The Impact of iBuying is About More Than Just Racial Disparities: Evidence from Mecklenburg County, NC

Isaac Slaughter
is28@uw.edu
University of Washington
Seattle, Washington, USA

Eva Maxfield Brown
evamxb@uw.edu
University of Washington
Seattle, Washington, USA

Nic Weber
nmweber@uw.edu
University of Washington
Seattle, Washington, USA

ABSTRACT

Instant buyers (iBuyers)—companies that buy and sell homes based on automated valuation models (AVMs)—now hold more than 5% market share in some USA cities. In this work, we investigate the fairness of iBuyers by constructing a dataset that links racial demographics from voter records with detailed property information on over 50,000 real estate transactions. Using Bayesian hierarchical modeling we find that: *1. iBuyers Decrease the Racial Sales Price Gap Between Black and White Home Sellers.* Controlling for over 50 property features we find that iBuyers reduce the racial price gap that otherwise exists between homes sold by Black and White homeowners. This is not, however, a result of equity achieved through proprietary AVMs, but rather a result of both Black and White homeowners being similarly disadvantaged by iBuyers' low purchase prices; and, *2. iBuyers Increase Property Conversion Rates from Individual to Institutional Ownership.* We trace iBuyers' purchases as well as their follow-on sales of homes in Mecklenburg County. In doing so, we show that iBuyers increase the rate at which properties are converted from being individually owned to institutionally owned. The eventual purchasers of iBuyer homes include national and international rental companies that have been tied to high eviction rates and poor property management. As with sale prices, we find that rather than reapportioning this social harm more equitably, iBuyers are simply increasing the rate at which homes bought from White homeowners are converted to institutional ownership. Ultimately, our analysis suggests that *iBuyers are Equalizing Housing Outcomes by Extending Real Estate Harms Typically Isolated to Black Homeowners to White homeowners as Well.*

CCS CONCEPTS

• **Social and professional topics → Technology audits**; *Sociotechnical systems*; *Race and ethnicity*; • **Applied computing →** *Economics.*

KEYWORDS

iBuying, Automated Valuation Models, Algorithm Auditing, Bayesian Hierarchical Modeling, Housing Discrimination

This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3659027

ACM Reference Format:
Isaac Slaughter, Eva Maxfield Brown, and Nic Weber. 2024. The Impact of iBuying is About More Than Just Racial Disparities: Evidence from Mecklenburg County, NC. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3630106.3659027

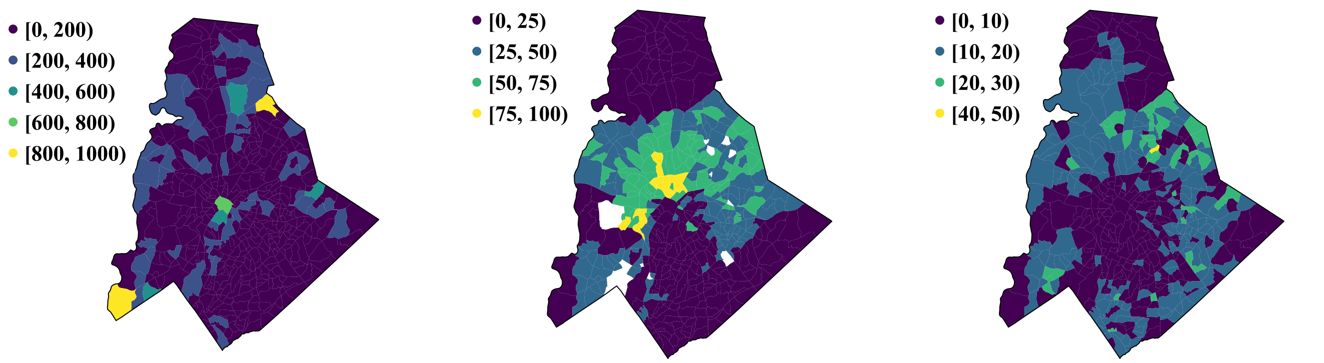
1 INTRODUCTION

Although housing policies that explicitly discriminate on race, like redlining and exclusionary zoning, have been rescinded in the United States, empirical research continually shows lingering differences in property values for Black and White homes of equivalent quality [4, 23, 34]. In this paper, we investigate whether racial disadvantages are perpetuated by instant buyers (iBuyers), companies that use proprietary Automated Valuation Models (AVMs) to purchase homes directly from individuals. Pitching themselves as a fast and easy alternative to brokerage-based home selling, iBuyers allow a homeowner to initiate a potential sale online, receive a quote within hours, and close within weeks [7, 35]. iBuyers' value proposition to investors is that through AVMs they can wipe out inefficiencies in a local housing market—acquiring large portfolios of homes cheaply, then reselling the properties at a scale that allows for small profit margins to pay large dividends.[1] Although iBuyers are a relatively recent innovation they currently hold more than 5% market share in major metropolitan areas like Phoenix, San Antonio, and Charlotte [20].

In our work, the first question we seek to address is this: *Do iBuyers tend to pay equitable prices for homes bought from Black and White homeowners?* Previous work has shown that AVMs may undervalue homes owned by Black home sellers relative to those owned by White home sellers [4], or may be less accurate for homes sold in Black neighborhoods than White neighborhoods [29, 40]. Using public records concerning over 50,000 individual property transactions in Mecklenburg County, North Carolina, however we first find that iBuyers are associated with a *smaller* racial sale price gap than individual purchasers. We estimate that on average individuals will pay a premium of over $36,000 to purchase a home from a White seller rather than a Black seller, while the average premium paid by iBuyers is approximately $4,000.[2] This race gap is

[1] Occasionally, iBuyers engage in property renewal and home repair or improvements, but these are a small component of the iBuying business model and should be considered quite separate from what are similar speculative real estate buyers and sellers like "home flippers" [13].

[2] This estimate is based on a model that controls for features of the home and neighborhood. In addition to estimating the racial sales price gap with these controls, we also fit a model that estimates the racial sales price gap while controlling for the assessed value of a home. Using this alternate model, the individual and iBuyer premiums are around $14,000 and -$2,000 respectively, rather than the $36,000 and $4,000 reported above. As we discuss in Section 4, we suspect that the appraisal-based estimates may contain biases themselves that could account for these differences.



(a) Number of sales in dataset

(b) Percent of population that is Black or African American, based on U.S. Census Bureau 2020 estimates

(c) Percent of sales in dataset that are to iBuyers

Figure 1: Locations of sales, Black population, and iBuying in Mecklenburg, shown per Neighborhood Profile Area. Dataset includes all iBuying transactions occurring in county between January of 2018 and October of 2023 as well as 30% Bernoulli random sample of transactions not involving iBuyers during this time period.

reduced to the extent that our model suggests a 16% probability that the effect is in the other direction, and on average iBuyers in fact pay a premium for Black-owned homes rather than White-owned homes.

While these results seem to paint iBuyers as an equalizing force in modern real estate, we caution against this interpretation. Instead of erasing disparities in sale prices by simply raising the prices paid to Black homeowners, iBuyers appear to be doing so by paying substantially less than individual buyers for White-owned homes. For Black home sellers, we estimate the average acquisition discount for homes sold to iBuyers—the amount the seller is missing out on by selling their home to an iBuyer rather than a personal purchaser—to be around -$4,000, with the probability that this discount is negative being 83%. This suggests that it is more likely than not that iBuyers are in fact paying slightly more than other purchasers to Black homeowners for equivalent homes. On the other hand, we estimate the iBuyer acquisition discount for White homeowners to be over $27,000 (around 8.8% of the $305,000 median home sale price in our dataset).

Beyond this, we also seek to shed light on the long-term effects of iBuyer mediation in sales, asking: *Do iBuyers tend to convert homes to institutional ownership at a higher rate than comparable direct home sales?* We believe large-scale conversions of residential property from individuals to institutions is potentially more harmful than short-term pricing harms introduced by iBuyers, as institutional investors are actively accumulating large swaths of single-family homes in the region and appear to be associated with large rent increases and eviction rates [14, 31].[3] We find that there is a meaningful chance that after a home is sold to an iBuyer, it will go on to be converted to institutional ownership regardless of the race of the home's original owner. In our dataset, 25% of sales made to iBuyers are resold to institutional investors (36% for Black homeowners and 17% for White homeowners), which is higher than a regional conversion rate of around 15% when homeowners sell their homes directly.[4] In the United States, homeownership remains one of the most important routes to financial security and wealth [19]. So, contrary to the positive spillovers that are suggested in previous work on iBuyers [20], our pricing and ownership conversion results paint a pessimistic picture of how unregulated AVMs might advance racial disadvantages in the housing market.[5]

2 BACKGROUND AND RELATED WORK

iBuyers such as Offerpad, Opendoor, and Knock have risen to popularity in real estate markets throughout the USA by providing home sellers a quick way to liquidate what is many people's most valuable and illiquid asset. After purchasing homes, iBuyers then try to sell them at appreciated prices, avoiding the added costs of repairs when they can [9]. In addition to making money through the returns on resold homes, iBuyers also charge sellers an explicit fee for their services.

iBuying is made possible by the development of Automated Valuation Models (AVMs), proprietary algorithms that iBuyers rely on to rapidly assess properties and their potential resale values. iBuyers tend to operate in markets where they can create a greater than 5% increase in sales price from purchase price [9], and where the housing stock is uniform and for that reason somewhat predictable [35]. Although relatively new, iBuyers are becoming increasingly popular amongst sellers and buyers in the USA—In 2022 iBuyers

[3]Some estimates suggest that just three SFR / REIT investors owned over 10% of single-family rental homes in Florida's Tampa-St. Petersburg-Clearwater Metropolitan Statistical Area in 2017, for example [11].

