Boston University Theses & Dissertations

2025

Essays in real estate and urban economics

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

ESSAYS IN REAL ESTATE AND URBAN ECONOMICS

by

LEI MA

B.S., University of Minnesota, 2015
M.A., University of Chicago, 2017
M.A., Boston University, 2023

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2025

Approved by

First Reader ______________________________
Adam Guren, PhD
Associate Professor of Economics

Second Reader ______________________________
Marc Rysman, PhD
Professor of Economics

Third Reader ______________________________
Kevin Lang, PhD
Laurence A. Bloom Professor of Economics

Fourth Reader ______________________________
Jihye Jeon, PhD
Assistant Professor of Economics

DEDICATION

To my parents

ACKNOWLEDGMENTS

I am deeply grateful to my advisors, Adam Guren, Marc Rysman, Kevin Lang, and Jihye Jeon, for their mentorship and support throughout my PhD. They have generously shared their knowledge and time at every step of my research and guided my growth as a scholar. I am also thankful to the many other professors and mentors who have offered me invaluable guidance and advice, including Pascal Noel, Peter Ganong, Christopher Palmer, Yuhei Miyauchi, Stefania Garetto, and Natalia Ramondo. Most importantly, I would like to thank my husband, parents, and friends for their love and unwavering support. They listened patiently to my complaints and encouraged me through some of the most stressful moments.

ESSAYS IN REAL ESTATE AND URBAN ECONOMICS

LEI MA

Boston University, Graduate School of Arts and Sciences, 2025

Major Professors: Adam Guren, Associate Professor of Economics
Marc Rysman, Professor of Economics

ABSTRACT

This dissertation consists of three chapters on real estate and urban economics. They examine how market forces and government policies affect distributional outcomes in the residential real estate market.

The first chapter examines the shift in new construction toward larger, more expensive homes. I study the causes of this pattern and evaluate the equilibrium impacts of proposed housing policies aimed at improving affordability at the lower end of the market. I develop an equilibrium model of segmented housing markets with two key features: (1) heterogeneous household preferences for housing quality by demographics, and (2) endogenous housing supply with heterogeneous development costs by housing quality. Using microdata on household housing choices and parcel-specific development costs for single-family homes in the Atlanta MSA, I find that the shift toward large home construction is partly driven by demand from high-income households, who are less price-sensitive and prefer larger homes, but zoning density restrictions play a more significant role in limiting the construction of smaller homes. Relaxing these restrictions could expand the supply of small homes and benefit low-income households, but such zoning reforms are often politically challenging. As an alternative, I evaluate the impact of recently proposed housing subsidies targeting first-time homebuyers and starter

homes. The model predicts that subsidizing young, low-income households provides substantial targeted welfare gains to recipients but hurts others due to rising prices. By contrast, subsidies to small home construction increase the supply of small homes but crowd out the construction of larger homes, resulting in modest welfare gains without effectively targeting those most in need.

The second chapter, co-authored with Leonel Diego Drukker, examines the financial disparities Black homeowners face during the home-selling process. Using repeat-sale transactions from 2003 to 2020, we document that Black homeowners earn, on average, 0.36% lower annualized unlevered returns than non-Black sellers for similar properties. These racial disparities in housing returns cannot be explained by seller characteristics, property renovations, the buyer's race, seller agent fixed effects, and appraisal measures. However, we find significant racial gaps in listing prices and time on market, which we attribute to intermediaries involved in housing transactions. Controlling for these factors reduces the racial gap in returns to effectively zero. Additionally, we find that when homes are sold to iBuyers, where human intermediary bias is removed, the racial gap in housing returns disappears. Our findings suggest that Black sellers experience higher search frictions, leading to worse selling outcomes.

The third chapter focuses on the entry of large institutional investors into the single-family rental market. Since 2012, institutional buy-to-rent (B2R) investors have entered the single-family rental market, converting a substantial number of owner-occupied homes into rental properties. This study examines the impact of B2R investors on nearby property values, providing reduced-form evidence on the size and origin of spillover effects resulting from their presence. I find that an additional property by B2R investors within 150 meters increases housing price growth by 2-3%. The impact is more pronounced in neighborhoods with a higher share of Black residents and lower property values. The reduced supply of owner-occupied housing and improvements in local amenities are key factors driving the positive spillover effects on housing prices.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ACS	American community survey
B2R	Buy-to-rent
BLP	Berry, Levinsohn, and Pakes (1995)
CBSA	Core-based statistical area
FE	Fixed effects
GSE	Government sponsored enterprise
HMDA	Home mortgage disclosure act
HUD	Department of housing and urban development
IO	Industrial organization
IV	Instrumental variable
MLE	Maximum likelihood estimation
MSA	Metropolitan statistical area
NAHB	National association of home builders
PDF	Probability density function
SFH	Single family homes
ZCTA	ZIP code tabulation areas

CHAPTER 1

Build What and for Whom? The Distributional Effects of Housing Supply

1.1 INTRODUCTION

Housing prices have increased significantly over the past decade, making homeownership increasingly out of reach for many low- and middle-income families. To address this, housing advocates and policymakers have pushed for policies to encourage new housing development. However, it is not just the quantity of new housing that matters. Housing markets are segmented by location, property characteristics, and prices, catering to the diverse needs and preferences of different households. This segmented market structure implies that the types of new homes being built have distributional implications, affecting different households differentially.

Most newly constructed single-family homes are larger and more expensive than existing homes. This tendency toward expanding supply in the higher-end market has raised doubts about the effectiveness of market-driven supply in addressing affordability at the lower end of the market. The filtering and "trickle-down" hypotheses (Rosenthal (2014), Sweeney (1974), and Nathanson (2023)) posit that building high-end housing could eventually benefit lower-income households through equilibrium adjustments or depreciation over time. However, critics argue that these effects are often limited, especially in the short term, and more evidence is needed to quantify the extent of these equilibrium forces across market segments. At the same time, there is limited empirical evidence on why private developers prioritize high-end construction. Is this supply-side response driven by differential demand or by supply-side restrictions on small home production, such as zoning regulations?

In this paper, I quantify the causes and distributional consequences of the shift toward larger homes in new construction and evaluate policy proposals aimed at improving affordability in the lower-end market segment. My analysis focuses on the single-family housing market in the Atlanta metropolitan area. I begin by documenting the growing size of newly constructed single-family homes and its correlation with local demand conditions and zoning regulations. Motivated by these descriptive facts, I build and estimate an equilibrium model of the housing market. This model incorporates heterogeneous demand by demographics over housing location and quality. It also includes an endogenous supply of new housing and accounts for variation in housing production costs. Using the model, I disentangle the roles of heterogeneous demand and zoning restrictions in affecting housing production and prices across different house sizes. Lastly, I evaluate the impact of two subsidy policies proposed by the Harris presidential campaign to address affordability issues at the lower end: a demand-side subsidy targeting first-time homebuyers and a supply-side subsidy incentivizing developers to build smaller homes.

I link several micro datasets to capture both the demand and supply sides of the housing market. I match property-level characteristics from CoreLogic with individual-level demographics from voter registration records using addresses. This matched household-property dataset allows me to observe each household's housing choice at a granular level. I also collect parcel-level construction costs for new residential construction from the City of Atlanta, zoning codes from CoreLogic, and neighborhood-level land value estimates from Davis et al. (2021). These cost data allow me to account for variation in housing production costs by location and housing quality on the supply side. The final matched sample includes over 1.3 million households and their residences in the Atlanta metropolitan statistical area (MSA), as of 2019.

I first document a significant shift toward larger homes in new construction. Since 2010, the median size of newly built homes has increased by nearly 30% compared to those built before 2010. Meanwhile, smaller homes experienced much higher price growth than larger ones between 2010 and 2019. This uneven price growth suggests that smaller homes are segmented from larger ones, limiting the impact of supply expansion in large homes on the lower-end market segment.

I also find that new construction is concentrated in high-demand neighborhoods with higher socioeconomic status. Moreover, new homes tend to be larger in areas with stricter zoning regulations. These descriptive facts suggest that local demand conditions and zoning rules are crucial in shaping developers' supply decisions. One possible explanation for the limited construction of smaller starter homes in neighborhoods with lower socioeconomic status is that buyers in these areas tend to be less willing to pay high prices, which results in lower revenue for developers. Additionally, exclusionary zoning may impose regulatory burdens that discourage developers from building small homes.

Based on these facts, I develop an equilibrium model that captures the segmented nature of the housing market and incorporates the roles of demand and zoning regulations in shaping developers' construction decisions. The model treats housing as a differentiated product characterized by their price, location, building age, and house size. On the demand side, households choose where and which type of home to live in within a metropolitan area, balancing the trade-off between housing price and quality. I use a mixed logit demand model, following Bayer et al. (2007) and McFadden (1978), to capture the rich heterogeneity in preferences for housing characteristics by household demographics. Incorporating this heterogeneity allows me to quantify the distributional impact of housing supply on different households. This framework also flexibly captures indirect substitution patterns across housing types and locations, which is crucial for quantifying the equilibrium forces.

On the supply side, competitive developers choose whether to develop and which type of new homes to build at the parcel level to maximize profits. I apply a discrete choice framework in which developers weigh the trade-off between expected housing prices and development costs for each potential house type. Development costs include land acquisition costs, physical construction expenses, and any additional costs imposed by local zoning regulations. Zoning costs capture fees paid to local governments, such as building permits and impact fees, as well as indirect costs, such as the time and effort required to comply with zoning rules and restrictions on permissible uses for each parcel. In equilibrium, households choose from the entire set of existing homes and new construction, with developers' supply decisions endogenously driven by households' willingness to pay for different types of new housing. Prices adjust to equalize demand and supply across all locations and house types.

I estimate the model using a revealed preference approach, matching model-predicted choices to observed choices in the data as closely as possible. To identify household preference parameters, I leverage variation in observed housing choices across different demographic groups. I address price endogeneity using an instrument based on the availability of similar housing nearby, following Berry et al. (1995) and Bayer et al. (2007). To estimate zoning regulatory costs, I apply the residual approach from Glaeser & Gyourko (2003) and Glaeser et al. (2005) to infer them from the gap between prices and observed production costs. I identify heterogeneous zoning costs by exploiting variation in development choices across parcels and house types.

The estimates reveal substantial heterogeneity in household preferences: Higher-income households are less price-sensitive and more willing to pay for newer larger homes in neighborhoods with better amenities. Moreover, there is significant variation in zoning costs, with higher costs in neighborhoods with higher housing prices, fewer minority residents, and closer proximity to the city center. Additionally, zoning costs depend on the interaction between house size and lot size. Building large homes, compared with small homes, on small parcels incurs higher zoning costs, primarily due to violations of setback and height restrictions. As lot sizes increase, the zoning costs for small homes rise, reflecting that neighborhoods favoring large parcels often impose restrictions, such as minimum house size requirements, to exclude smaller, more affordable developments.

Using the estimated model parameters, I simulate counterfactuals to examine how preference heterogeneity and endogenous supply contribute to the unequal distribution of housing supply and price growth. Specifically, I explore the role of rising income inequality. As the number of high-income households in a city grows, demand for larger homes increases, putting upward pressure on prices. Developers endogenously respond to this demand shift by constructing more larger homes to maximize profits. To quantify the extent to which rising top income growth contributes to the increase in house sizes, I simulate a counterfactual using the 2010 income distribution while keeping the total population and other demographic variables constant as of 2019. The results show that a 5 percentage point decrease in the share of households earning over $100,000 would result in a 4% decline in large home construction. However, the number of small homes constructed would increase by less than 1%, suggesting that supply-side constraints play a crucial role in restricting the production of these homes.

Next, I examine how zoning regulations, particularly density restrictions, contribute to the uneven growth in housing supply. These restrictions often limit the number of homes allowed per acre or impose minimum lot size requirements. I simulate a counterfactual that relaxes these density rules, letting developers build two properties on each vacant parcel. The results show that this alternative zoning environment shifts production toward smaller homes, resulting in a 79% increase in the number of new small homes compared to the baseline level. It also nearly doubles the number of parcels available for development, significantly increasing housing production at the extensive margin and raising the total housing stock by almost 6%. The additional supply lowers housing prices across all segments, with the largest reduction among small homes. These findings suggest that strict density restrictions discourage small home construction and contribute to the rising prices of small homes.

Last, I evaluate the impact of housing subsidies proposed by the Harris presidential campaign aimed at improving affordability in the lower-end segment in response to the limited production of small starter homes and rising prices. Providing a $25,000 subsidy to households under age 35 with income below $75,000 generates over $20,000 in welfare gains per subsidized household by enabling them to buy homes they otherwise could not afford. However, this demand-side subsidy leads to welfare losses of $1,500 per household for those not receiving the subsidy, as the positive demand shock drives up prices across all housing segments, with the most pronounced increase in smaller homes preferred by subsidized households. Notably, the demand shock results in only a 1% price increase and minimal supply response, as only 7.2% of households qualify for the subsidy.

In contrast, providing $25,000 to developers for small home construction increases the construction of small homes by 60% while reducing the construction of medium and large homes by 19% and 10%, respectively. As the subsidy makes small home construction more profitable, it crowds out the construction of larger homes and also encourages more development at the extensive margin. Despite the strong response in construction, the aggregate housing stock experiences small changes: The total number of small homes rises by 4%, leading to a price reduction of less than 1% for small homes. The modest changes in aggregate supply and prices translate to average welfare gains of approximately $600-$860 per household. These welfare gains are evenly distributed among all households, including high-income households with less need for government assistance, rather than being targeted toward a specific household group.

The positive analysis of the two subsidies highlights the trade-offs embedded in each policy. Under the demand-side subsidy, every dollar spent generates $0.83 in welfare gains for all subsidized households but leads to $0.73 in welfare losses for all non-subsidized households. The supply-side subsidy results in $0.56 in welfare gains for all households per dollar spent. The choice between these policies depends on policymakers' objectives — whether to directly support specific households, stimulate supply-side development without targeted assistance, or combine both approaches. This paper shows that expanding the housing supply, particularly by building more smaller starter homes, helps lower prices in the lower-end market segment. However, these supply expansions are often small relative to the aggregate housing stock. Therefore, combining them with complementary policy tools, such as demand-side subsidies, can better address affordability concerns at the lower end of the housing market.

Related literature This paper contributes to a growing literature on the impact of new construction on the housing market. Many papers have examined the local spillover effects due to new housing projects, such as Li (2021), Pennington (2021), Asquith et al. (2023), and Damiano & Frenier (2020). In this paper, I focus on the *equilibrium* impact of new construction across various housing segments. In this regard, it aligns more closely with studies that use quantitative models to measure equilibrium effects arising from changes in the housing stock or housing policies, such as zoning reforms (Anagol et al. (2023); Song (2021)), public housing demolitions (Almagro et al. (2023)), the entry of Airbnb (Calder-Wang (2021)), rent control (Favilukis et al. (2023)), credit supply shocks (Landvoigt et al. (2015)), and rental supply shocks (Anenberg & Kung (2020)). I complement this work by showing that expanding supply in one segment can lead to price changes in other segments, depending on the degree of segmentation.

This paper focuses on the cross-sectional variation in housing quality within a metro area. I go beyond merely assessing the impact of the *quantity* of new construction by further investigating the causes and consequences of building new homes at different *quality* levels. Nathanson (2023) also focuses on the quality of new housing, using an assignment model to summarize housing attributes into a single index. He uses the model to analyze the scarcity of high-quality houses in superstar metros. In contrast, I treat housing quality as a multidimensional object defined by location, year built, and house size, and I study the abundance of high-quality houses in high-growth metros like Atlanta. I also examine the supply decisions of developers regarding the number and characteristics of new homes.

The supply model builds on the literature on the micro-foundation of housing supply. I use a discrete choice framework to model how developers supply new housing at the parcel level. The setup is similar to Murphy (2018), Tokman (2023), and Diamond et al. (2024), where profit-maximizing developers decide whether to develop and what to build on each given parcel based on expected prices and costs. Different from these papers, I separate the components of housing production costs and estimate the costs attributed to zoning regulations. I estimate the zoning regulatory costs using the residual approach proposed by Glaeser & Gyourko (2003) and Glaeser et al. (2005), that is, inferring the implicit zoning tax as the gap between housing prices and physical construction costs. Most studies aiming to quantify the cost of local zoning regulations employ some form of the residual approach, such as Babalievsky et al. (2023), Tokman (2023), and Dobbels & Tavakalov (2023). Moreover, by integrating supply and demand models into one framework, I quantify the contributions of demand- and supply-side factors to the increase in home sizes in new construction.

This paper also relates to the literature studying the impact of housing subsidies on the housing market. On demand-side subsidies, Berger et al. (2020) find that the temporary tax credit for new homebuyers from 2008 to 2010 boosted housing demand and increased prices. Eriksen & Ross (2015) and Susin (2002) show that rental vouchers increase housing consumption of recipients but drive up prices in supply-inelastic areas and cause welfare losses to unsubsidized households. Consistent with these findings, my model predicts that demand-side subsidies benefit recipients but generate positive demand shocks and raise housing prices. On supply-side subsidies, Sinai & Waldfogel (2005), Eriksen & Rosenthal

(2010), and Soltas (2024) find that subsidized housing crowds out some unsubsidized housing, resulting in only a small net increase in total supply. Similarly, my model predicts that subsidizing the construction of small single-family home would crowd out the construction of larger homes. The results from my model are qualitatively consistent with existing empirical research and highlight the trade-offs in these subsidy policies.

The paper also contributes to the literature on the transmission of local shocks through connections between different housing segments. Piazzesi et al. (2020) uses housing search data to show that broad housing searchers spread local shocks to other connected segments. Mast (2023) and Bratu et al. (2023) use individual-level address history data to show that new construction in one neighborhood triggers moving chains, affecting housing prices in other neighborhoods. Schubert (2016) focuses on cross-region spillovers and finds that migrants transmit housing shocks from their origin cities to their destination cities. Greenwald & Guren (2021) focus on the degree of segmentation between rental and owner-occupied housing markets. Rosenthal (2014) and Liu et al. (2022) use a repeat-income model to measure the long-term filtering dynamics of housing, driven by house depreciation over time. I use a quantitative model to show how a supply shift in one segment affects prices in other segments and household welfare.

This paper is also related to papers examining the impact of zoning reforms. Many papers have empirically evaluated the effects of zoning reforms on housing supply, such as rezoning near public transportation in Chicago (Freemark (2020)), increased density in So Paulo, Brazil (Anagol et al. (2023)), affordability housing mandates in Seattle (Krimmel & Wang (2023)), the permission to build accessory dwelling units in Vancouver (Davidoff et al. (2022)), and upzoning in Auckland

(Greenaway-McGrevy & Phillips (2023)) and Zurich (Büchler & Lutz (2024)). I focus on how existing zoning regulations differentially affect housing production and prices for homes of different sizes.

The rest of the paper proceeds as follows. Section 1.2 describes the data used in the analysis. Section 1.3 presents the empirical facts that motivate the research question and the design of the structural model. Section 1.4 outlines the setup of the structural model, while Section 1.5 explains the estimation procedure and presents the estimation results. Section 1.6 presents the findings of counterfactual exercises. Section 1.7 concludes.

1.2 DATA

I compile a micro-dataset on the Atlanta-Sandy Springs-Marietta metropolitan statistical area (MSA), the sixth-largest metro area in the US. Throughout my analysis, I focus on single-family homes, which represent more than 72 percent of all housing units in this region (shown in Appendix Figure A.1) and the most common housing structure for homeownership.[1] Between 2010 and 2019, the Atlanta MSA experienced substantial new construction of single-family homes, with over 135,000 new homes built. These new homes make up two-thirds of all newly constructed housing units, including multi-family housing units (Appendix Figure A.1), and account for 10% of the total number of single-family homes as of 2019.

[1] It is important to study the single-family home market because it is the main pathway to homeownership and makes up the majority of the housing stock in the US. In the Atlanta MSA, about 80% of the total population lives in single-family homes. Rising housing prices have made it more difficult for many households to afford these homes. Unlike the multi-family housing sector, which receives government support through tax credits, public housing programs, and rental subsidies for extremely low-income households, the supply of single-family homes has largely been left to market forces. This paper focuses on the single-family sector to better understand the challenges in this market and the need for policy interventions.

Property characteristics and values Housing characteristics and values at the property level are from CoreLogic, a leading provider of real estate data in the US. CoreLogic compiles comprehensive property records from county assessors, which include detailed property characteristics such as property type, living area square footage, lot size, number of bedrooms, number of bathrooms, year built, street address, zoning district, and owner names. It also maintains a deeds database of property transaction histories, including sale prices and dates. To impute housing values of properties not transacted, I employ a machine learning approach using XGBoost[2] (details are in Appendix A.3.1). This method outperforms traditional hedonic pricing and random forest models, providing more accurate predictions of housing values.

Individual demographics and addresses I obtain individual-level demographics and residential addresses in 2019 from L2, a commercial data vendor that collects and standardizes voter registration records. The dataset includes each registered voter's exact residential address, name, age, gender, and race, as well as imputed income and education levels derived from L2's proprietary models.

To validate the quality of the dataset, I compare the demographics from L2 with those from the American Community Survey (ACS) at both the metro and Census tract levels in 2019 (as detailed in Appendix A.3.2). Overall, L2's data closely align with the Census data, indicating a high level of data quality. On average, individuals in the L2 voter sample tend to be older and have higher levels of education and income compared to those in the ACS sample. This demographic profile aligns well with households residing in single-family homes, making the voter sample

[2]XGBoost, or Extreme Gradient Boosting, is a machine learning algorithm that predicts outcomes by using input data features. It combines multiple sequential decision trees, where each tree corrects the errors of the previous ones.

well-suited for my analysis. It is also important to note that L2 does not include immigrants, resulting in an under-representation of Asians and Hispanics in the dataset. Therefore, my analysis classifies households as Black and non-Black without further differentiating Asians and Hispanics.

I merge L2 and CoreLogic data using property addresses to build a dataset capturing household-level housing choices. This micro-dataset allows me to capture the granular heterogeneity of housing preferences based on property characteristics and locations. I successfully match 89% of individuals from L2 to CoreLogic. The remaining unmatched individuals are due to misspelled addresses, non-residential addresses in L2 (such as PO boxes), and addresses in multi-family buildings. The merging process is described in detail in Appendix A.3.3. I group individuals into households, with each household defined as all individuals residing at the same property.

Neighborhood characteristics I compile a comprehensive set of neighborhood characteristics at the ZIP Code Tabulation Area (ZCTA) level from various sources. I obtain data on the socioeconomic composition from the American Community Survey (ACS), accessed from IPUMS NHGIS, including racial composition, population density, education, and household income. I also collect data from the National Neighborhood Data Archive (NaNDA), including the share of land developed at different density levels, the number of parks, public transit stops, and the proportion of primary and secondary road length. From County Business Patterns, I collect the number of restaurants and the number of arts and entertainment venues. From the Department of Housing and Urban Development (HUD), I obtain job proximity and school proficiency indices as proxies for labor market access and school quality.

Housing development costs I obtain property-level construction costs for new residential construction since 2010 through web-scraping the building permits database from the City of Atlanta. Each permit provides detailed information about the new construction, including the estimated construction cost, square footage, lot size, address, parcel number, parcel owner's name, and developer's name.[3]

To estimate land costs for new construction, I rely on land values calculated by Davis et al. (2021) (hereafter referred to as DLOS). DLOS estimate residential land values at the ZIP code level by calculating the difference between home appraisal values and the replacement cost of the housing structure, using home appraisals submitted to the Government Sponsored Enterprises (GSEs).

Final sample I merge property-level housing characteristics, housing values, house-hold demographics, housing costs, zoning regulations, and neighborhood charac-teristics to build a dataset covering both housing demand and supply informa-tion for each property and parcel in 2019. To limit the impact of extreme values, I winsorize housing price, square footage, lot size, and income at the 1st and 99th percentiles. I also cap the number of bedrooms and bathrooms at six.

The final sample includes over 1.3 million households and their residences in 2019. Table 1.1 reports the summary statistics of the sample. The median housing price in the Atlanta MSA is around $230,000. New homes built after 2010 have higher prices, with a median price of $326,508, whereas those built before 2010 have a median price of $218,847. A median home has three bedrooms and three bathrooms, with approximately 2,000 square feet of living space on a 0.5-acre lot. The housing stock is also relatively young, with a median construction year of

[3]The construction cost estimates are submitted by developers and compared with the estimates from Building Plan Reviewers, who oversee and regulate construction activities.

1996.

1.3 MOTIVATING FACTS

Before delving into the model, I present motivating facts about the changing characteristics of new homes, their impact on prices, and their correlation with local demand and zoning regulations.

Fact 1: Newer homes are larger and more expensive

As of 2019, about 10% of the housing stock consists of homes built after 2010, and 29% were built between 2000 and 2009 (as shown in Appendix Figure A.2). Although the homebuilding industry has scaled back its overall production since the Great Recession, there has been a noticeable shift toward building larger homes.

Figure 1.1 plots the empirical distribution of living area square footage for homes built before and after 2010 as of 2019, with the vertical lines representing the median house sizes. Homes built after 2010 are significantly larger, with the entire distribution shifting to the right for these newer homes. The median home size for homes built after 2010 is approximately 2,800 square feet, about 40% larger than the median size of homes built before 2010.

In Table 1.2, I regress a set of property characteristics of property j on an indicator variable for whether the home was built after 2010, controlling for ZCTA fixed effects:

$$PropertyCharacteristics_j = \beta \cdot \mathbf{1}(BuiltAft2010_j) + \alpha_{zcta(j)} + \epsilon_j. \tag{1.1}$$

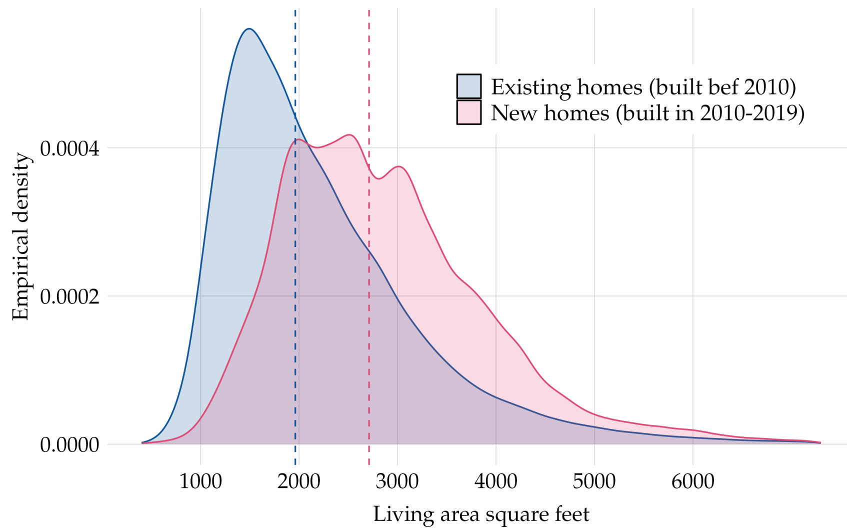
The results indicate that even after controlling for neighborhood fixed effects, newer homes are much larger, with more square footage, additional bedrooms and bath-

Table 1.1: Summary statistics

	Mean	Median	SD	Min	Max
Property characteristics					
Price ($)	272,573	230,673	137,184	80,445	862,240
New homes	365,824	326,508	153,853	147,387	862,240
Existing homes	262,234	218,847	131,173	80,445	852,059
Home size (sq ft)	2308	1977	729	1003	5015
Lot size (acre)	0.73	0.53	0.53	0.11	5.44
Bedroom (#)	3.46	3	0.8	1	6
Bathroom (#)	2.76	2.60	1.0	1	6
Year built	1992	1996	17	1780	2019
Household characteristics					
Income ($)	90,026	77,984	54,009	13,000	250,000
Black (%)	30.8	0	46.2	0	100
White (%)	59.8	1	49.0	0	100
College (%)	44.6	0	49.7	0	100
Married (%)	47.7	0	49.9	0	100
Child present (%)	62.53	1	48.5	0	100
Age <35 (%)	12.7	0	33.3	0	100
Age 35-50 (%)	27.9	0	44.9	0	100
Age 50-65 (%)	33.5	0	47.2	0	100
Age >65 (%)	25.9	0	43.8	0	100
Production costs					
Construction cost ($)	246,129	197,963	166,929	60,000	900,000
Const. cost per sqft ($)	82.1	72.3	33.4	27.2	213.4
Land value per acre ($)	93,731	34,650	188,432	7,881	1,268,411

Notes: This table presents the summary statistics for the analysis sample. Property and household characteristics are calculated from the merged CoreLogic and L2 data in 2019. New homes are defined as homes built after 2010, and existing homes are those built before 2010. Housing prices, square footage, lot size, and household income are winsorized at the 1st and 99th percentiles. The number of bedrooms and bathrooms is capped at six. Housing costs are calculated using data from web-scraped building permits and estimated land values from Davis et al. (2021), with all values winsorized at the 1st and 99th percentiles.

Figure 1.1: Empirical PDF of house size of the 2019 housing stock



Notes: This figure plots the empirical probability density distribution of living area square footage for single-family homes built before and after 2010 in the Atlanta MSA, based on the 2019 CoreLogic Assessor data. The density is plotted using a bandwidth of 100 square feet. The vertical lines represent the median square footage for new and existing homes, respectively. The median size of homes built after 2010 is approximately 2,800 square feet, compared to about 2,000 square feet for existing homes.

rooms, and sold at higher prices compared to older homes with similar characteristics.

Fact 2: Small homes experience higher price growth

Given that most new housing supply is geared toward larger homes, I explore whether the price growth rate differs across homes of different sizes. Figure 1.2 shows a binned scatter plot of annualized housing price changes for repeat-sale properties between 2010 and 2019 while controlling for ZCTA and sale year fixed effects.[4] Price changes are calculated as the log difference in nominal transaction prices divided by the number of years between sales. Smaller homes experience

[4] I exclude distressed sales, potential flips (properties sold within two years), and sales involving major renovations—homes with more than a 10% increase in square footage, with additional bedrooms, or with an annual price growth rate exceeding 50%.

Table 1.2: Characteristics of new homes

Year built	(1) Log sqft	(2) Bedroom	(3) Bathroom	(4) Log price	(5) Log price
Built aft 2010	0.20***	0.45***	0.59***	0.28***	0.10***
	(0.01)	(0.03)	(0.03)	(0.01)	(0.01)
ZCTA FE	✓	✓	✓	✓	✓
Property Attributes					✓
N	1,343,728	1,343,728	1,343,728	1,343,728	1,343,728
R^2	0.44	0.18	0.41	0.79	0.97
Within R^2	0.07	0.04	0.08	0.13	0.91

Notes: This table compares the characteristics of newly built homes to existing homes within ZIP Code Tabulation Areas (ZCTAs). It presents regression results where property characteristics are regressed on an indicator variable of whether the property was built after 2010, controlling for ZCTA fixed effects, as specified in Equation (1.1). The sample consists of all single-family homes in the Atlanta MSA from the 2019 CoreLogic Assessor-L2 linked data used in the main analysis. Columns (1) to (3) use log square footage, number of bedrooms, and number of bathrooms as the outcome variable. Columns (4) and (5) use log house prices as the outcome variable, without and with controls for property characteristics. Homes built after 2010 are bigger and more expensive than existing homes in the same neighborhood. Standard errors are clustered at the ZCTA level. Levels of significance: *p<0.1; **p<0.05; ***p<0.01.

