Attenuating Racial Price Differentials in the Housing Market: Evidence from iBuyers

Yuan Cheng [*,†], Chris Forman [‡], Peng Liu [§]

We study the implications of algorithmic buyers, specifically iBuyers, on racial price disparities in the U.S. housing market. iBuyers leverage algorithms and data analytics to automate the home buying and selling process, serving as a new type of housing market intermediary. With millions of housing transactions and mortgage data from markets with significant iBuyer presence, our analysis reveals that iBuyers significantly attenuate the price gap between Black and White homebuyers for comparable housing. To address potential selection bias, we employ coarsened exact matching to ensure comparable housing and neighborhood characteristics between iBuyer and non-iBuyer transactions, and our results remain robust. We document that iBuyers' market entry is associated with reduced racial price differentials between Black and White buyers. By separating iBuyers' market-level and transaction-level effects, we uncover iBuyers' role in correcting the information imbalance among racial groups. Notably, when comparing iBuyers to traditional housing market intermediaries, known as flippers, we find no evidence that flippers serve a similar function in addressing racial price disparities. In addition, our heterogeneity analysis explores how iBuyers' impacts on racial price differentials vary with neighborhood racial composition and buyer income levels. We find that iBuyers' mitigating effects remain strong across neighborhood racial compositions and buyer income levels.

Keywords: racial disparities, impacts of AI-driven technologies, housing market, fintech.

[*]Corresponding author, email: yc956@cornell.edu

[†]Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY 14850

[‡]Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY 14850

[§]Cornell Peter and Stephanie Nolan School of Hotel Administration, Cornell University, Ithaca, NY 14850

1 Introduction

Algorithms and digital technologies are playing an increasingly important role in real estate markets, providing efficiency and convenience. They also raise the concerns and discussions surrounding algorithmic fairness and disparate outcomes [Barocas and Selbst, 2016, Achiume, 2020]. Algorithms not designed to discriminate can also discriminate due to bias "baked-in" the data [Rambachan et al., 2020]. Studies on the impact of digital platforms show evidence that the design of such platforms could exacerbate racial disparities [Cui et al., 2020, Gunarathne et al., 2022]. This concern is particularly true in the U.S. housing market due to its historical discriminatory practices based on race and ethnicity [Bayer et al., 2020].

On the other hand, recent studies in fintech show positive evidence that automated technologies such as loan automation reduces racial disparities in financial services. Automation underwriting systems help reduce the processing time of loan decisions and racially biased credit decisions [Bhutta et al., 2022]. Fintech lenders with automated lending technology provide alternative options for more creditworthy borrowers regardless of race or ethnicity [Buchak et al., 2018].

In light of the conflicting evidence on innovative technologies in the housing market, we study whether a new type of housing market participant — iBuyers, which rely heavily on digital technologies and algorithms — exacerbate or attenuate racial price differentials. iBuyers rely heavily on big data and learning algorithms [Seiler and Yang, 2022, Buchak et al., 2020], they leverage automated valuation models (AVMs) to determine the house's worth, asking price, and potential profit [1]. Similar to comparative market analysis (CMA), which uses a handful of similar homes nearby that have sold recently, AVM pricing is more accurate as it combines the effort of analyzing individual features of hundreds of comparable homes and local pricing experts [2]. In addition, iBuyers offer self-tours for potential buyers to visit a home without human interaction [3]. There is potentially more room in negotiation as iBuyers don't have emotional attachment to the homes. iBuyers do not make profit from renting the homes,

[1] https://www.trulia.com/guides/ibuyer-what-it-is-and-how-it-works/
[2] https://www.opendoor.com/articles/what-is-an-ibuyer.
[3] https://www.redfin.com/blog/how-to-buy-home-from-ibuyer/

the longer holding periods indicate larger costs. As they flip with a sales target, there is more room in the negotiation process. Closing process tends to be simpler and faster as the homes are usually vacant and fewer parties involved for the escrow and financial process. This evidence suggests that iBuyers simplify and partially automate the home buying process via algorithmic intermediation, which suggests potentials of iBuyers in helping traditionally disadvantaged groups in the housing market.

With transaction-level owner-transfer data from CoreLogic and mortgage application data from the Home Mortgage Disclosure Act (HMDA) data disclosures, we are able to quantify the racial disparities in prices between different groups for comparable housing. Our primary data are assembled using property and owner-transfer records from CoreLogic, and the loan application registry records from Home Mortgage Disclosure Act (HMDA) data disclosures. We match data from two sources with a customized matching algorithm, to obtain race and income information of mortgage applicants for each transaction. We are able to observe the housing transaction prices, housing characteristics, purchase information, demographic information of the seller. Also, we derive information from the data about buyers in large markets across the US where ibuyers are active.

Our research design ensures that we are quantifying the racial price differentials for comparable housing. We estimate the racial price differentials for comparable housing with rich set of controls in a repeat sales framework, where multiple transactions for house over long time period are necessary to enable the estimation, in order to difference out house-specific time-invariant unobserved features.

We first document the Black-White price differentials in the Phoenix, Atlanta, Dallas, Houston, Austin, San Antonio metropolitan statistical areas (MSAs). It suggests the persistence of racial price differentials, which can not be explained by time-varying macroeconomic condition, local market neighborhood changes, quality of the homes, or buyers' financial standing. Black buyers pay 1%-4% more compared to White, which echoes with earlier findings in the literature [Bayer et al., 2017].

Most importantly, we find strong evidence of iBuyers attenuating racial price differentials in the housing market. We find that price differentials between Black and White racial groups are largely eliminated for purchases from iBuyers in all the studied metro areas. For robustness, we adopt coarsened exact matching (CEM) to obtain a more similar control group. By constructing a matched sample with similar observable home characteristics, we rule out the potential bias from the selection of homes by iBuyers. The significance of iBuyers' mitigating effects remains, with a slight decrease in magnitude. To understand iBuyers' mitigating effect, we separate the iBuyers' effect into market-level effect and transaction-level effect. The results show the significance of market-level effects for all three metro areas, while only Phoenix show significant transaction-level effect.

Do other market intermediaries have the same function of attenuating racial price differentials? To answer this question, we compare iBuyers with important housing market intermediaries, flippers. Flippers are individuals who invest in properties and resell within a short period of time. The comparison indicates a clear difference between iBuyers and flippers in their impacts on racial price differentials. The results for flippers exhibit a significant degree of heterogeneity across metro areas. In Phoenix flippers don't charge additional premia for Black buyers, while in Atlanta and Texas, Black buyers who purchase from flippers pay substantial premia compared to their White counterparts.

In the analysis of multiple heterogeneity sources, we explore how the racial and ethnic price differentials vary based on heterogeneity in local racial composition. Our results show that, while the Black premia remain high across neighborhoods with different racial compositions, the Black-White price premia are largely mitigated by iBuyers. Another source of heterogeneity is buyer income. As buyer income is the confounding factor between race/ethnicity and home selection. We add the interaction of buyer income and race/ethnicity, which controls for the confounding between buyer income and housing price. We find the mitigating effects of iBuyers persist across markets.

Our findings contribute to the literature in several important ways. Most salient, we bridge studies on the impacts of innovative technologies and those on racial price differentials in the residential real estate market. We contribute to the literature on the impacts of digital platforms in the U.S. housing market [Barron et al., 2021, Chen et al., 2022]. Earlier findings of the racial price differentials for Black buyers validates the persistence of racial disparities [Quillian et al., 2020, Bayer et al., 2017]. Our analysis adds to the literature studying the behaviors and impacts of housing market investors, specifically the discussion of the intermediary roles of different types of flippers [Bayer et al., 2020, Fu et al., 2016, Seiler and Yang, 2022]. We compare the similarities and differences between iBuyers and flippers and their role in attenuating racial price differentials. We focus more on the implications of intermediation on price variance across racial group, which is the major difference with prior studies on iBuyers [Buchak et al., 2020, Seiler and Yang, 2022]. Last, our findings add to the emerging body of literature in Fintech show evidence that the loan automation reduces racial bias in mortgage outcomes [Buchak et al., 2018, Bhutta et al., 2022, Howell et al., 2021, Bartlett et al., 2022] through the channel of removing face-to-face interactions and human judgment.