[4]For our purposes, institutional investors include both Single Family Rental (SFR) Trusts and Real Estate Investment Trusts.

[5]Code for this project is available online at https://www.github.com/isaacOnline/FAccT-Housing.

had a collective market share as high as 6.0% in Atlanta, GA and 5.7% in San Antonio, TX, [32].

Estimates of home value—whether given by humans or by AVMs—are known to discriminate against Black and Latine homeowners. For human-provided value estimates, this holds both for values assigned by appraisers [4, 22, 23, 34] and those assigned by homeowners themselves [6, 21, 30]. Audits of informational AVMs provided to consumers or researchers, such as Zillow's Zestimates, have also shown that when trained on historical transaction prices, the models can inadvertently learn and perpetuate racial biases [4, 27]. However, previous research into the impact of iBuying on minority homeowners has not found that this trend extends to the sale prices paid by the companies. Harrison et al. [20] show that the overall market effect of iBuyers is an increase in home prices by up to 2.8%. They attribute this rise in purchase price to sellers gaining an informational advantage from instant offers that help establish a home's minimum value. They also show that iBuyers crowd out potential local home buyers, forcing them into neighboring markets, and thereby causing home prices in these adjacent ZIP codes to increase as well. However, they find no direct evidence that the redirection of purchasing to neighboring markets is attributable to the race of either sellers or buyers. Similarly, Seiler and Yang [35] use real estate transaction and assessment records to show that iBuyers purchase homes at lower prices than individual owner-occupiers. They interpret iBuyers' lower purchasing price as as an 'acquisition discount' that sellers accept in return for reduced closing time and ease of transaction. Seiler and Yang [35], also find no support for the idea that iBuyers are disproportionately buying distressed homes, or that there is a racial preference in either home purchase volume or price.

We seek to build upon this previous work in two ways:

(1) Self-reported Identity: Previous work on the effects of race on iBuyers' sale prices has performed analysis at a ZIP code level—matching a home sale with a probability of the owner's race based on property location. We match property transaction records with voting records where individuals self-report their race. In doing so, we differentiate between effects based on the racial makeup of the neighborhood a home is in from the races of the actual buyers and sellers of the home.
(2) Property Conversion: Previous work has focused on iBuyer acquisitions and sales in terms of price. Here, we trace the purchase and sale of homes to understand rates and patterns of properties being converted from personal purchasers to institutional investors.

Together, we hope to provide fuller picture of how, and with whom, iBuyers operate.

3 DATA

To investigate the impact that iBuyers have on housing outcomes, we construct a dataset that includes information on properties sold in Mecklenburg, the neighborhoods those properties are in, and the people who bought or sold them. Our dataset was assembled from public records of property transfers occurring between January of 2018 and October of 2023, inclusive. Mecklenburg County is home to the City of Charlotte, where iBuyers held more than 8% market share in 2021 [35], the second largest share for any city studied by Seiler and Yang [35] in work on high-volume iBuying regions. Charlotte also has a relatively large share of African American residents, 35% according to 2022 ACS 5-year estimates, which behind Atlanta, Georgia is the second largest of the cities studied by Seiler and Yang [35]. We choose to focus on Mecklenburg for these reasons, as well as for the availability of person-level race data in the area, which in North Carolina is listed as a part of public voting records. The period we cover, starting in 2018, covers nearly all iBuying transactions that have occurred in the region.[6] Property transfers that we consider come from the Mecklenburg County Register of Deeds and include all transfers involving iBuyers, along with a 30% Bernoulli random sample of all other property transfers in the county in this period. Visualizations showing the geographic distribution of sales in the dataset property sales are shown in Figure 1. After processing, the data we use for analysis includes 50,344 recorded property transfers.

3.1 Labeling Property Transfers for Race and Identity

To study how home sales are affected by race, we first attempt to identify the race of each individual home seller and buyer in our dataset. Similar to work from Ali et al. [2] and Speicher et al. [37], we use race data from the North Carolina State Board of Elections (NCSBE) to do so. The demographic information in this dataset was originally collected upon registration for voting, and includes the voter's residential address, age, race, ethnicity, and gender, where the latter three were specified using optional discrete-choice selection boxes.

Of the humans identified as either property grantors (sellers) or grantees (buyers) in our property transfer dataset, approximately 47% were able to be matched to a voter record. Because property transfers often involve more than one buyer and more than one seller, buyer/seller groups with a single recorded race are identified as having that race, while buyer/seller groups with more than one recorded race are simply identified as having "multiple" races. Further details of the joining process are provided in Appendix A. For each transaction, we categorize the seller(s) as Black if all individual sellers associated with the transaction that could be joined to the voter data self-identified as Black/African American, White if all self-identified as White, Other if all self-identified as another race,[7] Multiple if the individual home sellers identified as different races, and Unknown if no individual home sellers could be joined to the voter data.

To determine how race affects sales made to iBuyers in a different way than it affects sales made to other purchasers, we also separate iBuyers from other types of entities. We follow Buchak et al. [9] and Harrison et al. [20] in identifying iBuyers using string matching. We classify any entities whose name contains the name of an iBuying affiliate as an iBuyer, where our list of iBuying affiliates is taken from public filings and subsidiary lists given on websites

[6] iBuying was available starting in 2017 in Charlotte, but only made up 0.01% of the market that year, compared to 3% or more in 2018 through 2021 [35].

[7] We choose to group people self-identifying as other races together both to center our work on discrimination against Black people, who have historically been affected worst by housing discrimination in the United States [4, 33], as well as due to the low frequency of people self-identifying as any other race in our dataset (around 3% of sellers, at the transaction-level) which would not have allowed for precise effects to be estimated.

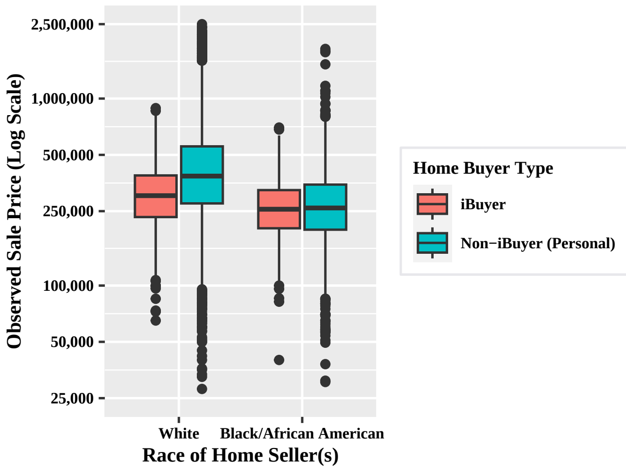


Figure 2: Observed sale prices for homes sold from White or Black/African American voters. iBuyers (shown in red) tend to pay more for homes to White home sellers than Black home sellers. The disparity for personal purchasers is more exaggerated. (Plot omits homes sold for more than $2,500,000 or less than $25,000.)

for iBuyers identified by Seiler and Yang [35]: Opendoor, Offerpad, Ribbon, Knock, Perch, RedfinNow, and Zillow Offers, along with an additional company, Orchard.[8] After string matching, we manually inspect entities, relabeling 38 which we believe were incorrectly labeled, (for example, for "Orchard" we relabel people with last name "Orchard" or entities such as "Orchard Terrace Estates, LLC" as non-iBuyers). We further classify non-iBuyers as being either personal or institutional, where personal entities include humans, executors, and personal trusts, and institutional entities include limited partnerships, limited liability corporations, and banks, (for example *Progress Residential*, or *Cerberus SFR Holdings*).

3.2 Control Variables

The United States has a long history of discriminatory housing practices, both at the government and individual levels, and homes owned by Black and White people in present day America tend to differ from each other in many ways [8, 15, 33, 34]. When examining racial disparities in home values or home value appreciation, previous work has largely tried to separate disparities that are attributable to characteristics of the home and neighborhood from disparities that are instead directly attributable to the race of the home's owner or the demographic makeup of the neighborhood the home is in [6, 21–23, 28, 30]. Control variables used in related work include those summarizing:

- **property characteristics**, like home age [6, 21–23, 28, 30], number of rooms [21–23, 28, 30], square footage [22, 28], or number of detached units [23, 28, 30],
- **economic conditions** in the neighborhood, like the typical commute time [6, 21, 22, 30], unemployment rate or income level [6, 21, 22],
- **real estate conditions** in the neighborhood, like the vacancy rate [21, 22] or owner occupancy rate [22, 30],
- **amenities** in the neighborhood, like the distance to parks [22] or number of libraries [30].

Like others, we control for a variety of features concerning the home and neighborhood in our work. Our data on **property characteristics** comes from the Mecklenburg County Assessor's Office, and includes 14 unique variables. We use the number of bedrooms/bathrooms in the the home, home age, and home square footage, however do not have access to the number of detached units for each home in our dataset. We also include additional characteristics not given above, like the heat source and foundation material of the home, or the quality of craftsmanship and construction materials as assessed by the county.