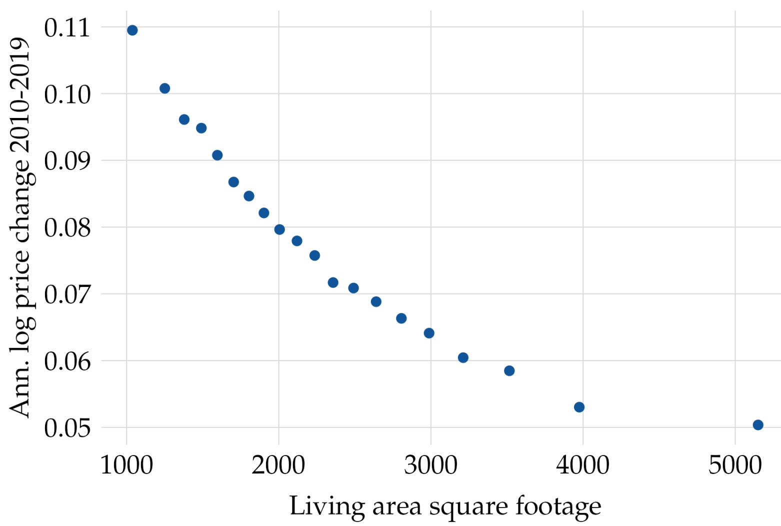
faster price appreciation. For instance, a 1,500-square-foot home experiences an annual price growth rate of approximately 9%, whereas homes larger than 3,000 square feet grow at around 6% annually. To further explore how the housing supply affects price growth, Appendix Table A.2 regresses the price change for each repeat-sale property on the number of new homes of the same housing type in the same zip code. All specifications show a negative relationship between new homes and housing prices.

The differential price changes *suggest* that the increase in newly built large homes exerts downward pressure on large home prices, and the limited construction of smaller homes is associated with higher price growth. However, they do not provide causal estimates on the magnitude of price elasticity, nor do they capture any potential cross-segment spillovers that extend beyond geographical regions and housing types. The goal of the structural model is to capture the degree of segmentation between housing sub-markets and understand how increasing supply in one segment differentially affects housing prices in other segments.

Fact 3: New homes are concentrated in neighborhoods with higher socioeconomic status

Next, I explore the geographic distribution of new homes and the demand factors associated with new construction activities. Panel (a) of Figure 1.3 presents a map of the number of new homes at the Census tract level. While new construction occurs in many neighborhoods, development is concentrated in suburban areas, especially in the wealthier northern and southwestern suburbs. To examine which types of neighborhoods are more likely to attract new construction, I regress the share of new homes in each neighborhood on their 2010 sociodemographic char-

Figure 1.2: House price change of repeat-sale properties from 2010 to 2019



Notes: The figure plots the annualized price growth for single-family homes with different square footage between 2010 and 2019 using CoreLogic transaction records. The sample uses repeat-sale properties sold at least twice between 2010 and 2019. Price changes are calculated as the log difference between the two transaction prices, divided by the number of years between sales. The binned scatter plot shows how price changes by home size, controlling for ZIP code sale year fixed effects. Distressed sales, flips, and properties with major renovations are also excluded from the analysis.

acteristics separately:

$$\frac{NewHome_n}{TotalHomes_n} = \beta \cdot X_n^{2010} + \epsilon_n. \quad (1.2)$$

Panel (b) of Figure 1.3 reports the coefficients on different neighborhood characteristics at the Census tract level.[5] New housing developments are more concentrated in areas with higher shares of White, college-educated, and high-income residents, as well as in areas with higher initial house values. This suggests that developers target areas with higher potential demand, where residents are financially capable and willing to pay for new housing. It also implies that developers may adjust housing features to cater to the preferences of households drawn to these neighborhoods with higher socioeconomic status.

This finding highlights the crucial role of demand factors in private developers' decisions. Neighborhoods with higher socioeconomic status tend to attract higher-income households, thereby providing developers with higher expected revenue in these areas. One goal of the structural model is to assess the extent to which demand-driven factors affect the quantity and quality of new housing developments.

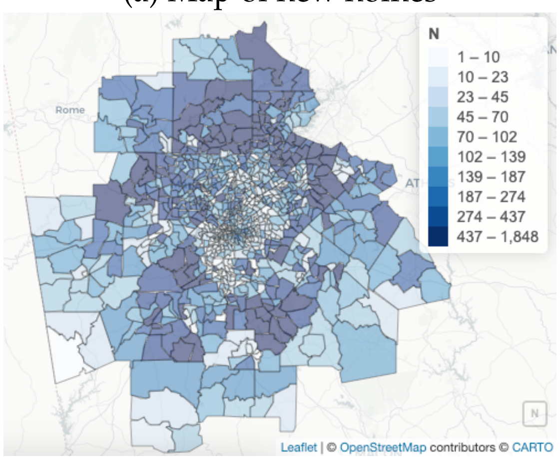
Fact 4: Larger homes are associated with more stringent zoning development standards

Housing markets are heavily regulated. Local zoning authorities determine the types of structures permitted in different zoning districts and often impose detailed requirements on housing characteristics, such as lot size, building height,

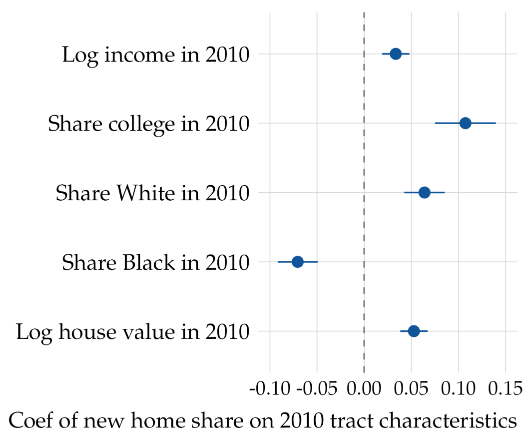
[5] Appendix Figure A.7 reports the coefficients measured at the ZCTA level. The results give the same qualitative implications as the Census tract-level results.

Figure 1.3: Geographic distribution of new homes

(a) Map of new homes



(b) New homes and Tract characteristics



Notes: These two figures show the geographical variation in new homes in the Atlanta MSA. Panel (a) shows a map of newly built single-family homes in each Census tract within the Atlanta MSA. New homes are defined as those built after 2010. Darker shades represent tracts with higher numbers of new homes. Panel (b) displays coefficients from separate regressions examining the relationship between the share of new homes in each Census tract and 2010 neighborhood characteristics, based on the regression equation specified in Equation (1.2). The horizontal bars represent the 95% confidence intervals calculated with robust standard errors.

setbacks, and densities.[6] Therefore, I explore how various zoning development standards shape house sizes. I focus on three common standards in the Atlanta MSA: minimum lot size, minimum house size, and maximum density allowed. I regress new home characteristics for each property j on each zoning standard applicable to its parcel while controlling for zoning jurisdiction fixed effects α_m:

$$NewHomeCharacteristic_j = \beta \cdot Zoning_j + \alpha_m + \epsilon_j. \tag{1.3}$$

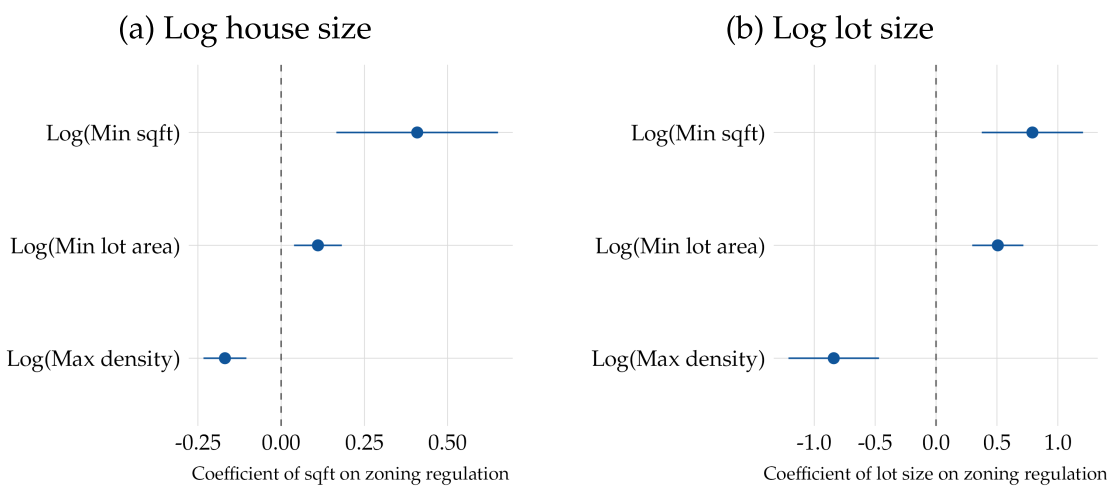
Panel (a) of Figure 1.4 uses the log of square footage of new homes as the dependent variable. The results show that houses are larger on parcels with higher minimum house sizes, higher minimum lot sizes, and lower density allowances. As house size and lot size are often positively correlated, panel (b) runs the same regression but replaces the outcome variable with the log of lot size. The results are consistent with panel (a) that parcels with more stringent zoning rules tend to be bigger and thus more likely to construct bigger homes, as building small homes on large parcels with higher land costs is less profitable.

The evidence suggests that zoning effectively regulates house sizes and plays a crucial role in shaping new house characteristics. It also suggests that zoning may impose different costs on developers for building homes of different sizes.[7] The goal of the structural model is to capture these heterogeneous zoning costs and assess their impact on housing supply.

[6] Bartik et al. (2023) apply Large Language Models (LLM) to zoning documents to analyze the key features of zoning regulations. They find that density restrictions, particularly minimum lot size requirements, are prevalent and strongly correlate with exclusionary zoning.

[7] Zoning regulations are not rigid rules that cannot be bypassed. In practice, developers can apply for exemptions or rezoning if the proposed project is not compliant with local zoning rules. However, this process requires additional effort and incurs higher development costs. For example, a developer may request a variance to build smaller homes on lots below the minimum lot size requirement.

Figure 1.4: Correlation between new home characteristics and zoning standards



Notes: These two figures show the correlation between the characteristics of new homes and zoning development standards in the Atlanta MSA. The analysis focuses on three common zoning regulations in the region: minimum lot size, minimum house size, and maximum density allowed. Panel (a) presents coefficients from regressions where the log of observed house size (in square footage) is regressed on these zoning standards, as specified in Equation (1.3). Panel (b) runs the same regression but uses the log of observed lot size as the dependent variable. All regressions include zoning jurisdiction fixed effects. Standard errors are clustered at the zoning jurisdiction level. The horizontal bars represent the 95% confidence intervals.

1.4 AN EQUILIBRIUM MODEL OF THE HOUSING MARKET

The previous facts suggest that building new homes at different quality levels affects housing prices and households differently across market segments. As new homes enter the market, households re-optimize their residential choices, triggering equilibrium effects that ripple through other housing segments. From the developers' perspective, the quantity and quality of new construction are influenced by local demand conditions and zoning regulations. Motivated by these facts, I build a structural model of the housing market to understand the causes and distributional consequences of new construction.

The model incorporates heterogeneity in both housing preferences and housing supply. On the demand side, I use a framework similar to Bayer et al. (2007) and McFadden (1978) to account for heterogeneous household preferences over housing quality. This approach allows me to capture realistic substitution patterns between different house types and evaluate the distributional consequences under alternative housing supplies. On the supply side, I endogenize the number of new homes built at each quality level. I explicitly model how developers balance the trade-off between housing prices and costs. The supply model enables me to conduct general equilibrium analysis in response to shifts in demand and costs.

1.4.1 Housing demand

I use a static discrete choice framework to model how households make housing choices. The city contains $\mathcal{I}$ households, where each household $i \in \mathcal{I}$ chooses a housing unit from a discrete set of housing segments $h \in \mathcal{H}$ available in the metropolitan area. Housing segments are categorized by their neighborhood location $n \in \mathcal{N}$, house type defined by bedroom count $k \in \mathcal{K}$, and year built $t \in \mathcal{T}$.

Houses in different segments are vertically differentiated by neighborhood characteristics, building age, number of rooms, living area square feet, and lot size. Houses within the same segment share the same characteristics and are no longer vertically differentiated. Households also face the outside option of living in the outlying counties of the metropolitan area.[8] The indirect utility of choosing the outside option is normalized to zero.

The utility of household i choosing a housing unit j from segment h is

$$u_{i,j,h} = \alpha_i p_h + \mathbf{X}_h \boldsymbol{\beta_i} + \xi_h + \epsilon_{i,j}. \tag{1.4}$$

Household utility depends on housing prices p_h, a K-dimensional vector of observed characteristics of segment $\mathbf{X}_h$, unobserved segment quality ξ_h, and household-specific idiosyncratic taste shocks $\epsilon_{i,j}$. I include both property and neighborhood characteristics in $\mathbf{X}_h$. Property characteristics include year built, house size in square footage, lot size, and the interaction between year built and house size. Neighborhood characteristics include population density, the share of Black residents, the share of college-educated residents, average household income, the share of incorporated areas, distance to the city center, school quality, local labor market access, land development intensity, access to highways and major roads, public transportation availability, park area, woods area, number of restaurants, and number of arts and entertainment establishments.

I allow households' preferences for housing to vary by their demographics $\mathbf{Z}_i$, including age, race, income, education, and whether children are present. This

[8]An alternative definition of the outside option is living in multi-family buildings, which are mostly renter occupied. The results are unlikely to change significantly if rentals are used as the outside option, as the demographics of households choosing these two options are very similar. In particular, these households tend to have lower income than those choosing the inside options.

introduces correlation in individual housing preferences based on demographic profiles. Specifically, the preference coefficients are decomposed into two terms,

$$\alpha_i = \alpha^0 + \sum_{r=1}^{R} \alpha^r z_i^r, \quad \boldsymbol{\beta_i} = \boldsymbol{\beta^0} + \sum_{r=1}^{R} \boldsymbol{\beta^r} z_i^r.$$

The terms $(\alpha^0, \boldsymbol{\beta^0})$ are linear coefficients that are common to all households, revealing how much households value prices and each housing attribute on average, and $(\alpha^r, \boldsymbol{\beta^r})$ represents heterogeneous coefficients varied by demographic characteristics z_i^r, capturing the extent to which each demographic group values prices and housing attributes differently from the average. The utility can then be rewritten as

$$u_{i,j,h} = \delta_h(p_h, \mathbf{X}_h, \xi_h; \alpha^0, \boldsymbol{\beta^0}) + \mu_{i,h}(p_h, \mathbf{X}_h, \mathbf{Z}_i; \alpha^r, \boldsymbol{\beta^r}) + \epsilon_{i,j},$$

where

$$\delta_h = \alpha^0 p_h + \mathbf{X}_h \boldsymbol{\beta^0} + \xi_h \tag{1.5}$$

represents the utility part of housing segment h that is common to all households, and

$$\mu_{i,h} = \left(\sum_r \alpha^r z_i^r \right) p_h + \left(\sum_r \boldsymbol{\beta^r} z_i^r \right) \mathbf{X}_h \tag{1.6}$$

represents the heterogeneous component for different households i. Household i chooses a housing unit from segment h if they derive the largest utility from the segment among all choices,

$$u_{i,h} \geq u_{i,h'}, \forall h' \in \mathcal{H} \cup \{0\}.$$

Assuming that the idiosyncratic shocks ϵ follow a type-I extreme value distribution, the probability of household i choosing segment h has a closed-form expression:

$$\Pr(i \text{ chooses } h) \equiv d_{i,h}(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, \mathbf{Z}_{\boldsymbol{i}}; \alpha^0, \boldsymbol{\beta}^0, \alpha^r, \boldsymbol{\beta}^r) = \frac{N_h \exp(\delta_h + \mu_{i,h})}{\sum_{h'} N_{h'} \exp(\delta_{h'} + \mu_{i,h'})}, \quad (1.7)$$

where N_h is the number of housing units in segment h.[9] The aggregate demand for segment h is

$$\mathcal{D}_h = \sum_i d_{i,h}(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, \mathbf{Z}_{\boldsymbol{i}}; \alpha^0, \boldsymbol{\beta}^0, \alpha^r, \boldsymbol{\beta}^r), \quad (1.8)$$

which depends on the empirical distribution of demographics.

1.4.2 Housing supply

Competitive developers choose whether and which type of home to build on each parcel $\ell \in \mathcal{L}$. Each parcel is characterized by its size, location, and development status, and it produces one housing unit. Parcel location is defined by the combination of neighborhood (ZCTA) n and the broader zoning jurisdiction m to which it belongs. Each parcel has one of two development statuses: vacant or already developed with an existing structure, each with distinct land acquisition costs. I assume all new homes are built on vacant parcels.[10]

[9] The inclusion of N_h accounts for the number of housing units available in each segment. This specification is equivalent to including $\ln N_h$ as one of the segment attributes in the utility function, with a coefficient of 1.

[10] It is reasonable to assume that all new homes in the Atlanta MSA are built on vacant land for two reasons. First, I calculate the number of homes built before 2010 twice using the CoreLogic assessor data: once using data from 2010 and then again from 2019. As shown in Appendix Figure A.5, the number of pre-2010-built homes with fewer than three bedrooms decreases by less than 2% between 2010 and 2019, and those with more than five bedrooms increase by 1%. This indicates that while some homes undergo renovations or demolitions, almost all existing homes have remained in place without teardowns. Second, because redevelopment is more likely to occur in areas with constrained land availability and old housing stock, I use the National Land Cover Database (NCLD) to map vacant barren land and the share of developed land at the ZCTA level (Appendix Figure A.8). These maps show that the region has sufficient vacant land for new development.

Developers make development decisions taking parcel characteristics as given.[11] Each developer is a price-taker and lacks the market power to affect market conditions. This assumption is reasonable given the competitive nature of the construction industry, characterized by a large number of small homebuilders.[12] On each parcel, the developer makes a choice s_ℓ — whether to develop the parcel, and if so, what type of new home $k \in \mathcal{K}$ to build:

$$s_\ell = \begin{cases} 0 & \text{not develop,} \\ k \in \mathcal{K} & \text{develop.} \end{cases}$$

If the developer chooses not to develop the parcel, they receive zero profit. If they choose to develop the parcel, the expected profit depends on the expected prices and costs associated with each specific housing type k built on it. The choice set of house types $\mathcal{K}$ is defined by the number of bedrooms, consistent with the classification used in the demand model. For each house type k, its characteristics are defined by the average number of bedrooms and the interior size of new houses belonging to that type in each neighborhood.

The developer expects to sell the newly built home to a household at the market price $p_\ell^{new}(k)$. To produce one housing unit, developers incur a physical construction cost $CC_\ell(k)$, a land acquisition cost LC_ℓ, and any additional costs due to local

[11]Here, I assume developers take parcel characteristics as given. There is an implicit first stage where developers choose the size of parcels. In Appendix A.2.1, I model the choice of parcel size as a reduced-form function of local zoning regulations, land costs, and land availability. Parcels are larger in neighborhoods with higher minimum lot size requirements, more limited land availability, and higher socioeconomic status.

[12]The 2017 Economic Census shows that Georgia has over 18,000 establishments and 178,000 employees in the construction industry. The national-level data show similar patterns. While a few large homebuilders, like D.R. Horton and Lennar Homes, exist, the vast majority of homes in the Atlanta MSA are constructed by small developers. The perfect competition assumption is also commonly used in the housing production function literature, such as Epple et al. (2010) and Combes et al. (2021).

zoning regulations $RC_\ell(k)$. The developer also faces an idiosyncratic profit shock $\nu_\ell(k)$ that varies by parcel and home type. The expected profit for building a home of type k on parcel ℓ is

$$\pi_\ell(d_\ell = k) = p_\ell^{new}(k) - CC_\ell(k) - LC_\ell - RC_\ell(k) + \nu_\ell(k).$$

Regulatory costs are inferred from the gap between actual housing prices and physical production costs. This residual approach has been used in the housing literature to quantify the cost of local land use regulations, such as Glaeser & Gyourko (2003) and Glaeser et al. (2005). These regulatory costs capture direct fees to local governments, such as building permits and impact fees for new residential construction, and indirect expenses, like the time and effort needed to comply with zoning rules. Additionally, these costs reflect restrictions on permissible land use imposed by zoning rules. Zoning regulations dictate what can be built and set development standards the structure needs to meet, which artificially makes certain structures *illegal*. Developers can seek rezoning or exemptions to bypass these restrictions, but doing so incurs additional costs.[13]

I assume regulatory costs vary by house type, parcel size, neighborhood attributes, house type, and zoning municipality:

$$RC_\ell(k) = \gamma_k L_\ell + \eta \mathbf{W}_{n(\ell)} + \rho_k + \tau_{m(\ell)}.$$

L_ℓ indexes parcel size. I allow the effect of lot size on regulatory costs to differ

[13]In practice, this residual term may capture the additional zoning taxes imposed by various land use regulations and additional construction costs due to terrain features that complicate land development or financing costs. Most of these land-related factors should be capitalized into land costs, and financing costs account for less than 2% of the total price, according to the home building industry report, so a significant portion of variation in the residual term can be attributed to variations in local land use regulations.

across different house types, captured by parameter γ_k. For example, building a large home on a small lot may incur higher regulation costs due to a higher likelihood of violating setback and height restrictions. $\mathbf{W}_{n(\ell)}$ is a vector of neighborhood characteristics that are potentially correlated with regulatory costs, including incorporation status, average housing price, share of Black population, land availability, and distance to city center. The term ρ_k represents the regulatory costs invariant to each house type, and $\tau_{m(\ell)}$ captures regulatory costs specific to each zoning municipality m that parcel ℓ belongs to.

Assuming the idiosyncratic profit shock follows a Gumbel distribution with scale parameter σ_ν, the probability of building a new home of type k on a given parcel ℓ can be written as follows:

$$
\begin{aligned}
\Pr(\text{build home } k \text{ on } \ell) &= s_{\ell,k} \\
&= \Pr(\pi_\ell(k) \geq \pi_\ell(k'), \forall k' \in \mathcal{K} \cup \{0\}) \\
&= \frac{\exp\left((p_\ell(k) - CC_\ell(k) - LC_\ell - RC_\ell(k))/\sigma_\nu\right)}{1 + \sum_{k' \in \mathcal{K}} \exp\left((p_\ell(k') - CC_\ell(k') - LC_\ell - RC_\ell(k'))/\sigma_\nu\right)}.
\end{aligned}
$$

The aggregate supply of new homes of type k in each neighborhood n is the sum of all parcels in neighborhood n, denoted as $\mathcal{L}_n$, that have been developed with type k homes:

$$
\mathcal{S}_{n,k} = \sum_{\ell \in \mathcal{L}_n} s_{\ell,k}(\mathbf{p}, \mathbf{L}, \mathbf{W}; \gamma, \eta, \rho, \tau).
$$

1.4.3 Equilibrium

In equilibrium, demand is equal to supply in every segment, where each segment h is defined by the combination of its location n, house type k, and year built t. Households choose from the entire set of homes, including both existing and newly built homes. Developers decide the number of new homes to build for each house type. The number and characteristics of existing homes are taken as exogenously given. The equilibrium consists of a price vector $\mathbf{p} = \{p_h\}_{h\in\mathcal{H}}$, a distribution of household housing choices $\mathbf{d} = \{d_{i,h}\}_{i\in\mathcal{I},h\in\mathcal{H}}$, and a distribution of developer construction choices $\mathbf{s}_{t=new} = \{s_{n,k,t=new}\}_{n\in\mathcal{N},k\in\mathcal{K}}$ such that (1) households maximize their utility, (2) developers maximize profits, and (3) markets clear. The equilibrium price vector $\mathbf{p}$ clears the market for every segment $h = (n, k, t)$,[14]

$$\mathcal{D}_{n,k,t}(\mathbf{p}; \mathbf{X}, \boldsymbol{\xi}, \mathbf{Z}_{\boldsymbol{i}}, \alpha^0, \boldsymbol{\beta}^0, \alpha^r, \boldsymbol{\beta}^r) = \mathcal{S}_{n,k,t}(\mathbf{p}; \mathbf{L}, \mathbf{W}, \gamma, \eta, \rho, \tau), \forall n \in \mathcal{N}, k \in \mathcal{K}, t \in \mathcal{K}.$$

1.5 MODEL ESTIMATION

I employ a revealed preference approach to estimate the model parameters. The demand model is estimated using the observed housing choices at the household-property level, and the supply model is estimated using the observed development decisions at the parcel level.

1.5.1 Defining housing segments

Given the large number of housing units, treating each unit individually in a discrete choice framework is computationally costly and introduces more noise to the estimates. To balance the bias-variance trade-off, I group properties into segments with similar characteristics. To capture cross-sectional variation in housing quality, I define each housing segment $h \in \mathcal{H}$ by a combination of three key quality

[14] An alternative set of equilibrium conditions can be defined by treating the demographic composition of each neighborhood as endogenous and including it as one of the equilibrium conditions. While the model allows each household's decision to be affected by the residential choices of others, I do not update the demographic composition or neighborhood amenities when calculating the equilibrium. This is based on the assumption that each household is too small to accurately predict the aggregate neighborhood composition resulting from the decisions of all other households.

attributes: neighborhood $n \in \mathcal{N}$, year built $t \in \mathcal{T}$, and house type $k \in \mathcal{K}$. I use ZCTA, which is the Census Bureau's equivalent of ZIP code, to define neighborhood.[15] I use four construction-year bins: built in 2010-2019, 2000-2010, 1980-2000, and before 1980. I classify house types based on the number of bedrooms: small homes (up to three bedrooms), medium homes (four bedrooms), and large homes (five or more bedrooms).[16]

To summarize, houses within each ZCTA are grouped into 12 types. Appendix Table A.1 gives an example of how housing units within one ZCTA are divided into different housing segments. This definition results in 1,755 distinct housing segments.[17] Each segment has 694 houses on average and 275 houses at the median. This granular segmentation allows me to capture the rich heterogeneity across different house types while keeping the model computationally feasible.

Houses in different segments are vertically differentiated by their year built, living area square footage, lot size, number of rooms, and neighborhood characteristics. For example, a three-bedroom home built in 2010-2019 in one neighborhood has different characteristics from a three-bedroom home built in 2010-2019 in another neighborhood. Appendix Figure A.4 presents the distribution of home size

[15]There are three main reasons to use ZCTA to define a neighborhood. First, ZCTAs closely match ZIP codes used by the US Postal Service, making them familiar geographic units for households during their housing searches. Second, neighborhood characteristics are readily available at the ZIP/ZCTA level compared to custom-defined neighborhoods such as Berkeley Park and East Lake. Third, it is computationally efficient to use ZCTAs. Alternatively, researchers can use Census tracts or smaller geographic units to define neighborhoods. However, the Atlanta MSA has over 900 Census tracts, and further grouping homes by year built and house size would yield over 7000 housing segments, requiring the estimation of 7,000 fixed effects in the demand model. Moreover, finer geographic definition might result in many small segments with very few housing units, increasing the variance of estimates. Using ZCTAs balances the bias-variance trade-off and maintains computational feasibility.

[16]I choose these bedroom cutoffs because they are the most common numbers of bedrooms. As shown in Appendix Figure A.3, very few homes have fewer than two or more than six bedrooms.

[17]Some ZCTAs do not have all house types available. I also drop small segments (those with fewer than 10 single-family homes).

in square footage by house type and year built. For existing homes, the interquartile range is 1,603–1,960 square feet for small homes, 2,601–3,036 square feet for medium homes, and 3,433–4,120 square feet for large homes. For new homes, the interquartile range is 1,942–2,270 square feet for small homes, 2,697–3,114 square feet for medium homes, and 3,485–4,073 square feet for large homes.

1.5.2 Demand model estimation

I use a standard two-step estimation procedure to estimate the demand model.[18] In the first step, I estimate the heterogeneous coefficients $(\alpha^r, \boldsymbol{\beta^r})$ and the mean indirect utilities δ_h using maximum likelihood estimation. In the second step, I recover the mean preferences for each housing and neighborhood attribute using an instrumental variable approach.

The first-step maximum likelihood estimation assumes that each household makes their own housing choices independently, taking prices, housing attributes, and neighborhood characteristics as exogenously given. It also implies that each household draws its idiosyncratic error term ϵ independently from other households' draws. This is a reasonable assumption given that each household is small enough not to directly influence other households' choices. The log likelihood function is

$$LL(\boldsymbol{\delta}, (\alpha^r, \boldsymbol{\beta^r})_r) = \sum_i \sum_h y_{i,h} \ln d_{i,h} = \sum_i \sum_h y_{i,h} \ln \frac{N_h \exp(\delta_h + \mu_{i,h})}{\sum_{h'} N_{h'} \exp(\delta_{h'} + \mu_{i,h'})}, \quad (1.9)$$

where y_{ih} indexes the observed housing choice: $y_{ih} = 1$ if household i chooses h, and 0 otherwise.

[18] The same estimation method has been widely used in the literature to solve this type of mixed logit demand models, such as Bayer et al. (2007); Calder-Wang (2021); Cook et al. (2024); Almagro et al. (2023).

Intuitively, the maximum likelihood estimation searches for parameters that match predicted choices with actual observed choices for each household as closely as possible. The identification of heterogeneous parameters relies on variation in housing choices across different demographic groups. For example, consider two households with identical demographic characteristics except for their income levels. If the wealthier household chooses a larger housing unit, a positive interaction coefficient between income and house size better matches the model-predicted choice with observed choices. I standardize all continuous variables by subtracting the mean and dividing by the standard deviation to improve computational efficiency, so the coefficients should be interpreted as the effect of a one-standard-deviation change in the variable.

In the second step, I estimate the mean preference coefficients $(\alpha^0, \boldsymbol{\beta^0})$ using the estimated $\hat{\delta}$ from step 1, $\hat{\delta}_h = \alpha^0 p_h + \mathbf{X}_h \boldsymbol{\beta^0} + \xi_h$. The identification of the mean preference parameters relies on variation in prices and characteristics across segments. A higher mean utility indicates a more attractive segment for all households. The identification challenge is that the unobserved segment quality ξ_h may potentially correlate with prices p_h.