2 Related Literature

The prevalence of transformative technologies raises concerns of disparate impacts based on race and ethnicity. A growing body of literature studies the racial and ethnic related impacts of digital platforms. Edelman and Luca [2014] show evidence that the rental charge by Black hosts is approximately 12% less than non-Black hosts for equivalent rental with all information visible controlled. In the setting of a business-to-customer (B2C) digital platform, Gunarathne et al. [2022] present evidence that African American customers are less likely to receive a response when they complain than comparable White customers. Experimental studies also present similar findings. Field experiments in [Edelman et al., 2017, Cui et al., 2020] document that distinctively African American names are less likely to be accepted compared to distinctively White names applications. Younkin and Kuppuswamy [2018] find African American men are significantly less likely to receive funding than similar White founders, and given lower rating for identical projects. These evidence suggest certain design in digital platform facilitate racial and ethnic discrimination.

In addition, fairness issues embedded in the algorithms underlying automated decisionmaking also raise concerns. Several studies show that the algorithms might inherit the biases and discriminative decisions from data. Even in the cases where algorithms for automated decisions do not utilize race/gender information, the bias and discrimination can be ”baked in” the data used for training. Strikingly, studies have shown that a simple race-blind or gender-blind design usually does more harm than good [Kleinberg et al., 2018]. Lambrecht and Tucker [2019] show that algorithm optimized for cost-effective ad delivery actually delivers ads disproportionately discriminative for female applicants.

Studies on the discrimination in U.S. residential real estate market examine the existence and trend of the racial inequality [Quillian et al., 2020, Ihlanfeldt and Mayock, 2009, Bayer et al., 2017][4]. While the racial inequality is showing signs of improvement since the enactment of the federal Fair Housing Act (FHA) in 1968, evidence suggests that the racial price difference is still persistent. As presented in [Bayer et al., 2017], around 2% premia are paid by Black and Hispanic buyers for comparable housing, and the premia are not explained by variation in buyer income or access to credit.

The racial price differentials largely originated from the frictions during the home buying process. Discrimination might be one source of these frictions, which can further be divided into statistical-based discrimination and taste-based discrimination. Statistical discrimination takes the form of stereotyping based on imperfect information, while taste-based discrimination results from decision maker’s prejudice [Guryan and Charles, 2013]. The home buying process is complicated, typically involves home search, bargaining, and mortgage application, discrimination might manifest in multiple forms during any part of the process. During the home searching and negotiation, discrimination can result in limited options and further lead to higher search cost, or disadvantaged position in bargaining. These two types of discrimination are usually intertwined and difficult to be separated.

Intrinsically, discrimination results from the search frictions in the housing market. Market intermediaries play a crucial role in providing liquidity and alleviating frictions when homeowner-to-homeowner matches are difficult. Home sellers could choose to sell directly

[4]For more discussion related to the history of racial discrimination in the housing market, refer to [Quillian et al., 2020].

to an intermediary [Buchak et al., 2020], which, in turn, resells the home using their market knowledge. iBuyers are new market participants that serve as market intermediaries. In this paper, we focus more on the implications of iBuyers' intermediation on price variance across racial group, which is the major difference with prior studies on iBuyers [Buchak et al., 2020, Seiler and Yang, 2022]. Our work distinguish from these two studies in that we focus more on iBuyers' reselling and implication of iBuyers' intermediation to racial price differentials in the housing market.

Lastly, our research contributes to recent work that explores implications of automation for reducing inequities in outcomes across groups. This emerging body of work suggests that automation reduces racial disparities in outcomes [Buchak et al., 2018, Bhutta et al., 2022, Howell et al., 2021, Bartlett et al., 2022] by removing face-to-face interactions and human judgment. Bhutta et al. [2022] shows evidence that discrimination may be less prevalent at fintech lenders where borrowers have no in-person contact with lenders. Howell et al. [2021] show Black applicants are more likely to get a Paycheck Protection Program (PPP) loan from fintech lenders compared to conventional lenders. Bartlett et al. [2022] find that finTech algorithms are 40% less than face-to-face lenders to discriminate in the consumer-lending context.

3 Context and Data

In this section, we first introduce the research context, answering questions about iBuyers' business model, technology, and role in intermediating housing markets. Second, we discuss market intermediation and compare iBuyers with another type of housing market intermediary, flippers. Third, we describe the sources for our primary data, the matching algorithm we developed to assemble the panel, and report summary statistics.

3.1 Background on iBuyers

iBuyers have a growing share of housing market transactions since 2014 when Opendoor started their iBuying business in Phoenix. Powered by AI valuation technologies, iBuyers determine the house's worth, asking price, and potential profit, then make an instant, all cash purchase offer. Unlike owner occupants, iBuyers don't occupy the homes they purchase. Unlike buy-to-rent institutions they also don't make profit from renting the homes. iBuyers carry out repairs and simple renovation after purchasing the homes. They usually resell the homes with a short holding period to minimize holding costs.

In the traditional home buying process, scheduling a tour is sometimes difficult if the home is occupied, as tours must accommodate schedules of the buyer and seller. Touring homes owned by iBuyers is much easier. iBuyers offer self tours for potential buyers, as homes resold by iBuyers are vacant and ready to be sold, buyers can schedule as many tours as needed and visit the homes during different times of the day. Especially when the buyer is in a hurry to purchase a home, the conveniences provided by iBuyers adds value and reduces buyers' search costs. In addition, iBuyers are less likely to exhibit taste-based discrimination as iBuyers flip the homes for a profit, usually with a sales target, and have no sentimental value attached to the homes as owner occupiers do. Some iBuyers also provide mortgage services, which further enhance buyers' access to credit. In short, iBuyers provide convenience, lower search cost, simplify negotiations, etc., and most importantly, iBuyers don't have the intention to discriminate.

3.2 Market Intermediation

Housing market intermediaries provide liquidity and reduce frictions in the housing market. Flippers — individuals who invest in properties and resell within a short period of time — are important housing market intermediaries[5]. iBuyers can be considered as a new type of market intermediary, serving similar economic function as flippers. Similar to flippers, iBuyers also resell the homes with a relatively short holding period, reducing search frictions and providing liquidity.

Interestingly, there are many similarities in the markets iBuyers choose to participate in compared to flippers. In Section 6, we document that the set of neighborhoods with iBuyer

[5]We note that there are small differences in the exact definition of flippers in the literature [Bayer et al., 2020, Buchak et al., 2020, Seiler and Yang, 2022].

entry is a subset of the neighborhoods with flipper activities, which aligns with the findings in [Buchak et al., 2020] that iBuyers enter markets with more active intermediation.

Flippers are known for acquiring distressed homes at reduced prices, conducting extensive renovations, and subsequently reselling at higher prices. In addition, flippers usually have better local market understanding, which makes it easier for them to sell for a premium based on market timing, Buchak et al. [2020] documented that flippers list more aggressively than ordinary buyers with a markup of 2.0%. In the meanwhile, iBuyers tend to buy homes that are already in good condition and don't require much renovation. We identify iBuyers and flippers from the transaction data with customized algorithms documented in Appendix B.

3.3 Data Description

Our primary data are assembled using property and owner-transfer records from CoreLogic proprietary data, and the loan application registry records from Home Mortgage Disclosure Act (HMDA) data disclosures. The CoreLogic data are derived from county deeds records. It contains sales transaction data such as property location, buyers/seller information, transaction type, sale date, sale price, and detailed mortgage information. It also contains property information such as year built, number of rooms, square footage, and lot size, etc.

The HMDA data are publicly-available loan application registry data collected from the lending institutions under the Home Mortgage Disclosure Act (HMDA). For the purpose of fair lending, HMDA documents mortgage lending practices and demographic information of applicants, in addition to the details of the mortgage[6]. We use loan application registry data to obtain the race/ethnicity information of the applicant(s).

We match transaction data from CoreLogic to mortgage application data from HMDA to obtain the race and ethnicity of the buyers. For privacy purposes, the mortgage application data neither disclose publicly the name of the applicant, nor the identifier for the property. We rely on tax year, census tract, lender company name, loan amount for the match. Matching procedure and matching statistics are detailed in Appendix C.