The data we use on neighborhood characteristics come from the Charlotte/Mecklenburg Quality of Life (QOL) Explorer, a public website that repackages data from sources such as the U.S. Census or Charlotte-Mecklenburg Schools, at the Neighborhood Profile Area (NPA) level. (NPAs are small geographic regions defined as a part of the QOL Explorer project, of which there are over 400 in the county.) We use variables summarizing:

- **economic conditions**, like the number of jobs per acre, foreclosure rate, or percent of land that is vacant,
- **real estate conditions**, like the percentage of single family residential housing units, percentage of units permitted for demolition, or percentage of single family residential units that are rentals,
- **amenities**, like the average proximity to parks or early childhood care facilities, or prevalence of library cards in the area,
- **school quality**, like the average proficiency of elementary and middle school students on standardized assessments or the rate of absenteeism in local schools,
- **other qualities**, like the population density, property crime rate, or average age of death.

We use 42 neighborhood-level variables in total. To ensure that we differentiate between differences in sales price that are attributable to the demographics of the homeowners and those that are attributable to the demographic makeup of the home's neighborhood, we also include the proportion of the population in the neighborhood that is Asian, White, Hispanic/Latino, and Black, respectively. A full list of home and neighborhood variables, as well as their descriptions, is provided with the code for this project.

4 PRICE DISPARITIES

To understand the extent to which iBuyers tend to pay equitable prices for homes bought from Black and White homeowners, we start by simply considering the raw prices paid to homeowners of different races. As shown in Figure 2, we find that iBuyers do tend to pay less to Black homeowners: The median sales price for homes sold from Black homeowners to iBuyers is $256,000, compared to $302,500 for those sold from White homeowners. In comparison, however, the gap between sale prices for homes sold to personal

[8]The full list of subsidiaries and affiliates consists of over 300 names, and is available as part of the code for this work.

purchases is over twice as large: For personal purchases, the median amounts paid are $260,000 and $385,000, respectively.

Due to historical housing disparities between Black and White homeowners [8, 15, 33, 34], these differences in raw prices are not necessarily due to any racialized differences in the decisions of personal purchasers or iBuyers. For example, it may be the case that they merely reflect the fact that the homes sold by Black homeowners tend to be smaller and have fewer bathrooms,[9] two variables tied to lower sale prices by Zietz et al. [41]. To assess this, we fit Bayesian hierarchical models that consider the sale price of a transaction to be dependent on the race of the seller and the identity of the buyer, while controlling for features of the home and neighborhood. We now describe the models that we use.

4.1 Sale Price Modeling Approach

Our primary model predicts sale prices based on the identities of the sellers and buyers, while controlling for features of the home and neighborhood. The model treats sale price as a random variable Y_i, drawn from a normal distribution whose mean is a linear function of the explanatory variables we consider. The model is given by:

$$Y_i \sim N(\mu = \beta_1 X_i + \gamma_{j[i]} + \beta_2 R_i + \beta_3 B_i + \beta_4 R_i B_i, \sigma^2 = \sigma_y^2)$$

where i is an index for the individual home sale, j is an index for NPAs (and $j[i]$ refers to the NPA in which transaction i occurs), X_i is a vector of variables characterizing the home and neighborhood as described in Section 3 (also including an overall intercept and a set of dummy variables corresponding to the year in which the home was sold), $\gamma_{j[i]}$ is a random intercept, which differs between different Neighborhood Profile Areas, R_i is a set of dummy variables corresponding to the race of the seller,[10] B_i is a set of dummy variables corresponding to the type of entity purchasing the property,[11] σ_y^2 is the variance of the random variable, which we treat as constant, and the βs are coefficients for these variables.

The model we use is hierarchical, by which we mean that beyond treating sale price Y_i as varying randomly, it also treats the explanatory variable $\gamma_{j[i]}$ as random as well, drawn from its own normal distribution. The model is given by:

$$\gamma_{j[i]} \sim N(\mu = \mu_\gamma, \sigma^2 = \sigma_\gamma^2)$$

where μ_γ and σ_γ^2 refer to a constant overall mean and variance specific to this variable. Hierarchical models are frequently used in regression analyses that involve categorical variables with a wide range of possible values, as by aggregating information across various categories they can offer more precise estimates compared to models that are fit separately for each individual category. We elect to use Bayesian models in our analysis as we consider them to be easier to interpret than equivalent frequentist versions. Fitting these models involves generating a large number of plausible coefficient values, which we then summarize to give both a best guess for what the coefficients might be, as well as a range of other credible values. We fit our model using the R package rstanarm 2.26.1 with default priors.

4.2 Sale Price Results

Based on the model described in Section 4.1, we find that not all differences in sale prices are attributable to features of the home and neighborhood: the identity of the seller and of the buyer both appear to contribute to sale price as well. As shown in Figure 3b, using a transfer from a Black homeowner to an iBuyer as a baseline, transfers involving homes with identical features and neighborhood characteristics that are sold by White homeowners earn higher prices on average by around $4,436 (95% Credible Interval: -$4,332, $13,236). In other words, our best guess is that, on average, for White homeowners and Black homeowners selling homes to iBuyers (with the same number of bedrooms, square footage, neighborhood crime rate, etc.) we would expect the homes from White homeowners to sell for an average of around $4,436 more. However based on our model, data, and priors, this is not the only plausible value for this coefficient. There is also a 95% chance that the difference would be between -$4,332 and $13,236. In fact we estimate the probability that the difference is in the other direction and that Black homeowners would earn more on average, to be around 16%.

The difference between the sale prices that White and Black homeowners receive when selling to personal purchasers appears to be much larger. We estimate the race gap when selling to personal purchasers to be $36,051 in favor of White sellers (95% Credible Interval: $29,380, $42,675). Again, based on the data, model, and priors, our best guess is that homes transferred to personal purchasers by White homeowners would be sold for an average of $36,051 more than equivalent homes sold by Black homeowners, with a 95% probability that the average is between $29,380 and $42,675. Using our model we can directly estimate the difference between these race gaps: our best guess is that the race gap when selling to personal purchasers is $31,615 higher than the race gap when selling to iBuyers (95% Credible Interval: $20,886, $42,277).

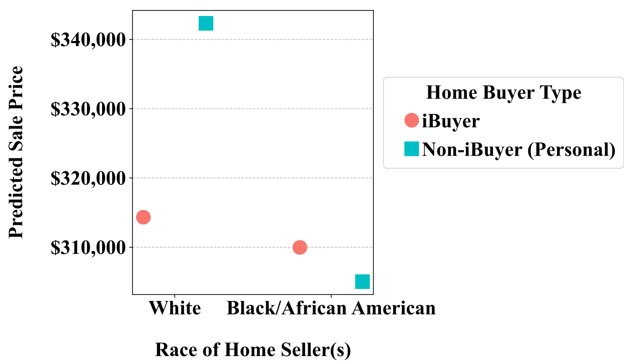
To put these figures in other terms, we can estimate the average acquisition discount for different sellers—the average amount that sellers would have their homes' sale prices lowered by if they sold to iBuyers rather than to a personal purchasers. For White sellers, we estimate the average acquisition discount to be $27,239 (95% Credible Interval: $21,472, $33,002). On the other hand, we estimate the acquisition discount for Black sellers to be -$4,376 (95% Credible Interval: -$13,445, $4,713): Our best estimate is that Black homeowners in fact earn *more* on average from selling to iBuyers than selling to individual purchasers, although as seen in the credible interval bounds, we also estimate that there is a more than 5% chance that this is not the case. The model we fit has a Bayesian R^2 of 0.78 [18], meaning that it explains around 78% of the variance in sale prices, a value we consider to be fairly good for this context.[12]

[9] The mean numbers of full bathrooms in our dataset for White and Black homeowners are 2.26 and 2.09, respectively, and the mean square footages are 2,987 and 2,493, respectively.

[10] As described in Section 3.1, this race can either be Black, White, Other (if the homeowners are all identified as another race), Multiple (if the homeowners have different races from one another), or Unknown (if no homeowners in the group could be identified for race).

[11] As described in Section 3.1, this entity type can either be iBuyer, Non-iBuyer (Personal), or Non-iBuyer (Institutional).

[12] Bayesian R^2 values typically range between 0 and 1, where an R^2 of 1 would indicate that a model is able to use the explanatory variables to exactly predict the target variable. While there is no universal cutoff for what a "good" R^2 is, the value we report is larger than or on par with R^2 values reported in similar work, e.g. the 0.23 from Fout et al. [16] or the 0.81 given by Howell and Korver-Glenn [22].



(a) Mean posterior predicted sale price for typical home in dataset. Personal purchasers pay significantly more for equivalent home when buying from White voter. iBuyers pay somewhat more for homes bought from White voter. (Typical home is one with median value for all numeric variables and modal value for all categorical values, e.g. 3 bedrooms and average grade.)

(b) For homes sold from Black/African American voters to iBuyers, the estimated amount that the sale prices would change if the homes had instead been sold by different buyer or to different seller. Neighborhood and property characteristics held constant. Shown with 95% Credible Intervals.

Figure 3: Model estimates for prices in Mecklenburg. Based on the model, data, and priors, there is a more than 95% probability that the average amount that White voters earn when selling to personal buyers is greater than the average amount they earn when selling equivalent homes to iBuyers. There is a greater than 50% probability that the average amount that Black/African American voters earn when selling to personal buyers is *lower* than the average amount they earn when selling equivalent homes to iBuyers (although less than 95% probability).