To address the potential endogeneity, I construct an instrument based on the methods of Berry et al. (1995) and Bayer et al. (2007), using a set of immutable property and neighborhood characteristics from segments located at least five miles from the segment. Specifically, the instrument includes house size, lot size, construction year, distance to the city center, and land usage. The instrument leverages the impact of market competition on pricing. Intuitively, a house with many competitors or close substitutes should have lower equilibrium prices to attract potential households. I use segments located five miles away because nearby segments may have spatial spillover effects that directly affect household utility and thus violate the exclusion restriction assumption. Following standard practice in the literature, I construct a single price instrument that approximates the optimal instrument to increase the power of the first stage. I first estimate Equation (1.5) to obtain mean preferences for exogenous segment features. With these estimated coefficients, I compute the market-clearing price vector assuming no unobserved segment quality by setting $\xi_h = 0$ for all h. This alternative price vector is the optimal instrument that isolates variations in housing prices attributable solely to exogenous housing and spatial characteristics. Then I use the simulated price vector as the instrument to estimate Equation (1.5).

Estimates of demand parameters

Figure 1.5 highlights the key heterogeneous preferences on prices and house size across different demographic groups. Panel (a) shows the average price sensitivity $\alpha_i = \alpha^0 + \sum_{r=1}^{R} \alpha^r z_i^r$ by demographic group. Households with lower incomes, or those that are Black, younger, or without a college degree are more price-sensitive. Panel (b) shows the marginal willingness to pay (calculated as $-\beta_i/\alpha_i$) for one standard deviation increase for house size. Households with higher incomes, non-Black, and college-educated have a higher willingness to pay for larger homes.

Appendix Table A.4 and Table A.5 present the mean and heterogeneous preference parameters. On average, households dislike paying more for housing and value living in newer larger homes. The mean willingness to pay for homes built after 2010 is about $39,293, and the willingness to pay for one standard deviation increase in house size[19] is $75,478. In terms of neighborhood characteristics, house-

[19]One standard deviation of house size is about 738 square feet.

holds prefer to live in neighborhoods with better amenities, such as incorporated areas and places with better job and transportation access.

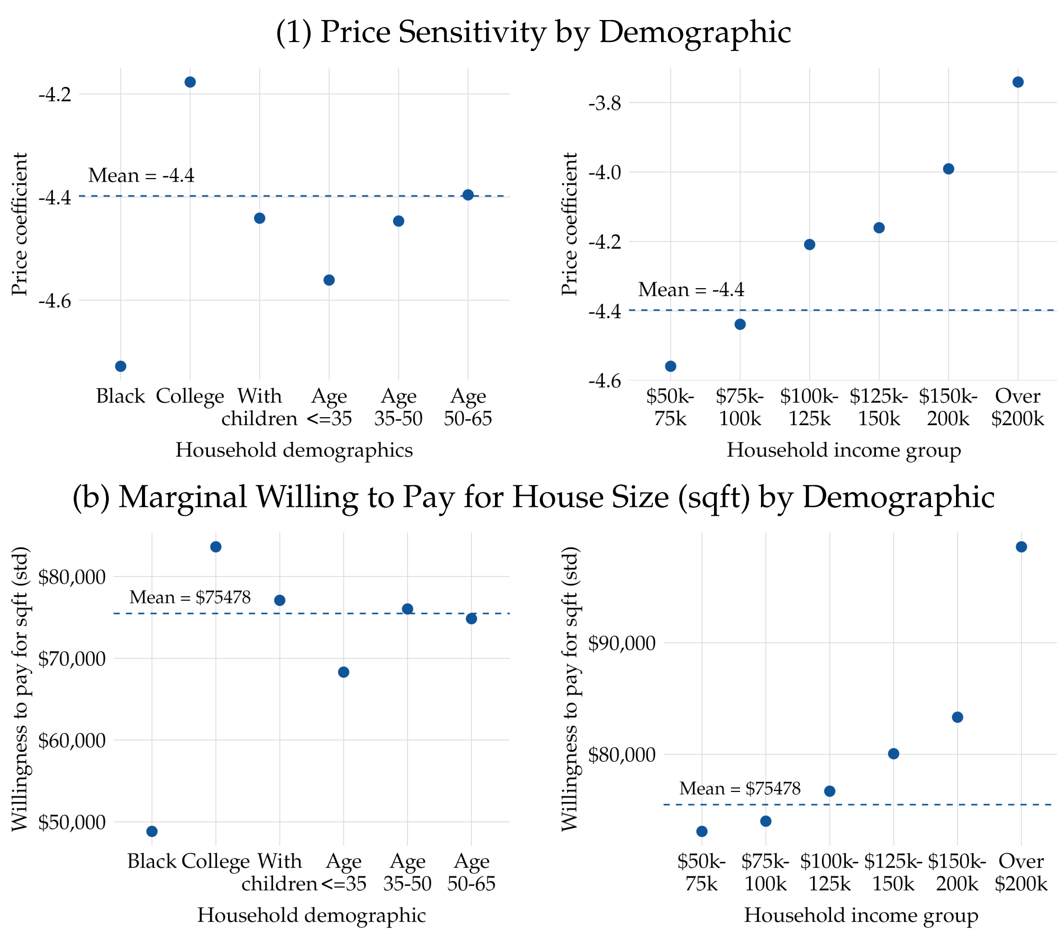
To measure the substitutability between different housing segments, I calculate the cross-price elasticities for each pair of segments (h, k). Appendix Figure A.9 shows the median estimated cross-price elasticities within the same house type h (row) if the prices of new homes of type k (column) change by 1%. Cross-price elasticities are higher among homes of the same type and construction period. For instance, when the prices of new large homes increase, there is a greater tendency to substitute toward other newer large homes than to shift toward older smaller homes. Appendix Figure A.10 shows the variation in cross-price elasticities by geographic distance between segments. As expected, segments in closer physical proximity are generally more substitutable.

Demand model validation

I conduct three exercises to evaluate the fit of the model. First, I simulate prices at the observed supply by setting the unobserved segment quality to zero, $\xi_h = 0$ for all h, and then I compare the simulated prices with observed prices. Panel (a) of Appendix Figure A.11 plots the simulated prices against observed prices. The predicted prices without accounting for unobserved segment quality match closely with observed prices, with the simulated prices explaining 94 percent of the variation in observed prices. This indicates that the included property and neighborhood covariates account for a large share of variations in equilibrium prices.

Second, I perform an out-of-sample exercise using the ACS data. Using the demographic characteristics of households residing in single-family homes from the ACS data and model parameters estimated from the L2 data, I solve for a set of

Figure 1.5: Estimated heterogeneous preferences for housing



Notes: These figures present the key parameters from the demand estimation. The top two plots in panel (a) show the price coefficient α_i for each demographic group. The bottom two plots in panel (b) show the marginal willingness to pay for a one-standard-deviation increase in house size for each demographic group, calculated as $-\beta_i/\alpha_i$, where β_i is the coefficient on house size.

prices that clear the market. I then compare these simulated prices to the observed prices, as shown in panel (b) of Appendix Figure A.11. The slope from regressing simulated prices using the ACS data on the observed prices is 1.029. This suggests that the L2 population data can fairly well capture variations in equilibrium prices and housing choices of different demographic groups.

In the third model validation exercise, I calculate the model-predicted demographic composition for each segment, $Z_h^r = (\sum_i d_{ih} \times z_i^r) / \sum_i d_{ih}, \forall h$, and compare them with the observed average demographic. Appendix Figure A.12 plots the model-predicted sociodemographic composition in each segment against the observed data. The model does a fairly good job of predicting the housing choices of each demographic group.

1.5.3 Supply model estimation

Similarly to the demand model, I use the maximum likelihood estimation to recover the parameters that determine regulatory costs. Before estimating these parameters, I first impute housing prices, construction costs, and land acquisition costs for new developments. To predict construction costs for each house type on a given parcel, I estimate a hedonic model using observed construction costs for newly built homes. The hedonic model is specified as

$$CC_\ell(k) = g(\mathbf{X}_k, L_\ell, \mathbf{W}_{n(\ell)}) + \xi_\ell,$$

where $\mathbf{X}_k$ is a vector of property characteristics, including the square footage and number of rooms. Square footage is the main dimension that affects construction costs. Appendix Figure A.13 shows a binned scatter plot of average construction cost per square foot for houses of different sizes. It shows that larger houses generally incur higher construction costs per square foot. This is likely because larger houses often include higher-quality appliances and better amenities, leading to higher construction costs. To capture the variation in construction costs by house size, I use indicator variables for different square-footage bins to estimate the model. The construction cost also depends on the parcel size L_ℓ and neighborhood characteristics $\mathbf{W}_{n(\ell)}$, including racial composition, income, and distance to the city center. The inclusion of neighborhood characteristics accounts for the variations in construction costs between neighborhoods. As shown in Appendix Figure A.13, the physical construction costs are higher in high-income neighborhoods because new construction in richer neighborhoods often uses better materials. The function $g(\cdot)$ includes polynomials and indicator variables for property and neighborhood characteristics. I then use the hedonic parameters to predict the construction cost for each house type on each parcel.

To predict housing prices for each house type, I assume that the expected prices equal the actual prices sold to households upon project completion.[20] To measure the land acquisition cost, I use the vacant land values per acre estimated by DLOS multiplied by parcel size as the land acquisition cost for vacant parcels. For already-developed parcels, I use the observed property prices, which include both land and structural values, as the land acquisition costs.

Taking expected prices, construction costs, and land acquisition costs as given, I estimate the regulatory cost parameters using maximum likelihood. The parameters to be estimated are $(\gamma_k, \eta, \rho_k, \tau_{m(\ell)})$ and the scale parameter σ_ν. The log likeli-

[20] I also predicted expected housing prices using a hedonic approach. The estimation results are similar.

hood function is

$$LL(\gamma_k, \eta, \rho_k, \tau_{m(\ell)}, \sigma_\nu) = \sum_\ell \sum_k y_{\ell,k} \ln \frac{\exp\left((p_\ell(k) - CC_\ell(k) - LC_\ell - RC_\ell(k))/\sigma_\nu\right)}{1 + \sum_{k \in \mathcal{K}} \exp\left((p_\ell(k) - CC_\ell(k) - LC_\ell - RC_\ell(k))/\sigma_\nu\right)},$$

where $y_{\ell,k}$ is the observed development choice on each parcel ℓ.

Regulatory cost parameters are identified by developers' development decisions using a revealed preference approach. The variation in the number and types of new homes across parcels, neighborhoods, and zoning jurisdictions identifies these parameters. Intuitively, if two parcels share similar characteristics but have different development outcomes, the differences can be attributed to different regulatory costs associated with different house types. The key assumption is that profits are uncorrelated with unobserved cost shocks. The main concern is that these unobserved cost shocks might capture cost differences associated with unobserved quality factors, such as higher material costs in high-amenity neighborhoods. However, the imputed construction costs already account for variation in neighborhood characteristics, and zoning costs are also allowed to vary by neighborhood, which supports the validity of the independence assumption.

Supply estimation results

Figure 1.6 presents the estimated regulatory costs for each potential house type by lot size.[21] Zoning regulatory costs tend to be higher for smaller parcels, mainly due to their location. Appendix Table A.7 reports that regulatory costs are higher in neighborhoods with limited land availability, higher property prices, fewer mi-

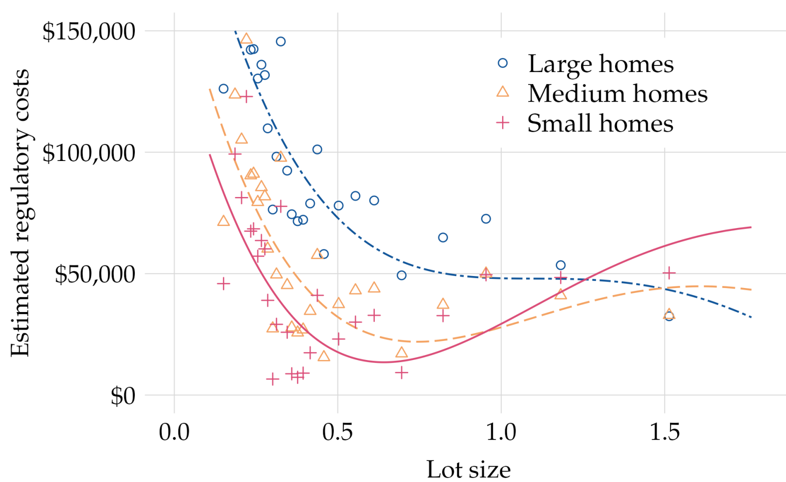
[21] Appendix Table A.7 reports the underlying estimated parameters for regulatory costs. I standardize all variables to improve computational efficiency. Thus, the coefficients represent the change in regulatory costs in dollar terms for a one-standard-deviation increase in each variable. The coefficients highlight the heterogeneity in regulatory costs by parcel and neighborhood characteristics.

norities, and greater proximity to the city center. These small parcels are often located near the city center or neighborhoods with constrained land availability, where zoning regulations are stricter.[22] Additionally, building larger homes on small parcels incurs higher regulatory costs. This aligns with the expectation that building large homes on small parcels will likely violate setback and height restrictions and thus requires rezoning. As parcel size increases, zoning costs for large homes decline due to reduced rezoning needs, while costs for small homes increase. This is partly due to the minimum house size requirements common in many Georgian municipal zoning regulations, which correlate positively with lot size, as shown in Figure 1.4. It may also result from other zoning measures that restrict the construction of small homes in neighborhoods that favor large parcels.

To put these estimates into perspective, Appendix Table A.8 reports the estimated regulatory costs for each house type. At the median, building a new small home incurs $49,507 in local zoning regulatory costs. For those opting for a medium home, this cost rises to $59,262, while a new large home incurs the highest regulatory cost of $98,755. The results highlight that developers pay substantial costs to comply with local zoning regulations. While building large homes incurs the highest regulatory costs, developers can sell them at higher prices. The optimal development decision depends on the trade-off between prices and total development costs, including regulatory and non-regulatory costs.

[22] Appendix Figure A.14 plots the same graph but is separated by distance from the city center. It shows that parcels within 10-miles of the city center face much higher regulatory costs and generally have smaller parcels.

Figure 1.6: Estimated regulatory costs by lot size and house type



Notes: This figure plots the estimated zoning regulatory costs from the supply estimation. It is a binned scatter plot showing how these costs vary by lot size and potential house type, not the actual house type built on each lot. The fitted lines are cubic global polynomials applied to the data for each house type. Blue circles with dot-dashed lines represent large homes, yellow triangles with dashed lines represent medium homes, and '+' symbols with solid lines correspond to small homes. Small homes are defined as homes with ≤ 3 bedrooms, medium homes have four bedrooms, and large homes have ≥ 5 bedrooms.

Table 1.3: Cost structure at observed house type

Cost-to-Price Ratio	Median	Mean	P25	P75
Construction Cost	63.2%	73.9%	49.2%	73.9%
Land Cost	18.5%	24.1%	15.8%	24.61%
Regulatory Cost	20.3%	22.2%	10.2%	32.3%

Notes: This table presents the cost structure for newly built homes from the supply estimation. It reports the median, mean, 25th percentile, and 75th percentile of the cost-to-price ratio for each cost component across all homes. The components include physical construction cost, land acquisition cost, and costs attributed to local zoning regulations. The sample includes only observed new construction, not all potential house types that could be built on each parcel.

Supply model validation

To assess the accuracy of cost estimation, I compare the estimated costs with industry data from the National Association of Home Builders (NAHB). Table 1.3 presents the cost-to-price ratio for each cost component for all newly built homes. At the median level, construction costs constitute 63.2% of the actual prices, land costs account for 18.5%, and regulatory costs make up 20.3% of the prices. The model-predicted cost structure closely aligns with the estimates from the surveys by NAHB.[23]

To evaluate the model fit, I compare the predicted number of new homes of each house type in each neighborhood to the observed data. Appendix Figure A.15 presents the binned scatter plots for each possible development decision at the neighborhood level. The model is able to accurately predict that the majority of

[23] See Table 2 in the NAHB's Special Study on the Cost of Constructing for a breakdown of the sale price by different cost components biennially from 1998 to 2022. The share of total construction costs between 2011 and 2019 ranges from 55.6% to 61.8%, and the share of lot acquisition cost ranges from 18.2% to 21.7%. To gauge the quality of my regulatory cost estimates, I compare them with another special study conducted by NAHB, which surveyed developers about the costs attributed to local land use regulations. They find that, on average, about 23% of the total costs of building a new home are due to local land use regulations, with the most significant expenses on various fees, development standards, changes to building codes, and land dedicated unbuilt. My structural estimation finds that regulatory costs account for about 20% of final housing prices, closely aligning with the industry report.

already-developed parcels in a given neighborhood are not profitable for redevelopment. Additionally, the model performs reasonably well in predicting the number of parcels with newly built small and medium homes at the neighborhood level. There are some discrepancies for large homes, suggesting that the model either over- or underestimates regulatory costs in some neighborhoods.

I also cross-validate the average regulatory costs for each ZCTA using two other measures. In panel (a) of Appendix Figure A.16, I plot these costs against the supply elasticities estimated by Baum-Snow & Han (2024). It shows that regulatory costs are higher in neighborhoods with lower supply elasticities. In panel (b) of Appendix Figure A.16, I examine the relationship between regulatory costs and the share of developed land from the 2016 National Land Cover Database. It reveals that more developed neighborhoods, which likely face limited land availability and a more stringent zoning environment, impose higher regulatory costs.

1.6 COUNTERFACTUALS

1.6.1 Causes and consequences of large home construction

In this section, I quantify how demand- and supply-side factors contribute to the shift toward increased large home construction, as well as its distributional impact on housing prices across different market segments. While the model incorporates rich heterogeneity in property characteristics and demographics, my analysis below focuses on variation by house size and income for expositional purposes.

Role of income growth The shift toward large home construction may reflect changes in household housing preferences, which leads developers to adjust their development decisions to meet changing demand. The demand estimation reveals substantial preference heterogeneity, particularly across income levels. Higher-income households have strong preferences for larger homes, while lower-income households prefer smaller, cheaper homes. As the number of high-income households in a metro area increases, developers may respond to this demand shift by building larger homes. Income inequality has increased over the years.[24] For example, Figure 1.7 shows that the share of households in higher-income brackets increased in 2019 compared to 2010, with the share of households earning more than $150,000 rising by 4.7 percentage points.[25] Therefore, in this section, I examine the role of income growth in driving the trend toward larger home sizes.

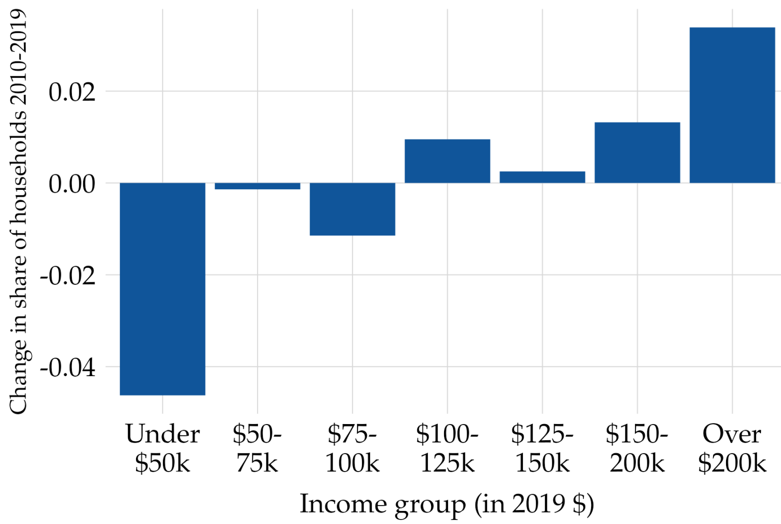
To assess the extent to which rising top income groups drive increases in house sizes, I simulate the model using the 2010 income distribution while keeping the total population and other demographic variables constant.[26] By holding the total population fixed, I can isolate the impact of income-distribution changes on supply and prices without introducing potential market expansion at the extensive margin. To compute the new equilibrium using the simulated data with the 2010 income distribution, I first solve for a new set of market-clearing prices under

[24] Appendix Figure A.17 shows changes in the income distribution from 1980 to 2019. There has been significant growth at the top income levels, while the 25th percentile of household income has experienced only a slight increase. Howard & Liebersohn (2021) emphasize the role of increased demand in driving up rents in supply-inelastic cities. I focus on how demand, particularly from high-income groups, affects the supply decisions of developers.

[25] Appendix Figure A.18 shows the number of households in each income group in 2010 and 2019. The number of households earning more than $150,000 increased by 116,285, or a 49% increase in 2019 compared to 2010. This growth in the number of high-income households can be attributed either to the in-migration of high-skilled workers to the city or to the wage growth of incumbent households. The Atlanta Regional Commission forecasts population growth, particularly growing numbers of high-income, high-skilled workers in the region driven by the booming technology, healthcare, and finance industries.

[26] Specifically, I calculate the share of households within each income bracket for each demographic group – defined by the combination of race, education, age, and children present status – using the 2010 ACS. Then I multiply these 2010 income shares with the total number of households in each demographic group from the baseline 2019 data. The step-by-step process of constructing the counterfactual dataset is described in Appendix A.3.4.

Figure 1.7: Change in household share by real income from 2010 to 2019



Notes: This figure shows the change in the share of households in each income group from 2010 to 2019. The data are from the American Community Surveys in 2010 and 2019. I convert the 2010 income to real 2019 dollars using the urban CPI in the Atlanta MSA.

this alternative income distribution, holding the observed supply constant. Next, I update developers' optimal development decisions using this new set of prices. Then, I solve for equilibrium prices again using the updated supply. I keep iterating these steps until both prices and supply converge.

Panel (a) of Figure 1.8 shows the change in new construction under the 2010 income distribution relative to the observed baseline. Panel (a) shows that developers endogenously adjust their development decisions in response to this shift in income distribution. A 5 percentage point reduction in the share of households earning over $100,000 leads to a 4% drop in large home construction. The shift in demand—fewer high-income households and more low-income households—reduces demand for large homes and boosts demand for smaller ones. To show this, Appendix Figure A.19 compares the simulated demand under the 2010 income distribution without any supply adjustment. Demand for homes bigger than

3,500 square feet decreases by 5% to 13%, while homes smaller than 2,000 square feet experience a demand increase of 5%. This shift puts downward pressure on the prices of larger homes and upward pressure on the prices of smaller homes. As a result, developers build fewer large homes because these constructions become less profitable. Meanwhile, the increased share of households with income less than $100,000 leads to less than a 1% increase in small home construction. This suggests that despite growing demand for smaller homes, prices remain insufficient to ensure profitability and stimulate a large supply response, largely due to high production costs.[27]

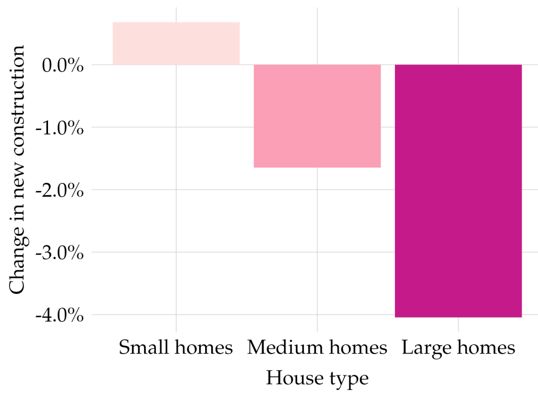
Panel (b) of Figure 1.8 shows the change in simulated equilibrium prices under the 2010 income distribution relative to baseline prices by house size. If the income distribution remains the same as in 2010, prices for homes larger than 3,000 square feet would decrease by 0.5% to 1%, and prices for homes smaller than 2,000 square feet would see a slight increase of up to 1%. The price decline among larger homes is primarily driven by reduced demand, which exerts downward pressure on prices. The slight price increase of small homes is due to the rising number of low-income households and a supply that has failed to meet the growing demand. Why are these overall price changes so modest? First, the simulation keeps the total population constant, so there is no change in aggregate demand. Second, new homes make up only 10% of the total housing stock, so changing the quality distribution of new homes has a limited impact on the whole price distribution.

The counterfactual shows that the model captures developers' endogenous supply response to shifts in demand, highlighting their responsiveness to changing

[27] To give a concrete example: To build a new home of 2,000 square feet, the construction cost alone is about $187,000 at the median level. The median price of an existing home of the same size is around $206,000. With the price premium on new homes, many low-income households would be unwilling to pay for these new homes, so the optimal choice for developers is not to build them.

Figure 1.8: Simulated equilibrium under 2010 income distribution

(a) Change in new construction



(b) Change in prices

Notes: These two figures show the difference between the simulated equilibrium under the 2010 income distribution and the baseline equilibrium. The left panel shows the percent change in new construction by house type under the 2010 income distribution compared to the baseline scenario with 2019 income. The right panel is a binned scatter plot showing the average percent change in housing prices by house size for both existing and new homes, relative to the baseline.

market conditions. Demand from high-income households for larger homes drives the increased construction of these homes. Additionally, the differential price growth by house size suggests a segmented housing market, with households from different income groups sorting into distinct market segments.

These results also indicate that while rising income inequality explains some of the changes in new construction, supply-side factors likely play a more important role.[28] Specifically, the limited supply response for small homes highlights the importance of these factors. The next subsection examines how supply constraints, particularly zoning regulations, contribute to the trend toward larger home construction.

[28] In Appendix A.2.2, I examine an extreme scenario in which all preference heterogeneity is eliminated. Even under this counterfactual with homogeneous preferences, developers reduce the construction of large homes by 20%, which points to the importance of supply-side factors.

Role of zoning costs Next, I explore the effect of zoning costs on supply and prices. The supply estimation shows that zoning costs are substantial and differ by house type and parcel characteristics. Many studies have shown that zoning regulations primarily focus on density restriction, and minimum lot size requirement is one of the most common and restrictive rules for single-family homes (for example, Bartik et al. (2023), Gyourko & McCulloch (2023), and Tokman (2023)). In this section, I examine how housing production and prices would change if density restrictions were relaxed.

I simulate an alternative zoning environment where developers now build two properties on parcels larger than the 25th percentile in lot size across all zoning municipalities between 2010 and 2019.[29] Mathematically, I split vacant parcels with sizes above the 25th percentile into two equal-sized parcels. This change nearly doubles the number of vacant parcels eligible for development, significantly increasing the potential housing supply at the extensive margin. I assume the total population remains unchanged under this counterfactual scenario.[30] To compute the new equilibrium, I calculate developers' decisions given the updated parcel sizes, the number of parcels, and the land acquisition cost per parcel. Given the

[29] The 25th percentile of lot sizes for all new homes is 0.348 acres. I choose this cutoff because parcels smaller than this might be difficult to subdivide further. As cities continue to develop, land is increasingly likely to become a scarce resource. Allowing smaller lots also improves land-use efficiency. In practice, this change can be achieved by various changes in zoning rules: lowering the minimum lot size requirement, which can reduce the actual lot size; eliminating single-family zoning and permitting duplexes; or increasing maximum density allowed per acre.

[30] I make this closed-economy assumption because I want to isolate the impact of lower density restrictions while holding demand fixed. An alternative assumption is that the total number of households grows at the same rate as the housing stock. The open economy would attenuate the price impact and potentially lead to a greater supply response. The model highlights that it is not just the number of new households that matters; the demographics of those households would also lead to different outcomes. For instance, a large influx of high-income households would likely drive up prices for larger homes, while an influx of low-income households would place greater upward pressure on the prices of smaller homes. The magnitude of price responses also depends on the endogenous supply adjustment.

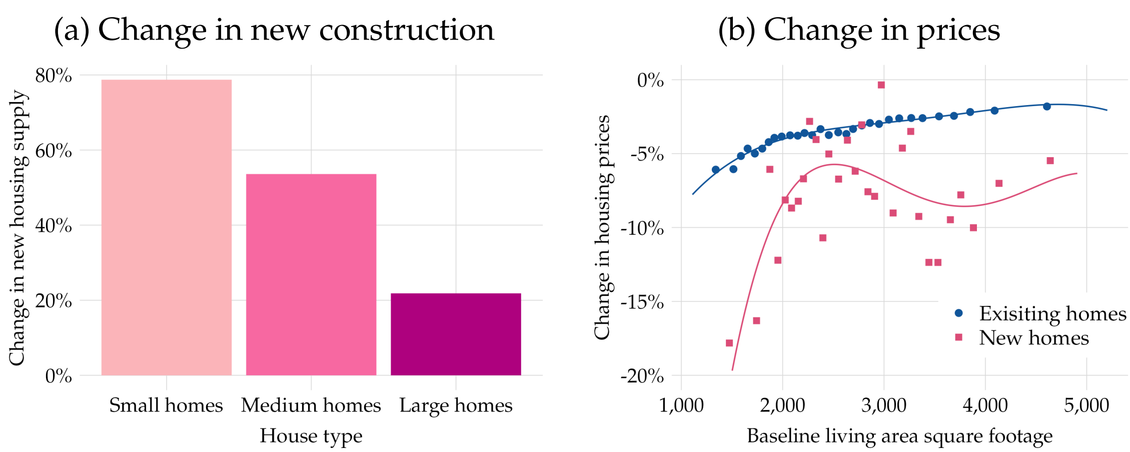
updated supply, I compute a new price vector that clears the market. Then I update developers' choices again using these new prices. I keep iterating this process until prices converge.

Figure 1.9 compares the resulting changes in housing supply and prices to the baseline scenario. Panel (a) shows that the supply of all types of new homes increases, with the largest growth seen in small homes: The supply of new small homes increases by 79%. Under the counterfactual, there are more smaller parcels under the simulated scenario, so developers find it more profitable to build smaller homes on these small parcels. Meanwhile, the supply of large homes also increases by 22%. Even after splitting parcels in half, many parcels remain sufficiently large to support the construction of large homes. At the aggregate level, new construction increases by over 67,000 units, expanding the total housing stock by nearly 6%.

Panel (b) of Figure 1.9 shows the average percentage change in house prices by house size relative to the baseline. The additional housing supply drives down prices for all homes. Newly built small homes experience the largest price reductions, with drops of up to 18%, while similarly sized existing homes see a 5% decline. Prices of existing homes fall because households substitute toward newly built homes. The disparity in price declines between existing and new homes occurs because newly built homes are segmented from existing homes. Although they are similar in size, new homes typically command a price premium and attract a different clientele due to preference heterogeneity.

This counterfactual shows that density restrictions not only reduce the total volume of new housing but also skew construction toward larger homes, contributing to rising prices of small homes. Stricter zoning regulations, particularly density re-

Figure 1.9: Simulated equilibrium under density relaxation



Notes: These two figures plot the simulated equilibrium under an alternative zoning environment with relaxed density restrictions. In this counterfactual scenario, developers build two properties on parcels larger than the 25th percentile in lot size across all zoning municipalities between 2010 and 2019. The left panel shows the percent change in new construction by house type in the metro under this counterfactual scenario compared to the baseline levels. The right panel is a binned scatter plot showing the average percent change in housing prices by house size for both existing and new homes, relative to the baseline.

strictions, penalize developers for building small homes and drive up price growth in the small home segment.

