Hispanic and Black buyers make up a larger share of buyers in periods with relatively high housing prices, accounting for between-group differences in buyer timing reduces the estimates of premiums for Hispanic and Non-Hispanic Black buyers. Nevertheless, premiums remain large and significant: 1.9% and 2.6% for Hispanic and Black buyers, respectively.

Table A1.10. Baseline Results: Price Differentials

	Dependent variable: ln(sales price)		
	(1)	(2)	(3)
Asian	−0.009***	−0.007***	−0.012***
	(0.0004)	(0.0003)	(0.0004)
Hispanic	0.047***	0.030***	0.019***
	(0.0003)	(0.0003)	(0.0003)
Non-Hispanic Black	0.035***	0.033***	0.026***
	(0.0004)	(0.0004)	(0.0004)
Comparison Mean ($)	248,555	248,555	248,555
Property FE	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes
Tract x Year FE	No	Yes	No
Year FE	Yes	No	No
Calendar Month FE	Yes	Yes	Yes
Race x Year dummies	No	No	Yes
Properties	16,456,528	16,456,528	16,456,528
Observations	39,470,293	39,470,293	39,470,293

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects and calendar month fixed effects. Column 1 presents OLS estimates without tract x year FEs but with year FEs, column 2 adds tract by year fixed effects (our main specification), column 3 is estimated by averaging the estimates in each two year period obtained by regressing sales prices on race by 2 year time period dummies as shown in equation 7. *p<0.1; **p<0.05; ***p<0.01

1.3.3 Outliers

Our baseline estimates exclude properties whose median sales price lies above/below the 99/1st percentiles of sales prices in the data. They also exclude transactions which are above/below the 99/1st percentiles of the ratio of the sales price of a given transaction to the median sales price for any given home.[29] This insures that, for any given property, we are excluding transactions that have a sales price far above or below the median sales price for this same home.

In table A1.11 we present the sensitivity of our results to changing these thresholds to the 95/5th percentiles. Using the more restrictive thresholds for overall prices or within property prices generally makes very little difference in our estimates. However, estimates are slightly attenuated when we apply the 95/5th percentile threshold to both property prices and within property median prices. Nevertheless, the general magnitude and sign of our estimates remain the same.

[29] i.e. $priceratio_{ij} = \frac{transactionprice_{ij}}{medianprice_j}$

Table A1.11. Robustness to outliers

	Dependent variable: ln(sales price)			
	Main	95/5% Threshold Overall	95/5% Threshold Within	95/5% Threshold Both
Asian	−0.007***	−0.009***	−0.004***	−0.012***
	(0.0003)	(0.0004)	(0.0005)	(0.0002)
Hispanic	0.030***	0.030***	0.036***	0.023***
	(0.0003)	(0.0002)	(0.0003)	(0.0002)
Non-Hispanic Black	0.033***	0.032***	0.032***	0.024***
	(0.0004)	(0.0004)	(0.001)	(0.0002)
Comparison Mean ($)	248,555	226,680	226,702	229,581
Property FE	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes
Race x Year dummies	No	No	No	No
Properties	16,456,528	15,244,446	15,191,800	14,372,978
Observations	39,470,293	38,034,095	38,170,926	35,073,798

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. Column 1 presents estimates of the baseline model, in which we trim the highest and lowest 1% of the overall sales price and ratio of the transaction to median property purchase price. In column 2 we restrict the sample to properties whose median sales price is below/above the 95/5th percentiles in our data. In column 3 we restrict the sample to observations below/above the 95/5th percentiles for the ratio of transaction to median sales price for a given home. In column 4 we combine the restrictions in columns 2 and 3. *p<0.1; **p<0.05; ***p<0.01

1.3.4 Loans and Lenders

The ZTRAX data contains information on loan characteristics for a large number of transactions. For 9M properties (20M observations) we observe loan information for more than one sale. The remainder of transactions are either for homes that are never bought with a loan or homes that switch from being bought with a loan to being bought with cash (or for which we are missing loan information) or vice-versa. Note that it is not clear whether homes with a loan amount entered as 0 in the ZTRAX data are cash sales or sales for which the loan amount is missing. For example, in Georgia, Maryland, Michigan, Minnesota, Oklahoma, New York, South Carolina and Tennessee a majority of home sales are reported without a loan amount. It seems unlikely that such a high proportion of sales would in fact not involve a loan.

In Table A1.12, we present results for regressions that control for the presence of a loan and lender names. Adding a loan control and lender name controls slightly attenuates our estimates for Hispanic and Black buyers but changes the sign on the coefficient for Asian buyers. Due to the lack of reliability of the loan measures we do not include this control in our main results.

Table A1.12. Loans

	Dependent variable: ln(sales price)			
	Main	Loan Control	Lender Name FE	Loan Control + Lender Name FE
	(1)	(2)	(3)	(4)
Asian	−0.007***	0.004***	0.004***	0.004***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Hispanic	0.030***	0.022***	0.017***	0.017***
	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Non-Hispanic Black	0.033***	0.026***	0.022***	0.022***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
loan		0.189***		0.100***
		(0.0002)		(0.002)
Comparison Mean ($)	248,555	248,555	248,555	248,555
Property FE	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes
Properties	16,456,528	16,456,528	16,456,528	16,456,528
Observations	39,470,293	39,470,293	39,470,293	39,470,293

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. Column 1 presents estimates of the baseline model. In the second column we add a loan control. In the third column we add lender name fixed effects and in the forth column we include both a loan and lender name fixed effect. *p<0.1; **p<0.05; ***p<0.01

2 Additional Results

2.1 Comparison to Bayer et al. (2017)

Table A2.1. Comparison to Bayer et al. (2017)

Panel A:	Premium for NH Black	
MSA	Bayer et al. (2017)	This Paper
Baltimore	1.6%	4.9%
Chicago	2.9%	7.0%
Los Angeles	1.3%	1.6%
San Francisco	1.1%	1.5%