A natural question to ask concerns the extent to which the model we have fit is correctly controlling for features of the home and neighborhood, and if rather than being attributable to seller race and buyer identity, the differences we've observed are in fact due to a feature we have not considered. For example, it is possible that homes sold from White homeowners tend to have pools more frequently than those sold by Black homeowners, a variable we have not included in our model, which has previously been found to be associated with sales price [41]. If this were the case, and if pools did in fact result in higher sales prices in Mecklenburg, it would lead our model to believe that differences in price were attributable to the race of the sellers, while in reality they were more directly attributable to the property itself. Similarly, it is possible that the homes sold to personal purchasers tend to differ more in this pools variable than homes sold to iBuyers, accounting for the differing disparities between iBuyers and personal purchasers.

To test whether there may be relevant housing or neighborhood features missing from our analysis, we also fit a model that uses assessed values, rather than features of the home and neighborhood, to predict prices.[13] The assessed values we use come from the Mecklenburg County Assessor, which estimates the market value for homes across the county in order to determine property taxes owed for each home. Assessed values are determined based on comparative sales, market trends, as well as individual home site visits, and we believe they are more likely to accurately predict sale prices than features of the home and neighborhood alone (as well as to be reflective of features, like kitchen counter-tops, that our main model has missed). Mecklenburg performs a county wide revaluation of all homes every four years, and individual homes may be revalued more frequently, for example if they add new features. Along with assessed value, we also include a variable indicating the time in between when the assessment occurred and when the home was sold.

When fitting the model based on assessed values, we find a similar story as when fitting a model based on features of the home and neighborhood. We estimate the racial sales price gap for iBuyers to be -$2,220 (95% Credible Interval: -$8,543, $4,064). This means that if we compare homes sold to iBuyers by White homeowners with homes with the same assessed values sold by Black homeowners, our best guess is that on average, the homes sold by Black homeowners would in fact be sold for $2,220 more, (this time with a 24% chance that the difference is in the other direction, and that those sold by White homeowners would earn more). For personal buyers, we estimate the racial sales price gap (when controlling for assessed value, rather than features of the home) to be smaller than before, however in the same direction: $14,212, (95% Credible Interval: $9,521, $18,908). As expected, the assessment based model is better able to predict sale prices than our main model, achieving a Bayesian R^2 of 0.88 [18].

While we are not aware of any recent work that estimates a gap in sale prices between homes sold by Black and White homeowners to which we can compare, we note that the effect we have found is consistent with what we would expect given similar work on appraised values and home appreciation. Using a national dataset of home valuations and appraisals from 2016 to 2020, for example, Rothwell and Perry [34] find that homes in neighborhoods that are

[13]We also provide further validation of the housing-and-neighborhood-based sale price model in Appendix B.

50% Black are valued more than 20% lower than homes in neighborhoods that are 0% Black, even after controlling for a wide variety of neighborhood-level features. Howell and Korver-Glenn [22] also use a variety of controls, and estimate that for homes sold in Harris County, Texas in 2015 (home to Houston) the average home in a White neighborhood would be sold for $289,000, compared to $127,000 if the same home were sold in a Black neighborhood. Given these estimates, we do not find the direction or size of the effect we have seen to be surprising. We also hypothesize that there may be biases in the assessed values themselves, similar to those that have been found in appraised values, which may account for the lower magnitude in the estimates from the assessment-based model than in the housing characteristics-based model.

5 CONVERSION TO INSTITUTIONAL OWNERSHIP

To help illuminate the effects of iBuying on neighborhood demographics and home ownership, we start by simply comparing the demographics of the groups that iBuyers are buying from to those they are selling to. This helps show whether, in aggregate, iBuyers tend to be facilitating the transfer of ownership between people of different identity groups. As a baseline, we compare iBuyers to direct sales.

We find that on average, those selling with iBuyers appear to be slightly younger than those selling directly to other purchasers. The average age for homeowner groups selling to iBuyers in Mecklenburg in this time period is 47.4, compared to 50.2 for those selling to non-iBuyers. Similarly, we find the homeowner groups who buy from iBuyers to be slightly younger than those who buy directly: 39.5 years compared to 41.1 years, respectively.[14] We also note that it appears iBuying is more popular with Black homeowners than with White homeowners. We find that 23% of sales to iBuyers started with Black homeowners. In comparison, 9% of direct sales start with Black homeowners. When it comes to purchasing, we find that 9% of sales intermediated through iBuyers end with Black homeowners, while 6% of direct sales end with Black homeowners.

Beyond differences in the demographic groups that iBuyers are buying from and selling to, we also find that the companies are taking part in a more basic change: the conversion of properties from personal to institutional ownership. Institutional investment in SFR (single-family rental) properties has risen in the United States since the great recession, with institutional SFR investors now owning a sizeable proportion of available homes in some markets [11]. Institutional SFR investors often operate as large scale landlords, and appear to be associated with difficult tenant experience, high levels of eviction, and high levels of rent inflation [14, 31]. In Figure 4, we show the percent of property transfers originating with Black and White homeowners that either end with personal purchasers or are converted to institutional ownership, both for direct transfers and for property transfers intermediated through iBuyers. We find that for homes sold by Black homeowners, 36% of those intermediated through iBuyers go on to be sold to institutional purchasers (compared to 33% of direct sales). For homes sold by White homeowners, 17% of those intermediated through iBuyers go on to be sold to institutional purchasers (compared to 9% of direct sales).

Due to the harms associated with institutional investment in SFR homes, we choose to further explore the relationship between iBuyers and institutional purchasers. As with sale prices, it is possible that, for example, Black-owned homes are for some reason more desirable to institutional purchasers, and that differences in Figure 4 between Black-owned homes and White-owned homes are more attributable to differences in the homes themselves than the race of the sellers. For this reason we again decide to perform an analysis that controls for features of the home and neighborhood.

5.1 Conversion Modeling Approach

Our model for property conversion estimates the probability that a property transfer will end with an institutional purchaser, based whether or not the transfer is intermediated through an iBuyer, as well as based on the race of the original seller. We again control for features of the home and neighborhood. We fit a Bayesian hierarchical logistic model, which treats the final owner of the home as a Bernoulli random variable Z_i taking on the value 1 if the transfer ends with an institutional buyer, and 0 otherwise. The probability that $Z_i = 1$ (that the home will end with an institutional buyer) is modeled as a function of the other variables we consider.

$$Z_i \sim Bernoulli(p = logit^{-1}(\beta_1 X_i + \gamma_{j[i]} + \beta_2 R_i + \beta_3 I_i + \beta_4 R_i I_i))$$

where i is an index for the individual home sale, j is an index for NPAs (and $j[i]$ refers to the NPA in which transaction i occurs), X_i is a vector of variables characterizing the home and neighborhood as described in Section 3 (also including an overall intercept and a set of dummy variables corresponding to the year in which the home was sold), R_i is a set of dummy variables corresponding to the race of the seller,[15] I_i is a dummy variable indicating whether the sale has been intermediated through an iBuyer, and the βs are regression coefficients for these variables. As before, the model includes a set of random effects $\gamma_{j[i]}$ that vary by neighborhood, and which have constant mean and variance:

$$\gamma_{j[i]} \sim N(\mu = \mu_\gamma, \sigma^2 = \sigma_\gamma^2)$$

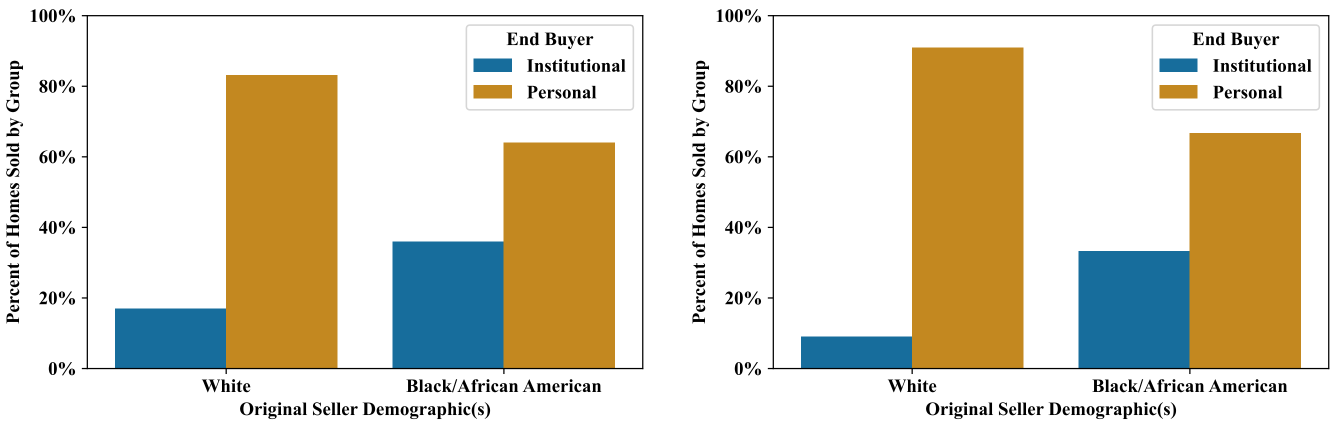
We omit sales where the original seller or final owner is an iBuyer from the conversion model.