Discussion In addition to zoning regulation, I also examine how rising construction and land costs affect the quality distribution of new housing construction. Between 2010 and 2019, the median construction cost per square foot increased by 11%, and the median land value per acre rose by 50%. Intuitively, developers would build larger homes when construction costs are lower. This is because construction costs act as variable costs in the model: When it becomes cheaper to add more floorspace, developers find it optimal to do so. On the other hand, higher land costs, which represent the fixed cost component, do not directly affect the optimal quality choice. To validate these theoretical predictions quantitatively, I simulate two equilibrium scenarios by reducing construction and land costs to their

2010 levels. Appendix Figure A.20 shows the composition of new homes under the baseline and the two simulated scenarios. As expected, reduced land costs lead to roughly the same share of new homes being built across different sizes without significant changes in the quality distribution, and lower construction costs result in more construction of large homes, pushing the overall quality distribution of new homes further toward the higher end.

To summarize, the counterfactual simulations show that demand from high-income households and strict zoning regulations are key drivers of the increase in house sizes. Notably, strict zoning regulations play a more significant role in restricting the production of small homes and driving up their prices. Land use reforms can significantly increase housing supply and expand the construction of smaller homes, which can help reduce distributional disparities. However, zoning reforms often face political resistance and typically result in slow, incremental changes.[31] Meanwhile, income inequality is expected to worsen, driven by a growing influx of high-skilled migrants and skill-biased technological change. Therefore, policymakers have started considering complementary approaches to stimulate housing production and provide targeted assistance to households struggling with rising housing prices. The following section examines two housing subsidy policies aimed at improving affordability at the lower end of the housing market.

1.6.2 Policy implications of housing subsidies

Given the limited production of small homes and rising housing prices, policymakers and housing advocates have proposed using subsidies to improve affordability, particularly at the lower end of the market. In this section, I evaluate the

[31] For example, Gyourko & McCulloch (2024) estimate that incumbent homeowners, on average, dislike density, highlighting the challenges of deregulating density on the demand side.

impact of subsidies targeting potential first-time homebuyers and starter homes on the housing market and household welfare.

Subsidizing young, low-income households I first consider a demand-side subsidy aimed at lowering young, low-income households' housing costs. President Biden and Vice President Harris have proposed providing $25,000 in federal down payment assistance to first-time homebuyers.[32] Many state and local governments have offered similar programs at different scales, such as the Georgia Dream Homeownership Program; Washington, DC's Open Doors Down Payment Assistance Loan; and Massachusetts' Downpayment Assistance Program.

I compute a counterfactual using a simple subsidy program, which provides a one-time $25,000 lump-sum transfer to households under age 35 and income below $75,000.[33] To compute the new equilibrium, I recalculate housing demand with subsidized households paying $25,000 below market prices. Using the updated demand, I then recalculate market-clearing prices and calculate developers' supply response at these new prices. I iterate this process until prices converge.

Figure 1.10 reports the impact of the demand-side subsidy on housing prices and new construction relative to the baseline scenario. Panel (a) shows that the subsidy leads to an increase in the prices of both existing and new homes. Prices for homes under 2,000 square feet rise by 1% (approximately $2,000), while larger

[32]For further details, see the White House Fact Sheet on President Biden's plan to lower housing costs for working families and the Harris-Walz campaign's Policy Book on how to create an Opportunity Economy.

[33]In practice, down payment assistance programs take various forms, such as grants, deferred-payment loans, or tax credits. These programs often restrict the types of homes eligible for purchase or require buyers to work with specific mortgage lenders. To simplify implementation and interpretation, I model the program as a one-time grant, avoiding the complexities of financing and future repayment. To determine eligibility, I aim to match the profiles of first-time homebuyers closely. Based on data from the Census Bureau's American Housing Survey, the average first-time homebuyer is 33 years old with an annual income of $81,000. Therefore, I provide the subsidy to all households under age 35 with incomes below $75,000, abstracting from potential take-up issues.

homes also experience price increases, though these are about one-third as large. As shown in the demand estimation, subsidized households have strong preferences for these smaller homes, and the subsidy significantly boosts their demand and drives up the prices of these homes. The price growth of medium and large homes reflects the cross-segment spillovers captured by the model. The magnitude of price growth diminishes with house size, as larger homes are more segmented from smaller ones.

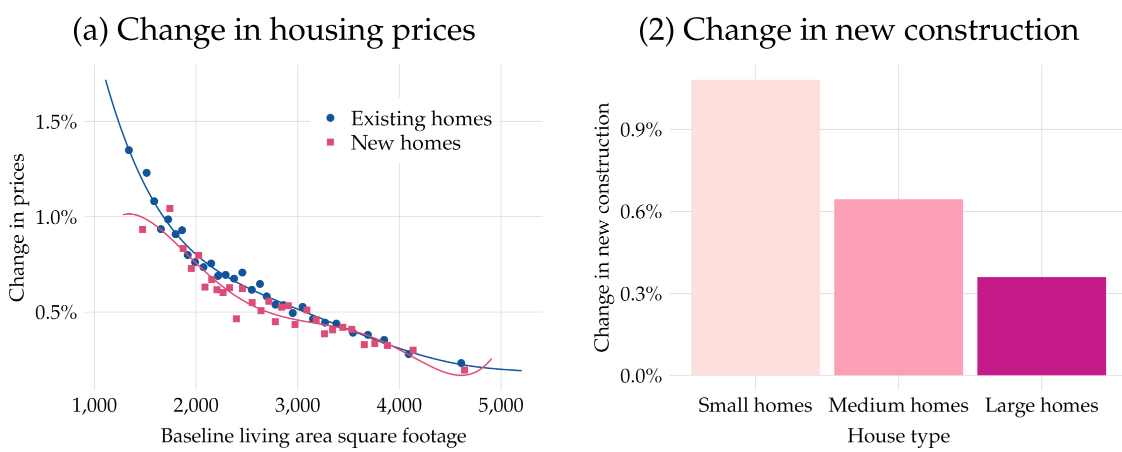
Panel (b) shows the endogenous supply response from developers to the positive demand shock at the lower end of the market. The demand shock incentivizes developers to build more new homes, especially smaller ones. However, the magnitude of the overall supply response is nearly muted, as the initial price increase without supply adjustment is small.

Why does the subsidy trigger such small equilibrium responses in housing markets? Under the counterfactual scenario, only 7.2% (96,962) of households are eligible for the subsidy, representing a small share of the total market. Moreover, these subsidized households mainly compete with older, unsubsidized households of similar incomes, so the aggregate impact on prices is not that large. Among subsidized households, the majority of them choose existing small homes rather than new ones[34] because the existing homes are typically cheaper compared to new homes of the same size in the same neighborhood. The split in demand between existing and new homes also explains the small price increase in new homes, which consequently leads to small supply responses.[35]

[34] Appendix Figure A.21 compares the housing choices of subsidized households under the baseline and the subsidy scenarios. It shows that most subsidized households switch to existing small homes. The subsidy also enables households who previously chose the outside option to enter the market.

[35] The equilibrium calculation assumes that all households re-optimize their housing choices under this policy change, regardless of whether they receive the subsidy. I consider an extreme sce-

Figure 1.10: Impact of demand-side subsidy on the housing market



Notes: These two figures plot the simulated equilibrium with a one-time $25,000 transfer to households under age 35 and income below $75,000. The left panel is a binned scatter plot showing the average percent change in housing prices by house size for both existing and new homes, relative to the baseline. The right panel shows the percent change in new construction by house type in the metro area under this counterfactual scenario compared to the baseline levels.

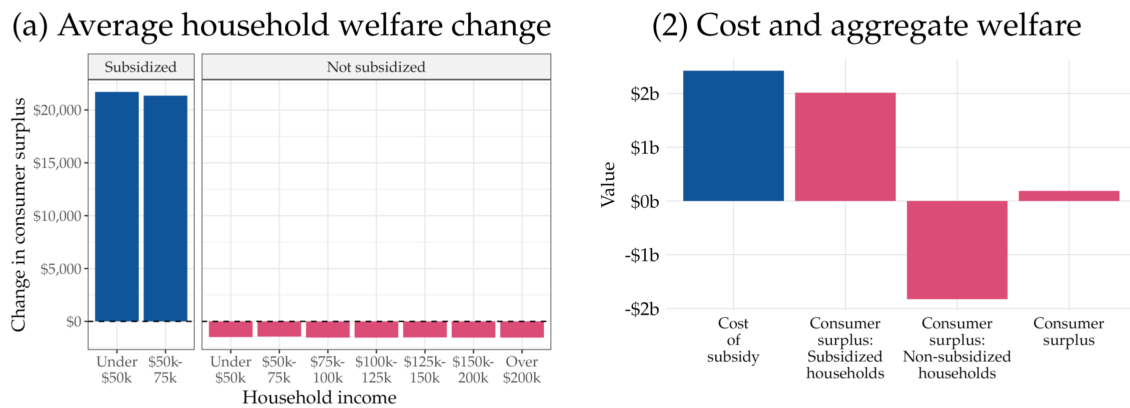
Next, I evaluate the distributional consequences of the subsidy on household welfare. For each household i, I measure the change in consumer surplus, which is the change in inclusive value normalized by the price coefficient,

$$\Delta CS_i = -\frac{1}{\alpha_i}\left(\ln \sum_h N_h^1 \exp(v_{ih}^1) - \ln \sum_h N_h^0 \exp(v_{ih}^0)\right),$$

where the subscripts 1 and 0 denote the counterfactual and baseline scenarios, respectively, and $v_{ih} = \delta_h + \mu_{ih}$ represents the indirect utility without idiosyncratic preferences. Both changes in prices and supply drive changes in household welfare.

nario where subsidized households can only purchase newly built homes. This restricted choice set gives an upper bound of the potential price impact of this subsidy. Appendix A.2.3 reports the price and supply responses. Under this scenario, prices for small homes rise substantially — about $20,000 for a 1,500-square-foot new house. The supply response is also stronger, with newly built small homes increasing by 27%, though the share of small homes in the total housing stock increases by less than 2%. The large price response suggests that an inelastic supply curve could amplify the impact of the demand shock.

Figure 1.11: Impact of demand-side subsidy on welfare



Notes: These two figures plot the household welfare change with a one-time $25,000 transfer to households under age 35 and income below $75,000. I measure welfare change as the difference in consumer surplus between the baseline and counterfactual scenarios. The left panel shows the average per-household welfare change by income and subsidization status. The right panel shows the aggregate cost of the subsidy and the aggregate household welfare change in billions of dollars.

Figure 1.11 presents the household welfare changes from the subsidy. Panel (a) shows the average per-household welfare change for subsidized and non-subsidized households. The policy generates substantial welfare gains for subsidized house-holds, amounting to over $20,000 per household, with part of the subsidy offset by higher housing prices. However, non-subsidized households experience welfare losses, as increased demand from subsidized households heightens competition, driving up prices and offsetting the modest supply response. Panel (b) shows the change in aggregate welfare. The subsidy costs approximately $2.4 billion, with nearly all benefits accruing to subsidized households. The demand-side subsidy is a highly targeted policy, with each dollar spent generating $0.83 in welfare gains to subsidized households. For a social planner who prioritizes the welfare of these households, the policy achieves its goal. However, it is important to note that the subsidy also imposes welfare losses of $0.73 per dollar spent for non-subsidized households.

Subsidizing developers for small development In addition to demand-side subsidies, many housing advocates and policymakers are pushing for supply-side subsidies. While most existing government supply-side programs focus on housing for households at the lowest end of the income distribution, there is growing interest in policies aimed at building homes for moderate-income households who do not qualify for these traditional programs.[36] For example, the Harris presidential campaign proposed a tax credit for homebuilders to build starter homes, Massachusetts' Chapter 40Y proposes the creation of starter home zoning districts, and Utah's First Homes program directs investment toward new starter homes.

In this counterfactual scenario, developers who construct small new homes receive a $25,000 subsidy. To compute the new equilibrium, I recalculate developers' supply decisions, with each developer receiving $25,000 above market prices when building small homes. Using the updated supply, I compute the market-clearing prices and allow developers to re-optimize their decisions based on the new prices. I iterate this process until prices converge.

Figure 1.12 presents the impact of the supply subsidy on the housing market. The subsidy increases the profitability of constructing small homes. Panel (a) shows that the subsidy triggers a significant construction response from developers: The number of new small homes increases by nearly 60% (25,487 units) compared to the baseline scenario. The supply growth is partly attributed to a shift in construction away from medium-sized and large homes, whose numbers decrease by 19% and 10%. Part of the supply expansion comes from the additional development of vacant parcels. In total, new construction rises by 12,273 units,

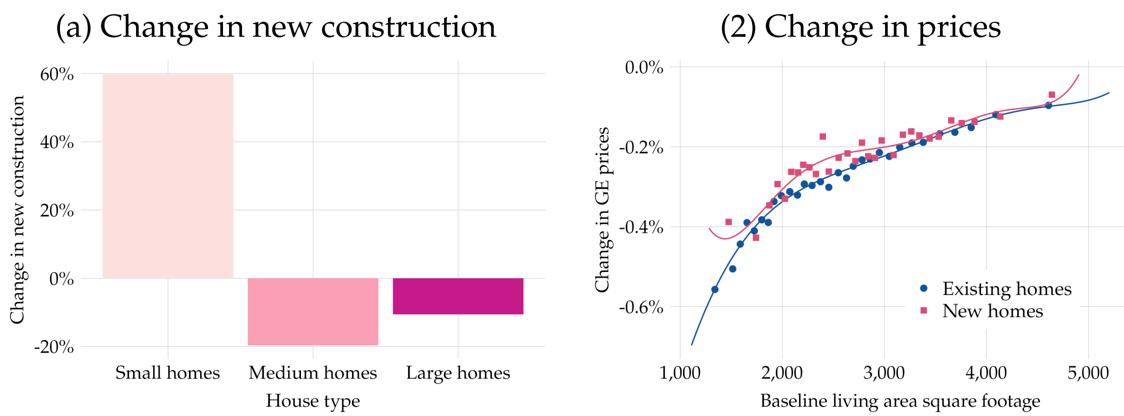
[36] Examples include public housing programs and the Low-Income Housing Tax Credit (LIHTC), which support the construction of multi-family buildings intended for individuals with extremely low incomes.

representing a 9.31% growth in new construction relative to the baseline. The results show that developers are responsive to profit incentives. As small home construction becomes more profitable, they adjust their development decision at the intensive margin and expand development at the extensive margin to capture the gains. However, in terms of the total housing stock, the number of small homes rises by less than 4%, and the total housing stock increases by only 1% (shown in Appendix Figure A.22).

Panel (b) shows the change in housing prices of existing and new homes. Prices have declined across all home sizes: Homes under 2,000 square feet experience a price decline of 0.4% to 0.6%, approximately three times greater than that for homes larger than 4,000 square feet. The price drop for smaller homes is primarily driven by the supply expansion of newly built small homes. Despite a reduction in the supply of newly built bigger homes, prices of medium and large homes also fall. It suggests that the effect of indirect substitution dominates the reduced supply. As shown in Appendix Figure A.23, households that previously chose larger homes are shifting toward newly built small homes. This shift in demand exerts downward pressure on larger home prices.[37] The small magnitude of price changes is partly due to the relatively modest increase in the aggregate housing stock and partly to demand from households shifting away from the outside option, and households who substitutes toward these newly built small homes. As shown in the bottom row of Appendix Figure A.23, the overall supply expansion attracts households who initially chose the outside option to enter the market.

[37]Theoretically, a reduction in the supply of larger homes could drive up prices if households exhibit a strong preference for these homes. However, model-implied preference parameters suggest that the demand elasticity for house size in new homes is not sufficiently inelastic to sustain higher prices. As a result, prices of large homes decline as households substitute toward newly constructed small homes that have become available.

Figure 1.12: Impact of supply-side subsidy on the housing market

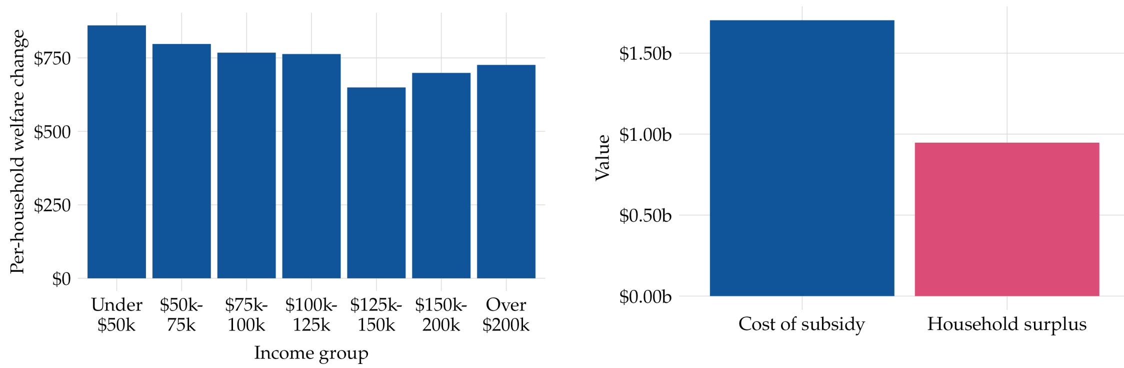


Notes: These two figures plot the simulated equilibrium with a one-time $25,000 subsidy to developers for building small homes. The left panel shows the percent change in new construction by house type in the metro area under this counterfactual scenario compared to the baseline levels. The right panel is a binned scatter plot showing the average percent change in housing prices by house size for both existing and new homes, relative to the baseline.

Figure 1.13 presents the household welfare change with the supply-side subsidy. Panel (a) shows that households gain $600 to $860 in consumer surplus on average, with lower-income households receiving slightly greater gains due to lower prices and supply expansion of smaller homes. The small magnitudes of average welfare gains are due to the small aggregate price and supply changes. Panel (b) shows the aggregate cost of subsidy and welfare changes. The total cost of subsidies to developers is $1.7 billion, lower than the $2.4 billion in demand-side subsidies to households. Each dollar given to developers generates $0.56 in welfare gains for households. However, these gains are evenly distributed among all households, with each household experiencing only modest welfare improvements. In contrast to the demand-side subsidy, characterized by redistribution, the supply-side subsidy functions as a universal policy that benefits all households, even those with less need for support.

Figure 1.13: Impact of supply-side subsidy on welfare

(a) Average household welfare change (2) Change in aggregate welfare



Notes: These two figures plot the simulated equilibrium with a one-time $25,000 subsidy to developers for building small homes. I measure welfare change as the difference in consumer surplus between the baseline and counterfactual scenarios. The left panel shows the average per-household welfare change by income. The right panel shows the aggregate cost of the subsidy and the aggregate household welfare change in billions of dollars.

Discussion The analysis presents a positive evaluation of the two subsidy policies rather than a normative assessment of the optimal policy. The decision on which policy to implement ultimately depends on the objectives of policymakers. The demand-side subsidy directly benefits recipients but also leads to price distortions that can offset the intended benefits. The magnitude of these distortions depends on the number of subsidized households and the restrictions on which homes can be purchased with the subsidy. The supply-side subsidy encourages the construction of smaller homes but lacks targeting. It is given to all developers who build small homes, including those who would have done so without the incentive, and it does not directly address the needs of specific households. Introducing restrictions that limit the subsidized homes to certain eligible households could improve targeting. The results also highlight that while expanding supply is crucial, the aggregate housing stock is unlikely to grow rapidly enough to deliver substantial benefits to households. Therefore, a combination of supply-side policies and demand subsidies can better address housing affordability challenges at the lower end of the market.

1.7 CONCLUSION

Housing markets consist of distinct segments, each with different characteristics that cater to the needs and preferences of different households. This paper focuses on the cross-sectional variation in the supply of single-family homes: Newly built homes tend to be larger and pricier. Using an equilibrium model with heterogeneous housing demand and supply, I disentangle the roles of demand and zoning regulation in the upsizing trend in new construction. The counterfactual simulations show that while high-income demand incentivizes developers to build larger homes for higher profits, restrictive density regulations is a more substantial factor driving the upsizing trend by raising overall production costs and reducing the profitability of smaller homes. Relaxing such restrictions could expand the supply of small homes, lower their prices, and benefit low-income households.

This paper provides a positive analysis of housing supply. I do not take a stance on the efficiency of zoning or the political economy considerations behind its design. I treat zoning regulations as given and let households and developers optimize their decisions within this environment. The findings show that density restrictions have distributional consequences for housing production and household welfare, with stricter rules penalizing the lower end of the market. I do not make a case for which zoning framework is optimal. Zoning is a complex issue that requires a more comprehensive analysis to account for its externalities and social impact. The current quality distribution of new housing is efficient under existing zoning regulations. However, it has unequal impacts on housing production and prices across market segments, leading to distributional implications for households of different demographics.

As housing prices continue to rise, lower- and middle-income families are becoming increasingly cost burdened. While the market tends to respond to the preferences of high-income households, government interventions can address these equity concerns. It might be politically challenging to reform land use regulations, but alternative policy instruments are available to promote housing equity and support more inclusive housing development. In recent years, policymakers and housing advocates have begun considering demand- and supply-side subsidies to improve affordability at the lower end of the market. The model predicts that subsidizing households generates substantial welfare gains for recipients, as they can purchase homes they could not otherwise afford. However, it causes slight welfare losses for others because the positive demand shock drives up prices of homes across the market, especially the small homes subsidized households prefer. In contrast, subsidizing developers to build small homes triggers a strong supply response for new small home production and reduces prices across the market. However, the aggregate impact on the total housing stock is small, leading to only slight price declines. While expanding housing supply is important, substantial increases in supply are unlikely in the short term. Targeted housing assistance is a complementary tool for policymakers looking to support specific households. Combining multiple policy instruments could more effectively address affordability challenges in the lower segment of the housing market.

CHAPTER 2

Racial Disparities in Home Selling

2.1 INTRODUCTION

Housing wealth is a crucial vehicle for households to build and accumulate wealth. Historically, various forms of structural discrimination have impeded Black households from entering homeownership and receiving its wealth appreciation benefits. While it might be expected that such discriminatory practices are no longer relevant today, recent research suggests that Black households continue to face higher mortgage rejection rates (Bhutta et al. (2022)), pay higher purchase prices (Bayer et al. (2017)), and realize lower total returns due to disparities in experiencing distressed property sales (Kermani & Wong (2024)).

In this paper, we focus on the home selling process and consider homes as financial assets that yield capital gains returns. One would expect that identical houses in the same neighborhood should sell for the same price and experience similar price growth over time on average. Conditional on the same property characteristics and market conditions, one would expect that the race of the homeowner should have no impact on the average selling outcomes. However, our findings suggest otherwise. We first show that racial disparities exist in the selling process and then provide empirical evidence for the underlying mechanisms driving these racial disparities. We provide several pieces of evidence that strongly suggest that the racial gap in housing returns is driven by racial differences in human intermediary practices. We believe this paper to be unique in that we demonstrate which mechanisms explain and do not explain these racial disparities.

We use repeat-sale transactions to estimate the racial gap in housing returns for regular, non-distressed sales. By comparing housing returns of properties within the same Census tract, sold in the same year and quarter, where one property is owned by a Black household and the other by a non-Black household, we effectively control for time-invariant property characteristics and market conditions. The repeat-sale approach mitigates the omitted variable bias that may be present in hedonic models. To identify the race of the sellers and buyers, we merge property transaction data with self-reported race from the Home Mortgage Disclosure Act (HMDA) data.

Using over 1.1 million repeat-sale transactions from ATTOM, a commercial data vendor that compiles nearly all property transaction data in the US, we find that Black households, on average, earn 1.74–1.88% lower total unlevered capital gains when selling their homes. To account for the varying duration of homeownership, we also calculate the annualized capital gains. Our results indicate that the annualized unlevered return gap for Black sellers ranges between -0.36% and -0.38%. These gaps are approximately 10% of the mean total and annualized returns to housing. Importantly, our analysis includes both Census tract and year-quarter sale date fixed effects, ruling out differences in neighborhood amenities and market timing as explanations for the observed racial gap.

We then explore the sources of the racial gap. Our main contributions to the existing literature are 1) identifying the specific factors that do and do not explain this racial gap and 2) demonstrating that this racial gap is systemic, caused by disparities stemming from human intermediary practices. The home-selling process involves multiple parties including sellers, potential buyers, and intermediaries such as realtors and appraisers. Discrimination can occur during interactions among these parties. Moreover, unobserved property characteristics, potentially correlated with race, can contribute to differential selling outcomes. We examine each of these factors to determine which ones contribute most to the existence of the gap.

We first examine the sellers' reasons for selling. Liquidity-constrained sellers may be willing to sell at lower prices to ease their financial constraints. To investigate this mechanism, we merge property transaction data with mortgage performance data from Black Knight McDash. Using the seller's mortgage delinquency status as a proxy for liquidity constraints, we find that the racial gap in annualized housing returns continues to exist.

We also explore whether unobserved housing quality can account for the differential returns. Although we observe a detailed set of structural property characteristics, there exist other property features that impact property values but are not observed in the data. Examples include the age of appliances, aesthetic appeal, and the number of parking spaces. To the best of our knowledge, a comprehensive dataset covering all these variables does not exist, so we focus on substantial renovations and remodeling to capture differences in housing quality. If Black homeowners systematically invest less in maintaining and renovating their homes, it would be reasonable to expect lower housing returns. We use property-level building permit data to obtain the renovation status of properties two years prior to the home sale. We find that property renovations do not explain the racial gap.

Next, we investigate whether potential racial prejudices by White buyers can explain the gap. One may speculate that White buyers hold prejudiced beliefs against Black sellers and thus offer lower prices. However, after controlling for the race of the buyer, we find that Black sellers experience lower housing returns regardless of the buyer's race. This finding suggests that any possible direct discrimination between the seller and the buyer is not a predominant factor driving the racial gap. It is also important to note that racial prejudices from potential buyers may indirectly affect this gap if they are unwilling to purchase homes from Black sellers or prefer homes owned by White sellers. Such biases would introduce additional search frictions for Black sellers to be matched with a buyer.

Given that none of the characteristics of the property, seller, or buyer explain much of the gap, we shift our focus to intermediaries. We begin by examining the role of real estate agents and brokers, who charge commission fees to sellers for their services in facilitating housing transactions. If the realtor holds prejudiced biases against Black sellers or believes that Black-owned homes are of inferior quality, they may recommend lower selling prices or exert less effort in selling the property. On the other hand, if Black sellers select agents who are less experienced or of lower quality, this can also lead to worse selling outcomes. To test the mechanism, we merge in property listing data to control for real estate agents and brokerage office fixed effects. Our findings show that controlling for real estate agents and brokerage offices only marginally reduces the racial gap.

Additionally, we examine the role of another key intermediary in the selling process: appraisers. Appraisers are expected to provide objective market valuations of properties, which can help lenders, sellers, and buyers determine a fair price. A report from Freddie Mac (Mac (2021)) finds that homes in predominantly Black neighborhoods often receive much lower appraisal values compared to similar homes in predominantly White neighborhoods. Motivated by this fact, we aim to control for the appraisal value at the time of sale. Unfortunately, we do not directly observe the actual appraisal estimates at the time of sale, so we calculate an implied appraisal value by multiplying the appraisal value at the time the seller refinances their mortgage with the tract-level house price growth rate. The implied appraisal values indicate that, on average, Black-owned homes are not valued lower than non-Black-owned homes – in fact, the implied impraisal values suggest that on average, Black-owned homes are valued more than non-Black-owned homes. Moreover, controlling for this appraisal value does not explain the racial gap.

Lastly, we examine the racial differences during the listing process. We find that Black sellers often set lower initial listing prices, reduce listing prices before the sale, offer greater price discounts between the final sale price and the listing price, and take longer to sell their properties. These patterns suggest that Black sellers may experience higher search frictions, compelling them to reduce listing prices to increase the probability of sale. Within the context of a search model framework, our findings suggest that Black sellers face a different level of market tightness than non-Black sellers, contributing to the price disparity. To test this hypothesis, we control for listing price and days on market when estimating the racial gap. Controlling for the listing price accounts for approximately three-fourths of the racial gap, and the racial gap reduces to almost zero after accounting for both factors.

However, it is important to note that the listing price is a noisy measure of a home's true market value, as it can be influenced by biases from various parties. Sellers may tie the listing price to the appraised value they observe, meaning any biases from appraisers would be reflected in the listing price. Additionally, real estate agents or sellers may lower listing prices if they believe that the market tends to discount Black-owned homes or that such homes tend to stay on the market longer. This suggests that listing prices are also influenced by differences in time on market, which are likely driven by differences in search frictions. Consequently, we argue that the estimated 25% direct contribution of search frictions in explaining the racial gap is a lower bound due to the indirect impact that longer time on market has on listing prices. When applying a stricter definition to identify potential short-sale transactions, we find that time-on-market explains 50% of the racial gap.

Overall, our results are consistent with previous studies using Department of Housing and Urban Development Discrimination Survey (HUDDS) data, which provide evidence of racial steering, agents showing Black home buyers fewer homes, and agents catering to prejudiced demands of White buyers (Ondrich et al. (1998), Yinger (1998), Galster & Godfrey (2005), Zhao et al. (2006), Christensen & Timmins (2022), and Christensen & Timmins (2023)). Each of these forces yield larger search frictions for Black sellers. To directly measure this, we combine our listings data with the publicly-available HUDDS data to construct a measure that is a form of market tightness which exhibits the severity of buyer agent discriminatory practices. Our results show that Black sellers no longer face lower housing returns relative to non-Black sellers in zip codes with higher Black buyer home visits relative to listings.

We provide additional evidence that intermediaries are responsible for the racial gap in housing returns by examining the impact of iBuyers, intermediaries who use algorithms to price and purchase homes with the goal of reselling, on housing returns. iBuyers use housing and property characteristics to price homes, removing real estate agents from the process and eliminating their influence on both search and pricing. We restrict the data to zip codes where iBuyers purchase homes, and find that homes purchased by iBuyers do not exhibit a racial gap in housing returns.

In our heterogeneity analysis, we find that the racial gap is larger in neighborhoods with longer seller search duration and when sellers face more within-neighborhood competition. Conditional on buyer demand, our results suggest that the racial gap widens when local markets are more slack. Additionally, we estimate larger racial gaps when Black sellers experience higher neighborhood-level price growth. We also provide estimates of the racial gap in neighborhoods with different socioeconomic characteristics. We categorize neighborhoods into quartiles based on their racial composition, education, income, and homeownership rate. Our findings reveal that the racial gap is economically significant in all types of neighborhoods. However, the racial gap is larger in neighborhoods with higher proportion of Black residents and lower household incomes.