[6]Refer to https://www.ffiec.gov/hmda for more details

We restrict our sample to contain only arms-length transactions for single-family homes purchased with mortgage loans. This also rules out all the interfamily transaction records, mortgage refinancing records, foreclosures, multi/split parcel sale, etc. We also remove the records with missing transaction price, missing sale date, incomplete buyer/seller information. Similarly, for mortgage application data, we only keep the originated purchase loan records, drop records with missing race information, census tract, mortgage information, as these are key information for the data matching. With regard to buyers, we only consider individual investors, ruling out all corporate and institutional buyers such as buy-to-rent institutions, as those transaction prices are systematically different from individual buyers.

Our study period spans almost two decades, from 2004 to 2020. While ibuying is comparatively new, we need the earlier data to recover home fixed effects and enable repeat sales estimation. We cover top iBuying markets[7], including the Phoenix, Atlanta, Dallas, Houston, Austin, San Antonio metropolitan statistical areas (MSA). Phoenix is one of the MSAs with the earliest iBuyer entry and largest share of iBuyer activities, Atlanta and Texas are also among the largest markets with iBuyer presence. We combine the four MSAs in Texas into one for ease of presentation, and due to the limited number of iBuyer transactions.

In our sample, based on a repeat-sales framework (as described in Section 4), we only include homes with more than one resale, excluding all homes with only one transaction record between 2004 and 2020. After the preprocessing, the summary statistics are shown in Table 1. In all six MSAs, we have a total of approximately 1.5 million transactions for our repeat sales research design. We observe that the proportions of different races are quite different among MSAs. In particular, we observe that Black buyers constitute 25.35% of transactions in Atlanta, compared to the 2.8% and 7.6% in Phoenix and Texas, respectively. Phoenix shows comparatively higher mean and median transaction prices, while Texas has higher mean and median income.

[7] https://www.ownerly.com/data-analysis/ibuyer-top-markets-report/.

Table 1: Summary statistics of the matched data samples

	MSA			
	Phoenix	Atlanta	Texas	All
No. observations				
After matching	806,862	834,862	2,101,735	3,743,459
Repeated sales	354,432	317,239	791,529	1,463,200
Buyer race				
White	0.696	0.605	0.658	0.656
Hispanic or Latino	0.222	0.076	0.194	0.175
Asian and other	0.053	0.067	0.072	0.066
Black	0.028	0.253	0.076	0.103
Transaction statistics ($)				
Mean transaction price	285,319	261,792	253,797	262,374
Median transaction price	235,000	207,500	200,687	211,000
Mean income	108,417	100,865	115,155	110,533
Median income	75,000	73,000	85,000	80,000

Notes: This table presents summary statistics for our primary data. The repeated sales samples exclude all homes with only single transaction record, and only include homes resold more than once. The buyer race and transaction statistics are based on the repeated sales samples. We note that Texas don't require full disclosure of home transaction prices, the reported statistics are only based on the disclosed ones.

4 Empirical Strategy

The price of a house not only varies by quality and its unique combination of features, but also exhibit significant fluctuations over time. To ensure that the estimated racial price differentials are for comparable housing, we construct the following research design that rules out the impacts of factors such as time-varying macroeconomic conditions, local market neighborhood changes, quality of the homes, and financial standing of the buyer. In addition to the observed home characteristics, unobserved home characteristics might also significantly affect housing prices. We adopt the repeat-sales framework to address this concern. The repeat-sales framework difference out the unobserved characteristics associated with each unique home. This approach is widely used in the literature to control for the unobserved characteristics of the homes [Case and Shiller, 1987].

The repeat sales method requires more than one observation of transaction for each home, consequently, we only keep homes with multiple sales transactions. Our baseline model is as follows,

$$\ln \text{Price}_{ijnt} = \alpha_r \text{Race}_{ir} + \beta \text{iBuyer}_i + \gamma_r \text{Race}_{ir} \times \text{iBuyer}_i + X_{ijnt} + \epsilon_{ijnt}, \tag{1}$$

where Price_{ijnt} is the ith transaction price for property j in neighborhood n at time t. iBuyer_i is an indicator variable that equals 1 if the seller of transaction i is iBuyer and else 0. Race_{ij} denotes the race of the buyer for transaction i of property j. X_{ijnt} which is specified in Equation (2) denotes the set of controls. ϵ_{ijnt} is the idiosyncratic error term.

$$X_{ijnt} = \theta_j + \delta_{nt} + \nu_j \text{HH}_{jt} + \lambda_i \text{BI}_{ij} + F_{jt}. \tag{2}$$

In Equation (2), θ_j denotes the house fixed effects which we use in the repeat sales framework to control home-specific unobserved heterogeneity. The neighborhood-by-time fixed effects δ_{nt} control for the time-varying neighborhood attributes such as local policy impacts, location-time sensitive appreciation (which includes the impacts of macroeconomic condition) for census tract n and year-quarter t. We also add house hedonics controls HH_{jt} for property j at time t that includes house age, land square footage, building square footage, assessed total value. Further, we add buyer income as a control for the potential differential sorting (i.e., statistical discrimination) based on buyer's financial standing. BI_{ij} denotes the income of buyer in ith transaction of property j.

The home buying process is complicated, typically involves home search, bargaining, and mortgage application, discrimination might manifest in multiple forms during any part of the process. Statistical-based discrimination and taste-based discrimination are usually intertwined and difficult to be separated. Instead of focusing on separating these two types of discrimination, and we focus on identifying the channels where iBuyers could help attenuate the racial disparities.

Ideally, we would like to observe if any home improvement is made and corresponding increased value. The major challenge for our strategy is that we don't observe the accurate amount of home-specific heterogeneous appreciation. The appreciation potentially comes from home renovation. Some homes are subject to within-home changes due to renovation, which intrinsically change the value of the home. We note that the repeated sales framework cannot address this issue, as it does not control for time-varying characteristics due to unobserved changes such as renovation [Case and Shiller, 1987]. Flippers are known to purchase distressed homes and renovate the homes for a profit, which adds value to the homes. Thus, we use whether the home is flipped by any flipper as a proxy for potential improvement and appreciation of the homes. F_{jt} denotes an indicator for whether the home j is ever flipped by any flipper in the transaction history before the time t of the transaction.

5 Empirical Results

We start by investigating impacts of iBuyers on existing racial disparities. We are interested in two questions. First, whether racial price differentials are persistent in the studied metro areas and, second, whether iBuyers attenuate or exacerbate existing racial price differentials. Further, we compare the roles of iBuyers and flippers in attenuating racial price differentials as market intermediary.

5.1 The Impacts of iBuyers on Racial Price Discrimination

Table 2 and Table 3 report results estimated from Equation (1). With the omitted group being White, the two tables report respectively the Black-White price differentials ($\alpha_{Black} - \alpha_{White}$), and the price differentials associated with the iBuyer resales ($\gamma_{Black} - \gamma_{White}$) as in Equation (1). We report the estimates based on six specifications to obtain deeper insights on the racial disparities[8]. We do not include results for other racial and ethnic categories as we do not have a strong prior about the direction, but results are available upon request.

By solely using time fixed effects that control for the macroeconomic condition that affect nationwide housing, the results in Column (1) show homes purchased by Black buyers have lower prices than White buyers. We then add census tract and time interactions to control

[8] The racial groups we control include “Black”, “Hispanic or Latino”, “Asian and other”, and “White”. “Black” denotes Black or African American, “White” denotes non-Hispanic White. “Asian and other” includes Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander.