5.2 Conversion Results

As is common with logistic regressions, we choose to interpret our model in terms of odds: in our case, the ratio between the probability that a home is sold to an institutional purchaser and the probability that a home is instead sold to a personal purchaser. Odds can range between 0 and infinity, where numbers above one indicate that the home is more likely to be sold to an institutional purchaser than a personal purchaser. As an example, an odds of 3 would indicate that a home is three times as likely to be sold to an institutional purchaser as to be sold to a personal purchaser—that there is a 3/4

[14] The age for a homeowner group is the average age for all homeowners in the group whose ages had been recorded. A group is the full set of homeowners listed on a deed, e.g. a couple who own a home together.

[15] As described in Section 3.1, this race can either be Black, White, Other (if the homeowners are all identified as another race), Multiple (if the homeowners have different races from one another), or Unknown (if no homeowners in the group could be identified for race).



(a) Property transfers mediated by iBuyers **(b) Direct transfers**

Figure 4: Observed percent of property transfers from White and Black sellers that end with institutional purchasers. Sales mediated through iBuyers, where the home is first sold to an iBuyer then sold to another entity, are more likely than direct sales to end with institutional purchasers. Sales from Black homeowners are also more likely to end with institutional purchasers than sales from White homeowners. Figures exclude sales where iBuyers are the original seller or end buyer.

probability that this will occur, compared to a 1/4 probability that it will not.

Based on the model described in Section 5.1, we find evidence that even when controlling for features of the home and neighborhood, homes sold by Black homeowners remain much more likely to be sold to institutional purchasers than homes sold by White homeowners. Considering homes sold directly from White homeowners, the odds of ending with an institutional purchaser would increase by a factor of 2.34 if the homes had instead been sold by Black homeowners (95% Credible Interval: 2.08, 2.64). This means that, in aggregate, even when controlling for a wide variety of home and neighborhood features, we expect homes bought from Black homeowners to go on to be sold to institutional purchasers at a much higher rate than homes bought from White homeowners.

We estimate that the predicted probability the typical home in our dataset would be sold to an institutional purchaser, if it were sold directly from a White homeowner, is 0.24.[16] The probability for a home with an identical number of bedrooms, square footage, neighborhood crime rate, etc., if sold directly by a Black homeowner, would be approximately 0.43.

Now considering the effect of iBuying, for Black homeowners, we do not find evidence that selling through an iBuyer changes the probability that homes will end with institutional purchasers. We estimate that when considering homes sold directly from Black homeowners, the odds of ending with an institutional purchaser would change by a factor of 0.97, if they were instead sold through iBuyers (95% Credible Interval: 0.83, 1.15). The fact that the credible interval is centered near 1 indicates that we do not have evidence that iBuying either decreases or increases the probability of Black homeowners' homes ending with institutional purchasers.

We do find evidence that for White homeowners, however, sales through iBuyers are more likely than direct sales to end with institutional purchasers. For White homeowners, we estimate that the odds of direct sales ending with an institutional purchaser would change by a factor 0f 1.33 (95% Credible Interval: 1.15, 1.54) if the homes had instead been sold through an iBuyer. Again, the predicted probability that the typical home in our dataset would be sold to an institutional purchaser, if sold directly by a White homeowner, is 0.24. In comparison, the predicted probability that the typical home in our dataset would end with an institutional purchaser, if sold through an iBuyer by a White homeowner, would be 0.31.

As with sale prices, we also fit an alternative model with assessment data and find similar results. All factors whose credible intervals did not include one (and so had either a more than 95% probability of increasing the odds or a more than 95% probability of decreasing the odds, depending on if their credible interval was above one or below one, respectively), still have effects in the same direction with credible intervals not including one.[17]

[16] Where typical is again defined as a home with median values for all numeric variables and modal values for all categorical variables.

[17] In the assessment based model, predicted probabilities of a typical property ending in institutional ownership are lower than in the model that controls for features of the home and neighborhood. The predicted probability for a White homeowner selling directly changes from 0.24 in the home/neighborhood model to 0.11 in the assessment model, for a Black homeowner selling directly changes from 0.43 to 0.24, and for a White homeowner selling through an iBuyer changes from 0.31 to 0.16. We suspect these differences may be related to differences in what the models consider a typical home to be—we define "typical" as a home with median values for all numeric variables and modal value for all categorical variables, and we suspect that the median/modal home in terms of number of bedrooms, square footage, and other housing and neighborhood features is not equivalent to the median/modal home in terms of assessed value. Odds ratios remain fairly similar between the models: The odds ratio between direct sales from White homeowners and direct sales from Black homeowners changes from 2.34 (95% Credible Interval: 2.08, 2.64) to 2.61 (95% Credible Interval: 2.33, 2.93); the odds ratio between direct sales from Black homeowners and iBuyer-mediated sales from Black homeowners changes from 0.97 (95% Credible Interval: 0.83, 1.15) to 1.12 (95% Credible Interval: 0.96, 1.32); and the odds ratio between direct sales from White homeowners and iBuyer-mediated sales from White homeowners changes from 1.33 (95% Credible Interval: 1.15, 1.54) to 1.56 (95% Credible Interval: 1.36, 1.80). We also evaluate the housing-and-neighborhood-based conversion model in Appendix B.

6 DISCUSSION AND CONCLUSION

Our results suggest that iBuyers close a sales price gap between Black and White home sellers, but do so by paying less money for White-owned homes. Controlling for over 50 variables (such as the size of the homes, or the crime rate and racial makeups in the containing neighborhood) we find that the racial sales price gap shrinks from $36,051 to just $4,436 when homes are purchased by an iBuyer rather than a personal purchaser. To put these results another way—when controlling for variables like home age, grade, and proximity to grocery stores, an iBuyer will pay around $27,000 less for a White-owned home than a personal buyer, and around $4,000 more for a Black-owned home than a personal buyer.

At face value, this appears to show that iBuyers' AVMs have found a strategy to significantly underpay for White-owned homes. Previous work has found that iBuyers, on average, do pay lower prices than individual owner-occupiers [35]. However, Seiler and Yang [35] also suggests that iBuyer sales are different in character from foreclosures and other distressed home sales, where homes are offloaded for low prices by sellers during times of financial distress; rather that these lower prices are instead attributable to impatient or motivated sellers. On this note, Seiler and Yang [35] conclude that an iBuyer's acquisition discount varies across market conditions and is "largely dependent on algorithm pricing accuracy and local demand for liquidation" [35]. Based on these results, our work would suggest that a 'demand for liquidation' is favorable for iBuyers when a homeowner is White.

While we do not have data that captures the extent or costs of any repairs undertaken by iBuyers, future work could investigate whether iBuyers are in fact underpaying for White-owned homes by considering the returns they earn. If iBuyers are indeed underpaying White homeowners, we would expect the properties sold to them by White homeowners to go on to earn large returns. If on the other hand the observed differences are in fact due to personal purchasers overpaying for White-owned homes, indicating that differences in sale prices are more attributable to a 'White privilege effect' being paid by personal purchasers, we would expect the markups to be smaller. The difference between iBuyers underpaying and personal buyers overpaying would speak to the nature of the effect, and to who is making money off it, but regardless, the aggregate harm to these White home sellers remains clear—an average of $27,239 lost by selling to iBuyers.

Despite the differences in averages we have shown, we also note that the effect on individual sellers requires further study. While we find that, on average, Black home sellers may earn more from selling to iBuyers than to personal purchasers, this does not necessarily mean that any given Black homeowner would be better off selling their home to an iBuyer, for example. One concern we have is that the lack of a difference between the prices being paid for Black homes by iBuyers and by personal purchasers is due to a higher overall rate of distressed home sales among Black sellers. Work by Kermani and Wong [25] suggests that differences in returns earned by Black and White home sellers may be partially attributable to differences in distressed sales: that the rate of distressed sales is higher among Black sellers than White sellers. Because distressed sales often result in direct sales to personal purchasers, it may be the case that the average prices we see for sales from Black homeowners to personal purchasers are so low because they reflect these distressed sales. If this were the case, it might also be the case that, like White homeowners, Black homeowners not taking part in distressed sales would tend to earn less money from selling to iBuyers as well.

Beyond this, the long-term impacts on sellers and neighborhoods remain unclear. Harrison et al. [20] find that the entrance of iBuyers into a region increases home prices by around 3%, indicating that sellers may have difficulty purchasing new homes once they have made an initial sale. It may also be the case that iBuying affects the long-term demographic composition of a neighborhood. Regardless, the crowding out of local buyers is an issue that should raise red flags for policymakers who are charged with protecting public interests in markets, including those susceptible to corporate capture like housing. Our findings indicate that iBuyers are funneling homes to institutional investors at a higher rate than comparable direct sales, and this should similarly raise concern for local and state policymakers.

In previous work, An [5] shows that in Atlanta, Georgia, institutional investors' large-scale acquisitions dampened home ownership, and this particularly harmed Black households. Similarly, Coven [12] shows that the entrance of institutional investors make it significantly harder to buy homes in USA suburbs, and increases effective rental prices throughout a region. Future research might further investigate institutional investors and their acquisition strategies based on race. Seymour et al. [36] find that corporate investors favor white neighborhoods while private equity firms have targeted Black homeowners. More accurately establishing how institutional investors acquire properties (e.g. direct purchase or through iBuyers) might better account for observed racial differences and suggest potential correctives.

ETHICAL CONSIDERATIONS

We consulted our university's Internal Review Board before launching our study, and due to the public nature of the dataset, it is exempt from IRB approval. Below, we note some of the considerations that were made when collecting and using data for this work.