Section 2 describes the data used in the analysis. Section 3 outlines our empirical strategy to identify the racial gap and presents the baseline results. Section 4 explores the sources of the racial gap. Section 5 provides heterogeneity analysis. Section 6 concludes.

Related literature Our paper contributes to the growing literature documenting racial disparities in housing transactions. Kermani & Wong (2024) find that minority home sellers are more likely to enter into distressed sales and when they do, they lose more of their home values compared to White homeowners. Furthermore, our baseline estimates of the annualized unlevered returns gaps for non-distressed sales are of similar magnitudes to their estimates. Bayer et al. (2016) find similar results that minority homeowners are less likely to sustain homeownership when facing adverse economic shocks. However, Diamond & Diamond (2024) find that Black homeowners face higher total housing returns, largely due to experiencing higher rental yields. Instead of analyzing racial gaps in total returns, our paper focuses on the racial gap in capital gains, namely, the home selling process for regular non-distressed home sales. We merge several micro datasets, including building permits, mortgage origination and performance, property listing data, and publicly-available survey datasets to gain a comprehensive understanding of the scope and sources of the racial gap in selling.

Bayer et al. (2017) and Couillard & Christensen (2024) focus on the home buying process and find that Black and Hispanic home buyers pay a premium when purchasing similar housing in the same neighborhood. Courant (1978) demonstrates how this premium can arise in a search model where White sellers are unwilling to sell to Black buyers. We provide additional results using a hedonic approach in Table B.1 to show that the gap we estimate comes from the discounting Black homes and not solely from Black buyer paying a premium for their homes. The main contribution of our paper is to investigate and explain where this selling gap comes from. We find that Black home sellers face substantial search frictions during selling.

Our paper is also related to the literature that examines the role of real estate agents in housing transactions. Several papers (such as Ondrich et al. (1998), Yinger (1998), Galster & Godfrey (2005), Zhao et al. (2006), Christensen & Timmins (2022), and Christensen & Timmins (2023)) use the audit studies from the HUDDS data to investigate the prevalence of discriminatory practices by real estate brokers. They find that real estate brokers show fewer homes to Black households, steer them toward less desirable neighborhoods, and cater to the prejudiced beliefs of White buyers. Additionally, Gilbukh & Goldsmith-Pinkham (2023) show that inexperienced real estate agents often cause worse selling outcomes for their clients. By using property listing data, we control for seller real estate agent fixed effects and investigate the extent to which seller agents contribute to the racial gap in housing returns. We find little evidence suggesting that seller agents are directly responsible for the racial gap, but we acknowledge that our sample size for this analysis is relatively small, which may potentially produce inaccurate estimates. We combine the publicly-available HUDDS data with our listings data to create a zip code and quarter date-level measure of buyer agent discriminatory practices in order to analyze role that buyer agents play in causing the racial gap in housing returns. We believe we are the first to create such a measure.

Our study adds to the growing body of literature that uses high-quality microdata to explore the persistent racial disparities in the housing and financial markets. Avenancio-León & Howard (2022) show that Black homeowners pay higher property taxes due to inflated property assessment values for tax purposes. Bhutta et al. (2022), Frame et al. (2021), Gerardi et al. (2023), and Zhang & Willen (2020) find that minority borrowers are less likely to get mortgage approvals and are charged higher interest rates. Butler et al. (2022), Argyle et al. (2023), Christensen et al. (2021), and Laouénan & Rathelot (2022) find that Black households face racial discrimination in the auto loan market, short-term rental market, and personal bankruptcy courts.

Our paper is also closely related to papers that study differential housing outcomes for different demographic groups. For example, Goldsmith-Pinkham & Shue (2023) find that single women experience a 1.5% lower annualized housing returns compared to single men. Similar to Goldsmith-Pinkham & Shue (2023), we explore various channels that can drive the gap by utilizing detailed property transactions and listing data.

2.2 DATA

We compile data from various sources to gain a comprehensive view of the sales process and the key players involved. In addition to sales records, we also gather information on the renovation status of the property, the sellers' mortgage payment history, and property listing details.

Property transactions and characteristics. The property-level transactions records are from ATTOM, a proprietary data provider that compiles national data from county deed and assessor records. For each property transaction, we observe the transaction date, sale price, deed type, property type, transaction type, property address, and distressed status of the sale (i.e., foreclosure, short sale). We restrict our sample to repeat-sale transactions, that is, properties that have been sold at least twice during our sample period. We exclude non-arms length transactions and partial-consideration sales, as well as sales involving non-individual sellers. We also merge the property transaction data with ATTOM's assessor panel to get detailed property characteristics, such as the number of bedrooms, bathrooms, and assessment values. This allows us to exclude homes that construct additional bed and bathrooms from the repeat-sales analysis.

Identifying sellers and buyers. While the property transaction records provide the names of the sellers and buyers, their race is not recorded in the dataset. We merge the self-reported race of mortgage applicants from the Home Mortgage Disclosure Act (HMDA) to the ATTOM dataset. We follow the same merging procedure used in previous studies, such as Bayer et al. (2017), Bartlett et al. (2022), Kermani & Wong (2024), and Avenancio-León & Howard (2022). We link each property transaction to its respective mortgage origination record using the time and amount of the property transaction, as well as the Census tract where the property is located[1]. To identify the race of the seller, we use the race of the buyer from the previous transaction, where the current seller was the prior buyer.

Given that multiple individuals may be involved in selling and purchasing the home, such as a married couple, we differentiate between mixed-race households and households consisting of members from a single race. We define a household to be black if either the main applicant or any of the co-applicants are Black, so Black households can include mixed-race households by definition. We define a household to be Black-only if both the main applicant and co-applicants are Black. White and White-only households are defined in a similar manner.

We identify iBuyers companies by using their names and their corresponding mailing addresses following the approach in Buchak et al. (2022) using the ATTOM dataset.

Borrowers and appraisals. We get loan-level mortgage performance data from Black Knight McDash. It provides monthly mortgage payment history for each mortgage, and appraisal values if the mortgage was refinanced. We merge the loan-level information to property transaction data from ATTOM following the same procedure used in Issler et al. (2023) and Kermani & Wong (2024). The matching algorithm uses a set of property and transaction attributes, such as loan amount, loan purpose, and interest rate, to find the k-nearest neighbors. We use the mortgage delinquency status as a measure of liquidity constraints for sellers. Specifically, we create an indicator variable to determine whether the seller was delinquent for 90 days within twelve months before selling their home. If the

[1]Note that we are only able to match properties that were purchased with a mortgage. Moreover, the merge is only up to 2016 because HMDA rounds property values to the nearest $10,000s which makes it more challenging to find perfect matches. So our sample does not include selling transactions where the seller purchased their home after 2016. Combined with the transaction data, we are able to observe the race of buyers between 2003 and 2016 and consequently, the race of sellers conditional on purchasing their home between 2003 and 2016.

seller's mortgage was refinanced within two years of the sale, we use the appraisal value at refinance to impute the appraisal value at the time of sale by multiplying it with tract-level housing price growth rate between the appraisal date and the sale date.

Building permits. We use national-level building permits data from BuildFax. We construct indicator variables for properties that have undergone significant improvements two years before the sale date. The improvements we consider include roof replacement, home remodeling, and other substantial renovations that require a permit. For the subsample used in the empirical analysis, we drop property transactions in Census tracts that have zero building permits.

Property listings. We use national property listing data between January 2011 and December 2020 from Altos Listing Intel. The property listing data provides weekly updates on the listing price, close price, days on market, and names of sellers' real estate agents and broker offices. We merge the listing data to the property transaction records using the property address, price, and sale date.

House price indices. We use the developmental Census tract and year-level FHFA House Price Index data to measure Census tract-level price growth between purchase and sale year.

Buyer agent showings. We use the latest HUDDS data published in 2012 to create an aggregate count of homes shown to Black buyers at the zip code and quarter date level. Merging this into our sample allows us to create a measure for Black showings per listing at the Census tract and quarter date level.

Neighborhood characteristics. We use the 2010 American Community Survey 5-year estimates from IPUMS to obtain sociodemographic characteristics at the Census tract level. The neighborhood characteristics we use include racial composition, median household income, median education level, and homeownership share.

Final data sample. Our main sample consists of 1.1 million repeat-sale transactions of single-family homes between 2003 and 2020. We restrict our sample to non-distressed, arms-length, and full-consideration sales. We drop properties that have undergone structural modifications, that is, properties with different numbers of bedrooms and bathrooms between the two transactions. We also exclude transactions with price differences that exceed 100% or fall below -100% to remove outliers. Finally, we exclude transactions in which the property was bought and sold in less than two years to remove potential flip sales.

2.3 THE RACIAL GAP IN HOUSING RETURNS

This section describes our empirical strategy to identify the racial gap in housing returns during property sales and presents our baseline results.

2.3.1 Empirical strategy

The ideal experiment is to compare the selling outcomes of two identical homes sold by sellers of different races in the same market. To replicate the ideal experiment as much as possible, we use a repeat-sales approach. The key advantage of the repeat-sales approach is that it controls for any time-invariant property characteristics between the two transactions, including unobserved property characteristics. We focus on comparing the growth in prices between homes within the same neighborhood while controlling for the timing of the sale and the mean house price growth between purchase and sale dates. This allows us to control for differences in neighborhood amenities and time-varying market conditions.

Our baseline specification estimates whether sellers of different races earn different housing returns. For property i located in neighborhood c sold in time t_1 and purchased in time t_0, we run the following regression

$$r_{i,c,t_0,t_1} = \beta \mathbf{1}(SellerRace_{i,c,t_1} = Black) + \gamma X_{i,c,t_0,t_1} + \alpha_{c,t_1} + \epsilon_{i,c,t}. \tag{2.1}$$

The dependent variable is the housing return on property i. The parameter of interest is β, which captures differential housing returns attributed to the race of the seller in the transaction occurring at time t_1. X_{i,c,t_0,t_1} denotes the control variable we use, specifically the Census-tract level housing price growth rate, calculated between the year the home was sold and the year the home was purchased. The control variable allows us to directly control for differential house price growth trends across different neighborhoods. α_{c,t_1} denotes neighborhood-by-time fixed effects, capturing unobserved shocks at neighborhood-time levels. For example, a neighborhood may start with few nice amenities but experience positive amenity shocks five years later, possibly due to the opening of a new school or a beautiful park. In our estimation, we use Census tracts to define neighborhoods c and year-quarter of sale date to define time t_1.

We use two measures of housing returns. The first is the total unlevered capital gains between the sale date t_1 and purchase date t_0, measured as the log price difference between the two transactions,

$$r^{tot}_{i,c,t_0,t_1} = \ln P_{i,c,t_1} - \ln P_{i,c,t_0}.$$

Our preferred measure is the annualized returns to standardize housing returns across transactions with varying holding periods. Unlike total capital gains, which do not differentiate the lengths of ownership, the annualized measure explicitly accounts for the variation in the lengths of asset holding. We define the annualized unlevered capital returns similar to Goldsmith-Pinkham & Shue (2023) and Kermani & Wong (2024). It is defined as

$$r^{ann}_{i,c,t_0,t_1} = \left(\frac{P_{i,c,t_1}}{P_{i,c,t_0}}\right)^{365/(t_1-t_0)} - 1$$

where the difference between the two transaction dates t_1 and t_0 is measured in days.

2.3.2 Baseline results

Table 2.1 presents the average racial gap in total and annualized capital gains from housing using the baseline regression from Equation 2.1. Intuitively, the regression compares the price growth of homes located in the same neighborhood and sold in the same year-quarter date while controlling for the price growth between purchase and sale dates, differing only in being owned by Black versus non-Black homeowners. In all specifications, we cluster standard errors at the Census tract level. As discussed in Section 2, we use two measures to categorize the race of the seller: whether at least one member of the sellers is Black, and whether all sellers involved in the transaction are Black.

Column (1) shows that on average, Black sellers realize 1.74% lower total capital gains when selling their houses compared to non-Black sellers. Additionally, Column (2) shows that sellers identified as Black-only, on average, realize total capital gains 1.88% lower compared to non-Black-only sellers. Columns (3) and (4) present the average gap in annualized capital gains for Black and Black-only sellers, respectively. We find that the average gap in annualized capital gains of a Black seller is 0.36% lower than that of non-Black sellers, and 0.38% lower for Black-only sellers than that of non-Black-only sellers.

Table 2.1: Racial gap in returns to housing

	Capital Gains Returns		Annualized Returns	
	(1)	(2)	(3)	(4)
Black seller	-0.0174***		-0.0036***	
	(0.0011)		(0.0003)	
Black-only seller		-0.0188***		-0.0038***
		(0.0013)		(0.0003)
Observations	1,151,718	1,151,718	1,151,718	1,151,718
R^2	0.701	0.701	0.680	0.680

Notes: This table presents the average racial gap in total and annualized capital gains using the baseline regression Equation 2.1. Columns (1) and (2) report the gap in total capital gains, while Columns (3) and (4) report the gap in annualized returns. Black sellers are defined as sellers that include at least one Black individual, and Black-only sellers are defined as sellers that are all Black individuals. All regressions include the Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. Standard errors are clustered at the Census tract level. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

To assess the magnitude of the gap in dollar terms, we conduct a simple back-of-envelope calculation. Imagine two homes, both in the same neighborhood and each bought for $250,000 in February 2012, but one was purchased by a Black-only household, while the other one was purchased by a non-Black-only household. Furthermore, imagine these two homes were sold after eight years, approximately the average duration of homeownership. In this time period, mean house price growth in the US was approximately 60%. If the non-Black household's home was sold for $400,000, our baseline annualized results suggest that the Black-only household would have received $11,323 less when selling the home.

One potential concern with our findings is the possibility that the gap solely arises from Bayer et al. (2017)'s results that Black home buyers tend to pay more for their homes, thereby mechanically reducing their housing returns. If this were the case, we would expect to find no racial gap on the seller's transactions when using a hedonic approach. We shed light on this matter by estimating a simple hedonic regression by regressing the log sale price on a binary variable for the race of the seller along with characteristics of the seller's home and interacted Census tract and year-quarter sale date fixed effects. The characteristics of the home controlled for include the building square footage and lot area, age, heating and cooling systems, pools, garages, and categorical dummies for the number of bedrooms and bathrooms. Because the specification is only conditional on a home's characteristics at the time of sale, we do not need to exclude transactions with changes in the home's characteristics. This significantly increases the sample size by a factor of two.

Appendix Table B.1 displays the regression results. Column (1) presents the results comparing Black and non-Black sellers and Column (2) presents the results for Black-only and non-Black-only sellers. The hedonic regression specification finds approximately a -4.1% racial gap in sales price for Black sellers and a -4.6% racial gap in sales price for Black-only sellers. Overall, we interpret these results as evidence that the racial gaps in housing returns that Black sellers experience are not solely explained by a racial gap on the purchasing side.

2.4 EXPLANATIONS OF THE RACIAL GAP

In the last section, we document that, on average, Black households receive lower returns on their homes relative to non-Black households. We explore the underlying mechanisms driving the gap in this section. We use the annualized housing return as the outcome variable to account for any differences in holding periods.

We find that this negative gap in returns is not explained by liquidity constraints, nor by observable differences in housing renovations. On average, the race of the buyer does not appear to explain the racial gap either. Using proxy measures for appraisals, we also find no evidence that appraisal values explain away the gap. Additionally, controlling for seller agent and brokerage office fixed effects does not rationalize the racial gap.

However, controlling for listing price and time on market, which is likely influenced by real estate agents involved in a housing transaction, effectively explains the entire gap. Additionally, less exposure to buyer agent discriminatory practices mitigates the returns gap. Finally, we observe no racial gap among housing transactions that involve an iBuyer purchasing the property, i.e., transactions where buyer real estate agents are not directly involved in the matching process. The results strongly suggest that Black sellers' disparities in search frictions, caused by real estate agent practices, are the main driver of the racial gap.

2.4.1 Seller and property characteristics channel

We begin by examining the role of the seller's liquidity position on sale price differences. If the seller is liquidity-constrained, they may be willing to accept lower prices to sell the house faster. To test the mechanism, we bring in additional data from Black Knight McDash's mortgage performance data. We use the seller's mortgage delinquency status as a proxy for the seller's liquidity position. If a seller has been delinquent on their mortgage payments for more than 90 days, they are nearing default and face the high risk of losing their home soon. Consequently, a liquidity-constrained seller might prioritize a quick sale over finding a buyer with a higher offer to ease their financial burden. We construct an indicator variable of whether the seller was at least 90 days delinquent on their mortgage within 12 months of selling the home.

On the other hand, the disparities in housing returns might be attributable to differences in unobserved property characteristics correlated with race. Any relation between differential investment in a seller's home and their race could potentially generate a racial gap. We test the validity of this claim by merging building permit data from BuildFax. We create three indicator variables to capture the renovation status of a property: the installation of a new roof, the presence of home remodeling, and other types of replacements that require a permit. We measure these variables within two years before the sale. We acknowledge that there exist many property features that can impact property values but are not observed in the data. Examples include the aesthetic appeal of the house, the availability of a central air conditioning system, the age of appliances, and the status of the basement finishing. To the best of our knowledge, a complete dataset including all property features does not exist. Therefore, we focus on significant renovations and remodeling to capture differences in housing quality, as these factors are likely to have a more pronounced impact on property values compared to other minor property features.

Table 2.2 presents estimates of the annualized racial gap, after controlling for the seller's liquidity status and the renovation status of the properties. Since we merge the main transactions sample with mortgage performance and permits data from the Black Knight McDash and BuildFax, respectively, the size of the merged dataset drops from 1.1 million to about 600,000 observations. In Columns (1) and (4), we rerun the baseline regression Equation 2.1 using the newly merged sample. The coefficients estimated from the newly merged samples have similar magnitudes to those in the baseline sample, shown in Columns (3) and (4) of Table 2.1. The similarity suggests that the characteristics of the newly merged samples closely align with those in the baseline sample.

Table 2.2: Seller and property characteristics channel

	Annualized housing returns					
	(1)	(2)	(3)	(4)	(5)	(6)
Black seller	-0.0032***	-0.0029***	-0.0029***			
	(0.0003)	(0.0003)	(0.0003)			
Black-only seller				-0.0032***	-0.0029***	-0.0029***
				(0.0003)	(0.0003)	(0.0003)
Delinquent 90d+		-0.0098***	-0.0098***		-0.0098***	-0.0098***
		(0.0003)	(0.0003)		(0.0003)	(0.0003)
Roof			0.0030***			0.0030***
			(0.0005)			(0.0005)
Replacement			0.0028***			0.0028***
			(0.0004)			(0.0004)
Remodel			0.0114***			0.0114***
			(0.0008)			(0.0008)
Observations	593,895	593,895	593,895	593,895	593,895	593,895
R^2	0.723	0.723	0.724	0.723	0.723	0.724

Notes: This table presents the estimates of the racial gap in annualized returns when we control for seller's mortgage delinquency status and property renovations. All regressions follow the same specification in Equation 2.1. All regressions include the Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. The first three columns show the estimates for Black sellers, while the last three columns show the estimates for Black-only sellers. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

In Columns (2) and (5), we control for the mortgage delinquency status of the seller. The estimated racial gap in annualized housing returns is -0.29%, which is very similar to the baseline estimates of -0.32% in Columns (1) and (4). This suggests that the differences in need for liquidity between Black and non-Black sellers do not substantially account for the gap in final housing returns. We further add controls for property improvements in Columns (3) and (6). The estimated annualized racial gap remains at -0.29%, the same as the estimates in Columns (2) and

(5). The results indicate that neither the seller liquidity nor property renovations explain an economically significant part of the observed racial gap.

2.4.2 Race of buyers

Next, we examine the role of the buyer, with a specific focus on the buyer's race. Given the historical prevalence of prejudice against Black households by White individuals, we assess whether Black sellers receive lower housing returns when selling their homes to White buyers. If White buyers hold prejudiced beliefs against Black sellers or their homes, they may offer lower prices. The race of the buyer is determined by the self-reported race from HMDA. We classify a buyer as White if the household is exclusively White, meaning the household is not mixed-race.

Table 2.3 presents the results while accounting for the race of the buyer. The specifications are similar to our baseline regressions, with an additional interaction term between the seller's race and the buyer's race. Columns (1) and (3) follow the same specifications as the baseline regressions shown in Columns (3) and (4) of Table 2.1, respectively. The sample including the buyer's race is smaller because we are not always able to identify the race of the buyer when we merge the baseline sample with HMDA. The magnitudes of the racial gap are slightly smaller in this subset of the data, but they remain economically meaningful.

Columns (2) and (4) present results where we explore the role of White buyers in explaining the racial gap by interacting the race of the seller with the race of the buyer. We find that the interaction does not absorb the racial gap estimate for Black sellers. If the racial gap was primarily caused by potentially prejudiced White buyers, then the interaction term between a Black seller and a White buyer would yield a statistically significant negative estimate that absorbs the coefficient of the Black seller indicator. However, the estimate on the interaction term is close to zero and is neither statistically nor economically significant. Also, the estimated racial gaps in annualized housing returns are -0.23% for Black sellers and -0.30% for Black-only sellers. The magnitudes are similar to the baseline regressions without accounting for the race of the buyer. Overall, Black sellers experience lower housing returns regardless of the race of the buyer, suggesting that any potential direct discrimination between the seller and the buyer is not a dominating force driving the racial gap.

Table 2.3: Race of buyers

	Annualized housing returns			
	(1)	(2)	(3)	(4)
Black seller	-0.0027***	-0.0023**		
	(0.0005)	(0.0008)		
Black seller × White buyer		-0.0007		
		(0.0011)		
Black-only seller			-0.0032***	-0.0030***
			(0.0006)	(0.0009)
Black-only seller × White buyer				-0.0002
				(0.0012)
White buyer		0.0005**		0.0005**
		(0.0002)		(0.0002)
Observations	412,203	412,203	412,203	412,203
R^2	0.737	0.737	0.737	0.737

Notes: This table presents the estimates of the racial gap when we control for buyer's race. We use the subsample that we can identify the buyer's race for this table. In Columns (1) and (3), we rerun the baseline regression for Black and Black-only sellers to get the baseline racial gap. We include an interaction term between the race of the seller and the race of the buyer in Columns (2) and (4). All regressions include the Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

2.4.3 Seller real estate agent and broker office

The majority of real estate transactions in the US are facilitated by real estate agents. Specifically, the listing agent, who represents the seller, plays a crucial role in shaping the selling outcomes. The listing agent provides a range of services, including advising on the list price, staging and advertising the property, showing homes to potential buyers, and negotiating the close price on behalf of the seller. If Black sellers work with realtors who exert lower efforts due to racial prejudice, inexperience, or lower intrinsic quality, Black sellers may experience worse selling outcomes. In this section, we use detailed property listing data from Altos Listing Intel to investigate how much of the gap can be attributed to seller real estate agents.

Occasionally, real estate agent and office names are shortened, re-arranged, or misspelled. We first manually standardize these names to minimize instances of the same agent/office being assigned multiple id numbers, such as instances like "Century 21" and "Century twenty-one". We then use a bigram decomposition algorithm to fuzzy-match similar agent name string and siminlar office name strings within a CBSA.

We rerun our baseline regression using the repeat-sale sample of the matched transaction-listing data. To isolate the role of real estate agents, we control for listing agent fixed effects. The names of real estate agents are often missing in the data, so we exclude listings and associated transactions without agent names. The sampling criteria shrink our sample size significantly, with the subsample with real estate agent information dropping to less than 50,000 observations. This reduction in sample size is due to the lack of data reporting issues in Altos, rather than a lack of representations of real estate agents in these transactions. So we create another subsample where we observe brokerage offices and include them as fixed effects. The results for annualized returns are presented in Table 2.4. In Columns (1) to (4), we use the sample in which we can identify the real estate agents on the listings. Analogously, Columns (5) to (8) use the sample where we can identify the brokerage office. To ensure the subsamples are comparable to our primary sample, we rerun the baseline regressions without agent and brokerage controls. The results are shown in Columns (1), (3), (5), and (7). The estimated racial gaps in the subsamples are around -0.30% to -0.38%, which are very similar to our baseline results shown in Table 2.1.

Table 2.4: Role of listing agent and broker offices

	Annualized housing returns							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black seller	-0.0030**	-0.0024			-0.0038***	-0.0035***		
	(0.0011)	(0.0014)			(0.0005)	(0.0005)		
Black-only seller			-0.0036**	-0.0033*			-0.0038***	-0.0036***
			(0.0013)	(0.0017)			(0.0006)	(0.0006)
Observations	48,852	48,852	48,852	48,852	231,727	231,727	231,727	231,727
R^2	0.735	0.846	0.735	0.846	0.714	0.743	0.714	0.743
FE		agent		agent		office		office

Notes: This table presents estimates of the racial gap in annualized returns when we control for real estate agents and brokerage office fixed effects. Columns (1) to (4) use the matched transaction-listing sample in which the real estate agent information is available. Columns (5) to (8) use the matched transaction-listing sample in which the brokerage office is available. All regressions include the Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Columns (2) and (4) estimate the racial gap including listing agent fixed effects. We find that adding listing agent fixed effects attenuates the racial gap to some extent, reducing the annualized gap by 0.06% for Black sellers and 0.03% for Black-only sellers. However, they still do not fully explain the racial gap. Columns (6) and (8) estimate the racial gap including brokerage office fixed effects. We find that adding brokerage office fixed effects only reduces the gap marginally, reducing the annualized gap by 0.03% for Black-sellers and 0.02% for Black-only sellers. We conclude that on average, any direct discrimination from real estate agents towards their Black clients does not contribute significantly to the gap in housing returns.

2.4.4 Appraisal measures

Appraisals play a crucial role in real estate transactions by providing objective accurate evaluations of a property's market value. The appraisal estimates are widely used in housing transactions. For example, sellers use appraisal values as benchmarks for setting listing prices, lenders use them to assess the collateral value, and buyers use them as references when negotiating the final sale prices with sellers.

When appraising a home, appraisers fill out a form detailing the characteristics of a home and provide an estimate of the fair market value. The appraisers record structural features such as the number of bedrooms and bathrooms. These objective measures leave little room for any sort of bias. However, other subjective measures such as the perceived quality of the home's feature, opens the possibility for subjective racial biases to influence fair evaluations. Recent newspaper stories have reported instances in which the appraisal value rose by over $30,000 after a Black couple removed family photos and other personal items (Romo (2023)). It should be noted that some subjective measures, such as the drive to a home, are more likely to affect neighborhood-level differences in the racial gap as shown in FHFA blog posts (Broadnax & Wylie (2021) and Vajja (2023)), but they are less likely to affect the within-neighborhood gap that we estimate.

Although we do not have data on the actual appraisal estimates received by home sellers before listing their homes on the market, as well as the appraisal estimates requested by lenders for buyers, we impute a measure that is likely highly correlated with these actual appraisal estimates.

We impute an implied appraisal value based on the appraisal values the seller receives when they refinance their property. For households who refinance their mortgages, the lender re-appraises the property value in order to determine the loan terms of the new mortgage. Using loan characteristics data from Black Knight McDash, we are able to obtain the appraised property value ordered by the lenders for refinancing purposes. We then multiply the appraisal value by the FHFA's house price growth rate at the Census tract level to impute the appraisal value at the time of sale. The key assumptions made are that the appraisal values grow at the same rate as the average price growth rate and that the racial gap in the appraisals does not vary between refinancing a mortgage and selling a home.

We first re-estimate the racial gap in housing returns if sellers were to receive sale prices equal to the imputed appraisal values. Table 2.5 presents the results. We recalculate annualized returns using the implied appraisal value as the numerator, and rerun the baseline regressions. Columns (1) and (2) report a positive racial gap of 0.44% for Black sellers and 0.48% for Black-only sellers if sale prices equal to the implied refinance appraisal values. This result is the opposite of the racial gap in returns to housing which we have been finding and strongly suggests that, on average, the value of homes Black owners sell is not less than the value of homes White owners sell. We view this as strong evidence that homes sold by Black sellers should, on average, grow in value by at least as much as, if not more than, homes sold by non-Black sellers; in other words, these results strongly suggest that racial differences in unobserved home investment or depreciation are likely not valid explanations for the racial gap in housing returns. It is important to note that this data sample restricts sellers to those who refinanced their mortgage. The next step is to first see if a racial gap exists with this subset of the data, and then see if controlling for these implied appraisal values absorbs the racial gap.

Table 2.5: Racial gap in appraisals

	Annualized returns implied from appraisals	
	(1)	(2)
Black seller	0.0044*	
	(0.0019)	
Black-only seller		0.0048*
		(0.0021)
Observations	117,027	117,027
R^2	0.584	0.584

Notes: This table presents the estimates of the racial gap in annualized housing returns if the transaction price were the imputed appraisal value based on the appraised value the seller received when they refinance their mortgage. In other words, the left-hand side variable is the annualized return from purchasing at the actual purchased value and selling at the imputed appraisal value. The appraised value received at the time of refinancing is adjusted by house price growth between the appraisal date and sale date. All regressions include Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

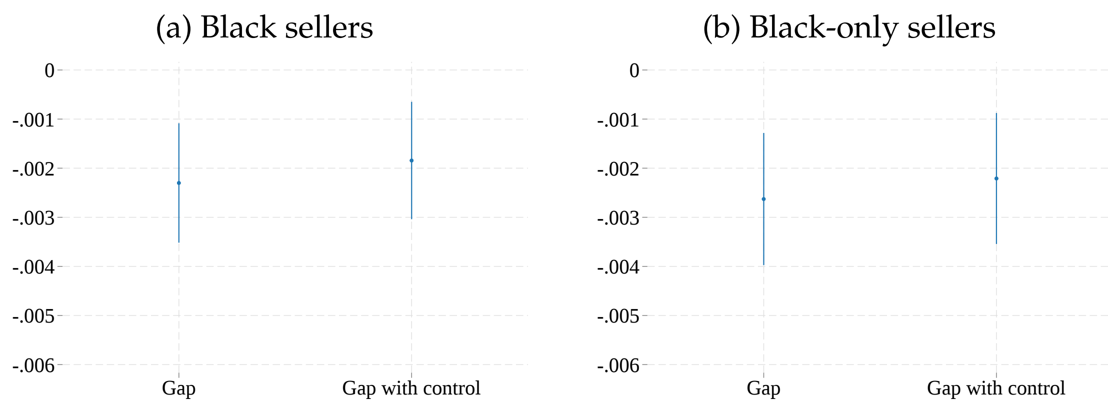
Figure 2.1 presents the point estimates and 95% confidence intervals for the annualized racial gap after controlling for the implied appraisal measure. We do not control for house price growth in the plots.[2] The left panel shows the results for Black sellers, and the right panel shows the results for Black-only sellers. In each plot, the first coefficient provides the annualized gap without controls, and the second coefficient provides the annualized gap after controlling for the appraisal measure.