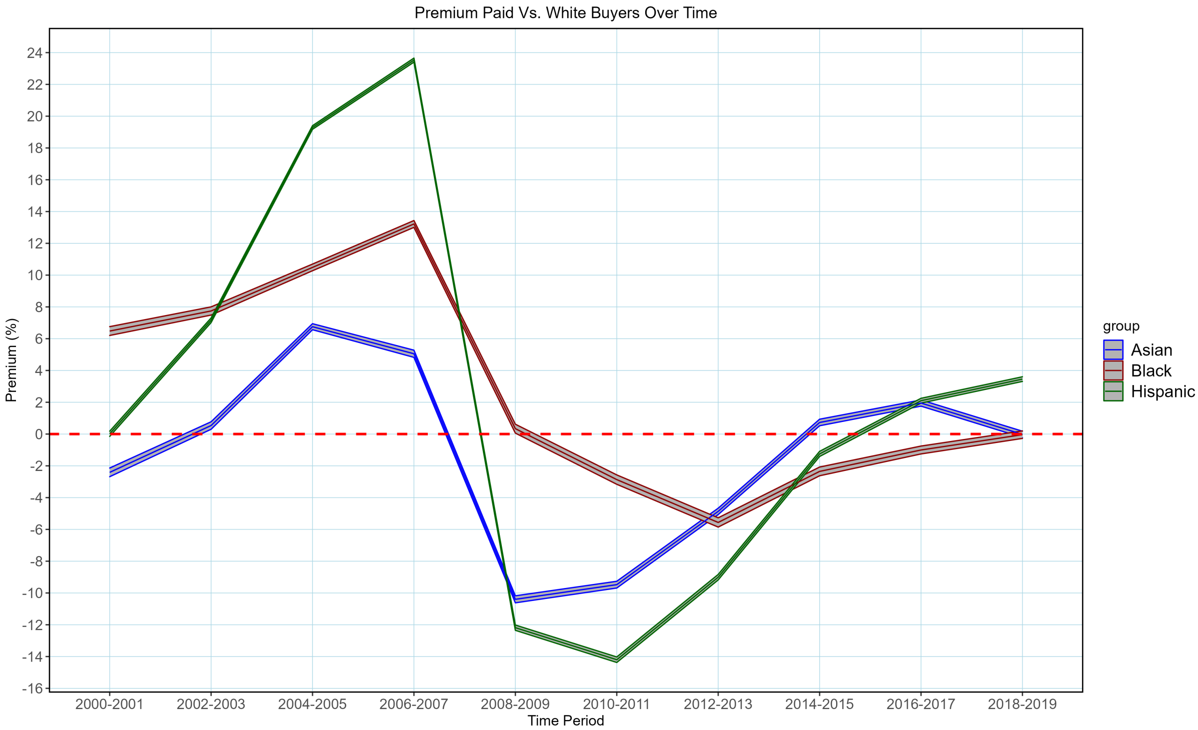
Panel B:	Premium for Hispanic	
MSA	Bayer et al. (2017)	This Paper
Baltimore	2.4%	0.9%
Chicago	1.5%	3.7%
Los Angeles	1.1%	2.5%
San Francisco	2.8%	2.9%

Panel C:	Observations	
MSA	Bayer et al. (2017)	This Paper
Baltimore	278,221	551,888
Chicago	382,389	1,456,520
Los Angeles	925,622	1,690,009
San Francisco	535,286	714,819

Notes: Panels A and B of the above table present a comparison between the estimates of the price premium paid by Black and Hispanic homebuyers in our sample compared to the same estimates for the same MSAs in Bayer et al. (2017) using the model specification that includes tract x year FEs (Table A1.10 column 2). Panel C presents the number of observations used in the estimations in both papers.

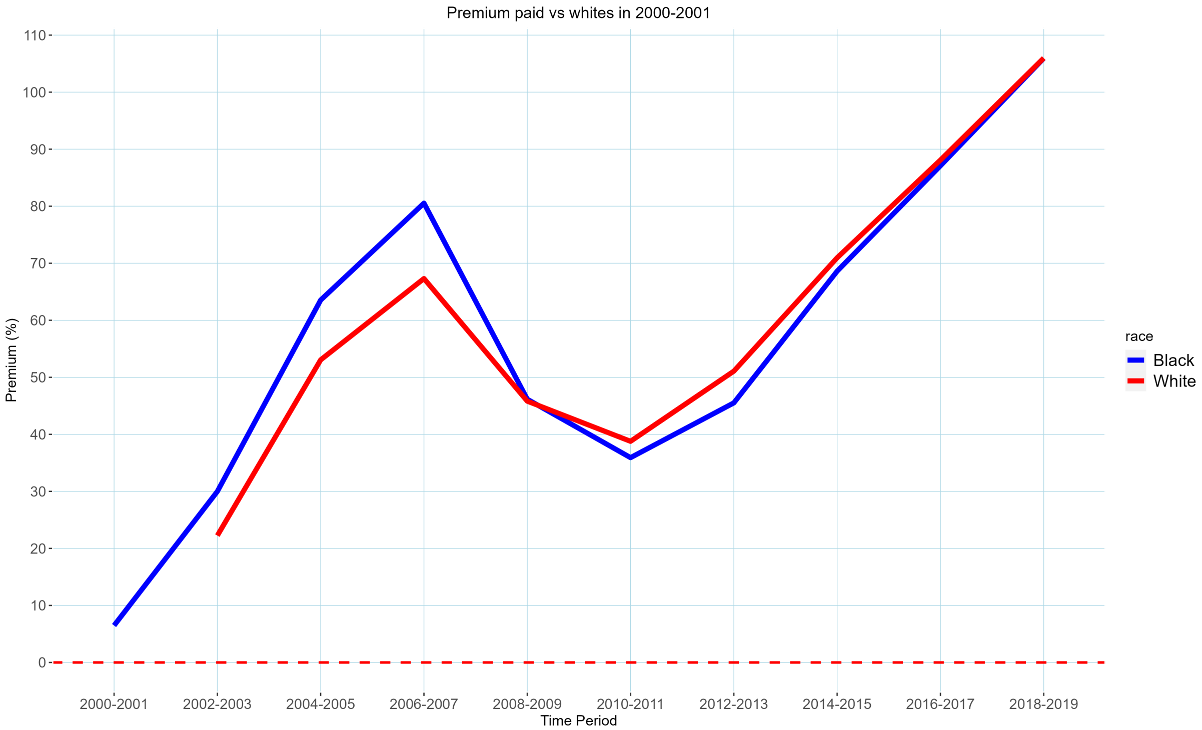
2.2 Time-varying Estimates

Figure A2.1. Racial Housing Price Differentials Over Time



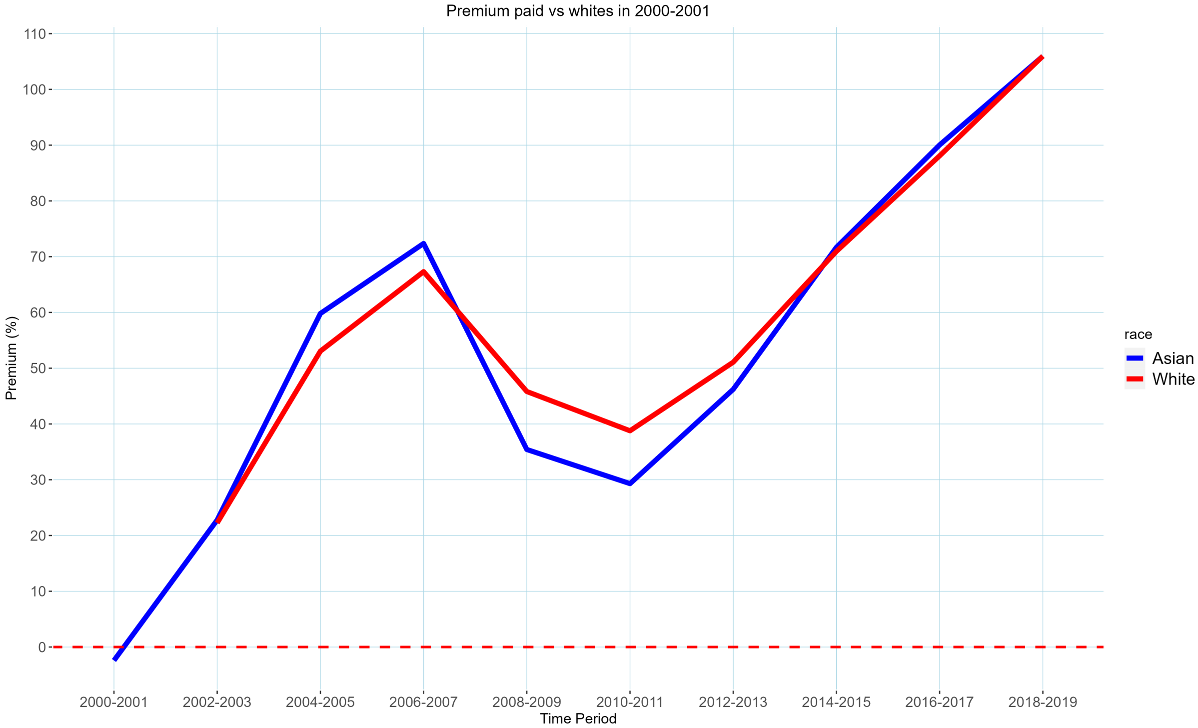
Notes: Figure shows point estimates and 95% confidence intervals (in %) for the housing price premium paid by Asian, Black and Hispanic buyers relative to white buyers in 34 US states and the District of Columbia over the period 2000-2020. These estimates are based on a specification that includes house age controls, property fixed effects, race by year dummies and calendar month fixed effects as in Table A1.10 column 3.