As discussed by Buck and Ralston [10], a key question when dealing with "public" data is the degree to which the information contained is in fact freely and easily available. As of submission, the data we use is accessible to anyone with an internet connection and featured prominently on local government websites. With that said, the data is not available in the consolidated format that we have created: over 50,000 records showing people's names, races, and home sale prices together in one place. As such, we have chosen to restrict releasing the data publicly for fear that doing so may subject people found in the data to unintended harms, for example solicitation, or mass harassment from someone unhappy with how much Black homeowners earned from property sales.

In weighing the harms and benefits of conducting this research, we give high weight to the Belmont Principles, which state that research should be respectful of persons, beneficence, and justice. We believe that the benefits to persons, namely an improved public understanding of the behavior of iBuyers, significantly outweigh the risks that are posed, which we consider to be both of minimal likelihood and of minimal potential impact. Our decision to use a dataset of this sort follows similar decisions by previous FAccT authors, for example Franchi et al. [17], who rely on police dashboard camera data, or Abbasi et al. [1], who rely on the same set of public voter records that we use from North Carolina. To help mitigate risks of harassment and solicitation, we have released our dataset on Harvard Dataverse with restricted access. Access will be granted to any researcher who provides contact information and a brief description of their research plans.[18]

Another ethical consideration in the usage of any dataset is the extent to which the data accurately represents the target population, or whether drawing conclusions from the data constitutes erasure of certain subpopulations [10]. To ensure the accuracy of race information in our dataset, and that the race labels we use align with homeowners' own self-identifications, we chose to label race by joining to voter records. We performed this joining rather than using a name-based model due to established inaccuracies in separating Black and White names [26], however this also means that our results are most applicable specifically to Black and White voters. As mentioned in Appendix A, of the 52,544 property records that we attempted to join to voter records, 40,737 (around 78%) were successfully able to be labeled for race. With that said, sampling bias is certainly possible, and for example could occur in our sale price results if the prices paid by voters, when broken down by seller race and purchaser identity, tended to differ from the prices paid by non-voters, when broken down by these same categories. We would not be surprised if this were to occur—for example if there is more between-group heterogeneity in sale prices for non-voters than for voters, our work could be underestimating effects. We think that further explorations using race data from another source is worth consideration in the future.

Finally, we note that while we control for features of the home and neighborhood in our statistical models, we do not mean to imply that these features are in reality easily separable from race. Historical segregation has meant the homes owned by Black people tend to be smaller and older than those owned by White people [23]. While we attempt to delineate between historical discrimination and discrimination by individual actors in today's housing market using these adjustments, we do not mean to say that the impacts of historical discrimination are not still present and impactful.

ACKNOWLEDGMENTS

We would like to thank Gregg Colburn for providing feedback on our work, and Aviva Furman and Julia Alcaraz-Gamez for labeling data. This work was funded, in part, by a grant from New America's program on Public Interest Technology (grant no. 015906), and the University of Washington's Information School. Data processing was facilitated by the advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington. ChatGPT 4 was used to create initial drafts for some paragraphs in the paper's introduction, as well as to provide feedback on clarity for Sections 4 and 5.

[18]The data is available at https://doi.org/10.7910/DVN/3DS7BT.

REFERENCES

[1] Mohsen Abbasi, Calvin Barrett, Kristian Lum, Sorelle A. Friedler, and Suresh Venkatasubramanian. 2023. Measuring and mitigating voting access disparities: a study of race and polling locations in Florida and North Carolina. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1038–1048. https://doi.org/10.1145/3593013.3594061

[2] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 199:1–199:30. https://doi.org/10.1145/3359301

[3] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow, Julien Launay, Quentin Malartic, Daniele Mazzotta, Badreddine Noune, Baptiste Pannier, and Guilherme Penedo. 2023. The Falcon Series of Open Language Models. https://arxiv.org/abs/2311.16867v2

[4] Brent W Ambrose, James N Conklin, N Edward Coulson, Moussa Diop, and Luis A Lopez. 2021. Do Appraiser and Borrower Race Affect Valuation? https://ssrn.com/abstract=3951587

[5] Brian Y. An. 2023. The Influence of Institutional Single-Family Rental Investors on Homeownership: Who Gets Targeted and Pushed Out of the Local Market? *Journal of Planning Education and Research* (June 2023), 0739456X231176072. https://doi.org/10.1177/0739456X231176072 Publisher: SAGE Publications Inc.

[6] Katrin B. Anacker. 2010. Still paying the race tax? Analyzing property values in homogeneous and mixed-race suburbs. *Journal of Urban Affairs* 32, 1 (Feb. 2010), 55–77. https://doi.org/10.1111/j.1467-9906.2008.00437.x

[7] Jackson T. Anderson, Franz Fuerst, Richard B. Peiser, and Michael J. Seiler. 2023. iBuyer's Use of PropTech to Make Large-Scale Cash Offers. *Journal of Real Estate Research* 0, 0 (2023), 1–22. https://doi.org/10.1080/08965803.2023.2214467

[8] Patrick Bayer, Marcus Casey, Fernando Ferreira, and Robert McMillan. 2017. Racial and ethnic price differentials in the housing market. *Journal of Urban Economics* 102 (2017), 91–105. https://doi.org/10.1016/j.jue.2017.07.004

[9] Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2020. Why is Intermediating Houses so Difficult? Evidence from iBuyers. https://doi.org/10.3386/w28252

[10] Amber M. Buck and Devon F. Ralston. 2021. I didn't sign up for your research study: The ethics of using "public" data. *Computers and Composition* 61 (Sept. 2021), 102655. https://doi.org/10.1016/j.compcom.2021.102655

[11] Gregg Colburn, Rebecca J. Walter, and Deirdre Pfeiffer. 2021. Capitalizing on Collapse: An Analysis of Institutional Single-Family Rental Investors. *Urban Affairs Review* 57, 6 (Nov. 2021), 1590–1625. https://doi.org/10.1177/1078087420922910

[12] Joshua Coven. 2023. The Impact of Institutional Investors on Homeownership and Neighborhood Access. https://doi.org/10.2139/ssrn.4554831

[13] Craig A. Depken, Harris Hollans, and Steve Swidler. 2009. An Empirical Analysis of Residential Property Flipping. *The Journal of Real Estate Finance and Economics* 39, 3 (Oct. 2009), 248–263. https://doi.org/10.1007/s11146-009-9181-3

[14] Liam Dillon. 2024. Nation's largest single-family home landlord to pay $3.7 million in California rent-gouging case. https://www.latimes.com/homeless-housing/story/2024-01-08/invitation-homes-tenant-rent-gouging-settlement

[15] Jacob W. Faber and Ingrid Gould Ellen. 2016. Race and the Housing Cycle: Differences in Home Equity Trends Among Long-Term Homeowners. *Housing Policy Debate* 26, 3 (May 2016), 456–473. https://doi.org/10.1080/10511482.2015.1128959

[16] Hamilton Fout, Nuno Mota, and Eric Rosenblatt. 2022. When Appraisers Go Low, Contracts Go Lower: The Impact of Expert Opinions on Transaction Prices. *The Journal of Real Estate Finance and Economics* 65, 3 (2022), 451–491. https://doi.org/10.1007/s11146-020-09800-6

[17] Matt Franchi, J.D. Zamfirescu-Pereira, Wendy Ju, and Emma Pierson. 2023. Detecting disparities in police deployments using dashcam data. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 534–544. https://doi.org/10.1145/3593013.3594020

[18] Andrew Gelman, Ben Goodrich, Jonah Gabry, and Aki Vehtari. 2019. R-squared for Bayesian Regression Models. *The American Statistician* 73, 3 (July 2019), 307–309. https://doi.org/10.1080/00031305.2018.1549100

[19] Laurie S. Goodman and Christopher Mayer. 2018. Homeownership and the American Dream. *Journal of Economic Perspectives* 32, 1 (Feb. 2018), 31–58. https://doi.org/10.1257/jep.32.1.31

[20] David M. Harrison, Michael J. Seiler, and Liuming Yang. 2023. The Impact of iBuyers on Housing Market Dynamics. *The Journal of Real Estate Finance and Economics* (May 2023). https://doi.org/10.1007/s11146-023-09954-z

[21] John R. Hipp and Amrita Singh. 2014. Changing Neighborhood Determinants of Housing Price Trends in Southern California, 1960-2009. *City and Community* 13, 3 (2014), 254–274. https://doi.org/10.1111/cico.12071

[22] Junia Howell and Elizabeth Korver-Glenn. 2018. Neighborhoods, Race, and the Twenty-first-century Housing Appraisal Industry. *Sociology of Race and Ethnicity*