The annualized gaps without controlling for appraisal values are -0.22% and -0.27% for Black and Black-only sellers, respectively, smaller than the estimates pro-

[2]We exclude the FHFA house price growth control because the price growth from our appraisal measures and the seller's purchase price serve the same role.

vided in Columns (2) and (4) of Table 2.1, but they are still statistically significant at the 95% level. The estimated gaps after including the implied appraisal values have similar magnitudes, with values -0.19% and -0.22%, respectively. The results suggest that, on average, appraisers likely do not play a major role in explaining the racial gap.

Figure 2.1: Racial gap in housing returns controlling for appraisals



Notes: This figure plots the point estimates and 95% confidence intervals for the racial gap in annualized housing returns after controlling for appraisal measures. The implied appraisal values are derived from the appraisal value at mortgage refinancing multiplied by average price growth rate at the Census tract level. Each plot shows the estimates without and with appraisal controls for Black and Black-only sellers. All regressions include the Census tract by sale year-quarter fixed effects. Standard errors are clustered at the Census tract levels.

2.4.5 Search frictions

This section investigates the role of listing prices and time on market in explaining the gap Black homeowners face when selling their homes. We find that accounting for both the time on market and listing price can explain nearly all of the racial gap.

We first document a few key facts about racial differences in listing outcomes shown in Table 2.6. The listing price and time on market information are merged in from Altos Listing Intel data. We use the baseline regression specification shown in Equation 2.1, but replace outcome variables with listing characteristics.

Listing price is an estimate of the market value of a home, and it is often closely tied to an appraised value sellers receive. A seller's agent may also provide input on the listing price – for instance, if an agent believes that it would be difficult to find a buyer for a house, and thus be on the market for an extended period of time, they may suggest lowering the listing price. Additionally, sellers themselves may have their own views on what the listing price should be – potentially lowering the listing price themselves if their home is listed for a long period of time without receiving a sufficient offer. As a result, a listing price may be influenced by the bias of several players. If these biases are correlated with race of the seller, they are likely to influence the relative difference in sale prices between Black and non-Black sellers.

The regression specification in Columns (1) and (2) is similar to that in Table 2.5 where we re-estimate the racial gap in housing returns if sellers were to receive sale prices equal to the last listing price we observe prior to sale. The results indicate that Black sellers would face lower annualized returns implied by our listing price measure. Given our findings from Section 2.4.4, we suspect that racial differences in listing price are likely due to the influence of biased parties and not due to unobserved quality of the home.

Columns (3) and (4) regress the log change between the final listing price and the first listing price on the race of the seller. We find that Black and Black-only sellers lower their listing prices by 0.06% more than non-Black and non-Black-only sellers, respectively. Columns (5) and (6) regress the log difference between the final sale price and the final listing price on the race of the seller. The regression re-

Table 2.6: Listings characteristics gaps

	Listing price gap		Listing price change		Sale-listing gap		Days on market	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black seller	-0.0033***		-0.0006**		-0.0024**		8.4589***	
	(0.0005)		(0.0002)		(0.0008)		(0.8008)	
Black-only seller		-0.0032***		-0.0006*		-0.0031***		9.4490***
		(0.0005)		(0.0003)		(0.0009)		(0.9259)
Observations	296,052	296,052	595,242	595,242	595,245	595,245	595,249	595,249
R^2	0.705	0.705	0.348	0.348	0.411	0.411	0.458	0.458

Notes: This table presents the racial differences in listing outcomes using the matched transaction-listing sample. The first two columns show the racial difference in listing price adjustment, measured by the log difference between the first and final listing prices. The third and fourth columns show the racial difference in the sale-listing price gap, measured as the difference between the sale price and the final listing price. The last two columns show the racial difference in the number of days on market. All regressions include the Census tract by sale year-quarter fixed effects. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

sults indicate that Black and Black-only sellers sell their homes at larger discounts of -0.24% and -0.31% relative to non-Black sellers.

Columns (7) and (8) regress the number of days it takes to sell the house on the race of the seller. We find that Black homes are listed on the market for over a week longer than non-Black homes. Given that we include interacted neighborhood and time of sale fixed effects, the difference in days on market between Black and non-Black sellers is particularly stark. The results show that Black sellers face lower probabilities of matching with a potential buyer, suggesting that Black sellers may face higher search frictions during the selling process. Additionally, this racial difference in time on market likely explains much of the racial difference in listing prices we display in Columns (1) and (2). Within the context of a search model, our findings suggest that Black sellers face a different level of market tightness than non-Black sellers. This result is consistent with the results from the literature that uses the HUDDS data to show that real estate agents have racially steered buyers, shown Black home buyers fewer homes, and catered to prejudiced preferences of home buyers (Galster & Godfrey (2005), Ondrich et al. (1998), Yinger (1998), Zhao et al. (2006), Turner et al. (2013)).

Several factors can explain the racial gap in listing outcomes. First, Black sellers learn about their true property values during the listing process. If the house receives few visits and offers from buyers, the Black seller may learn that they need to adjust prices to attract buyers. Second, the listing agent may put less effort in facilitating the sale of the Black-owned properties, lowering the probability of sale. As a result, the Black seller may respond by lowering listing prices to increase the probability of sale. Third, Black sellers may have less negotiating power when closing the sale, and this can explain why there is a gap between the final sale price and final listing prices. The difference in negotiation skills can be due to bad recommendations by the listing agent, inexperience by sellers, or property defects found during home inspections. Results in the following subsections suggest the former.

Our next set of results tests whether listing behaviors can explain the racial gap. Table 2.7 uses our baseline specification, and we incrementally add controls for listing price growth and time on market.[3] The sample is smaller than the baseline sample because we are not able to find matches for all transactions when merging with the listings data.

In Columns (1) and (4), we rerun the baseline regression using the new subsamples. The estimated racial gaps are of similar magnitudes as the results in Table 2.1. We control for the annualized price growth between the final listing price and purchase price in Columns (2) and (5). Here, the racial gap in annualized returns

[3] We exclude the FHFA house price growth control because the price growth from listing price and the seller's purchase price serves a similar role.

Table 2.7: Search frictions regressions

	Annualized housing returns					
	(1)	(2)	(3)	(4)	(5)	(6)
Black seller	-0.0041***	-0.0010*	-0.0004			
	(0.0004)	(0.0004)	(0.0004)			
Black seller × 1/Time on mkt			-0.0013**			
			(0.0005)			
Black-only seller				-0.0043***	-0.0010*	-0.0004
				(0.0004)	(0.0005)	(0.0004)
Black-only seller × 1/Time on mkt						-0.0015**
						(0.0005)
Listing price growth		0.7793***	0.7773***		0.7793***	0.7773***
		(0.1031)	(0.1030)		(0.1031)	(0.1030)
1/Time on mkt			0.0079***			0.0079***
			(0.0007)			(0.0007)
Observations	440,386	440,386	440,386	440,386	440,386	440,386
R^2	0.543	0.905	0.907	0.543	0.905	0.907

Notes: This table presents the estimates of the racial gap in annualized returns after controlling for listing price difference and weeks on market. All regressions include the Census tract by sale year-quarter date fixed effects. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

decreases substantially, from -0.4% to -0.1%. Including the listing price reduces the gap for two main reasons. First, any bias inherent in the listing price can transmit to the racial gap in the final returns. Second, the listing price reflects unobserved heterogeneity of the property itself. Although we cannot disentangle these two forces from the listing price, our results from Section 2.4.4 suggest the latter explanation is less likely to be legitimate. Additionally, listing price may be lower precisely because Black sellers' homes spend more time on the market – suggesting that search frictions may directly play a role in the lower listing prices Black sellers set.

Columns (3) and (6) include the inverse of the total weeks the home is on the market as well as its interaction with the race of the seller and the listing price control. We include the inverse of time on market because it directly maps to the probability of matching with a buyer. We find that accounting for both the listing price and time on market reduces the racial gap by a factor of ten, and it removes the statistical significance of the estimate as well. We interpret these results as evidence that accounting for the differential search frictions Black sellers face explains the price disparity. When the time on market decreases, that is, it takes a short time to find a buyer, non-Black sellers see a higher increase in their annualized returns than Black sellers due to the negative coefficient in the interaction term. For instance, the median number of weeks on market in our data is approximately 5 weeks, and the results in Table 2.6 give the average within-neighborhood difference estimates in time on market between Black and non-Black sellers and Black-only and non-Black-only sellers of 8.5 and 9.4 days or close to one and half weeks. If it took a Black seller 5 weeks to match with a buyer, a difference in time on market of one and a half weeks between Black and non-Black sellers and Black-only and non-Black-only sellers yields a difference in annualized returns of approximately 0.09% and 0.1%, respectively. This equates to roughly one-quarter of the total racial gap we estimate. However, longer time-on-market can lead sellers to lower their listing price – thus, racial differences in time on market are likely to impact listing prices. Consequently, we believe our estimates of search frictions explaining the racial gap to be a lower bound.

2.4.6 Buyer agent discriminatory practices

In this section, we construct a measure of exposure to discriminatory practices by real estate agents, using data from the HUD Discrimination Study (Turner et al. (2013)). We then assess its impact on the racial gap in housing returns. The HUD experiment uses paired testing to detect discriminatory practices among real estate agents. It arranges pairs of potential home buyers of different races to contact real estate agents and inquire about a randomly selected set of homes. This experiment conducted over 8,000 tests in 28 metropolitan areas nationwide in 2011 and 2012. The study finds that real estate agents practice racial steering and show Black home seekers 17.7% fewer homes than White buyers, despite both groups being equally qualified. These discriminatory practices directly impact home sellers' sale prices, especially for those more exposed to these practices, as they reduce the probability of matching with a potential buyer.

Therefore, we construct a zip code-quarter date-level measure to quantify the extent of these discriminatory practices. Using HUDDS data, we compute the number of showings to Black home seekers for each zip code and divide it by the number of available listings in the same area, as recorded in the Altos listings data. This measure is constructed analogously to market tightness. A lower value indicates a higher prevalence of discriminatory practices, not only reflecting greater difficulty for Black home buyers to find a match, but combined with the racial steering results directly impacts the match probability for Black sellers as well, consequently lowering house prices because of greater search frictions. Additionally, these discriminatory practices may have larger consequences for Black sellers in states that allow for dual agents.

Table 2.8 illustrates the impact of such discriminatory practices on the racial disparities in annualized housing returns. The regression is the same as our baseline specification in Table 2.1 except that we interact the measure with the race of the seller. Due to the data limitations of the HUDDS study as well as the time period having relatively few transactions not involving a distressed sale, the sample size for this analysis is much smaller. The sample includes 6,095 repeat-sale transactions across 831 zip codes in 25 CBSAs. All regressions include zip-code and sale year-quarter fixed effects, with standard errors clustered at the zip-code level.

Table 2.8: Controlling for buyer agent's discriminatory practices

	(1)	(2)	(3)	(4)
Black seller	-0.0101*	-0.0162**		
	(0.0044)	(0.0053)		
Black seller × Black showings/listings		0.0936*		
		(0.0457)		
Black-only seller			-0.0114*	-0.0200**
			(0.0058)	(0.0064)
Black-only seller × Black showings/listings				0.1350**
				(0.0475)
Black showings/listings		0.0003		0.0001
		(0.0044)		(0.0045)
Observations	6095	6095	6095	6095
R^2	0.361	0.363	0.361	0.364

Notes: Regressions include zipcode and year-quarter date of sale fixed effects. Due to the nature of the HUD Discrimination Study experiment, regressions are weighed by the number of tester-home pairs within a zipcode. Regressions include control for tract-level annualized house price growth between year of purchase and year of sale. Sample period spans 2011-2012 – the period of the HUD Discrimination Survey. Reported standard errors are clustered at the zipcode level. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Column (1) reports the racial gap in annualized housing returns for this analysis sample. Consistent with our baseline results, Black sellers earn about 1% lower returns compared to non-Black sellers. In Column (2), we control for the measure of buyer agent discriminatory practices. The key coefficient of interest is the interaction term between the Black seller indicator variable and the discriminatory measure constructed from HUD data. The positive coefficient on the interactive term indicates that increasing the number of showings to Black home seekers reduces the racial gap in housing returns. Specifically, increasing the Black showing-to-listing ratio by 1 is associated with a 9% increase in housing returns, offsetting the negative racial gap for Black sellers. Columns (3) and (4) replicate the analyses from Columns (1) and (2), but replace the Black home seller indicator with a Black-only seller indicator. Although the magnitudes are larger relative to those previously estimated in this paper, likely reflecting both contemporary market conditions with high price volatility and some measurement error, the stark results are in line with those in Section 2.4.5 suggesting that Buyer agents play a large role in explaining the racial gap in housing returns.

2.4.7 iBuyers

In this section, we examine the racial gap when a buyers' agent is removed from the selling process. In order to do this, we examine homes purchased by iBuyers – companies that use algorithms to make instant offers to buy and sell properties. By removing buyers' agents from the selling process, we eliminate key historical discriminatory practices which lead to differences in search frictions and directly impact prices like racial steering, black homes being shown to fewer buyers, and catering to prejudiced demands of buyers as documented in the literature using the HUD Discrimination Survey (Ondrich et al. (1998), Yinger (1998), Galster & Godfrey (2005), Zhao et al. (2006), Turner et al. (2013), Christensen & Timmins (2022), and Christensen & Timmins (2023)).

We identify iBuyer purchases by name and mailing address following the approach described in Buchak et al. (2022). iBuyers typically purchase homes that they believe are relatively easy to sell. As described in Buchak et al. (2022), we focus on a data sample of homes that are in an accessible price range that are relatively young. Consequently, we restrict the data sample to include transactions where the transaction value is within \$150,000–\$350,000 and where the age of the home is no more than 40 years old.

The regression specification is the same as Equation 2.1 except that we interact the race of the seller with a binary indicator for whether the home was purchased by an iBuyer. Due to the relatively small number of iBuyer home purchases, we use zip code by year of sale fixed effects and control for zip code-level annualized mean house price index growth between purchase year and sale year. Table 2.9 presents the results for Black sellers in Column (1) and Black-only sellers in Column (2). The coefficient estimates for the black seller indicator and the iBuyer indicators are negative in both columns. The magnitude of the racial gap is in line with previous results, and the magnitude of the iBuyer coefficient denotes the annualized discount sellers observe from selling to an iBuyer. However, the coefficient for the interaction is positive and offsets the negative racial gap suggesting that when we remove players who potentially impact prices, such as buyers' agents, the racial gap is, on average, eliminated. Overall, this set of results combined with the results from Tables 2.7 and 2.8 provide strong evidence for the role that real estate agents have in creating this racial disparity in housing returns.

2.4.8 Robustness

Alternative delinquency definitions As a robustness check, we replicate the results from Table 2.2 with alternative measures of seller delinquency status. Table B.2 reports the results where we proxy liquidity constraints with any 30 day or longer delinquency on the sellers' mortgage within a year of selling their property, Table B.3 reports the results where we proxy liquidity constraints with any 90 day or longer delinquency on the sellers' mortgage within one month of selling their property, and Table B.4 reports the results where we proxy liquidity constraints

Table 2.9: iBuyer impact

	(1)	(2)
Black seller	-0.0025***	
	(0.0004)	
Black-only seller		-0.0029***
		(0.0004)
iBuy	-0.0055***	-0.0054***
	(0.0005)	(0.0005)
Black seller × iBuy	0.0037*	
	(0.0019)	
Black-only seller × iBuy		0.0036
		(0.0020)
Observations	218,324	218,324
R^2	0.695	0.695

Notes: Regressions include zip code-level annualized mean house price index growth between purchase year and sale year as well as zip code tract × year of sale date fixed effects. Reported standard errors are clustered at the zip code level. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

with any 30 day or longer delinquency on the sellers' mortgage within one month of selling their property. Overall, the results do not change from those in Table 2.2, demonstrating the robustness of our liquidity constraint result.

Restricting additional potential short-sales The data sample used in our main analysis identifies and excludes short sales using a proprietary algorithm from ATTOM. In this section, we perform the same analysis from the main sections of the paper with more restrictive exclusion of property transactions that are potentially short-sales. Figure B.1 presents these results. First, we exclude transactions that ATTOM identifies as a short sale. Second, we add an additional exclusion rule by excluding transactions identified by an algorithm first used in Ferreira & Gyourko (2015). This algorithm identifies transactions as a short sale if the transaction value is less than 90% of the initial unpaid principle balance. Like Kermani & Wong

(2024), we find that the definition of short sale does not qualitatively change the results. However, the results in this potentially cleaner sub-sample suggest time-on-market explains 50% of the racial gap for Black sellers.

First, we replicate the baseline regression from Table 2.1 in the first plotted coefficient and estimate a racial gap in housing returns of -0.28% annually. Next, we estimate the racial gap after controlling for delinquency status and renovations in the third plot using the same specification from Table 2.2 with the second plot being the racial gap on the same sample without the seller characteristics controls. The fifth coefficient plots the racial gap after controlling for white buyers using the same specification from Table 2.3 with the fourth plot being the estimated racial gap on the same sample without controlling for the buyers' race. The seventh coefficient plots the racial gap after controlling the implied growth imputed from the values observed up refinancing one's home using the same specification from Figure 2.1 with the sixth plot being the estimated racial gap on the same sample without the appraisal control. Finally, the tenth and last coefficient plots the racial gap after controlling for the inverse of listed time-on-market and listing price using the same specification from Table 2.7 with the ninth plot being the estimated racial gap on the same sample without the listing price control and the eighth plot being the estimated racial gap on the same sample without either controls.[4]

To summarize, we find that controlling for seller characteristics, the race of the buyer, and appraisal values, do not explain any large portion of the racial gap in unlevered housing returns. Additionally, we see that accounting for time-on-market reduces the racial gap by approximately 50% in this sub-sample.

[4]There was very little power in the regressions with the agent fixed effects so the results are not included.

2.5 HETEROGENEITY

2.5.1 Market heterogeneity

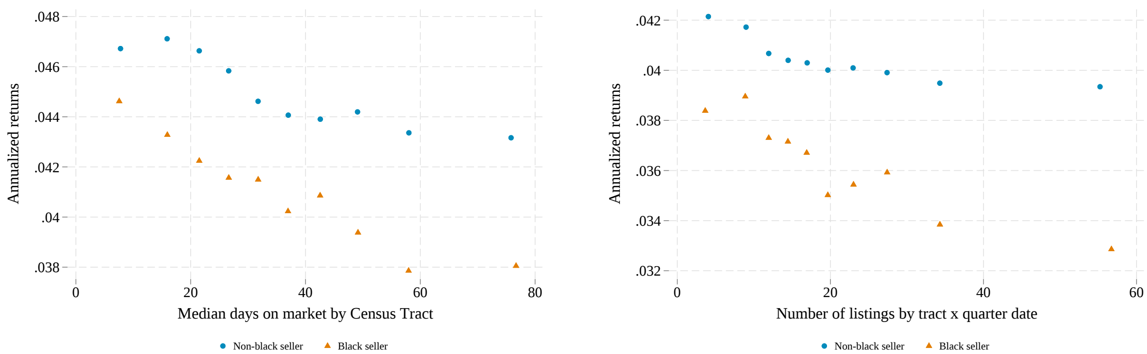
In Figure 2.2, we present binscatter plots showing annualized housing returns for Black and non-Black sellers in Census tracts with different levels of market characteristics that reflect search. We use two measures: Panel (a) uses median days on market of housing transactions at the Census tract level, while panel (b) uses the number of active regular, non-distressed listings by Census tract and sale year-quarter. Both plots consistently show that Black sellers experience lower housing returns than non-Black sellers. Conditional on housing demand, both Black and non-Black sellers tend to experience lower returns in neighborhoods with more slack markets, characterized either by lower probabilities of sellers matching with a buyer or increased competition in the local housing market. Crucially, the plots show that as local markets become more slack, the racial gap not only persists but also widens. This trend suggests that Black sellers encounter greater search frictions in more competitive housing markets for sellers. These findings highlight that the racial disparities in housing returns are exacerbated under unfavorable market conditions for sellers.

2.5.2 Price-growth heterogeneity

In Figure 2.3, we present a binned regression plot showing the annualized returns gap for Black and Black-only sellers over twelve quantiles of Census tract-level annualized house price growth between purchase and sale year. The x-axis value refers to the median value within its respective quantile. The plot shows a negative returns gap at each quantile, increasing in magnitude as annualized Census tract-

Figure 2.2: Annualized housing return gaps by market characteristics

(a) Median days on market (b) Number of active listings



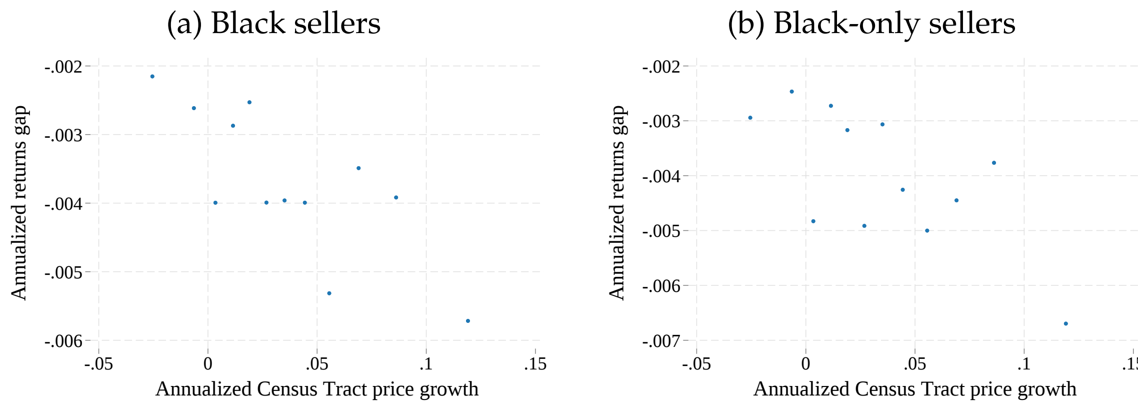
Notes: This Figure presents a binscatter of the annualized housing returns by Census tract market characteristics for Black and non-Black sellers. Panel (a) Plots the annualized returns by the median number of days on market within a home's Census tract and Panel (b) plots the annualized returns by the number of active listings by Census tract and year-quarter date of sale. All regressions control for Census tract-level house price index growth between purchase year and sale year and include year-quarter date of sale fixed effects.

level price growth increases. Black and Black-only sellers in the highest do-deciles of price growth experience an average negative returns gap of -0.57% and -0.67%, respectively, significantly larger than the baseline estimates in Table 2.1. Overall, these results demonstrate that racial disparities in housing returns worsen when Black sellers experience higher neighborhood house price growth.

2.5.3 Neighborhood heterogeneity

This section provides heterogeneity analysis at the Census tract level. We define neighborhoods using Census tracts, a common definition used in the literature. Our neighborhood-level analysis provides estimates of the racial gap in annualized unlevered returns to housing across neighborhoods with different socioeconomic characteristics. These estimates give us an understanding on which types

Figure 2.3: Annualized returns gap by annualized price growth



Notes: This figure plots the annualized housing returns gap between Black and non-Black sellers for each do-decile of Census Tract-level annualized house price growth between purchase and sale where the x-values are the median annualized house price growth within each quantile. Regressions include Census tract-level annualized mean house price index growth between purchase year and sale year and Census tract $\times$ year-quarter of sale fixed effects.

of neighborhoods have higher or lower estimated racial gaps.

We use neighborhood characteristics from the 2010 American Community Survey 5-year estimates. We group Census tracts into quartiles based on the share of Black residents, the share of White residents, the homeownership rate, the share of residents with a college degree, and the median household income. Figure B.2 displays the racial gap estimates in annualized unlevered housing returns across neighborhoods for each of our socioeconomic characteristics. Overall, we find that the annualized gap persists in each quartile group of Census tracts for every socioeconomic variable we examine.

Figure B.2 Panel (a) displays the racial gap estimates across neighborhoods with different Black population shares. Neighborhoods with a higher share of Black population experience larger racial gaps in housing returns. In neighborhoods at the bottom quartile of Black population, the racial gaps are approximately

-0.29% for Black sellers and -0.32% for Black-only sellers, and the estimates are statistically significant. However, although we estimate negative gaps in the second quartile, the estimates are statistically insignificant. The statistical insignificance is probably due to the mechanical fact that these neighborhoods have very few Black sellers. In neighborhoods at the top quartile of Black population, the racial gap increases to -0.39% for Black sellers and -0.42% for Black-only sellers. We believe this is due to search frictions negatively impacting Black sellers more in neighborhoods with higher Black populations. Additionally, we find that the racial gaps are generally higher for households identified as Black-only.

Figure B.2 Panel (b) displays the racial gap estimates across neighborhoods with different shares of White residents. We find lower gaps in annualized housing returns in neighborhoods with the lowest shares of minority residents, though the relationship does not appear to be linear. The results show that the estimated gaps are consistently negative.

Figure B.2 Panels (c) and (d) display the racial gap estimates by the quartile groups of neighborhood homeownership rates and of median household income, respectively. We find that the largest gaps are in neighborhoods in the bottom quartiles of median household income with estimates as large as -0.39% and -0.42% for Black and Black-only sellers, respectively. This is likely because neighborhoods with low levels of income typically have higher minority populations. Given this relationship, it is likely that search frictions negatively impact Black sellers more in neighborhoods with lowest income levels than in neighborhoods with higher income levels. We find a flatter relationship between homeownership rates and the racial gap in housing returns. Overall, the results show that the estimated gaps are consistently negative across these types of neighborhoods.

Finally, Figure B.2 Panel (e) displays the racial gap estimates by the quartile groups of neighborhood shares with college education. In short, we find a flatter relationship between college education levels of a neighborhood and the racial gap in housing returns. We find that Black sellers realize lower housing returns if their properties are in neighborhoods with lower education levels ranging from -0.43% to -0.26%. However, the gaps for Black-only sellers do not follow a similar pattern. One consistency is that the estimated gaps are slightly larger for Black-only sellers compared to Black sellers realizing an average racial gap of up to -0.46%.

To summarize, we find that both Black and Black-only sellers have realized lower housing returns in all quartiles of neighborhoods for each of the neighborhood characteristics we examine. This indicates that the racial disparity in housing returns is not confined to certain neighborhoods, but is broadly evident across various geographical areas. We also find that the gap tends to be larger in neighborhoods with higher shares of Black residents and lower incomes.

2.6 CONCLUSION

This paper investigates whether there exists a racial gap in unlevered housing returns for regular non-distressed sales. Although many papers have studied racial disparities in the housing market, no one has focused on the within-neighborhood disparities during the home-selling process for non-distressed home sales nor provided a detailed empirical explanation of the main causes for the disparities. We use over 1.1 million repeat-sale transactions between 2003 and 2020 in our analysis. When estimating the differences in housing returns between Black and non-Black sellers, we control for Census tract by sale year-quarter fixed effects and tract-level house price growth between purchase and sale dates. Our main findings are that

Black and Black-only sellers on average realize 1.7% and 1.9% lower total capital gains, and 0.36% and 0.38% lower annualized returns, respectively, when selling their homes, and our explanations of what does and does not explain these gaps. We also find that the gap exists in all types of neighborhoods.

To explore the sources of the gap, we bring in additional data from various sources to examine the roles of seller's liquidity, property renovations, race of the buyer, real estate agents, and appraisers on the racial gap. We find that none of these factors can explain much of the gap. However, our appraisal analysis strongly suggests that, on average, Black homes are not valued less than non-Black homes.

We find that Black sellers tend to set lower listing prices and experience longer time on market. These results suggest that Black sellers face different levels market tightness than their non-Black neighbors. Motivated by this new set of facts, we investigate the role of search frictions. Within the context of a simple search model, different levels of market tightness implies different levels of prices. Accounting for listed prices is key because it is a measure of a home's market value that can be influenced by the biases of several players such as home appraisers, real estate agents, and sellers who likely internalize the racial differences in time on market. As a result, we control for listing prices and days on market, and find that the racial gap drops to near zero, suggesting that both the biases behind listing prices and the search frictions faced by Black sellers are the dominating factors contributing to the gap. In a complementary set of analyses, we find that Black sellers living zip codes less exposed to buyer agent discriminatory practices realize a lower returns gap relative to non-Black sellers. Finally, we find that when intermediaries such as buyer agents, whose involvement we argue causes racial differences in search frictions and consequently in housing returns, are absent, the racial gap in housing returns is eliminated. Overall, the results in this paper provide strong evidence for the significant role that human intermediaries, such as buyers' real estate agents, in creating this racial disparity in housing returns.

CHAPTER 3

When Institutional Investors Come To Town: The Local Effects of Buy-to-Rent Properties

3.1 INTRODUCTION

Large institutional buy-to-rent (B2R) investors entered the single-family rental sector since 2012, disrupting the market that has long been dominated by small mom-and-pop investors. This shift in housing market structure has led to large public disapproval. A statement from the White House in 2021 stated that "Large investor purchases of single-family homes and conversion into rental properties speeds the transition of neighborhoods from homeownership to rental and drives up home prices for lower cost homes, making it harder for aspiring first-time and first-generation home buyers, among others, to buy a home. At the same, these purchases are unlikely to meaningfully boost supply in the lower-cost portions of the rental market, as investors charge more for rent to recoup higher purchase costs."

However, the institutional investors have voiced strong disagreement with these assertions, arguing that they are unfairly blamed for the broader housing crisis primarily driven by housing supply shortages. They point out that they only own a small share of single-family homes at the aggregate level and thus have a minimal impact on the housing market. Moreover, their rental business simply responds to the growing market demand for more rental options, rather than creating barriers to the path of homeownership.

The center of the debate hinges on a key empirical question: Does the entry of large buy-to-rent investors drive up local housing prices, thereby raising the barriers of homeownership? To answer these question, this paper provides reduced-form evidence on the spillover effects of buy-to-rent investment properties on housing prices of nearby properties. Furthermore, it distinguishes the underlying mechanisms that drive the changes in housing prices, offering insights into how housing market dynamics are affected by large-scale investment activities.

This paper examines the spillover effects using a comprehensive dataset of property transactions from CoreLogic in the Atlanta metropolitan area from 2012 to 2019. I focus on the Atlanta region because it represents other similar cities in the Sunbelt region that experience a significant influx of institutional buy-to-rent investors. High-growth cities in the Sunbelt region are characterized by their relatively newer housing stock, an appealing feature for institutional investors because these newer properties likely face lower maintenance costs. Moreover, the Sunbelt cities were hit particularly hard by the foreclosure crisis, resulting in undervalued properties that reduce acquisition costs for these investors.