Figure A2.2. Housing Price Trends for Black vs. White Buyers relative to whites in 2000-2001



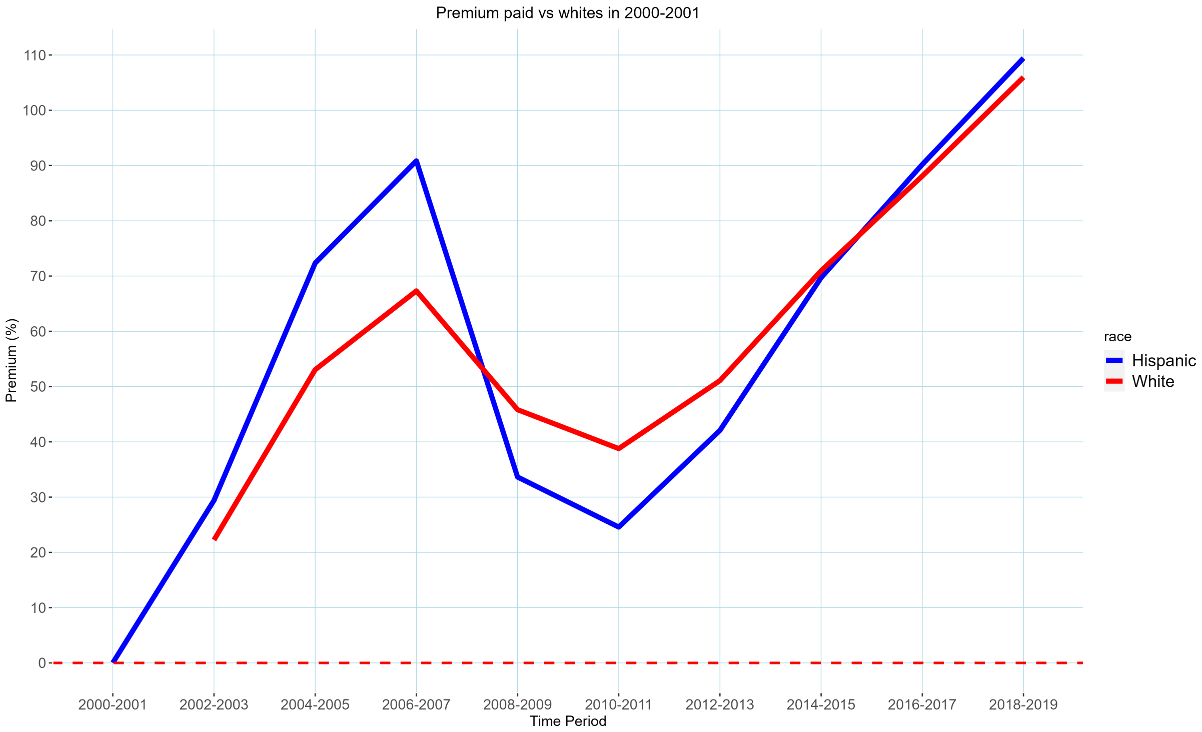
Notes: Figure shows point estimates for the housing price premium paid by Black (blue line) and White (red line) buyers relative to white buyers in 2000-2001 over the period 2000-2020. These estimates are obtained by plotting the coefficients from equation 7. Specifically, for whites each point is a coefficient β_0^t, for Blacks each point is the sum of β_1 (the intercept) and β_1^t. The difference between the red and white lines is the premium paid by Blacks in each year relative to whites in each year, as shown in figure A2.1.

Figure A2.3. Housing Price Trends for Asian vs. White Buyers relative to whites in 2000-2001



Notes: Figure shows point estimates for the housing price premium paid by Asian (blue line) and White (red line) buyers relative to white buyers in 2000-2001 over the period 2000-2020. These estimates are obtained by plotting the coefficients from equation 7. Specifically, for whites each point is a coefficient β_0^t, for Asians each point is the sum of β_2 (the intercept) and β_2^t. The difference between the red and white lines is the premium paid by Asians in each year relative to whites in each year, as shown in figure A2.1.

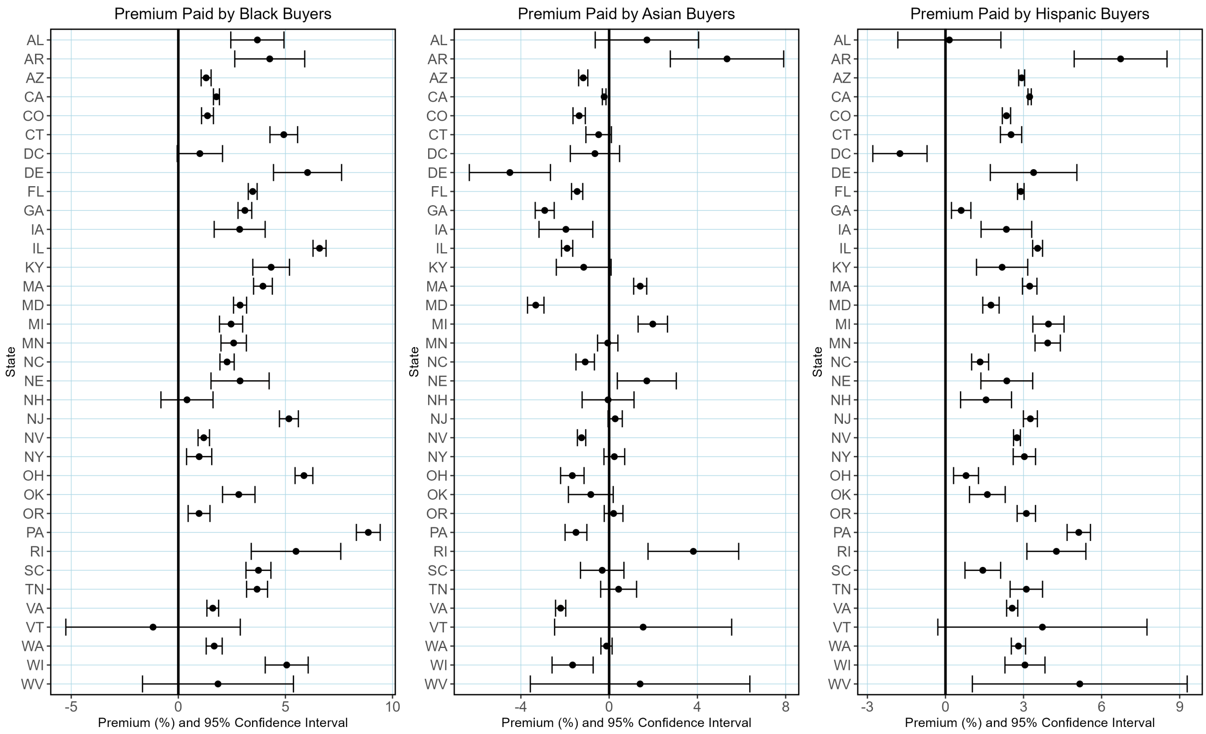
Figure A2.4. Housing Price Trends for Hispanic vs. White Buyers relative to whites in 2000-2001