4, 4 (Oct. 2018), 473–490. https://doi.org/10.1177/2332649218755178
[23] Junia Howell and Elizabeth Korver-Glenn. 2021. The Increasing Effect of Neighborhood Racial Composition on Housing Values, 1980–2015. *Social Problems* 68, 4 (Oct. 2021), 1051–1071. https://doi.org/10.1093/socpro/spaa033
[24] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. LoRA: Low-Rank Adaptation of Large Language Models. https://arxiv.org/abs/2106.09685v2
[25] Amir Kermani and Francis Wong. 2021. *Racial Disparities in Housing Returns.* Technical Report. National Bureau of Economic Research. http://www.nber.org/papers/w29306
[26] Jeffrey W Lockhart, Molly M King, and Christin Munsch. 2023. Name-based demographic inference and the unequal distribution of misrecognition. *Nature Human Behaviour* 7, 7 (2023), 1084–1095. https://doi.org/10.1038/s41562-023-01587-9
[27] Guangli Lu. 2019. *How Machine Learning Mitigates Racial Bias in the U.S. Housing Market.* Technical Report. https://ssrn.com/abstract=3489519
[28] Richard Moye. 2014. Neighborhood racial-ethnic change and home value appreciation: Evidence from Philadelphia. *Urban Geography* 35, 2 (Feb. 2014), 236–263. https://doi.org/10.1080/02723638.2013.856194
[29] Michael Neal, Sarah Strochak, Linna Zhu, and Caitlin Young. 2020. *How Automated Valuation Models Can Disproportionately Affect Majority-Black Neighborhoods.* Technical Report. Urban Institute. Publication Title: Washington, DC: Urban Institute.
[30] Andre Perry, Jonathan Rothwell, and David Harshbarger. 2018. *The Devaluation of Assets in Black Neighborhoods.* Technical Report.
[31] Elora Lee Raymond, Richard Duckworth, Benjamin Miller, Michael Lucas, and Shiraj Pokharel. 2018. From Foreclosure to Eviction: Housing Insecurity in Corporate-Owned Single-Family Rentals. *Cityscape* 20, 3 (2018), 159–188. https://www.jstor.org/stable/26524878
[32] Zillow Research. 2022. Zillow Q1 2022 iBuyer Report. https://www.zillow.com/research/zillow-q1-2022-ibuyer-report-31105/
[33] Richard Rothstein. 2017. *The color of law: A forgotten history of how our government segregated America* (1st ed.). Liveright Publishing Corporation, a division of W.W. Norton & Company, New York.
[34] Jonathan Rothwell and Andre Perry. 2022. How racial bias in appraisals affects the devaluation of homes in majority-Black neighborhoods. https://www.brookings.edu/articles/how-racial-bias-in-appraisals-affects-the-devaluation-of-homes-in-majority-black-neighborhoods/
[35] Michael J. Seiler and Liuming Yang. 2022. The burgeoning role of iBuyers in the housing market. *Real Estate Economics* (2022). https://doi.org/10.1111/1540-6229.12384
[36] Eric Seymour, Taylor Shelton, Stephen Averill Sherman, and Joshua Akers. 2023. The metropolitan and neighborhood geographies of REIT- and private equity–owned single-family rentals. *Journal of Urban Affairs* 0, 0 (2023), 1–25. https://doi.org/10.1080/07352166.2023.2276766
[37] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Potential for Discrimination in Online Targeted Advertising. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency.* PMLR, 5–19. https://proceedings.mlr.press/v81/speicher18a.html
[38] William E. Winkler. 1990. *String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage.* Technical Report. https://eric.ed.gov/?id=ED325505
[39] Antonia Zapf, Stefanie Castell, Lars Morawietz, and André Karch. 2016. Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology* 16, 1 (Aug. 2016), 93. https://doi.org/10.1186/s12874-016-0200-9
[40] Linna Zhu, Michael Neal, and Caitlin Young. 2022. *Revisiting Automated Valuation Model Disparities in Majority-Black Neighborhoods.* Technical Report. Urban Institute.
[41] Joachim Zietz, Emily Norman Zietz, and G Stacy Sirmans. 2008. Determinants of House Prices: A Quantile Regression Approach. *The Journal of Real Estate Finance and Economics* 37, 4 (2008), 317–333. https://doi.org/10.1007/s11146-007-9053-7

A FURTHER DETAILS ON DATA COLLECTION AND PROCESSING

The data that we use comes from a variety of public sources, and as such requires us to perform standard joining and pre-processing to prepare it for analysis. We start by scraping property transfers from the web portal for Mecklenburg's Recorder of Deeds. The portal allows one to search either by date or keyword. To ensure that we collected all iBuyer transactions, we search for the affiliates and subsidiaries list we had collected for iBuyers and download all deeds returned by the search. To gather information on non-iBuying transactions, we iteratively search over all days between January 1st of 2018 and October 31st of 2023, and download each deed returned by the search with probability 30%. Between the iBuying search and Bernoulli random sampling, we gather 78,062 deeds.

Each observation returned from the Recorder of Deeds contains names for the property grantor (seller) and property grantee (buyer) in text, along with a PDF file documenting other details of the transfer. The identification numbers for any parcels involved in a deed are contained only in the PDF documents. We use the Optical Character Recognition (OCR) library Tesseract to extract Parcel IDs from these PDF files, and are successfully able to read at least one Parcel ID for 76,898 property transfers.

The next data source we collect is property-level characteristics from the Mecklenburg County Assessor, such as square footage. This data is scraped from the Assessor's web portal. We search for each individual Parcel ID mentioned in each deed, and are able to find characteristics for parcels in 70,127 deeds. Based on manual inspection, we believe the most common reason deeds are not able to found is due to issues with the OCR engine. We next remove 14,683 deeds that had sale prices of $0.00 listed, leaving 55,444 records. Upon manual inspection, these instances largely appear to be transfers from people to personal trusts (often where the personal trust takes the person's name) or from people to family members. As such, we did not feel that they provide an appropriate point of comparison for sales to and from iBuyers. To further ensure iBuying and non-iBuying deeds are comparable, we also remove sheriff, trustees, quit-claim, corrected and commissioners deeds, any deeds not occurring within our specified time period, deeds that involve more than one parcel ID, and deeds not having a finished area listed, indicating vacant land. After these steps, we are left with 52,544 records.

Next, we collect race data for voters in Mecklenburg County. The NCSBE releases regular snapshots of their voter database, and because our list of deeds spans nearly six years, we utilize the complete set of snapshots occurring within this time frame. Each snapshot lists all active and inactive voters in the state at the time, along with residential addresses and optional ethnicity, race, and sex information provided by the voter. Because a person can appear in multiple snapshots, we elect to de-duplicate the dataset. When doing so, we handle cases where one name appears with multiple addresses by keeping both records, and cases where one name-address pair appears with multiple sets of demographic information by listing the changed piece of demographic information as "Unknown." After de-duplicating records and filtering to only Mecklenburg County, the voter dataset contains 2,003,414 name-address pairs.

When joining property transfers to the NCSBE data, we find that names and addresses listed on deeds often have clear similarities with voter records, despite not being exact matches. Common patterns we see are initials, nicknames, and middle names being used in one dataset but not the other, or addresses in the two datasets using alternative city names (e.g., "Charlotte" vs. "Unincorporated"). Rather than requiring identical strings between the datasets, we choose to instead perform entity matching to merge them. To do so, we start by fuzzy matching each name-address pair mentioned in a deed to the closest name-address pair in the NCSBE data, measured in Jaro-Winkler distance [38]. We then randomly select 1,000 records, (where each record consists of a name-address pair from the deed dataset along with the closest name-address pair from the voter dataset) and ask three annotators to label each record for whether they believe the name-address pairs from the two datasets refer to the same person. We find what we consider to be substantial inter-rater agreement: complete agreement between annotators for 91.1% of records, a Krippendorff's alpha value of 0.88 (95% Bootstrap CI: 0.85, 0.90), and a Fleiss's K value of 0.88 (95% Bootstrap CI: 0.85, 0.90) (sic) [39].

Using these labeled records, we fine-tune a classifier for identifying whether name-address pairs from the two datasets match. We use 700 records for fine-tuning, 100 for validation, and 200 for testing. We use the majority vote from the raters as the target for the model. The base model we fine-tune is a Falcon-7B [3] language model, which we adapt using Low-Rank Adaptation [24]. The model achieves 97% accuracy on the held out test set. After labeling each of the 164,897 humans mentioned in deeds using the fine-tuned model, we find 77,101 matches to voter records.

When joining back to the housing records, of the 52,544 housing records, we find that 17,648 property transfers have a buyer whose name could not be associated with race information and 11,807 have a seller whose name could not be associated with race information. If there are multiple buyers or multiple sellers who have different identified races involved in a transaction, we classify the buyer/seller as having "multiple" races. Otherwise the race of a buyer/seller is the race for all people whose races were able to be identified.

To gather neighborhood data, we use information from the Charlotte/Mecklenburg Quality of Life (QOL) Explorer. The QOL Explorer lists data for individual Neighborhood Profile Areas (NPAs), non-overlapping geographic regions of which there are over 400 in the county. Of the 52,544 records with both buyer and seller information, 52,211 could be geocoded and matched to an NPA. Finally, to further ensure that transactions from iBuyers and non-iBuyers are comparable, we filter out 1,867 records for housing units not similar to those bought and sold by iBuyers (but that had not previously been removed in our data processing), for example those involving sales of shopping centers or warehouses, or those that involve properties with more than 10,000 square feet. After these steps, we are left with 50,344 recorded transfers.

B MODEL VALIDATION

The models that we use rely on a number of statistical assumptions: for example in the main sale price model that the relationship between mean sale price and the number of bedrooms in a home is linear, or that sale prices are conditionally normally distributed. To validate our models, we turn to posterior predictive checking, a technique in Bayesian analysis whereby the outputs in observed data (for example the sale prices for the home sales we use in our models) are compared to a model's predictions for these observed data (the predicted sale price for the homes given by the models). This analysis gives a sense of types of data where the model is successfully able to map the relationship between inputs and output, as well as areas where the model struggles.