Table 2: Black-White Price Differentials

Location	Black-white differential					
	(1)	(2)	(3)	(4)	(5)	(6)
Phoenix	-0.1367***	0.0144***	0.0116***	0.0113***	0.0115***	0.0110***
	[0.0044]	[0.0028]	[0.0019]	[0.0019]	[0.0019]	[0.0019]
Atlanta	-0.3067***	0.0200***	0.0229***	0.0229***	0.0241***	0.0234***
	[0.0021]	[0.0020]	[0.0015]	[0.0015]	[0.0015]	[0.0015]
Texas	-0.2103***	0.0424***	0.0379***	0.0379***	0.0400***	0.0390***
	[0.0022]	[0.0016]	[0.0016]	[0.0016]	[0.0015]	[0.0015]
All MSAs	-0.2672***	0.0312**	0.0287**	0.0287**	0.0300**	0.0294**
	[0.0014]	[0.0060]	[0.0054]	[0.0054]	[0.0057]	[0.0056]
Controls						
Year-quarter FE	Y					
Neighborhood x time FE		Y	Y	Y	Y	Y
House FE			Y	Y	Y	Y
House Hedonics				Y	Y	Y
Buyer Income					Y	Y
Flipped by flipper						Y

Notes: (1) This table reports the estimates of Black-White differentials in Phoenix, Atlanta and Texas markets (we combine the 4 metro areas in Texas into one). (2) House hedonics include house age, building square footage, land square footage. We exclude total rooms due to large fraction of missing values. (3) "Flipped by flipper" is a proxy we use to control for potential home renovation. (4) In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

for time-varying neighborhood-level policy impact, market trend and economic conditions. The importance of controlling for time-varying neighborhood fixed effects can be seen from Column (2) in Table 2, where we observe the Black-White differentials change from negative to positive values. To investigate whether the premia paid by Black buyers are associated with higher quality of the homes, we introduce house hedonics such as building square footage and house age, estimated in the repeat sales framework with unobserved characteristics controlled. By comparing Column (3) and (4), the estimates for Black-White price differentials are almost the same, ruling out the house hedonics as the reason for the premia. We further control for buyer income as shown in Column (5). The premia increase slightly. In Column (6) we control for potential renovation which is proxied based on whether the home is flipped by a flipper. It is worthnoting that Column (6) are the results for our baseline model defined in Section 4. We observe no significant changes in the estimates, which implies that the premia paid by Black buyers are not due to the differences in buyer income or potential renovation.

Next we discuss the second part of Equation (1), i.e., racial price differentials in iBuyer resales as shown in Table 3. When only controlling for time or neighborhood-time fixed ef-

Table 3: Black-White Price Differentials For iBuyer resales

	Black x iBuyer interaction					
Location	(1)	(2)	(3)	(4)	(5)	(6)
Phoenix	0.1042***	0.0201	-0.0300**	-0.0297**	-0.0305**	-0.0309**
	[0.0306]	[0.0123]	[0.0106]	[0.0105]	[0.0105]	[0.0105]
Atlanta	0.1099***	0.0553***	-0.0240*	-0.0240*	-0.0241*	-0.0251*
	[0.0299]	[0.0148]	[0.0120]	[0.0120]	[0.0119]	[0.0119]
Texas	0.1628***	-0.014	-0.0590***	-0.0590***	-0.0598***	-0.0591***
	[0.0340]	[0.0157]	[0.0165]	[0.0165]	[0.0164]	[0.0164]
All MSAs	0.1320***	0.0142	-0.0412**	-0.0413**	-0.0422**	-0.0423**
	[0.0177]	[0.0149]	[0.0085]	[0.0085]	[0.0086]	[0.0086]
Controls						
Year-quarter FE	Y					
Tract x time FE		Y	Y	Y	Y	Y
House FE			Y	Y	Y	Y
House Hedonics				Y	Y	Y
Buyer Income					Y	Y
Flipped by flipper						Y

Notes: (1) This table reports the estimates of Black-White differentials for iBuyer resales in Phoenix, Atlanta and Texas markets. (2) The specifications are the same as in Table 2. (3) In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

fects, homes purchased by Black buyers are more expensive as shown in Column (1). This may be a selection effect: homes purchased by Black buyers are different on average from those purchased by White buyers. Once we control the unobserved and observed house characteristics as shown in Column (3) and (4), we observe price premia paid by Black buyers are significantly reduced compared to Column (2). The mitigating effects of iBuyer remain significantly by adding the control for buyer's income as in Column (5), and the control for potential renovation as in Column (6).

If we compare the discounts shown in Column (6) of Table 3 and the premia in Column (6) of Table 2, we note the premia associated with Black buyers are largely offset. To examine whether the premia paid by Black buyers are mitigated in iBuyer resales sample, we report the sum of the price premia and the price discount associated with iBuyer resales from Column (6), i.e., $(\alpha_{Black} - \alpha_{White}) + (\gamma_{Black} - \gamma_{White})$. These findings, summarized in Table 4 confirms the marginal effects of iBuyers on elimination of racial price discrimination for all the metro areas.

Table 4: Premium/Discount for Black buyers in iBuyer resales

	MSA			
	Phoenix	Atlanta	Texas	All
Black	0.0110***	0.0234***	0.0390***	0.0294**
	[0.0019]	[0.0015]	[0.0015]	[0.0056]
Black x iBuyer	-0.0309**	-0.0251*	-0.0591***	-0.0423**
	[0.0105]	[0.0119]	[0.0164]	[0.0086]
Black + Black x iBuyer	-0.0199*	-0.00115	-0.0206	-0.0129*
	[0.0103]	[0.0119]	[0.0164]	[0.00515]
Controls				
Tract x time FE	Y	Y	Y	Y
House FE	Y	Y	Y	Y
House Hedonics	Y	Y	Y	Y
Buyer Income	Y	Y	Y	Y
Flipped by flipper	Y	Y	Y	Y

Notes: (1) This table represents the sum of the price premia in Column (6) of Table 2 and the price discount associated with iBuyer resales in Column (6) of Table 3. (2) In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

5.2 Coarsened Exact Matching

We adopt the coarsened exact matching (CEM, [Blackwell et al., 2009]) procedure to ensure the racial price differentials are for comparable housing. CEM reduces imbalance in housing characteristics between treated and control group, i.e., homes sold by iBuyers and similar homes that are resold by other sellers.

The covariates utilized for the matching include housing characteristics (e.g., log transaction price, house age, log building square feet, log assessed total value, log land value, total number of rooms) and census tract level characteristics (FFIEC median family income[9], percentage of tract median family income compared to MSA/MD median family income, number of owner occupied units, etc.) For each covariate used for the matching, we coarsen it into multiple strata. Then we assign the original data to strata, and based on which stratum each data point is assigned to, we perform exact matching. For example, if house age is one of the covariates used for matching, and the house age in the sample is between 0 to 100 years, we coarsen house age into 4 different strata 0–25, 25–50, 50–75, and 75–100. If a home is a decade

[9]The FFIEC (Federal Financial Institutions Examination Council) Median Family Income (MFI) Report shows the estimate MFI that corresponds to the year when loan application data are collected.

Table 5: Statistics of Housing Characteristics for iBuyer vs. non-iBuyer resales after CEM

	Phoenix				
	iBuyer		non-iBuyer		t- test
	Mean	SD	Mean	SD	p-value
Buyer race					
White	0.621	0.485	0.691	0.462	0.000
Hispanic or Latino	0.266	0.442	0.214	0.41	0.000
Asian and other	0.062	0.24	0.06	0.238	0.674
Black	0.052	0.222	0.035	0.184	0.000
Transaction statistics					
Log Transaction price	12.46	0.253	12.456	0.301	0.321
Log Building square	7.533	0.309	7.543	0.303	0.043
Log Land value	10.438	0.394	10.435	0.418	0.624
Total rooms	1.864	0.222	1.868	0.216	0.285
Log Assessed total	8.136	0.394	8.133	0.419	0.649
House Age	21.027	12.371	20.87	12.458	0.404

Notes: (1) We show the statistics comparison for iBuyer and non-iBuyer resales in Phoenix. (2) After CEM, there are 4521, 1693, 2725 iBuyer resales and 82521, 83110, 223314 non-iBuyer resales in Phoenix, Atlanta, Texas respectively. (3) In the brackets are the standard errors, p-value is from the two-sample t-test.

old, it is assigned to the first stratum. By exact matching, we consider the treated and control homes matched if they are in the same stratum. The stratum that does not contain at least one treated and one control unit is dropped.

We note an inherent trade-off of CEM. More coarse binning result in fewer strata, fewer strata result in more diverse observations within the same strata and, thus, higher imbalance (Blackwell et al. [2009]). More refined binning result in more strata, but likely fewer matched treated and control groups due to the exact matching procedure. Also, different combination of the covariates generate different matched groups.

We balance the trade-off between getting the control group that are closer to the treated group and maintaining sufficient number of records, also we experiment with different combinations of the covariates for matching. We obtain much more similar treated and control groups with CEM. Table 5 reports the statistics of housing characteristics for iBuyer (treated) and non-iBuyer (control) resales with CEM. We further conduct two sample t-tests to see if there is a significant difference between the means of the treated and control group. By comparing the p-value in Table 5, we can see that the treated and control are more similar in Phoenix and

Atlanta compared to Texas. This could be due to the variation among the 4 MSAs in Texas.