Notes: Figure shows point estimates for the housing price premium paid by Hispanic (blue line) and White (red line) buyers relative to white buyers in 2000-2001 over the period 2000-2020. These estimates are obtained by plotting the coefficients from equation 7. Specifically, for whites each point is a coefficient β_0^t, for Hispanics each point is the sum of β_3 (the intercept) and β_3^t. The difference between the red and white lines is the premium paid by Hispanics in each year relative to whites in each year, as shown in figure A2.1.

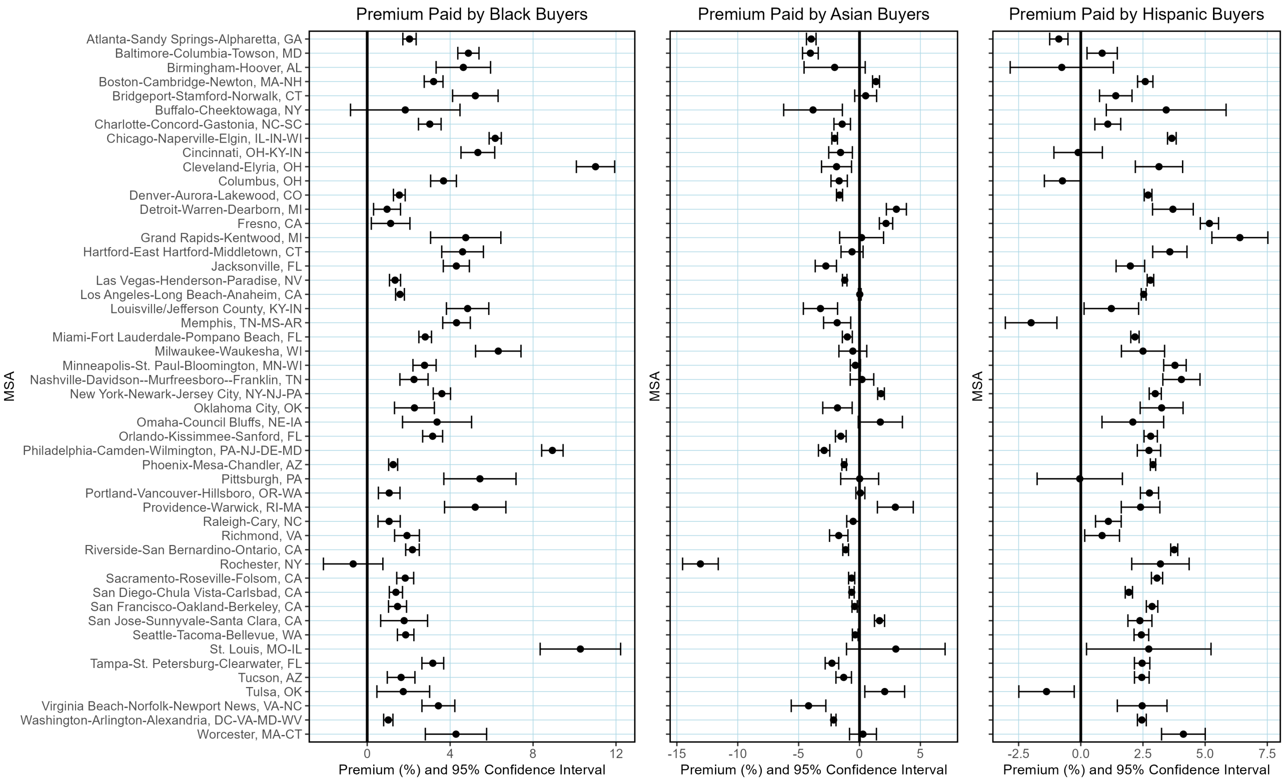
2.3 Geographic Variation

Figure A2.5. Racial Housing Price Differentials by State



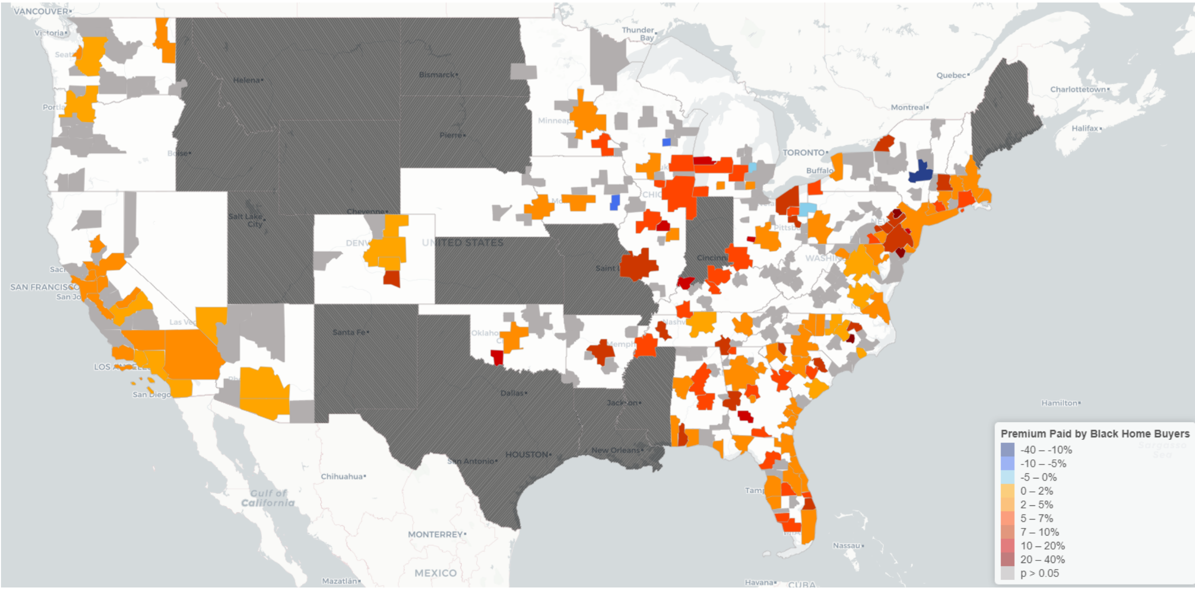
Notes: Figure shows point estimates and 95% confidence intervals (in %) for the housing price premium paid by Black, Asian and Hispanic homebuyers relative to white homebuyers in 34 US states and the District of Columbia over the period 2000-2020. These estimates are based on a specification that includes house age controls, property fixed effects, census-tract by year and calendar month fixed effects as in Table A1.10 column 2.