To estimate the spillover effects of B2R properties, I exploit the quasi-random variation in the timing and location of buy-to-rent properties. I use a hedonic regression to compare housing prices of two properties in the same neighborhood sold at the same time but differ in the number of nearby buy-to-rent properties. The first identification assumption is that institutional investors cannot perfectly control which specific house to purchase in a given neighborhood because they cannot force the incumbent owner to sell the property to them. The second identification assumption is that houses within a small area are similar enough to experience similar housing price growth. I find that the presence of one to three buy-to-rent properties within a 150-meter radius increases the sale price of the focal property by two to three percent. Buy-to-rent properties located more than

1200 meters away have negligible impact on the focal property, highlighting the localized nature of the spillover effects.

The main threat to the baseline hedonic approach is that houses within a small neighborhood can experience differential price growth, driven by a correlation between housing demand shocks and unobserved property and nearby neighborhood characteristics. For example, one house near a beautiful lake might experience higher price appreciation compared to another house slightly further away if there is a positive demand for houses with closer access to natural amenities. To mitigate the influence of such confounding factors, I use a repeat-sale specification to difference out any time-varying factors that are common to properties in the same neighborhood. Econometrically, this is equivalent to include a triple neighborhood-purchase year-sale year fixed effect. The repeat-sale results suggest that the presence of buy-to-rent properties within a 150-meter radius increases the housing price growth rate by one to two percentage point.

I further explore the variation in the spillover effects across different neighborhoods. I find that neighborhoods with low socioeconomic status experience stronger spillovers from the entry of buy-to-rent investors. For instance, in neighborhoods with a higher share of black residents, the presence of buy-to-rent properties within 1200 meters increases the sale price of the focal property by as high as ten percent. In neighborhoods in the bottom quartile of housing prices in 2010, the entry of buy-to-investors increases the housing prices by ten to fifteen percent. The results are consistent with the business strategy of B2R investors targeting entry-level homes in undervalued communities, particularly those more affected by the foreclosure crisis.

To understand the underlying causes of the positive spillovers, I distinguish the roles of different potential mechanisms through which buy-to-rent properties influence local housing prices. I identify three main channels. First, buy-to-rent investors increase the level of competition in the local housing market. The rationale is intuitive: assume the pool of houses for sale is fixed, an influx of additional prospective home buyers pushes up housing demand and push prices upward as a result. Second, buy-to-rent investors reallocate housing units from the for-sale market to the for-rent market, lowering the inventory of homes available for sale in subsequent years. This reduction in housing stock shifts the supply curve in the for-sale market to the left and leads to higher prices. Third, buy-to-rent properties can change local amenities. On one hand, investors often undertake renovations and updates to standardize and modernize acquired properties. These improvements can enhance the overall appeal of the local area, potentially attracting more homebuyers to the neighborhood. On the other hand, the increase in rental properties might be perceived negatively by some potential homeowners, which could the lower the desirability of adjacent properties.

To isolate these three potential channels, I leverage the timing of buy-to-rent property acquisitions. The competition channel is expected to operate contemporaneously because it is directly related to the immediate increase in housing demand, whereas the supply and amenity channels would typically take place after some delay. I find that the competition channel matters less and is hyper-local. An additional purchase by buy-to-rent investors within the same year increases the housing price of properties within 150 meters by one to three percent, and it has no significant effect on more distant properties. On the contrary, buy-to-rent property acquired one to two years ago increases the housing prices of properties within 1200 meters by one to four percent. The results suggest that the supply and amenity channels are more influential in driving housing price changes.

Related literature This paper is related to the literature that investigates the emergence and influence of buy-to-rent investors on the housing market. Studies such as Mills et al. (2019), Oosthuizen (2023), Chilton et al. (2018), Christophers (2023), Colburn et al. (2021), and Fields & Vergerio (2022) have documented the birth and growth of buy-to-rent investors following the foreclosure crisis. These institutional investors, with substantial financial resources, capitalized on the low property prices during the foreclosure crisis to create large portfolios of single-family homes, particularly in the Sunbelt regions. Allen et al. (2018) note that these institutional investors often purchase properties at a discount rate. Lambie-Hanson et al. (2019) and Mills et al. (2019) find that the entry of investors improves local housing market conditions in the immediate aftermath of the foreclosure crisis, but the long-term implications of their presence are debatable. As buy-to-rent investors continue to expand their portfolios, they become major figures in some local housing markets, with some even owning entire blocks of single-family homes, highlighting their growing footprint and potential influence on the local market.

This paper aims to evaluate the longer term impact of buy-to-rent investors on the local housing markets, an area that haw drawn increasing attention in recent years. Gurun et al. (2023) and Austin (2022) use the merger and acquisition decisions of institutional investors as the exogenous variation to estimate the impact of ownership concentration on local housing prices and rents. DLima & Schultz (2022) and Ihlanfeldt & Yang (2021) analyze the price growth in neighborhoods with a higher share of buy-to-rent properties with other neighborhoods, yet they do not address the potential selection issue that investors endogenously select into neighborhoods with higher potential price growth. Although these studies attempt to answer similar questions, I use a different empirical strategy to assess the price spillover effects of buy-to-rent properties. I consider my study as complementary to the existing literature. I use property-level housing price data and exploit the quasi-random location of buy-to-rent properties to evaluate the spillover effects at a much granular level.

The paper also contributes to the literature on the externalities of local housing markets. The value of a property not only depends on its own features, but also on the characteristics of nearby properties and environment. A substantial body of literature has documented the negative externality from locating near foreclosed properties, poorly maintained homes, and toxic factories (Anenberg & Kung (2014), Fisher et al. (2015), Gerardi et al. (2015), Campbell et al. (2011), Autor et al. (2014), and Currie et al. (2015)). This paper applies a methodology similar to this strand literature and examines the externalities from buy-to-rent properties. It finds that the presence of buy-to-rent investors imposes positive externalities on nearby properties by increasing their property values. While this is good for incumbent homeowners who sell their properties, it also escalates the housing costs of potential individual home buyers.

3.2 CONCEPTUAL FRAMEWORK

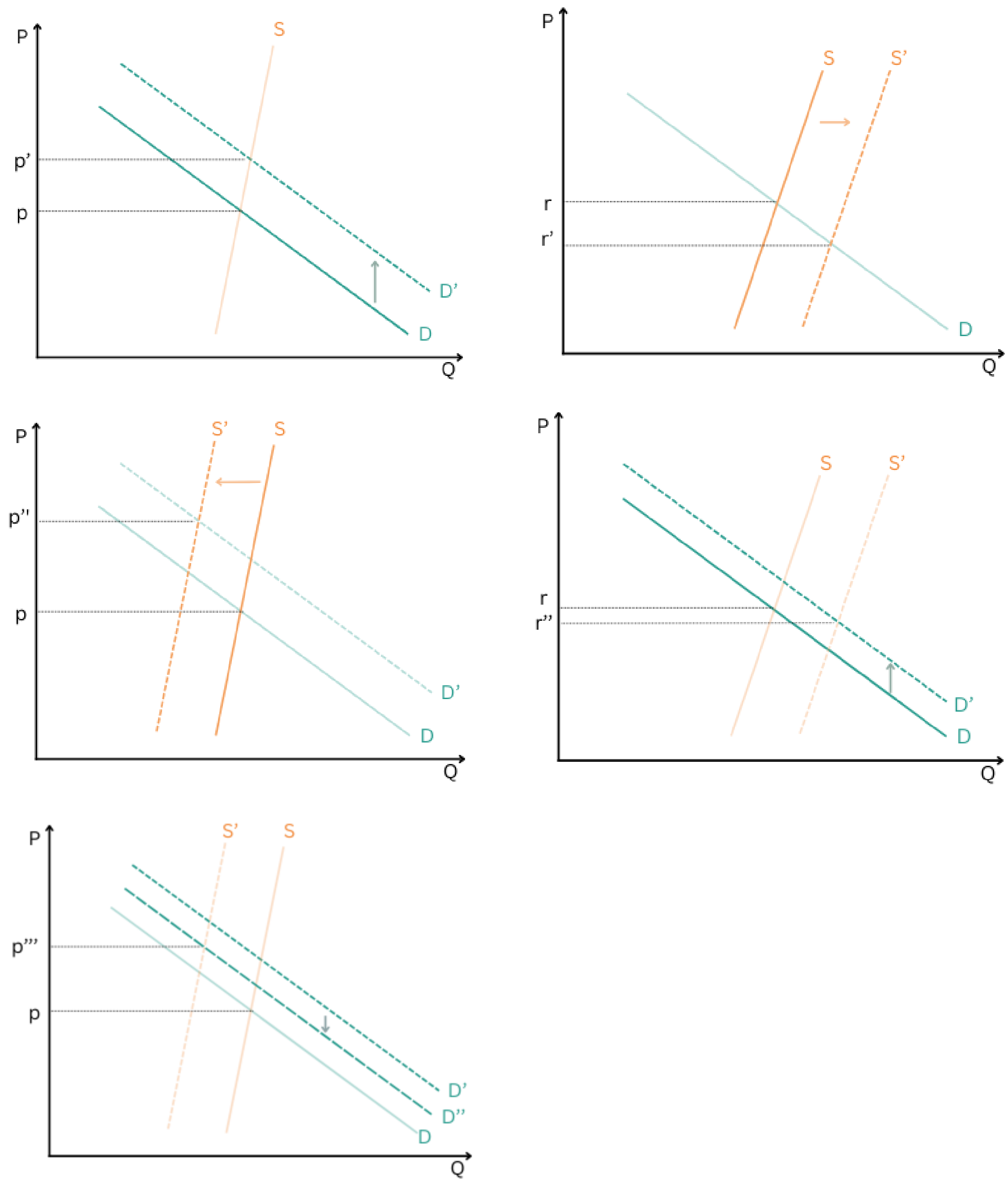
In this section, I present a simple model to illustrate the possible effects and mechanisms instigated by buy-to-rent investors. Consider a scenario where two distinct markets exist for single-family homes: one catering to owners and another to renters. Due to technological developments and financial advantages, buy-to-rent investors reallocate housing units from the for-sale market to the for-rent market.

I first consider the impact of buy-to-rent investors on the for-sale market. As shown in the left panel of Figure 3.1, the initial equilibrium price in the for-sale market before the entry of buy-to-rent investors is p at time $t = 0$. In period $t = 1$, B2R investors enter the market and shift the demand curve to the right (shown in the top left panel of Figure 3.1), causing the price to increase to p'. This captures the contemporaneous competition channel of investors. In period $t = 2$, there are fewer homes available on the for-sale market because buy-to-rent investors reallocate these housing units to the rental market and thus lower the housing stock in the for-sale market. The supply curve shifts to the left (shown in the middle left panel of Figure 3.1), and causes the prices to rise even further to p''. In addition to the competition and reallocation channel, the presence of buy-to-rent investors can change the values of local amenities and shift the demand curve to either direction. On one hand, most homeowners do not like living near rental properties because of the belief that renters are less invested in the local community, so the influx of renters may make the area to be less attractive for potential individual homebuyers. On the other hand, buy-to-investors often undertake substantial renovations on the properties, enhancing the aesthetic appeal of the local neighborhood and increasing the average home value. Depending on the relative magnitude of these two amenity effects, the demand curve can further shift to either direction and change prices. If the net amenity effect is negative, the demand curve will shift downwards (as shown in the bottom left panel of Figure 3.1) and thus lower prices below p''. In the long term, there can also be supply responses from real estate developers and homebuilders to increase housing supply in the for-sale market. The supply response would depend on local supply elasticity and is unlikely to materialize in the short term.

Then I consider the impact in the for-rent market. In period $t = 0$, the rental market is in equilibrium at price r. The equilibrium rent is higher than the initial equilibrium price in the for-sale market, $r > p$. Buy-to-rent investors are financially motivated to capitalize on the arbitrage opportunity and reallocate housing units to the rental market. The reallocation channel is reflected by the right-shifted supply curve in the top right panel in Figure 3.1. Given the supply increase in the rental market, rent would be lower under a perfect competitive environment. These institutional investors create a new market for professionally-managed single-family rentals, expanding the housing options of households. The upgrades they carry out also affect the attractiveness of the rental option to residents. As a result, the demand curve can shift rightward. The magnitude depends on the elasticity of substitution between owning and renting, and also the substitution between renting from small mom-and-pop landlords and professionally managed corporations. Moreover, the entry of these large corporations can transform the rental market towards a more monopolistic competitive environment given that they often own a large number of homes in one local area. It is possible that these large corporations can leverage their market power to increase rent. Therefore, the net impact on rent depends on the relative magnitude of these competing forces.

Figure 3.1: A stylized model of buy-to-rent investors

(a) For-sale market (b) For-rent market



3.3 DATA AND DESCRIPTIVE FACTS

In this section, I describe the data sources used for the empirical analysis and the procedure to identify large institutional buy-to-rent investors. I also show the descriptive facts of the growth and spatial distribution of buy-to-rent properties.

3.3.1 Data

The main dataset used in this study are property transaction records between 2012 and 2019 from CoreLogic. The CoreLogic property transaction data provides detailed information for each property sale, including the transaction date, property address, property type, buyer name, seller name, and sale price. I restrict the sample to all single-family homes that are arms-length transactions. I also merge each transacted property with CoreLogic Assessor files to obtain detailed property characteristics, including the number of bedrooms, bathrooms, square footage, lot size, and year built.

I use the buyer name on each transaction to identify properties bought by large buy-to-rent investors. I first identify a set of large corporate buy-to-rent investors from industry reports, newspapers, the internet, and academic papers (such as Mills et al. (2019) and Gurun et al. (2023)). I find 14 largest institutional investors in the Atlanta metro area, with four of them as Real Estate Investment Trust (REITs) and ten as private equity and investment firms. The list of corporations includes Invitation Homes, Main Street Renewal, Progress, Tricon, American Homes 4 Rent, Front Yard, FirstKey, Home Partners of America, Sylvan, Roofstock, VineBrook, Promise Homes, Lafayette, and RESICAP.

The challenge in identifying property acquisitions by each corporation is that each corporation operates through a myriad of subsidiary entities, which are the entities actually responsible for acquiring properties. Moreover, these subsidiaries have many name variations, adding to the difficulty in identifying property acquisitions. For example, the company Invitation Homes rarely purchases properties directly under its corporate name. Instead, it uses subsidiaries like Tabert LLC, SWH 2017-1 Borrower, CAH 2014 1 Borrower, IH3 Property Borrower, and SFR ATL Owner 5 LP for property acquisitions. To find the subsidiaries linked to each corporation, I search each corporation in Mergent Online, a database that provides information on business descriptions and subsidiaries, to find a list of subsidiaries. I then employ an iterative searching algorithm to find all subsidiaries to address potential missing data in Mergent. I search each subsidiaries found on Mergent in the CoreLogic transaction data using buyer's name, find the common buyer's mailing addresses associated with each subsidiary, and then use the mailing address to identify other subsidiaries associated with the corporation. To confirm that these subsidiaries indeed belong to the corporation, I verify the names and addresses on Georgia's business registration records from the Office of the Secretary of the State. I also randomly choose properties and check them on Zillow and Realtor.com to verify these properties are rental properties owned by these corporations.

After identifying properties acquired by these large buy-to-rent investors, I calculate the number of homes owned by these investors (shown in Table 3.1). If the company is public, its annual reports provide the count of properties they own at the city level. To validate my identification approach, I compare the number of properties I identify from CoreLogic with the numbers stated in the annual reports for public companies. Only three out of the total fourteen corporations are public – Invitation Homes, Tricon, and American Homes 4 Rent. Among these three public companies, I am able to find over 96 percent of all properties owned by them. The coverage rate for Invitation Homes is over 100 percent. This may be attributed to the timing difference in reporting records from the two data sources, or the fact that some of newly acquired properties are not yet in operation.

Table 3.1: Number of properties owned by large institutional investors as of 2019, Atlanta CBSA

Company	CoreLogic	Annual Report	Coverage
Invitation Homes	13759	12555	109.59%
Main Street Renewal	5507	-	-
Progress	5484	-	-
Tricon	5050	5253	96.14%
American Homes 4 Rent	4960	4977	99.66%
Front Yard	4173	-	-
FirstKey	3406	-	-
Home Partners of America	2413	-	-
Sylvan	1490	-	-
Roofstock	775	-	-
Others	2086	-	-

Notes: This table presents the total number of properties owned by large buy-to-rent investors as of 2019. The second column gives the number of properties I identify from CoreLogic property transaction data. The third column gives the official numbers from the annual reports of public companies. Only three companies, Invitation Homes, Tricon, and American Homes 4 rent are public, so the rows are empty for non-public companies in the third column. The last column gives the ratio of the second column to the third column.

I use the 2010 American Community Survey (ACS) from the IPUMS NHGIS to measure sociodemographic characteristics at the Census block group and tract levels. Specifically, I retrieve data on the age composition, racial composition, share of college educated individuals, median household income, homeownership rate, median housing value, and housing stock composition.

3.3.2 Descriptive facts

In this section, I present four descriptive facts about the large institutional buy-to-rent investors. First, since the creation of the single-family rental business model in 2012, institutional buy-to-rent investors have significantly increased the size of their portfolio. Figure 3.2 shows that the number of buy-to-rent properties has increased from zero in 2011 to over 45,000 in 2020. The investors keep expanding their portfolio and show no sign of a slowdown in their acquisition. At the aggregate level, the institutional buy-to-rent investors only own 2.7 percent of all single-family homes in the region, but this masks significant geographical variation in the location of buy-to-rent properties.

Figure 3.2: Growth of large institutional buy-to-rent investors in Atlanta CBSA

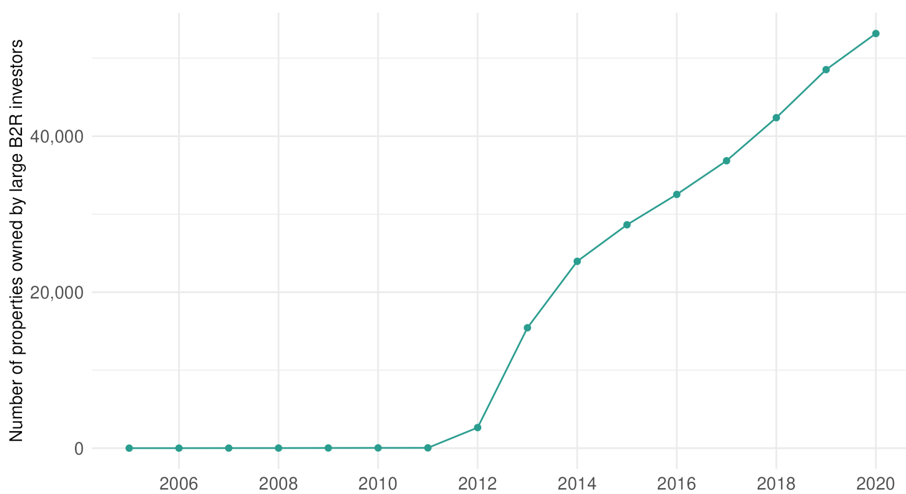
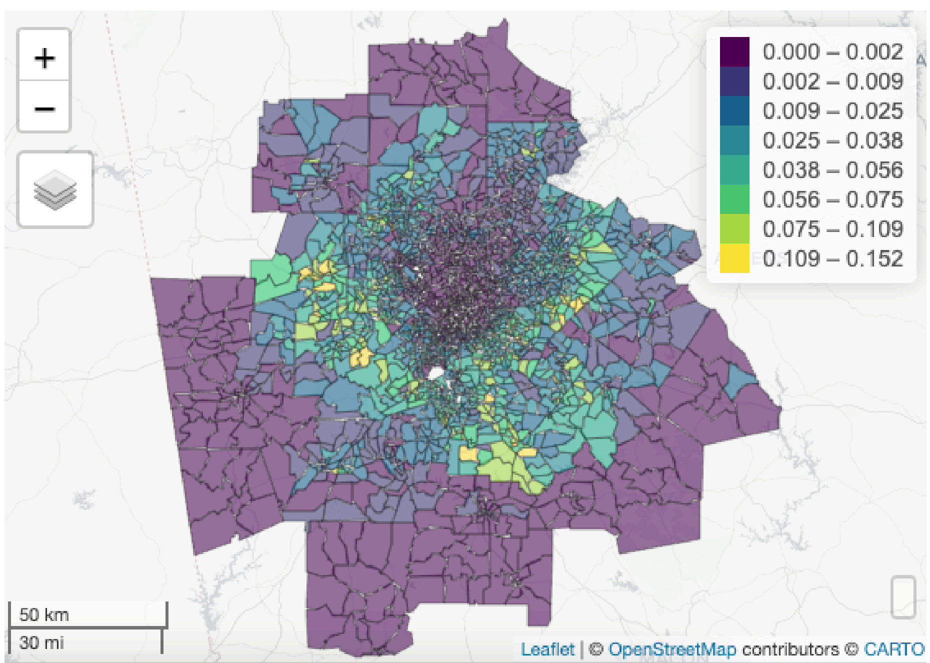


Figure 3.3 shows the share of buy-to-rent properties owned by institutional investors at the Census block group level in 2019 in the Atlanta CBSA. There is substantial geographical variation in ownership concentration. While the majority of downtown areas and outskirt suburbs does not have any footprint of institutional investors, investors own a large share of single-family homes in neighborhoods

right outside the central business districts. Their ownership share can be as high as ten percent in some neighborhoods. It is important to note that neighborhoods with a high share of buy-to-rent properties are in the southern, western, and eastern regions. These areas are characterized by a high minority population and were hit hard by the foreclosure crisis. In contrast, the northern neighborhoods that are occupied by more white affluent residents are barely touched by institutional investors. The geographical distributional pattern confirms the endogenous acquisition process of investors. Instead of adopting a random acquisition strategy, the investors strategically target properties in undervalued communities with high growth potentials.

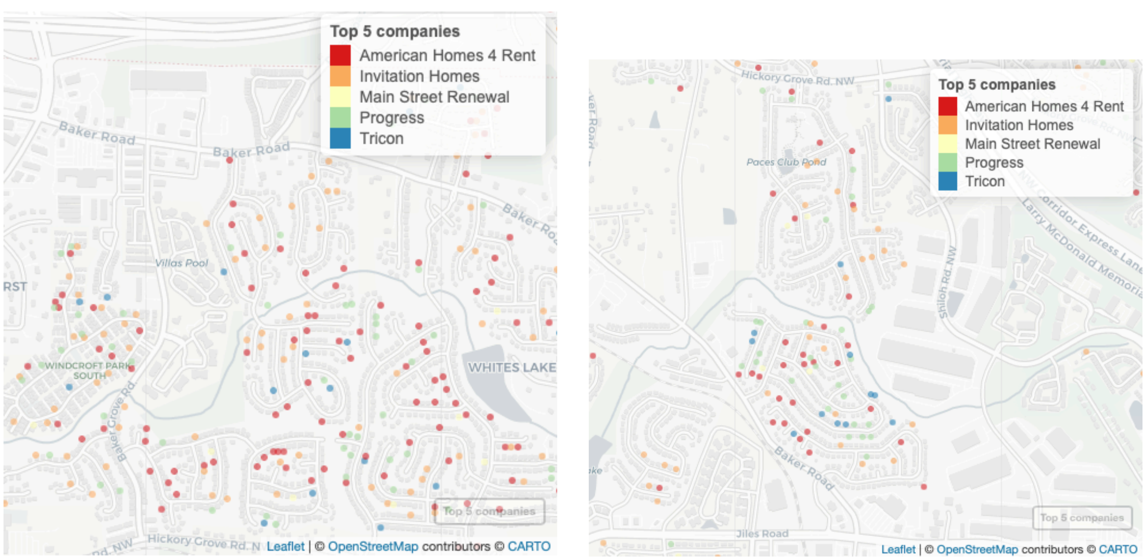
Figure 3.3: Geographical variation of buy-to-rent properties as of 2019, Atlanta CBSA



Notes: This figure shows the geographic distribution of buy-to-rent properties in the Atlanta CBSA as of 2019. Each polygon represents a Census block group. The color represents the share of buy-to-rent properties calculated as ratio of the number of buy-to-rent properties to all single family homes. Darker purple color represents a lower share, while lighter green and yellow colors represent higher shares.

While institutional investors strategically target certain neighborhoods, they cannot perfectly control which properties they can acquire. The reason is simple: they cannot force incumbent owners to sell their properties to them. Ideally, the investors would prefer to acquire homes that are geographically close to each other because such clustering would lower management and maintenance costs. Figure 3.4 shows two examples of the exact locations of buy-to-rent properties. Each dot represents a property owned by one of the five largest institutional investors. There is indeed a lot of clustering happening at the neighborhood level, but there also exist substantial hyper-local variations at the street level in the number of buy-to-rent properties and the companies that own them. This lends support to the empirical strategy that the exact location of buy-to-rent properties is as good as randomly assigned.

Figure 3.4: Examples of the exact locations of B2R-owned properties



Notes: This figure gives two examples of the exact locations of buy-to-rent properties in the Atlanta CBSA. Each dot represents a property owned by one of the largest five buy-to-rent corporations.

Lastly, I look at the property characteristics of single-family homes acquired by institutional buy-to-rent investors, and compare them with the average characteristics of all single-family homes in Atlanta. Table 3.2 shows that the average interior size of B2R-owned homes is 1,945 square feet, which is approximately 400 square feet smaller than the average single-family home. The lot size for B2R-owned homes also tends to be smaller, consistent with the fact that buy-to-rent properties are closer to the city center and properties in these regions have smaller lots compared to homes in outskirt suburbs. Buy-to-rent properties also tend to be newer properties that are built after 1995. Perhaps the biggest difference between buy-to-rent properties and average homes is in their market values. The average value of buy-to-rent properties is around \$160,000, about \$60,000 less than the average value of all properties. All the evidence combined suggests that buy-to-rent investors are more likely to acquire starter homes. These homes are exactly the type of homes that are ideal for first-time homeowners. As a result, this acquisition pattern can have a significant impact on the path to homeownership.

Table 3.2 also shows that the standard deviations of all property characteristics are much smaller for buy-to-rent properties. This pattern suggests that buy-to-rent investors intentionally acquire properties with similar features. The uniformity of properties aligns with the conjecture that it is more efficient to manage and advertise similar properties.

Table 3.2: Property characteristics of single-family homes in Atlanta CBSA

	B2R-owned SFH			All SFH		
	Median	Mean	Sd	Median	Mean	Sd
Living square feet	1,836	1,945	616	2,018	2,354	1,384
Lot size in sqft	11,761	14,302	10,957	15,814	48,978	119,976
Num of bedrooms	3	3.4	0.68	3	4	4.137
Num of bathrooms	3	2.6	0.67	3	2.7	1.36
Num of rooms	6	6.65	1.42	6	7	2.41
Year built	1999	1995	14.67	1994	1988	22.77
Market value ($)	154,370	160,815	65,866	180,496	225,305	188,151

Notes: This table shows the summary statistics of single-family homes owned by buy-to-rent investors (the left panel) and all single-family homes (the right panel) in the Atlanta CBSA.

3.4 LOCAL SPILLOVER EFFECTS OF B2R PROPERTIES

In this section, I describe the empirical strategy to recover the causal impact of buy-to-rent properties on nearby housing prices, and present the baseline results.

3.4.1 Baseline hedonic approach

The goal of the empirical analysis is to measure the local price spillovers stemming from the entry of buy-to-rent investors on neighboring properties. The identification challenge is that buy-to-rent investors endogenously select into certain neighborhoods and purchase properties that align with their business model. Ideally, we want to compare two identical properties within the same neighborhood, differing only in their exposure to buy-to-rent properties nearby. The empirical strategy aims to replicate the ideal experiment as much as possible.

To address the endogeneity issue, I utilize quasi-random variations in the exact locations of buy-to-rent properties. As discussed in the last Section, buy-to-rent investors do not have perfect control over which properties they can acquire at the hyper-local street level. Similarly, incumbent homeowners cannot dictate who purchases their neighbors' houses. This creates quasi-random variations in the number of buy-to-rent properties within these hyper-local areas. The key identification assumption is that the exact location and timing of buy-to-rent entry within a homogenous neighborhood is as good as random.

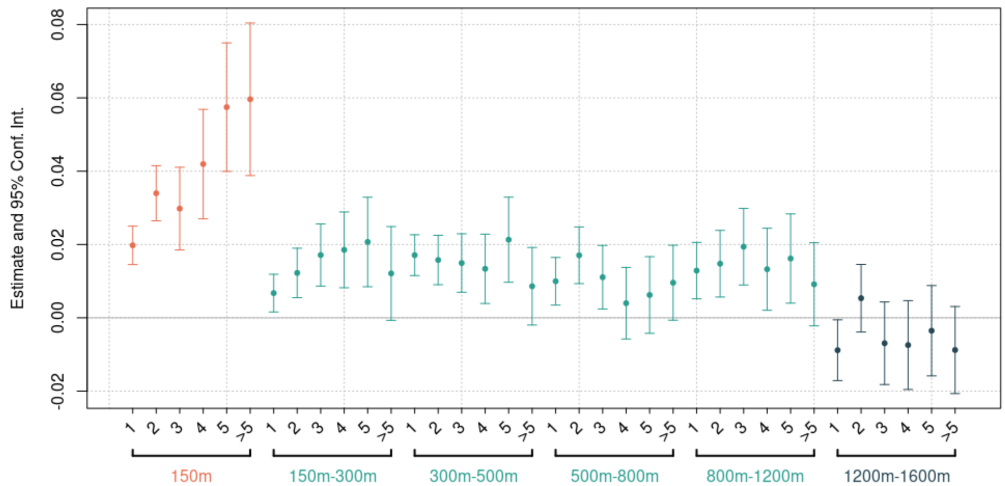
For property i in neighborhood c that was sold in time t, I run the following regression:

$$\log p_{ict} = \sum_d \sum_j \beta_{dj} \mathbf{1}(N_{it}^d = j) + \delta X_{it} + \alpha_{ct} + \epsilon_{ict}. \tag{3.1}$$

p_{ict} indexes the sale price of property i located in neighborhood c sold at time t. N_{it}^d is the total number of properties owned by buy-to-rent investors within distance d when property i was sold in t. I discrete the distance to six non-overlapping bins, $d = \{$150m, 300m, 500m, 800m, 1200m, 1600m$\}$. j indexes the number of buy-to-rent properties, and it takes discrete values $j \in \{0, 1, 2, 3, 4, 5, > 5\}$. The discretization allows me to flexibly capture the nonlinear effect associated with different numbers of buy-to-rent properties. X_{it} represents a set of control variables. It includes the property characteristics of property i, such as the number of bedrooms, bathrooms, square footage, lot size, and age of the building. X_{it} also includes the total number of properties within each distance bin to capture density variations in the adjacent neighborhood. To control for potential within-neighborhood variations in housing prices, I include the average housing prices within 500 meters of property i. α_{ct} are neighborhood-time fixed effects to control for unobserved demand shocks that vary by neighborhood and time of sale. Additionally, I restrict the analysis sample to all non-distressed sales (i.e., exclude foreclosures and short sales) and sales to buy-to-rent investors. The parameters of interests are β_{dj}, which capture the spillover effects from different numbers of buy-to-rent properties j in each distance bin d.