Table 6: Black-White Price Differentials For iBuyer resales after CEM

	Panel A			Panel B		
	Phoenix	Atlanta	Texas	Phoenix	Atlanta	Texas
Black	0.00823***	0.0138***	0.0245***	0.0084***	0.0096***	0.0280***
	[0.00240]	[0.00177]	[0.00215]	[0.0022]	[0.0016]	[0.0020]
Black x iBuyer	-0.0204*	-0.0251**	-0.0434**	-0.0216**	-0.0195*	-0.0439**
	[0.0112]	[0.0119]	[0.0178]	[0.0110]	[0.0118]	[0.0175]
Black + Black x iBuyer	-0.0122	-0.0113	-0.0188	-0.0132	-0.0099	-0.0159
	[0.0109]	[0.0119]	[0.0177]	[0.0108]	[0.0118]	[0.0174]
Tract x time FE	Y	Y	Y	Y	Y	Y
House FE	Y	Y	Y	Y	Y	Y
House Hedonics	Y	Y	Y	Y	Y	Y
Flipped by flipper	Y	Y	Y	Y	Y	Y

Notes: (1) This table reports the regression results with the same specification as in Column (6) in Table 2. (2) The sample in the left panel shares the same sample with Table 5. (3) The sample size after CEM is smaller than the original sample. (4) In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

We match the homes that are sold by non-iBuyers to the homes resold by iBuyers. Once we obtain the matched homes, we retrieve the whole transaction history of the homes for repeat sales estimation. Table 6 reports the regression results for Equation (1) using two versions of the samples matched by CEM. CEM offers an automatic binning algorithm, user can also specify the coarsening in which case balance of samples in each stratum can be better controlled by the researcher. The matched sample used in Panel A is from the automatic binning algorithm. The panel B shows results from cut-points chosen based on the histogram of each matching covariate. In addition, we experiment with different combinations of the covariates for matching. Due to the differences in local regulations, average home characteristics and density, we develop customized sets of matching covariates for each metro areas respectively. Compared to the Black-White differentials from the baseline regression in Column (6) of Table 2 and Table 3, the findings align with the discussion in Section 5.1, and the significance stay the same, with slight decrease in magnitude.

5.3 Transaction-level or Market-level Effect

Having documented the impact of iBuyers on the racial price differentials in the housing market, we next investigate the impact of iBuyer entry on the racial price differentials. The mitigating effect of iBuyers can be transaction-level effect, market-level effect, or both.

The transaction-level effect quantifies the impact of iBuyers on the individual transactions. iBuyers set resale price with automated valuation models (AVM), which valuate the homes based on the historical sales data for the homes in the neighborhood. If the transaction-level effect is significant, it suggests that iBuyers could correct the price of the homes that are overpriced. On the other hand, the market-level effect denotes the impact of iBuyers on the housing market as a whole. The entry of iBuyers could change the equilibrium of the local housing market, also amending the information imbalance among racial groups. Thus, it is especially beneficial for the racial groups that might previously receive incorrect or limited information regarding housing market. If the market-level effect is significant, it suggests iBuyers' correction of the market-level equilibrium with more transparent pricing.

To separate the transaction-level and market-level effect, we add the interaction term between iBuyer entry and racial groups. The coefficients of the added interaction term captures the market-level effect of iBuyers. The model is specified in Equation (3).

$$\begin{aligned} \ln \textit{Price}_{ijnt} =& \alpha_r \textit{Race}_{ir} + \beta_1 \textit{iBuyer}_i + \beta_{2k} \textit{iBuyerentry}_{ikt} + \gamma_{1r} \textit{Race}_{ir} \times \textit{iBuyer}_i \\ &+ \gamma_{2ir} \textit{Race}_{ir} \times \textit{iBuyerentry}_{ikt} + X_{ijnt} + \epsilon_{ijnt}, \end{aligned} \tag{3}$$

where γ_{1r} captures the iBuyers' effect on the transaction-level price for race/ethnicity r, while γ_{2r} captures the iBuyers' effect on the market-level price for race/ethnicity r. $\textit{iBuyerentry}_{ikt}$ is an indicator variable indicating by the time t of the transaction i, whether iBuyers have entered the market k where the purchased home locates. X_{ijnt} is defined same as in Equation (2). The identification of the transaction-level and market-level effects replies on the observations of more than one home. We report results for iBuyer entry to each market. The results for iBuyer entry to each neighborhood (census tract) are similar. It is worth noting that if there are neighborhoods with only one home, it will create challenges identifying between the transaction-level and market-level effects.

Table 7: Market-level vs. Transaction-level effect

	MSA		
	Phoenix	Atlanta	Texas
Black	0.0175***	0.0323***	0.0455***
	[0.0024]	[0.0016]	[0.0017]
Black x iBuyer	-0.0196*	0.0085	-0.0272
	[0.0107]	[0.0120]	[0.0167]
Black x iBuyerentry	-0.0170***	-0.0536***	-0.0418***
	[0.0038]	[0.0034]	[0.0041]
Controls			
Tract x time FE	Y	Y	Y
House FE	Y	Y	Y
House Hedonics	Y	Y	Y
Buyer Income	Y	Y	Y
Flipped by flipper	Y	Y	Y

Notes: (1) We report $(\gamma_{1,Black} - \gamma_{1,White})$ and $(\gamma_{2,Black} - \gamma_{2,White})$, where respectively captures the transaction-level and market-level effect of iBuyers for the racial price differentials. (2) We use the same set of controls as the baseline, which control for the time-varying neighborhood characteristics, house fixed effects, house hedonics, buyer income, and potential renovation. (3) In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

Table 7 shows the market-level effect exists for all three MSAs, while transaction-level effect is significant in Phoenix but not Atlanta and Texas. The results suggest the role of iBuyers changing the market equilibrium, and correcting the information imbalance among racial groups, and thus mitigating the racial price differentials in the housing market. The corrected premium for Black buyers implies the importance of iBuyers in correcting the racial price differentials in the housing market.

6 How do iBuyers compare to flippers

Recall the comparison of iBuyers and flippers as market intermediary in Section 3. The set of neighborhoods with iBuyer entry is a subset of neighborhoods with flipper activities. Naturally, we are interested in whether flippers play a similar role in attenuating racial price differentials as iBuyers.

We investigate the impacts of flippers on the racial price differentials and further compare with iBuyers. The research design expands the baseline regression Equation (4) by including the interactions between flipper and seller race. In addition, we group flippers into different categories based on their flipping volume. Following [Bayer et al., 2020], we define flippers based on the number of flips in the study period, 1–2 flips for low-volume flippers, 3–5 flips for mid-volume flippers, and 6 or more flips for high-volume flippers. The low-volume and mid-volume flippers are likely individual investors, while the high-volume flippers are more likely to be professional flippers, which are similar to iBuyers in terms of the business model. The regression setup is as follows,

$$\ln \text{Price}_{ijnt} = \alpha_r \text{Race}_{ir} + \xi_k \text{Flipper}_{ik} + \eta_{kr} \text{Race}_{ir} \times \text{Flipper}_{ik} + \gamma_r \text{Race}_{ir} \times \text{iBuyer}_i + X_{ijnt} + \epsilon_{ijnt}, \tag{4}$$

where flipper$_{ik}$ is a categorical variable indicating whether the seller of the i^{th} transaction is a flipper and if so, it further denotes the type k of flippers[10]. X_{ijnt} is defined same as in Equation (2).