Figure A2.6. Racial Housing Price Differentials by MSA



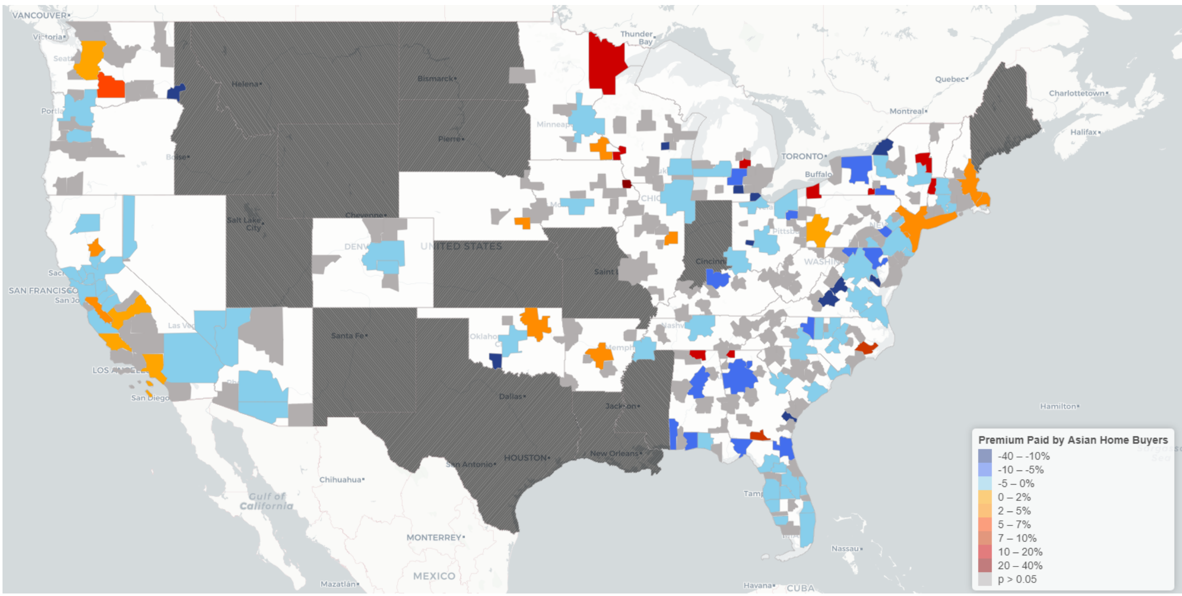
Notes: Figure shows point estimates and 95% confidence intervals (in %) for the housing price premium paid by Black, Asian and Hispanic homebuyers relative to white homebuyers in the 50 largest MSAs in our sample between 2000-2020. These estimates are based on a specification that includes house age controls, property fixed effects, census-tract by year and calendar month fixed effects as in Table A1.10 column 2.

Figure A2.7. Racial Housing Price Differentials for Black buyers by MSA



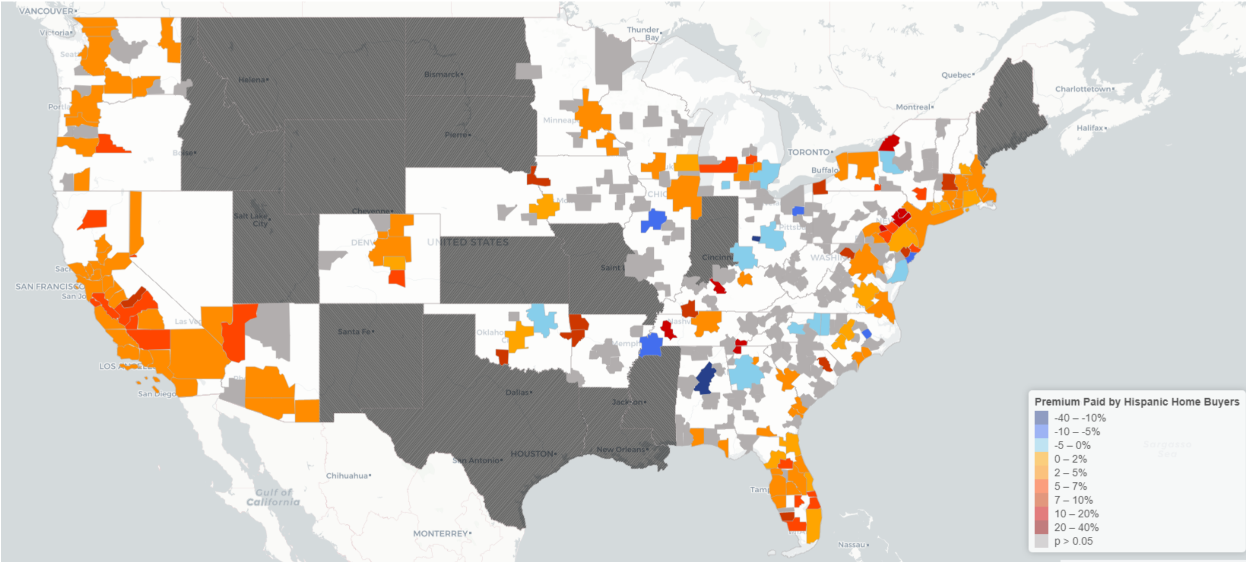
Notes: Figure shows point estimates for the housing price premium (in %) paid by Black homebuyers relative to white homebuyers in each MSA in our sample over the period 2000-2020. These estimates are based on a specification that includes house age controls, property fixed effects, census-tract by year and calendar month fixed effects as in Table A1.10 column 2. Darker shades of orange/red indicate higher premiums, blue shades indicate negative premiums relative to whites. States in dark grey are non-disclosure states for which we have no data. MSAs in light grey are MSAs for which estimates are not significantly different from 0. Note that certain MSAs straddle disclosure and non-disclosure states. For these MSAs, the results shown are obtained using only data from the portion of the MSA which is in a full-disclosure state (ie. for which we have data). For similar figures for Asian and Hispanic buyers, see appendix section 2.3.

Figure A2.8. Racial Housing Price Differentials for Asian buyers by MSA



Notes: Figure shows point estimates for the housing price premium (in %) paid by Asian homebuyers relative to white homebuyers in each MSA in our sample over the period 2000-2020. These estimates are based on a specification that includes house age controls, property fixed effects, census-tract by year and calendar month fixed effects as in Table A1.10 column 2. Darker shades of orange/red indicate higher premiums, blue shades indicate negative premiums relative to whites. States in dark grey are non-disclosure states for which we have no data. MSAs in light grey are MSAs for which estimates are not significantly different from 0. Note that certain MSAs straddle disclosure and non-disclosure states. For these MSAs, the results shown are obtained using only data from the portion of the MSA which is in a full-disclosure state (ie. for which we have data).

Figure A2.9. Racial Housing Price Differentials for Hispanic buyers by MSA



Notes: Figure shows point estimates for the housing price premium (in %) paid by Hispanic homebuyers relative to white homebuyers in each MSA in our sample over the period 2000-2020. These estimates are based on a specification that includes house age controls, property fixed effects, census-tract by year and calendar month fixed effects as in Table A1.10 column 2. Darker shades of orange/red indicate higher premiums, blue shades indicate negative premiums relative to whites. States in dark grey are non-disclosure states for which we have no data. MSAs in light grey are MSAs for which estimates are not significantly different from 0. Note that certain MSAs straddle disclosure and non-disclosure states. For these MSAs, the results shown are obtained using only data from the portion of the MSA which is in a full-disclosure state (ie. for which we have data).