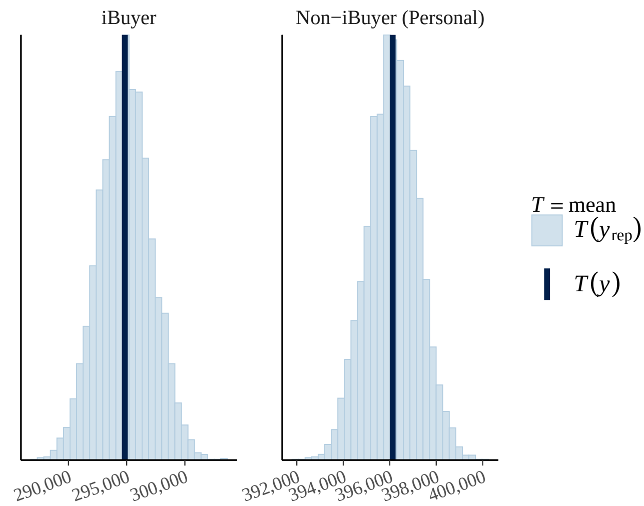
For the main sale price model, we find that average sale prices for different racial groups tend to align between the model's predictions and the observed data (Figure 5a), that average sale prices for different types of purchasing entities tend to align between the model's predictions and the observed data (Figure 5b), and finally that average sale prices for different combinations of buyers and sellers tend to align between the model's predictions and the observed data (Figure 5c).

We do find however that predictions from the model appear to be marginally overdispersed relative to the observed data, as shown in Figure 6a. After fitting a log-transformed model we find that this over-dispersion is somewhat alleviated (Figure 6b), implying that the linearity relationship used by our main model may be less suited to the data than, for example, an exponential relationship. Below, we consider how our conclusions change based on a log-transformed model.

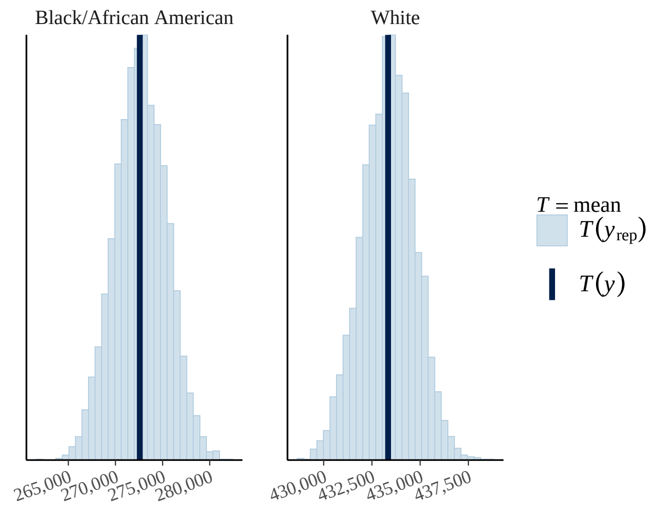
Based on the linear model we fit, we find the race gap between homes sold from Black and White homeowners to be $36,051 (95% CI: $29,380, $42,675). For the log-transformed model, which predicts log sales price, we find the race gap to be 9.2% (95% CI: 6.9%, 11.7%), indicating that the log model's best estimate is that on average homes sold from Black home sellers to individual buyers would have sold for 9.2% more had the sellers instead been White. For context, the median sales price in our dataset is $305,000, for which a 9.2% increase would amount to $28,184. Similarly, whereas in the linear model the iBuyer race gap was estimated to be $4,436 (95% CI: -$4,332, $13,236), the log model's estimate is 1.3% (95% CI: -1.7%, 4.3%). The iBuyer discount for White homeowners from the linear model was $27,239, (95% CI: $21,472, $33,002), and from the log model is 5.1%, (95% CI: 3.1%, 7.0%). The iBuyer discount for Black homeowners from the linear model was -$4,376 (95% CI: -$13,445, $4,713), while for the log model was -2.5% (95% CI: -5.6%, 0.5%).

As shown above, the conclusions that follow from the log-based model are roughly equivalent to those that follow from the linear model: that iBuyers tend to pay equivalent prices for homes from Black and White homeowners, but do so by paying less money than individual buyers for homes bought from White homeowners. The main difference is in the interpretation of the magnitude of the effects: the log-based model suggests that instead of buyer and seller identity having a constant real dollar effect on sale price (the same effect on a home sold for $200,000 as a home sold for $500,000), as suggested by the linear model, it may be the case that the size of the effect is larger for more expensive homes. We chose to rely primarily on the linear model for ease of interpretation.

We also perform posterior predictive checking for the logistic regression model used to predict whether a home will be converted to institutional ownership. For the logistic model, we calculate 5,000 posterior draws, then compare the proportion of positive and negative predictions from the posterior draws to the proportion of positive and negative predictions in the observed data, as shown in Figure 7. We find that the proportions align well between the observed data and posterior predictions.



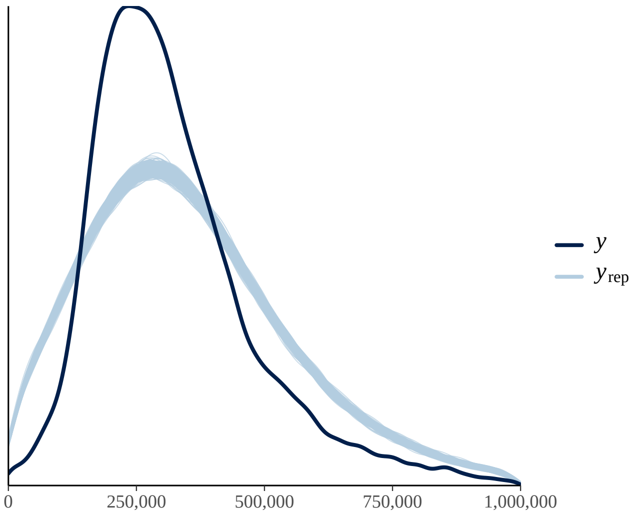
(a) Homes sold to iBuyers and individual buyers



(b) Homes sold from Black and White home sellers

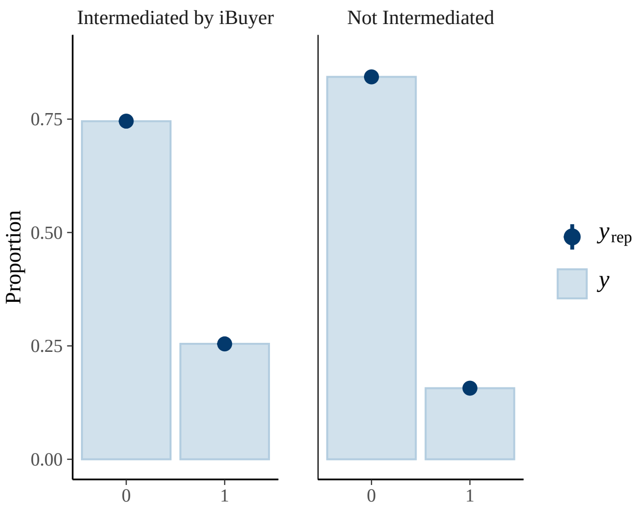
(c) Homes sold from Black and White home sellers to iBuyers and individual buyers

Figure 5: Mean posterior predictions for buyers and sellers of different types. Histogram shows means across 5,000 simulations of the posterior, dark blue line shows mean of observed data. Alignment between histogram and observed mean indicates model provides good summary of the data.

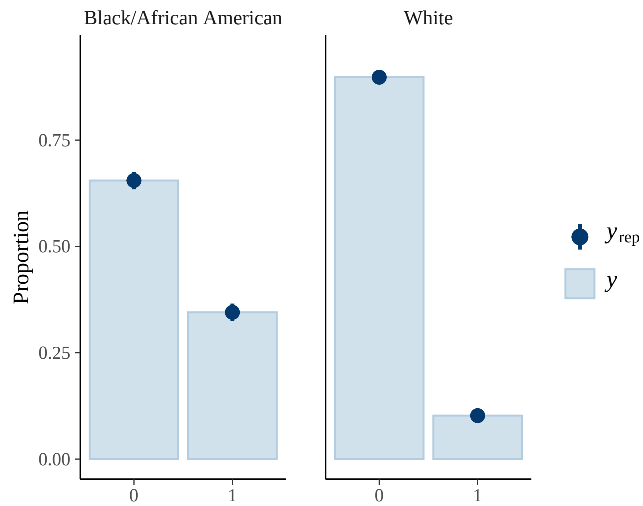


(a) Linear model **(b) Log-transformed model**

Figure 6: Posterior densities for linear and log-transformed sales price model. Light blue lines show densities of 500 simulated posterior draws, while dark blue lines show densities of observed data.



(a) Home sales intermediated and not intermediated by iBuyers



(b) Home sales from Black and White home sellers

(c) Home sales from Black and White home sellers that were intermediated and not intermediated by iBuyers

Figure 7: Posterior predictive proportions for different types of sellers and buyers. Light blue boxes show proportion of home sales that ended with institutional owners in observed data (where 1 indicates institutional ownership) and dark blue point shows mean proportion that ended in institutional ownership across 5,000 posterior draws (along with 95% credible intervals). Alignment between bars and dots indicates model provides good summary of the data.