Table 8: Flipper premium and Black-White price differentials for flipper resales

	Panel A: Flipper Premium			*Panel B: Black x Flipper interaction*		
	Flipper Type			*Flipper Type*		
Location	Low	Med	High	Low	Med	High
Phoenix	0.0243***	0.0561***	0.0526***	-0.019	-0.0009	-0.0011
	[0.0023]	[0.0051]	[0.0050]	[0.0255]	[0.0089]	[0.0214]
Atlanta	0.0251***	0.0563***	0.0393***	0.0761***	0.0810***	0.0666**
	[0.0051]	[0.0099]	[0.0118]	[0.0182]	[0.0079]	[0.0208]
Texas	0.0217***	0.0791***	0.1098***	0.0874**	0.0648***	0.0364
	[0.0048]	[0.0115]	[0.0153]	[0.0297]	[0.0103]	[0.0273]
All MSAs	0.0234***	0.0610***	0.0601**	0.0700***	0.0696**	0.0481***
	[0.0033]	[0.0065]	[0.0106]	[0.0081]	[0.0130]	[0.0062]
Controls						
House FE	Y	Y	Y	Y	Y	Y
Tract x time FE	Y	Y	Y	Y	Y	Y
Buyer Income	Y	Y	Y	Y	Y	Y
Flipped by flipper	Y	Y	Y	Y	Y	Y

Notes: We classify flippers as follows: low-volume flipper if <= 2 flips, medium-volume flipper if 3-5 flips, and high-volume flipper if >=6 flips. In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

Panel A in Table 8 reports the coefficients ξ_k for different types k of flippers, and Panel B

[10]We note that this variable is different from the proxy for renovation F_{jt} as in Equation (2).

reports the differences ($\eta_{k,Black} - \eta_{k,White}$) associated with flipper resales. We note Panel A shows that there are premia associated with all types of flippers, with the premia associated with low-volume flippers relatively lower. This aligns with our analysis that flippers purchase and renovate home, which contributes to home appreciation. Additionally, we note that more experienced flippers achieve larger margins compared to low-frequency flippers. The price premia in Panel A of Table 8 are expected, as flippers usually add value to homes through renovation. The appreciation of the homes in this process reflect the value added in renovation. In addition, flippers are considered to have better market trend understanding and sell for a premium based on market timing.

Panel B of Table 8 shows the Black-White price differentials for the flipper resales. The results across metro areas exhibit a significant degree of heterogeneity. In Phoenix flippers don't charge additional premia for Black buyers, while in Atlanta and Texas, Black buyers who purchase from flippers pay substantial premia compared to their White counterparts. However, as shown in Panel B, the premia are smaller when the sellers have higher flipping frequency. As discussed in [Bayer et al., 2020] which classify low-frequency flippers as speculators, and higher-frequency flippers as middlemen, middlemen are more experienced and do not rely on market timing as much as speculators, thus providing better intermediation.

How do iBuyers compare to flippers? By comparing Table 3 and Panel B of Table 8, we see a clear difference between iBuyers and flippers in their impacts on price differences between racial groups. While the goals of both iBuyers and flippers are to maximize their profits, the iBuyers set the price for the flipped homes based on an automated valuation model, while flippers have better local market understanding to aim for a higher price. iBuyers prefer homes that are in good condition without much renovation needed, while flippers tend to purchase distressed homes and make significant renovations that contribute to home appreciation. Finally, recall the discussion in Section 3.2, flippers are individual investors instead of corporate or institutional buyers, so idiosyncratic tastes might contribute to the premia associated with Black buyers.

Table 9 reports the composition of buyers who purchase from iBuyers and flippers. The

Table 9: Racial composition of homes buyers in iBuyer and flipper resales

	MSA			
	Phoenix	Atlanta	Texas	All
Buyer race for iBuyer resales				
White	0.622	0.444	0.477	0.544
Hispanic or Latino	0.266	0.145	0.311	0.257
Asian and other	0.061	0.098	0.104	0.081
Black	0.051	0.314	0.109	0.118
Buyer race for flipper resales				
White	0.669	0.509	0.63	0.615
Hispanic or Latino	0.251	0.072	0.206	0.191
Asian and other	0.051	0.051	0.06	0.054
Black	0.029	0.368	0.105	0.14

Notes: The composition is calculated based on the repeated sales sample.

percentages of Black buyers in the flipper resales and iBuyer resales samples are both higher compared to the main sample in Table 1, which implies the effective intermediation of flippers and iBuyers. We also note the heterogeneity across metro areas. For example, among iBuyer resales, Black buyers constitute 31.4% of buyers in Atlanta compared to the 5.1% and 10.9% in, respectively, Phoenix and Texas. See Section 7.1 for a more in depth discussion of the heterogeneity by neighborhood racial composition.

7 Heterogeneity

In this section, we explore the racial price differentials based on multiple heterogeneity sources, including the racial composition of the neighborhoods, and the buyer income. We examine the implications of heterogeneity results, which shed light on uncovering the mechanisms of iBuyers' role in attenuating racial price differentials in the housing market.

7.1 Heterogeneity by Neighborhood Racial Composition

Our results in Section 5 show significant heterogeneity across different metro areas. As shown in Table 1, the racial composition of different metro areas differs significantly. For example, Atlanta has relatively high aggregated percentage of Black population compared to Phoenix and Texas. The heterogeneity in racial composition across neighborhoods within the same metro area is also substantial. In this section, we explore how racial price differentials and iBuyers' mitigating effect vary based on the heterogeneity in racial composition in local neighborhood.

In addition to our primary data, we obtain census tract level racial composition data from American Community Survey (ACS). We merge ACS data to our primary data. We note that, the census tract information only dates back to 2010, also due to the change of census tracts boundaries overtime, some earlier records in our primary data have no corresponding information in ACS, and thus have to be excluded.

Table 10: Heterogeneity by neighborhood racial composition

	Panel A: Black Premium			*Panel B: Black x iBuyer interaction*		
Racial Composition	Phoenix	Atlanta	Texas	Phoenix	Atlanta	Texas
White > 0 %	0.0104***	0.0145***	0.0340***	-0.0231**	-0.0141	-0.044
	[0.0022]	[0.0017]	[0.0033]	[0.0107]	[0.0128]	[0.0242]
White > 25 %	0.0123***	0.0148***	0.0363***	-0.0340***	-0.0316**	-0.0488
	[0.0023]	[0.0017]	[0.0040]	[0.0115]	[0.0141]	[0.0401]
White > 50 %	0.0123***	0.0138***	0.0390***	-0.0388***	-0.0528***	-0.0465
	[0.0026]	[0.0021]	[0.0044]	[0.0133]	[0.0174]	[0.0214]
White > 75 %	0.0080*	0.0146***	0.0570***	-0.0206	-0.0688**	-0.0824**
	[0.0047]	[0.0038]	[0.0053]	[0.0254]	[0.0324]	[0.0253]

Notes: (1) Each row show results from local neighborhoods with certain racial composition of White population. (2) The "$White > 0\%$" sample is the primary data we use for the baseline results, the estimates are from Column (6) in Table 2 and Table 3 respectively. (3) We use the same set of controls as the baseline, which control for the time-varying neighborhood characteristics, house fixed effects, house hedonics, buyer income, and potential renovation.

Results are presented in Table 10. Estimated in Row "$White > 0\%$" show the estimate from Column (6) in Table 2 and Table 3. Comparing the estimates in Panel A, we find that the premia paid by Black buyers versus White buyers remain high across neighborhoods with different percentages of White residents, which aligns with the observations in [Bayer et al., 2017]. In Panel B, we observe iBuyers' mitigating effects for the Black-White price differentials are present in almost all neighborhoods with different racial compositions. Comparing the magnitude of Panel A and Panel B, the Black-White price premia are largely mitigated. The results indicate the mitigation of premia by iBuyers does not depend on the fraction of Whites.

7.2 Heterogeneity by Buyer Income

Income is an important indicator of a home buyer's financial standing and significantly influences their selection of homes. The home buyers who choose to purchase from iBuyers might have different income levels. The potential differences in between-group income distribution raise concerns about spurious correlations and biased estimation of racial price differentials, resulting from the confounding between iBuyers and racial price. If certain income groups are more likely to purchase from iBuyers, the racial price differentials could be driven by the income differences rather than iBuyers' effects.

To control for the confounding effects of buyer income on racial price differentials, we further add the interaction term between iBuyer and buyer income. This approach addresses the concern of attributing racial price differentials solely to iBuyers when the differences might be influenced also by buyer income disparities. The model is specified in Equation (5).

$$
\begin{aligned}
\ln \mathit{Price}_{ijnt} =& \alpha_r \mathit{Race}_{ir} + \beta \mathit{iBuyer}_i + \gamma_r \mathit{Race}_{ir} \times \mathit{iBuyer}_i \\
&+ \phi \mathit{iBuyer}_i \times \mathit{BI}_i + X_{ijnt} + \epsilon_{ijnt}
\end{aligned}
\tag{5}
$$

BI_i denotes the income of the home buyer in the ith transaction. ϕ is the coefficient for the interaction term of buyer income and iBuyers, which captures the iBuyers' effects for buyers with different income levels. With the additional control.