2.4 Additional Outgroup Results

Table A2.2. Racial Price Differentials, Outgroup Interactions and Neighborhood Composition

	Dependent variable: ln(sales price)			
	Own-Group Share	Outgroup	Outgroup by Race/Eth.	Outgroup by Own-Group Share
Asian	−0.012***		−0.048***	−0.090***
	(0.0005)		(0.001)	(0.002)
Hispanic	0.016***		0.0002	−0.051***
	(0.001)		(0.001)	(0.002)
NH Black	0.017***		−0.049***	−0.086***
	(0.001)		(0.002)	(0.004)
≥ 5% Share Asian * Outgroup	0.013***			0.062***
	(0.001)			(0.003)
≥ 5% Share Hispanic * Outgroup	0.019***			0.062***
	(0.001)			(0.003)
≥ 5% Share NH Black * Outgroup	0.028***			0.052***
	(0.001)			(0.005)
Outgroup		0.020***	0.002***	−0.005***
		(0.0003)	(0.0004)	(0.001)
Asian * Outgroup			0.055***	0.096***
			(0.001)	(0.002)
Hispanic * Outgroup			0.031***	0.074***
			(0.001)	(0.002)
NH Black * Outgroup			0.075***	0.102***
			(0.002)	(0.004)
Asian * ≥ 5%				0.001**
				(0.001)
Hispanic * ≥ 5%				0.012***
				(0.001)
NH Black * ≥ 5%				0.002***
				(0.001)
Asian * ≥ 5% Share * Outgroup				−0.059***
				(0.003)
Hispanic * ≥ 5% Share * Outgroup				−0.051***
				(0.003)
NH Black * ≥ 5% Share * Outgroup				−0.031***
				(0.005)
Comparison Mean ($)	249,430	253,255	253,255	254,564
Property FE	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes
Census Tract x Year FE	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes
Properties	16,005,732	9,627,718	9,627,718	9,380,871
Observations	38,446,690	21,370,397	21,370,397	21,359,674

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. In column 1 we interact buyer race/ethnicity with own-group race share. In column 2, we regress prices on an outgroup indicator that is equal to one if a transaction is between members of different groups. In column 3 we further break out outgroup premiums by interacting buyer race and the outgroup dummy. In column 4 we add interactions between buyer race/ethnicity, outgroup seller status and own-group census tract population share. *p<0.1; **p<0.05; ***p<0.01

Table A2.3. Racial Price Differentials and All Group Interactions

	Dependent variable: ln(sales price)
	(1)
Asian -> Asian	−0.045***
	(0.001)
Hispanic -> Hispanic	0.004***
	(0.001)
NH Black -> NH Black	−0.054***
	(0.002)
Hispanic -> Asian	0.009***
	(0.001)
NH Black -> Asian	−0.014***
	(0.002)
NH White -> Asian	0.012***
	(0.001)
Asian -> Hispanic	0.050***
	(0.001)
NH Black -> Hispanic	0.022***
	(0.002)
NH White -> Hispanic	0.035***
	(0.0004)
Asian -> NH Black	0.034***
	(0.003)
Hispanic -> NH Black	0.049***
	(0.002)
NH White -> NH Black	0.021***
	(0.001)
Asian -> NH White	0.005***
	(0.001)
Hispanic -> NH White	0.011***
	(0.001)
NH Black -> NH White	−0.017***
	(0.001)
Comparison Mean ($)	254,564
House Age Control	Yes
Month FE	Yes
Census Tract x Year FE	Yes
Properties	9,627,718
Observations	21,370,397

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on seller->buyer race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. Variable labels should be read as sold from group x to group y.
*p<0.1; **p<0.05; ***p<0.01

2.5 Additional Market Conditions Results

Table A2.4. Racial Price Differentials and Market Thickness

	Dependent variable: ln(sales price)					
	Main	P75/25	P80/20	P85/15	P90/10	P95/5
Asian	−0.016***	−0.014***	−0.016***	−0.016***	−0.017***	−0.017***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Hispanic	0.031***	0.033***	0.032***	0.032***	0.033***	0.033***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Non-Hispanic Black	0.047***	0.051***	0.051***	0.049***	0.049***	0.047***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Thick		0.007**	0.007**	0.006	0.001	0.016***
		(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
Thin		0.001	0.0001	−0.004	−0.010**	−0.022***
		(0.003)	(0.003)	(0.004)	(0.004)	(0.006)
Black*Thick		−0.021***	−0.023***	−0.023***	−0.024***	−0.014
		(0.006)	(0.006)	(0.007)	(0.008)	(0.011)
Black*Thin		0.011	0.013*	0.017**	0.011	0.035**
		(0.007)	(0.007)	(0.008)	(0.008)	(0.015)
Hispanic*Thick		−0.013***	−0.015***	−0.015***	−0.014***	−0.031***
		(0.004)	(0.004)	(0.005)	(0.005)	(0.007)
Hispanic*Thin		0.013***	0.018***	0.015***	0.009	0.015
		(0.005)	(0.005)	(0.006)	(0.008)	(0.012)
Asian*Thick		−0.005	−0.001	−0.004	−0.0003	0.010
		(0.006)	(0.006)	(0.006)	(0.008)	(0.011)
Asian*Thin		−0.0003	0.003	0.009	0.018*	0.036**
		(0.006)	(0.006)	(0.007)	(0.010)	(0.016)
Comparison Mean ($)	250,376	250,376	250,376	250,376	250,376	250,376
Property FE	Yes	Yes	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Properties	675,387	675,387	675,387	675,387	675,387	675,387
Observations	1,348,606	1,348,606	1,348,606	1,348,606	1,348,606	1,348,606

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. We add thick/thin market dummies and race*thick/thin dummy interactions using Realtor.com's measure of listings by ZIP code. Note that due to the limited availabilty of Realtor.com's measure, only properties that sold twice between July 2016 and December 2019 are included in this analysis. *p<0.1; **p<0.05; ***p<0.01

Table A2.5. Outgroups -> Buyer Race and Market Thickness

	Dependent variable: ln(sales price)					
	Main	P75/25	P80/20	P85/15	P90/10	P95/5
Outgroup	0.039***	0.041***	0.041***	0.040***	0.040***	0.039***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Thick		0.005	0.010**	0.002	0.0004	−0.002
		(0.004)	(0.004)	(0.004)	(0.005)	(0.006)
Thin		−0.002	0.001	0.004	0.002	−0.020**
		(0.004)	(0.004)	(0.005)	(0.006)	(0.008)
Outgroup*Thick		−0.004	−0.006	−0.002	−0.00002	0.006
		(0.004)	(0.004)	(0.004)	(0.005)	(0.007)
Outgroup*Thin		−0.005	−0.004	−0.012**	−0.021***	−0.011
		(0.004)	(0.005)	(0.005)	(0.007)	(0.010)
Comparison Mean ($)	251,289	251,289	251,289	251,289	251,289	251,289
Property FE	Yes	Yes	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Properties	437,784	437,784	437,784	437,784	437,784	437,784
Observations	873,661	873,661	873,661	873,661	873,661	873,661

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. We add hot/cold dummies and seller -> buyer race*race*hot/cold dummy interactions using Realtor.com's measure of number of listings.