Table 11 show iBuyers' mitigating effects for Black-White price differentials are robust to the inclusion of buyer income. Notably, in Atlanta and Texas, the Black-White price differentials associated with iBuyers become nearly twice as negative after accounting for buyer income, compared to the results in Table 4. This rules out the concern that the mitigating effects of iBuyers for the premia paid by Black buyers are driven by the spurious correlation caused by buyer income.

Further, we investigate the mitigating effects of iBuyers for buyers with lower income compared to those with higher income with the specification in Equation (6). For each MSA, we categorize buyers based on their income level. LI_i indicates whether the income of the home buyer in the ith transaction falls below the 25th percentile within that MSA. χ is the coefficient

Table 11: iBuyer's mitigating effects after accounting for buyer income

	MSA		
	Phoenix	Atlanta	Texas
Black	0.0828***	0.0662***	0.0712***
	(0.00437)	(0.00208)	(0.00165)
Black x iBuyer	-0.0350*	-0.0697***	-0.0989***
	(0.0188)	(0.0141)	(0.0165)
iBuyer x Buyer Income	-0.172***	-0.130***	-0.0945***
	(0.00899)	(0.0106)	(0.0123)
Controls			
Tract x time FE	Y	Y	Y
House FE	Y	Y	Y
House Hedonics	Y	Y	Y
Buyer Income	Y	Y	Y
Flipped by flipper	Y	Y	Y

Notes: (1) We use the same set of controls as the baseline, which control for the time-varying neighborhood characteristics, house fixed effects, house hedonics, buyer income, and potential renovation. (2) In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

for the interaction term of buyer income and iBuyers, which captures the iBuyers' effects for buyers with lower income compared to those with higher income. With the additional control, γ_r captures the Black-White price differentials for buyers whose income is above 25th quantile.

$$\begin{aligned}\ln \textit{Price}_{ijnt} =& \alpha_r \textit{Race}_{ir} + \beta \textit{iBuyer}_i + \gamma_r \textit{Race}_{ir} \times \textit{iBuyer}_i \\ &+ \chi \textit{iBuyer}_i \times \textit{LI}_i + X_{ijnt} + \epsilon_{ijnt}\end{aligned} \tag{6}$$

Comparing the results in Table 12 and the baseline results in Table 4, the Black-White price differentials are similar for the buyers with larger than 25th quantile income and the buyers in all income levels. iBuyers' mitigating effects on Black-White price differentials are also similar, with differences of 0.2%, 0.3%, 0.36% for Phoenix, Atlanta, Texas, respectively. Table 12 also shows that buyers with lower income pay more compared to higher income buyers in iBuyer resales, which aligns with the intuition that lower income buyers face higher borrowing costs due to limited access to credit.

Table 12: iBuyer's mitigating effects for lower income buyers

	MSA		
	Phoenix	Atlanta	Texas
Black	0.0109***	0.0239***	0.0389***
	[0.00188]	[0.00142]	[0.00152]
Black x iBuyer	-0.0288***	-0.0248**	-0.0627***
	[0.0104]	[0.0117]	[0.0163]
Low Income x iBuyer	0.0344***	0.0490***	0.0202
	[0.00574]	[0.0123]	[0.0132]
Controls			
Tract x time FE	Y	Y	Y
House FE	Y	Y	Y
House Hedonics	Y	Y	Y
Buyer Income	Y	Y	Y
Flipped by flipper	Y	Y	Y

Notes: (1) Here we define low-income buyers as those with an income below the 1st quartile of the buyer income distribution. (2) We use the same set of controls as the baseline, which control for the time-varying neighborhood characteristics, house fixed effects, house hedonics, buyer income, and potential renovation. (3) In the brackets are the cluster-robust standard errors, ***, **, * respectively denote statistical significance at 0.01, 0.05 and 0.1 levels.

8 Conclusion

We examine the implications of iBuyers intermediation on racial disparities in the U.S. residential housing market. While documenting the persistence of racial price differentials for Black buyers for comparable housing in the studied metro areas, we show strong evidence that iBuyers attenuate racial price differentials. The Black-White price premia are largely mitigated across neighborhoods in metro areas with top iBuyer presence. These findings are robust with coarsened exact matching, which ensure the similarity of the housing characteristics between iBuyer and non-iBuyer resales, and rule out the potential bias from the selection of homes by iBuyers. We further compare iBuyers with traditional market intermediaries, flippers, but we find no evidence of similar function of flippers in attenuating racial price differentials like iBuyers.

We separate iBuyers' market-level effect and transaction-level effect. The results suggest the role of iBuyers changing the market equilibrium, and correcting the information imbalance among racial groups, and thus mitigating the racial price differentials in the housing market.

Further, we explore the racial price differentials based on multiple heterogeneity sources, including the racial composition of the neighborhoods, and the buyer income. The heterogeneity results based on neighborhood racial composition show iBuyers' mitigating effects remain strong across neighborhoods with different racial compositions. The potential differences in between-group income distribution raise the concern for bias in estimation of racial price differentials due to composition effect. We alleviate the concern of the composition effect and controls for the confounding between buyer income and housing price. The results show that iBuyers' mitigating effects remain strong, which rule out the concern that iBuyers' mitigating effect is driven by the spurious correlation caused by buyer income.

As a new market intermediary powered by digital technologies and algorithms, iBuyers show considerable potential in attenuating racial price differentials. Our findings unveil the value of digitization and automation on reducing frictions, and relieve the concerns of entrenched racial price differentials in the housing market. Our work contributes to a broader study on the impacts of innovation and technologies on racial disparities, adds to the literature demonstrating the positive effects of financial technologies in advancing equity.

References

E Tendayi Achiume. Racial discrimination and emerging digital technologies: a human rights analysis: report of the special rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance. *United Nations Digital Library, Genebra*, 18, 2020.

Solon Barocas and Andrew D Selbst. Big data's disparate impact. *California law review*, pages 671–732, 2016.

Kyle Barron, Edward Kung, and Davide Proserpio. The effect of home-sharing on house prices and rents: Evidence from airbnb. *Marketing Science*, 40(1):23–47, 2021.

Robert Bartlett, Adair Morse, Richard Stanton, and Nancy Wallace. Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*, 143(1):30–56, 2022.

Patrick Bayer, Marcus Casey, Fernando Ferreira, and Robert McMillan. Racial and ethnic price differentials in the housing market. *Journal of Urban Economics*, 102:91–105, 2017.

Patrick Bayer, Christopher Geissler, Kyle Mangum, and James W Roberts. Speculators and middlemen: The strategy and performance of investors in the housing market. *The Review of Financial Studies*, 33(11):5212–5247, 2020.

Neil Bhutta, Aurel Hizmo, and Daniel Ringo. How much does racial bias affect mortgage lending? evidence from human and algorithmic credit decisions. *FEDS Working Paper*, 2022.

Matthew Blackwell, Stefano Iacus, Gary King, and Giuseppe Porro. cem: Coarsened exact matching in stata. *The Stata Journal*, 9(4):524–546, 2009.

Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483, 2018.

Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru. Why is intermediating houses so difficult? evidence from ibuyers. Technical report, National Bureau of Economic Research, 2020.

Karl E Case and Robert J Shiller. Prices of single family homes since 1970: New indexes for four cities, 1987.

Wei Chen, Zaiyan Wei, and Karen Xie. The battle for homes: How does home sharing disrupt local residential markets? *Management Science*, 2022.

Ruomeng Cui, Jun Li, and Dennis J Zhang. Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science*, 66(3):1071–1094, 2020.

Benjamin Edelman, Michael Luca, and Dan Svirsky. Racial discrimination in the sharing economy: Evidence from a field experiment. *American economic journal: applied economics*, 9 (2):1–22, 2017.

Benjamin G Edelman and Michael Luca. Digital discrimination: The case of airbnb. com. *Harvard Business School NOM Unit Working Paper*, (14-054), 2014.