*p<0.1; **p<0.05; ***p<0.01

Table A2.6. Racial Price Differentials and Market Hotness

	Dependent variable: ln(sales price)					
	Main	P75/25	P80/20	P85/15	P90/10	P95/5
Asian	−0.014***	−0.013**	−0.013**	−0.012**	−0.013***	−0.012***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
Hispanic	0.038***	0.043***	0.043***	0.042***	0.039***	0.039***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Non-Hispanic Black	0.049***	0.054***	0.052***	0.055***	0.056***	0.052***
	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)
Hot		0.007*	0.006	0.004	0.007	−0.004
		(0.004)	(0.004)	(0.004)	(0.005)	(0.006)
Cold		−0.003	−0.008*	−0.005	0.001	−0.010*
		(0.004)	(0.004)	(0.004)	(0.005)	(0.006)
Black*Hot		0.009	0.016	0.001	−0.001	−0.008
		(0.011)	(0.011)	(0.013)	(0.015)	(0.020)
Black*Cold		−0.021**	−0.021**	−0.035***	−0.057***	−0.043**
		(0.010)	(0.010)	(0.011)	(0.013)	(0.018)
Hispanic*Hot		0.010	0.011	0.021***	0.031***	0.012
		(0.007)	(0.007)	(0.008)	(0.009)	(0.013)
Hispanic*Cold		−0.029***	−0.036***	−0.042***	−0.034***	−0.023*
		(0.007)	(0.007)	(0.008)	(0.010)	(0.013)
Asian*Hot		0.017*	0.015	0.013	0.019	0.014
		(0.009)	(0.010)	(0.010)	(0.012)	(0.017)
Asian*Cold		−0.013	−0.015*	−0.018**	−0.016*	−0.025**
		(0.008)	(0.008)	(0.009)	(0.010)	(0.012)
Comparison Mean ($)	223,736	223,736	223,736	223,736	223,736	223,736
Property FE	Yes	Yes	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Properties	291,003	291,003	291,003	291,003	291,003	291,003
Observations	586,184	586,184	586,184	586,184	586,184	586,184

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, house age, property fixed effects, census-tract by year and calendar month fixed effects. We add hot/cold dummies and race*hot/cold dummy interactions using Realtor.com's measure of market hotness. Note that due to the limited availabilty of Realtor.com's measure, only properties that sold twice between August 2017 and December 2019 are included in this analysis. *p<0.1; **p<0.05; ***p<0.01

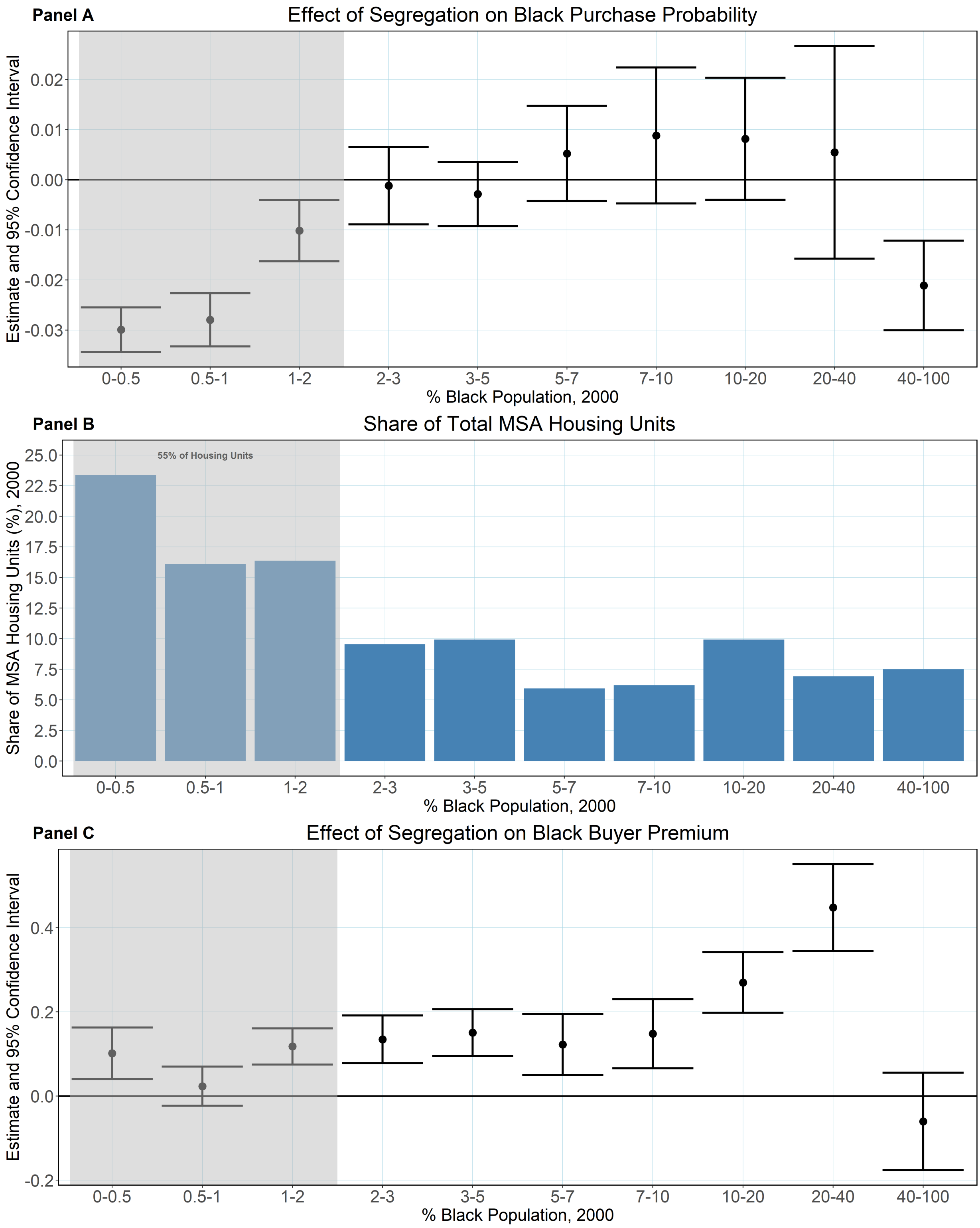
Table A2.7. Outgroups -> Buyer Race and Market Hotness

	Dependent variable: ln(sales price)					
	Main	P75/25	P80/20	P85/15	P90/10	P95/5
Outgroup	0.045***	0.053***	0.054***	0.053***	0.051***	0.048***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Hot		0.004	0.004	−0.001	0.005	−0.010
		(0.005)	(0.005)	(0.005)	(0.006)	(0.008)
Cold		−0.002	−0.002	0.0003	0.001	0.004
		(0.005)	(0.005)	(0.006)	(0.006)	(0.008)
Outgroup*Hot		0.018***	0.017**	0.018**	0.008	0.008
		(0.007)	(0.007)	(0.008)	(0.009)	(0.012)
Outgroup*Cold		−0.039***	−0.047***	−0.048***	−0.046***	−0.042***
		(0.006)	(0.006)	(0.007)	(0.007)	(0.010)
Comparison Mean ($)	222,597	222,597	222,597	222,597	222,597	222,597
Property FE	Yes	Yes	Yes	Yes	Yes	Yes
House Age Control	Yes	Yes	Yes	Yes	Yes	Yes
Tract x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Properties	193,165	193,165	193,165	193,165	193,165	193,165
Observations	385,680	385,680	385,680	385,680	385,680	385,680