Yuming Fu, Wenlan Qian, and Bernard Yeung. Speculative investors and transactions tax: Evidence from the housing market. *Management Science*, 62(11):3254–3270, 2016.

Priyanga Gunarathne, Huaxia Rui, and Abraham Seidmann. Racial bias in customer service: Evidence from twitter. *Information Systems Research*, 33(1):43–54, 2022.

Jonathan Guryan and Kerwin Kofi Charles. Taste-based or statistical discrimination: the economics of discrimination returns to its roots. *The Economic Journal*, 123(572):F417–F432, 2013.

Sabrina T Howell, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong. Racial disparities in access to small business credit: Evidence from the paycheck protection program. Technical report, National Bureau of Economic Research, 2021.

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

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Algorithmic fairness. In *Aea papers and proceedings*, volume 108, pages 22–27, 2018.

Anja Lambrecht and Catherine Tucker. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management science*, 65(7): 2966–2981, 2019.

Lincoln Quillian, John J Lee, and Brandon Honoré. Racial discrimination in the us housing and mortgage lending markets: a quantitative review of trends, 1976–2016. *Race and Social Problems*, 12(1):13–28, 2020.

Ashesh Rambachan, Jon Kleinberg, Jens Ludwig, and Sendhil Mullainathan. An economic perspective on algorithmic fairness. In *AEA Papers and Proceedings*, volume 110, pages 91–95, 2020.

Michael J Seiler and Liuming Yang. The burgeoning role of ibuyers in the housing market. *Real Estate Economics*, 2022.

Peter Younkin and Venkat Kuppuswamy. The colorblind crowd? founder race and performance in crowdfunding. *Management Science*, 64(7):3269–3287, 2018.

A Name Matching

Name matching serves the basis for identifying corporate and institution names in data pre-processing, identifying iBuyer names variants, identifying roundtrip transactions for finding flippers, and the matching between the owner transfer data and mortgage application data.

Due to the prevalence of spelling errors in the names, we adopt a fuzzy name matching procedure to calculate the similarity of names based on Levenshtein distance. The Levenshtein similarity ratio based on the Levenshtein distance measures the difference between two sequences of words. We combine partial string match with full string match. For example, JP MORGAN CHASE and CHASE should be perfect match, but the similarity score based on full string match would generate a similarity score of 50, while the partial string match is 100.

B Market players Identifications

B.1 Identification of iBuyers

The iBuyers we identify include Opendoor, Offerpad, Zillow, Knock, and Redfin. We identify iBuyers using a set of subsidiaries company names of iBuyers from public sources, mailing addresses, and a fuzzy name matching procedure.

Due to the spelling errors in the transaction records (e.g., 'QPENDOOR PROPERTY N LLC'), we use an interactive method to obtain as many variants of iBuyer names as possible. We start with the fuzzy name matching to compare buyer/seller names and addresses of the sales transaction with the base set of iBuyer company names and addresses, and put the similar names into the base set. In the end of every round of search, our base set would have more variants of iBuyer names, and eventually the base set stops growing, and we obtain the full set of iBuyer name variants. We conduct the iBuyer identification market by market as the deed documentation convention varies among MSAs. Take Opendoor as an example, our algorithm for identifying iBuyer names find 440 variants in Phoenix, 189 variants in Atlanta, and 271 variants in Texas.

B.2 Identification of Flippers

As mentioned in section 3.2, we define flipper as individual investors who purchase and then resell homes within short period of time. To identify flippers, we need to obtain the number of resales made by each seller and the corresponding holding period. Roundtrip transaction identification is necessary for calculating holding period. We sort the transaction records for each CLIP (unique identifier for a property) based on sale date, and compare the buyer names of ith transaction with seller names of $(i+1)$th transaction, if they match, we identify a roundtrip transaction[11]. In the algorithm for flipper identification, we first collect information of the number of roundtrip transactions for each investor. To avoid the measurement error related to common names in the same local market, we use mailing addresses in addition to the individual investors' names. In this procedure, we adopt a fuzzy name matching procedure to account for the measurement errors due to spelling errors in the names and addresses. With algorithm, we are able to observe how many roundtrip transactions there are for each individual investors, calculate the holding period of each roundtrip, and identify flippers from owner occupants of the set of buyers in our data. Based on the volume of the flipping, we further classify flippers into low-frequency, medium-frequency, and high-frequency flippers.

C Matching SBL to HMDA

To get the race and ethnicity information of the buyers, we need the matching between the transaction data from CoreLogic and the mortgage application data from HMDA. For privacy

[11] We note that if any transaction is dropped due to missing key information such as sale date, and transaction price, the roundtrip is also lost.

purposes, the mortgage application data neither disclose publicly the name of the applicant, nor the identifier for the property. The availability of detailed mortgage information in both data sources enable the matching, specifically, we rely on tax year, census tract, lender company name, loan amount[12].

Before developing the matching algorithm, we first randomly select several census tracts and perform the matching manually in order to have a better understanding the data. During the manual matching, we notice that (i) spelling errors of lender company name are prevalent, (ii) the naming recording convention in CoreLogic and HDMA are different, e.g., FINAM LLC vs. FINANCE AMERICA, LLC, (iii) the loan amount information recorded in CoreLogic is often different from the loan amount in HMDA, and simple minimization of the absolute value of the loan amount difference does not lead to the perfect match.

In the design of the matching algorithm, we take into account the observations in the manual matching. The developed matching algorithm has several stages. First, for each transaction record, we retrieve all the loan origination records in the same tax year and census tract as candidates. We filter the candidates based on whether the loan amount difference exceeds \$10,000, and whether the loan term matches exactly (only for tax years from 2018 to 2020). Second, we calculate the similarity of the lender company name between the transaction record and the candidate HMDA record, and rank each candidate based on the similarity score. The candidate with the highest similarity score larger is considered the best candidate. If the similarity score of the best candidate exceeds 90 on a scale of 1-100, the best candidate passed the test and considered as the match, and it is automatically removed from the candidates of the other transaction records. Due to the name recording convention and spelling errors, usually the best candidate can't pass the final test. So we further include a procedure to remove the strings such as "MTG/MORTGAGE" or "CORP/CORPORATION" in the lender company names and keep the strings that has distinctive power, e.g., we use SWBC in "SWBC MTG CORP" and "SWBC MORTGAGE CORPORATION". The matching algorithm runs recursively until no left candidates pass the test.

[12]We also use loan term as additional information for matching (only for 2018-2020). HMDA data disclosures have additional information such as loan term starting in 2018, we have loan term information in HMDA (potentially giving more details about the changes of the reporting requirement of HMDA).

Table 13: Matching rates for different MSAs in different tax years

	MSA					
	Phoenix	Atlanta	Dallas	Houston	Austin	San Antonio
Tax year						
2004-2013	0.682	0.655	0.642	0.655	0.625	0.646
2014-2017	0.799	0.789	0.698	0.722	0.701	0.674
2018-2020	0.767	0.759	0.672	0.703	0.678	0.626

Note: The matching rates vary over years due to the change in the format/quality of HMDA files.

We note that the unmatched is usually no match due to removal of records due to missing information. During the data preprocessing, a lot of records are lost due to missing key information, such as sale date, sale amount, etc.

While the ideal matching rate is calculated by the matched number of records divided by the maximum number of matched records, our reported matching rates are the approximate estimate as we use the total number of transactions to approximate the maximum number of matched records [13] We compare the results of our matching algorithm with manual matching for several census tracts in different metro areas and tax years, our algorithm recover an average of 98% of the matches. Table 14 shows the home characteristics for the matched and unmatched sample. The statistics for the matched and unmatched samples are close within MSA as shown in Table 14.

Table 14: Comparison of statistics for matched and unmatched samples

	Phoenix		Atlanta	
	Matched	Unmatched	Matched	Unmatched
Housing characteristic				
Year built	1991.6	1990.9	1991.1	1991.5
Square footage	2032.7	2016.1	2427	2336.9
Lot Size	12920.7	11469.6	32072.1	32001.1
Total rooms	6.6	6.5	7.1	7
Assessed value	208,195	202,802	281,971	259,812

Notes: We report the summary statistics separately due to the heterogeneity across metro areas. We exclude Texas here as most of the information in the Texan records are missing.

[13] Recall that we lost a lot of records during data preprocessing.