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on an indicator for whether the seller and buyer races are different, house age, property fixed effects, census-tract by year and calendar month fixed effects. We add hot/cold dummies and seller -> buyer race*hot/cold dummy interactions using Realtor.com's measure of market hotness. *p<0.1; **p<0.05; ***p<0.01

2.6 Additional Segregation Results

Figure A2.10. Segregation as a Constraint



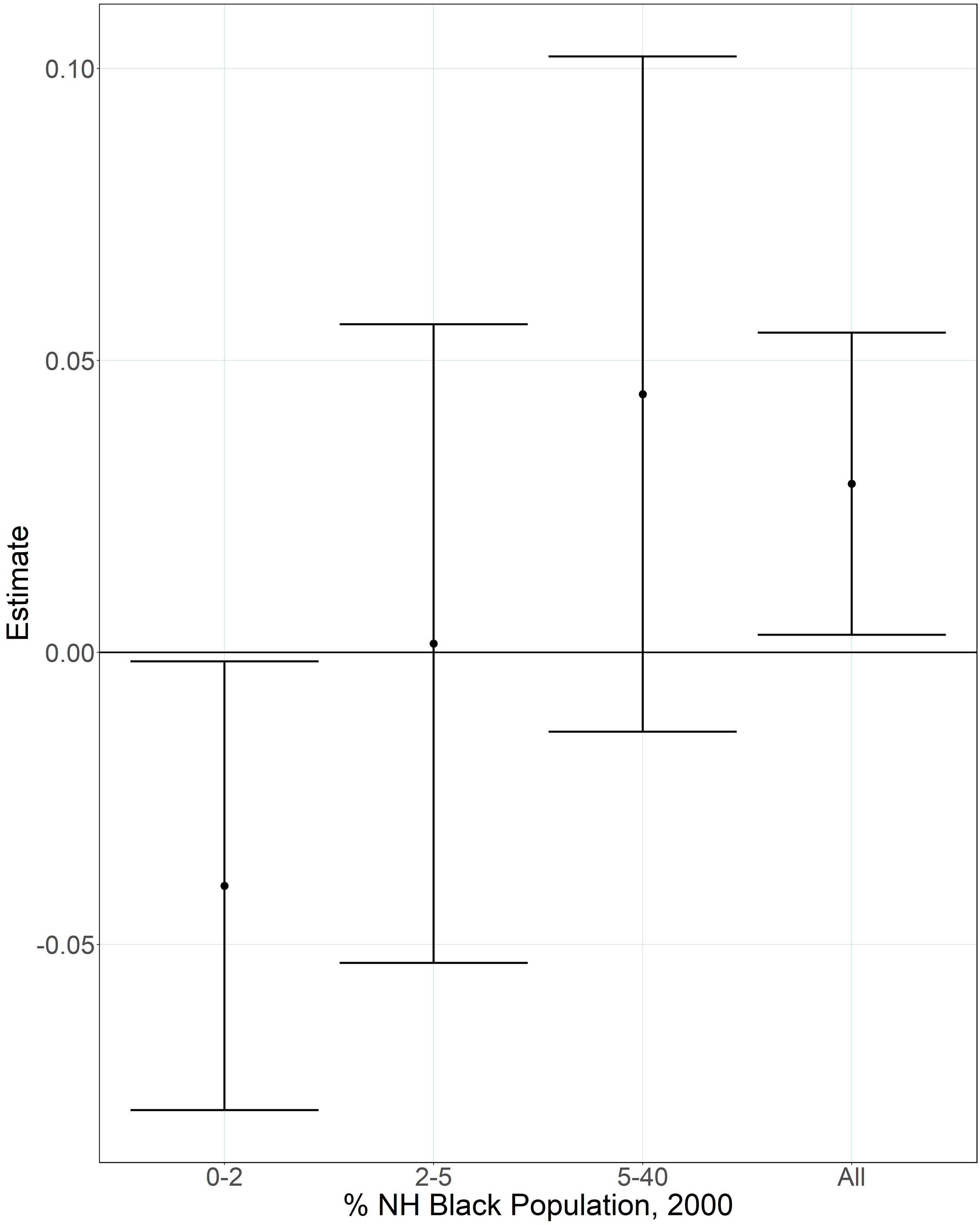
Notes: Panel A shows point estimates and 95% confidence intervals from a regression of the probability that a Black buyer in an MSA purchases in any given tract on dissimilarity (instrumented by the RDI), by bin of % Black population. Panel B shows the average share of total housing units in an MSA in each % Black population bin. Panel C shows the effect of segregation on price differentials for Black buyers as estimated in table 3 with the sample split into bin of % Black population.

Table A2.8. Outgroup Segregation and Price Differentials

	Dependent variable: ln(sales price)				
		IV			
	Outgroup	Outgroup*Seg	0-2% Black Share	2-5% Black Share	5-40% Black Share
Asian	−0.003***	−0.001**	0.006***	−0.0001	−0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	0.027***	0.025***	0.020***	0.021***	0.035***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
NH Black	−0.023***	−0.255***	−0.196***	−0.311***	−0.186**
	(0.004)	(0.035)	(0.072)	(0.081)	(0.075)
Outgroup	0.007***	0.111***	0.062***	0.154***	0.155***
	(0.0004)	(0.005)	(0.006)	(0.009)	(0.010)
NH Black*Outgroup	0.044***	0.154***	0.168**	0.178**	0.047
	(0.004)	(0.036)	(0.072)	(0.082)	(0.076)
Outgroup*Seg		−0.179***	−0.102***	−0.251***	−0.242***
		(0.008)	(0.010)	(0.016)	(0.018)
NH Black*Seg		0.359***	0.208*	0.464***	0.281**
		(0.055)	(0.124)	(0.135)	(0.118)
NH Black*Outgroup*Seg		−0.151***	−0.146	−0.212	0.005
		(0.056)	(0.125)	(0.137)	(0.120)
Comparison Mean ($)	272,951	272,951	292,042	277,898	233,233
Comparison Mean (Dissimilarity)	NA	0.59	0.58	0.59	0.61
House Age Control	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Census Tract x Year FE	Yes	Yes	Yes	Yes	Yes
1st Stage F-Stat	NA	203.7	33.5	90.5	72.8
Properties	3,353,215	3,353,215	1,711,855	817,357	742,739
Observations	7,810,672	7,810,672	3,948,321	1,933,798	1,743,650

Notes: This table presents repeat-sales estimates from a regression of the log of the sales price on race/ethnicity dummies, outgroup indicator, house age, property fixed effects, census-tract by year and calendar month fixed effects for the sample of transactions in cities for which the RDI is available. In column 1, we present baseline estimates of the outgroup differential for Black buyers for this more limited sample. In the second column, we interact the NH Black indicator with the dissimilarity index and outgroup indicator. In the last three columns we split the sample based on the Black population share in a tract. *p<0.1; **p<0.05; ***p<0.01

Figure A2.11. Effect of Segregation on Outgroup Premiums



Notes: Figure shows the effect of segregation on outgroup price differentials for Black buyers as estimated in table A2.8. Coefficients and 95% confidence intervals are calculated using the delta method.