Figure 3.5 plots the coefficients by distance bin and number of buy-to-rent properties. I use census tract to define neighborhoods and calendar year to capture time fixed effects. It shows that buy-to-rent properties within 150 meters have the most significant impact on property values. Having one buy-to-rent properties within 150 meters increases the sale prices by 2 percent, and having five or more buy-to-properties within 150 meters increases the prices by 6 percent. Buy-to-rent properties located 150 meters away but within 1200 meters still have positive spillover effects on property values, but the effects are more muted. Having five buy-to-rent properties within a 150 to 300-meter radius increases property values by 2 percent, the same magnitude as having one buy-to-rent properties within 150 meters. Buy-to-rent properties that are located 300 to 1200 meters have similar effects on property values, with the average impact ranging from 1 to 2 percent. Buy-to-rent properties located 1200 meters away do not have a significant nor meaningful impact on property values.

Figure 3.5: Hedonic price spillover effects of B2R properties by distance and number of B2R properties



Notes: This figure plots the coefficients and 95% confidence intervals of the hedonic regression Equation (3.1). It shows the results by distance bin and number of adjacent buy-to-rent properties. The top numbers in the x-axis represents the number of buy-to-rent properties, while the bottom numbers in the x-axis represents the six non-overlapping distance bins.

Table 3.3 shows the baseline results with different specifications. Give that there is a large number of coefficients to be estimated, I cluster the number of buy-to-rent properties into three bins [1-3], [4-5], and over 5. To highlight the most important results, I only show the coefficients where $d = 150m$. Column (1) uses census tract by year fixed effects, and control for property characteristics. The results are largely similar to Figure 3.5. The immediate proximity to buy-to-rent properties increases the property value by 2.7 to 6.3 percent. Column (2) adds control for the total number of properties within 500 meters to capture density variations in the local area. The coefficients are similar to Column (1). Column (3) adds property values within 500 meters to control for potential variations in property values, and the coefficients remain at similar magnitudes.

Since housing markets exhibit strong seasonality patterns, Column (4) uses calendar year and quarter as the measure for time fixed effects. Comparing two properties transacted at different times of the year may capture the seasonality effects and confounds the actual price spillovers from buy-to-rent investors. After adjusting for seasonality, the coefficients are still of similar magnitudes to the baseline specification. Column (5) uses zip codes as a proxy for neighborhoods. This specification compares properties within a broader area than census tracts. Still, the coefficients are statistically significant and remain similar to previous specifications.

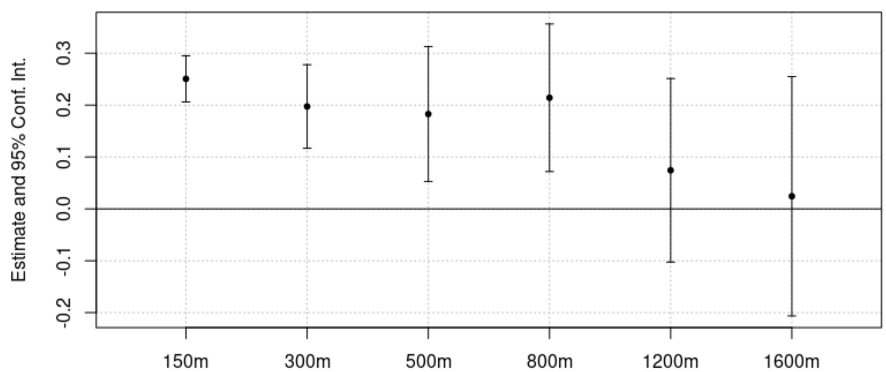
Table 3.3: Hedonic spillover effects of B2R properties

	(1)	(2)	(3)	(4)	(5)
N(B2R in 150m)					
$[1, 3]$	0.0267^{***}	0.0269^{***}	0.0343^{***}	0.0265^{***}	0.0178^{***}
	(0.0041)	(0.0042)	(0.0031)	(0.0026)	(0.0052)
$[4, 5]$	0.0462^{***}	0.0493^{***}	0.0588^{***}	0.0501^{***}	0.0422^{***}
	(0.0095)	(0.0100)	(0.0077)	(0.0067)	(0.0126)
> 5	0.0627^{***}	0.0740^{***}	0.0736^{***}	0.0610^{***}	0.0693^{***}
	(0.0159)	(0.0164)	(0.0126)	(0.0104)	(0.0179)
Controls:					
Number of B2R in 150-1600m	×	×	×	×	×
Property characteristics	×	×	×	×	×
Number of SFHs within 500m		×	×	×	×
Average log price within 500m			×	×	×
Fixed effect	Tract-Year	Tract-Year	Tract-Year	Tract-YearQuarter	ZIP-Year
N	817,511	817,511	803,774	803,774	799,316
Overall R^2	0.53	0.53	0.60	0.63	0.59
Within R^2	0.15	0.15	0.27	0.25	0.35

Notes: This table reports regression results from the hedonic regression Equation (3.1). Each column uses a different specification with different controls and fixed effects. All specifications include the number of buy-to-rent properties in various distance bins, but I only report the results on the buy-to-rent properties within 150 meters to highlight the results. I also include property characteristics in all specifications.

Figure 3.6 presents the coefficients using the share of properties owned by buy-to-rent investors as the independent variable. Using the continuous share of buy-to-rent properties directly captures the density variation in the adjacent neighborhood. Consistent with the baseline finding, buy-to-rent properties located further away have no effects on property values. When the share of buy-to-rent properties within 150 meters increases by 10 percent, the property values increases by 2.5 percent. The effects are slightly smaller for buy-to-rent properties located 150 meters away but still statistically significant and positive.

Figure 3.6: Hedonic price spillover effects of share of B2R-properties



Notes: This figure plots the coefficients and 95% confidence intervals of the hedonic regression Equation (3.1) using the share of buy-to-rent properties in various distance bins as the independent variable.

3.4.2 Repeat-sale approach

The hedonic approach relies on the identification assumption that two properties within the same neighborhoods are similar to each other enough that they should experience the same price growth over time. Although I control for observed property characteristics, it is possible that there exist other unobserved property characteristics that drive differential price growth trends. For example, the house near a local grocery store or Starbucks might experience higher price growth compared to another house that is located slightly further away if potential homebuyers place a greater value on the proximity to business amenities. There also exist many property characteristics that are not observed in the data but can have a big impact on property values. For example, properties with central air conditioning system and newer roofs can have higher values.

Therefore, I employ a repeat-sale specification to control for unobserved property and nearby neighborhood characteristics. The method is similar to the repeat-sale specification in Gerardi et al. (2015). In the repeat-sale analysis, I only include properties that are transacted at least two times during the sample period. For property i in neighborhood c that was transacted in time t and $t+k$, the hedonic regressions in these two periods are

$$\log p_{i,c,t} = \sum_d \beta_{d,t} N^d_{i,t} + \delta X_{it} + \alpha_{c,t} + \epsilon_{i,c,t},$$
$$\log p_{i,c,t+k} = \sum_d \beta_{d,t+k} N^d_{i,t+k} + \delta X_{it} + \alpha_{c,t} + \epsilon_{i,c,t+k}.$$

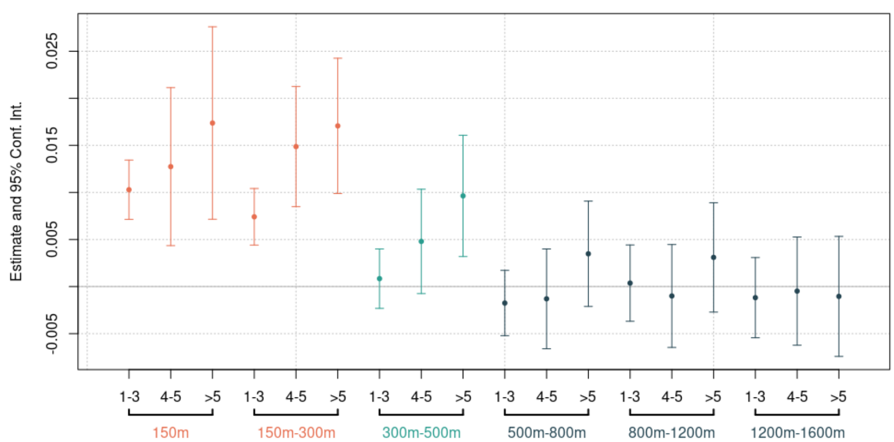
Taking the difference of these two equations gives the repeat-sale specification,

$$\log \frac{p_{i,c,t+k}}{p_{i,c,t}} = \sum_d \beta_d \cdot \Delta N^d_{i,t,t+k} + \alpha_{c,t,t+k} + \epsilon_{i,c,t,t+k} \tag{3.2}$$

The parameter of interests β_d measures the difference in housing price appreciation caused by the differential exposure to buy-to-rent properties. The repeat-sale specification difference out all property characteristics that affect housing prices if they do not vary over time, as well as any time-invariant neighborhood characteristics. The triple fixed effects control for unobserved time-varying shocks that are common to properties transacted in t and $t+k$ in neighborhood c. Intuitively, the repeat-sale approach compares the price growth of two properties within the same neighborhoods that were purchased and sold in the same years, but differing in the growth of nearby buy-to-rent properties between the two transaction dates.

Figure 3.7 shows the coefficients using the repeat-sale specification. Having buy-to-rent properties within 500 meters increases the housing price appreciation rate. When institutional investors purchase one to three more properties within 150 meters during the two transaction dates, the sale price growth increases by 1 percentage point. The spillover effects on price appreciation rate increase when institutional investors purchase more properties nearby. Similar to the hedonic results, the strength of the spillover effects decreases with distance. Buy-to-rent properties located further than 500 meters do not appear to affect price growth rates anymore.

Figure 3.7: Repeat-sale price spillover effects of B2R-properties



Notes: This figure plots the coefficients and 95% confidence intervals of the repeat-sale regression Equation (3.2). It shows the results by distance bin and number of adjacent buy-to-rent properties. The top numbers in the x-axis represents the number of buy-to-rent properties, while the bottom numbers in the x-axis represents the six non-overlapping distance bins.

3.5 MECHANISM AND HETEROGENEITY

In this section, I run additional empirical analysis to distinguish different mechanisms driving the spillover effects, and explore neighborhood heterogeneities.

3.5.1 Mechanism

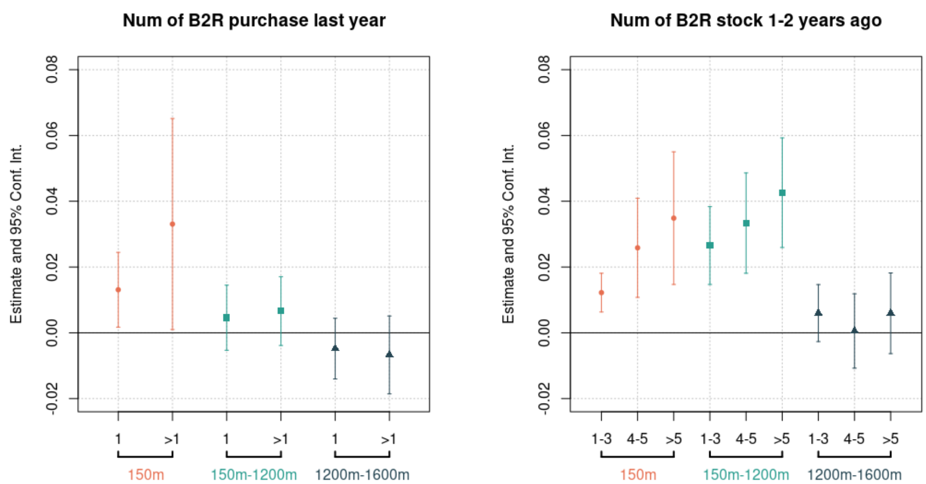
As illustrated in Section 2, there are three forces through which positive externalities from the buy-to-rent properties operate: competition channel, supply reduction due to reallocation, and changes in local amenities. I exploit the timing of buy-to-rent property acquisitions to distinguish these mechanisms.

The competition channel is expected to take effect at the time of purchase. When buy-to-rent investors shop for properties, they directly compete with other homebuyers who look for housing in the same area in the same time. So I use the number of buy-to-rent purchases in the same year to capture the competition effect. Mathematically, I run the baseline hedonic regression Equation (3.1) with the number of newly acquired buy-to-rent properties in the same year as the independent variable.

On the contrary, the supply and amenity channels are likely to take effect with a time lag. Buy-to-rent investors reallocate housing unit to the rental market and do not put these homes back on the for-sale market. This leads to a decrease in the housing stock available for sale in subsequent years. It also takes time for any renovations and residential compositional change to take place. Therefore, I use the number of buy-to-rent properties acquired one to two years ago as the independent variable to capture the supply and amenity channels. In the future, I plan to obtain building permits data from the local governments to directly measure renovations.

Figure 3.8 plots the coefficients and 95% confidence intervals from the two hedonic regressions using the number of buy-to-rent properties acquired within last year and one to two years ago as the independent variables, respectively. The left panel shows that contemporaneous buy-to-rent property acquisitions within 150 meters increases property values by one to three percent, and contemporaneous buy-to-rent property acquisitions further than 150 meters have close to zero impact on property values. The right panel shows that the presence of buy-to-properties acquired one to two years ago within 1200 meters increases property values by one to four percent, while buy-to-rent properties further than 1200 meters have minimal impact on property values. The results show that all three channels – competition, supply, and amenity – are at work in affecting housing prices in the presence of buy-to-rent properties. However, the changes in available housing stock and perceived neighborhood amenities induced by buy-to-rent properties are more influential on price appreciation.

Figure 3.8: Mechanisms driving the spillover effects of B2R properties



Notes: Figure 3.8 presents the coefficients and confidence intervals of hedonic regressions examining the impact of buy-to-rent properties on housing prices. Each dot represents the coefficients, and each bar represents the corresponding 95% confidence intervals. The left panel uses the buy-to-rent properties acquired within the last year of property sales as the independent variable, and the right panel uses buy-to-rent properties acquired one to two years before property sales. Each panel also distinguishes buy-to-rent properties within different distance bins around the sold property.

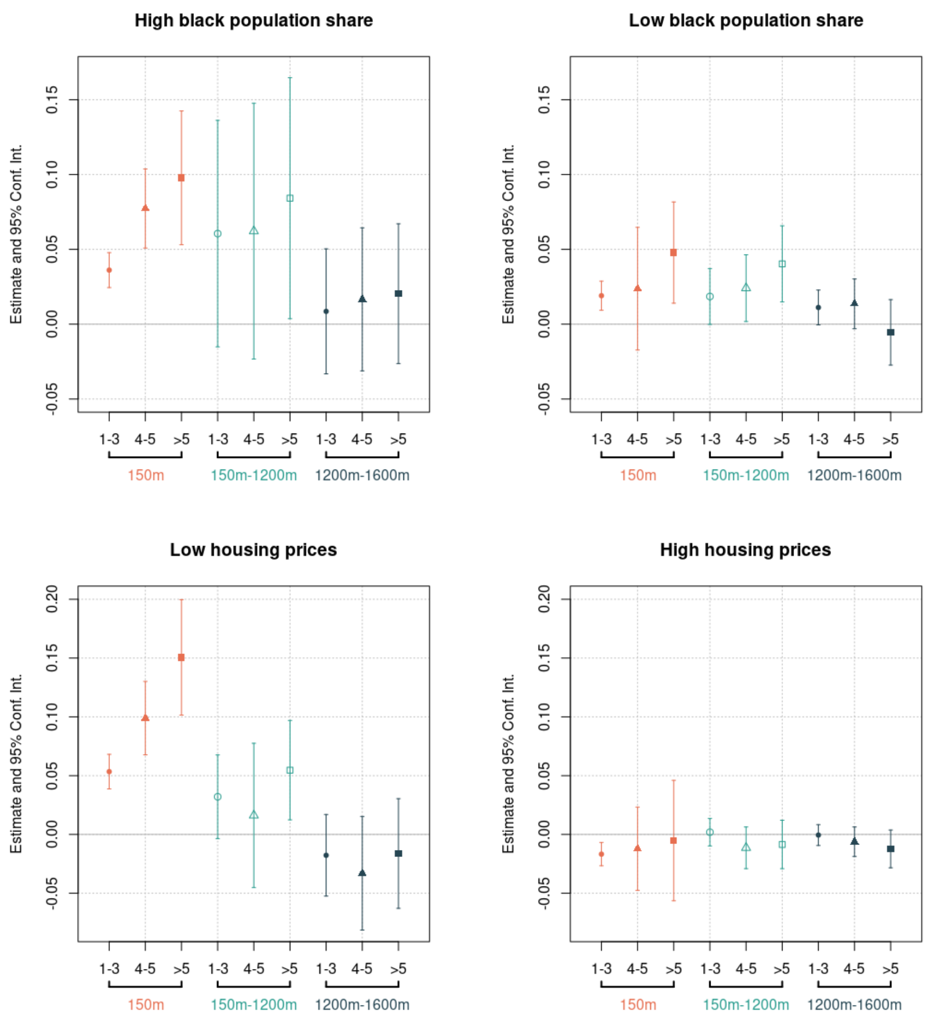
3.5.2 Neighborhood heterogeneity

Having examined the average spillover effect of buy-to-rent properties on property values of nearby houses, I will now investigate whether properties in different neighborhoods are affected differently by the entry of buy-to-rent investors. I define neighborhoods using Census tracts. I use neighborhood characteristics from the 2010 American Community Survey, and group neighborhoods into quartiles based on their share of black residents and housing prices.

Figure 3.9 displays the coefficients and confidence intervals of the spillover effects from buy-to-rent properties in different neighborhoods. I rerun the baseline regression Equation (3.1) for neighborhoods with different racial composition and housing prices separately. The top two panels show the results in neighborhoods at the top and bottom quartiles based on their share of black population. Neighborhoods with a higher share of black residents, that is, neighborhoods with more than 43% black residents, tend to experience larger positive price spillovers from buy-to-rent investors. The magnitudes of positive spillovers are about twice as large in neighborhoods with a higher black population share compared to neighborhoods with a lower black population share.

The bottom two panels of Figure 3.9 plot the results for neighborhoods in the highest and lowest quartiles based on their housing prices. The bottom right panel shows that properties in neighborhoods with lower initial housing prices in 2010 experience large positive spillover effects from the entry of buy-to-rent investors. Specifically, in neighborhoods below the 25th percentile in housing prices, the presence of buy-to-rent properties within 150 meters increase property prices by five to fifteen percent. Conversely, neighborhoods above the 75th percentile in housing prices show no significant changes in response to the presence of buy-to-rent properties. This pronounced difference reveals that cheaper neighborhoods are more susceptible to price spillovers caused by the influx of buy-to-rent investors, raising concerns about housing affordability for low- and middle-income households.

Figure 3.9: Neighborhood heterogeneity of the spillover effects



3.6 CONCLUSION

This paper investigates the impact of buy-to-rent investors on local housing markets. By leveraging the quasi-random variation in the location and timing of buy-to-rent property acquisitions, this paper examines the spillover effects the buy-to-rent properties have on nearby property prices. The main finding is that buy-to-rent properties contribute to local housing price appreciation, particularly in neighborhoods with more black population and lower housing prices. The positive price spillovers are mainly driven by supply reduction and positive amenity improvements. While existing homeowners benefit from property appreciation, it also raises barriers to homeownership for prospective buyers.

In recent years, many policy discussions have focused on restricting the property acquisitions of institutional investors, driven by concerns that these investors deter homeownership and raise local rents. The reduced-form results show that regulating investor behavior could help mitigate price increases. However, it is unclear whether limiting investor behavior would simply pass these properties to small mom-and-pop landlords, rather than returning them to the for-sale market. Moreover, to evaluate the impact of such regulations, it is important to consider their effects on the rental market and housing supply, which I deter to future research.

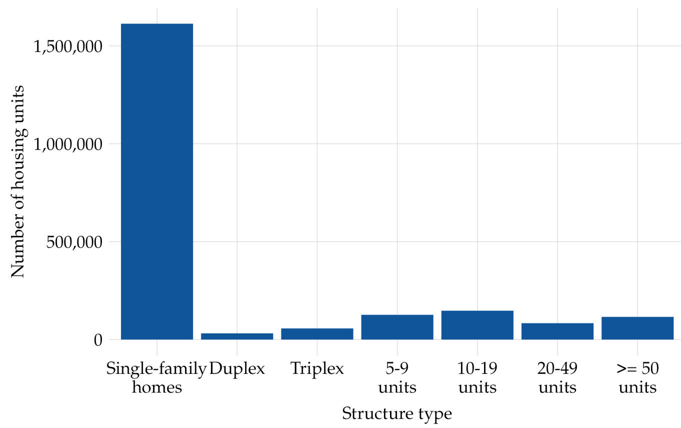
APPENDIX A

Appendix to Chapter 1

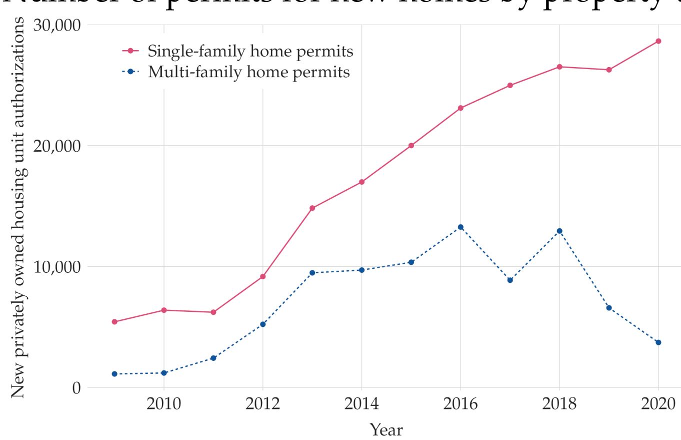
A.1 ADDITIONAL FIGURES AND TABLES

Figure A.1: Number of housing units by structure type

(a) Total number of housing units by structure type

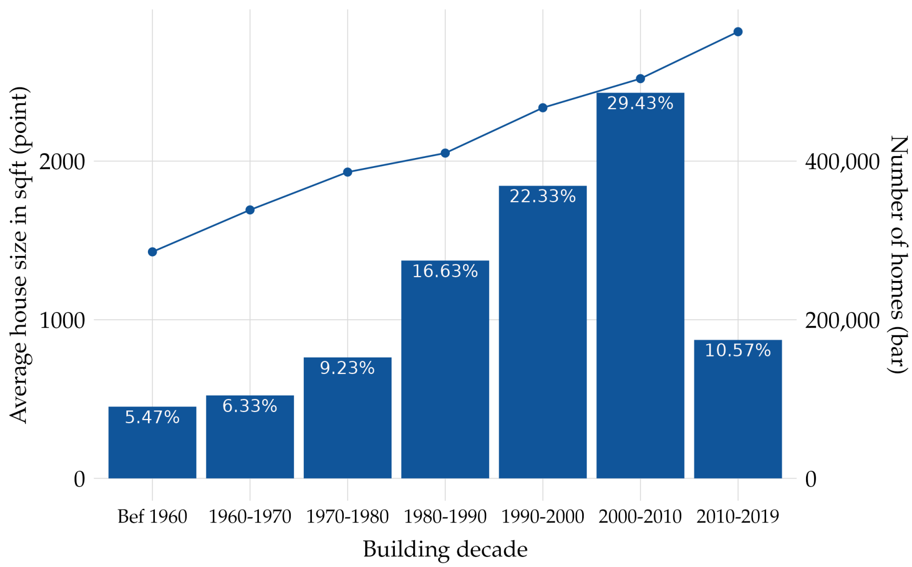


(b) Number of permits for new homes by property type



Notes: Panel (a) presents the number of housing units by structure type in 2010 and 2020, based on data from the American Community Survey 5-year estimates provided by IPUMS NHGIS. Panel (b) shows the number of building permits issued for single-family and multi-family homes between 2009 and 2020, calculated from the Census Bureau's Building Permits Survey. All number pertain to the Atlanta metropolitan area.

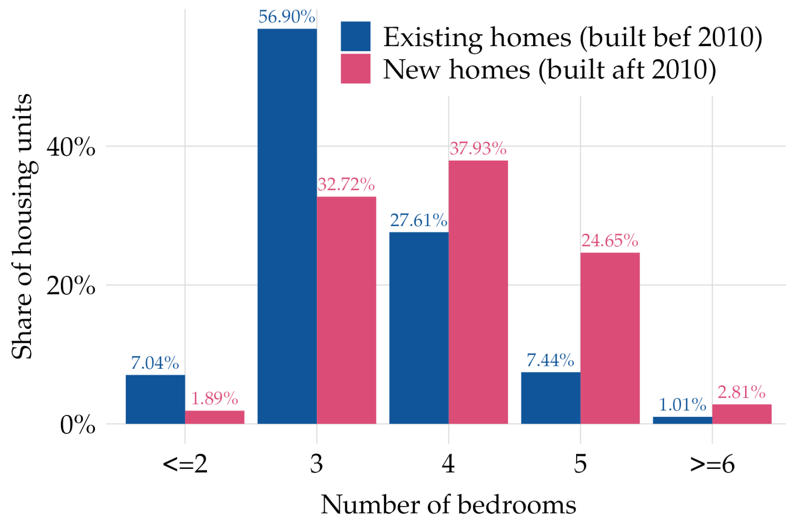
Figure A.2: Housing stock composition as of 2019



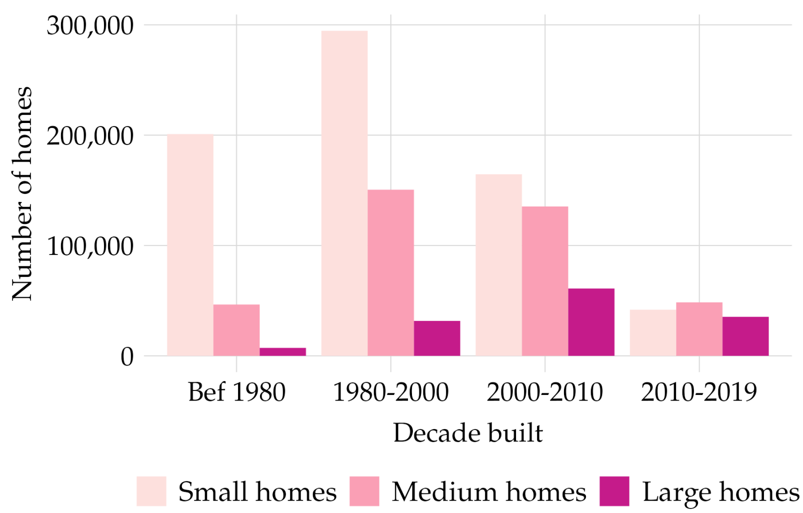
Notes: This graph shows housing stock composition by year built in the 2019 CoreLogic Assessor data. The points and line correspond to the left axis and show the mean house size for houses built in different years. The bars correspond to the right axis and show the number of houses built in different years. The numbers at the top of each bar show the share of houses built in different years to the total housing stock.

Figure A.3: Composition of single-family homes by house type and year built

(a) Share of homes by number of bedrooms and decade built

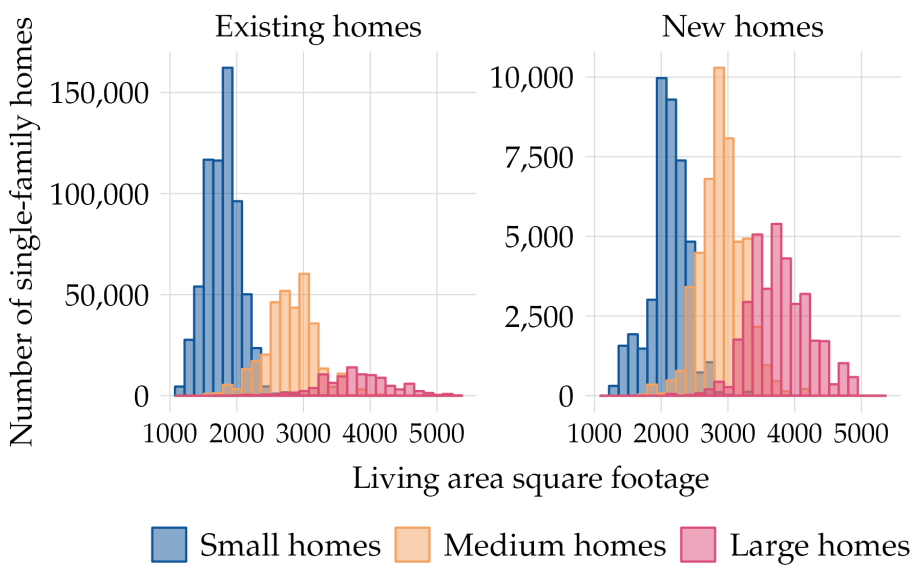


(b) Number of homes by decade built and house type



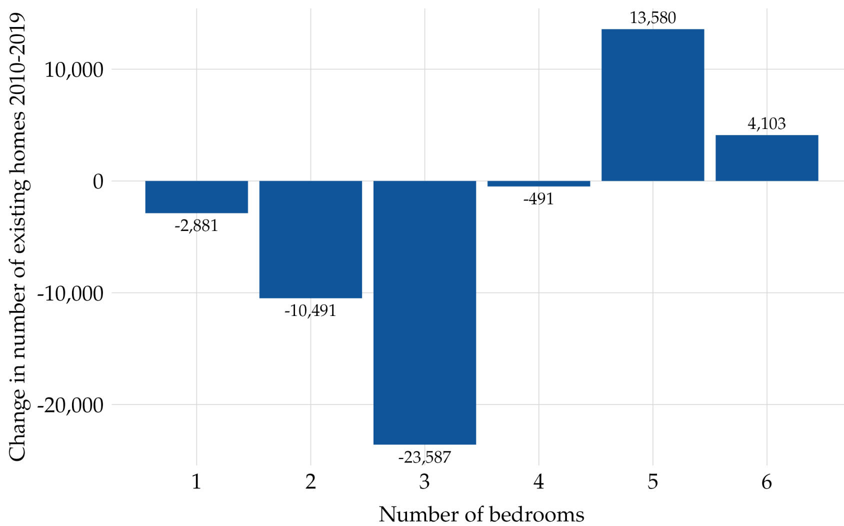
Notes: These graphs show housing stock composition by decade built and house type in the 2019 CoreLogic Assessor data. Small homes are defined as homes with fewer than three bedrooms, medium homes are defined as homes with four bedrooms, and large homes are defined as homes with more than five bedrooms.

Figure A.4: Single-family home size distribution by house type



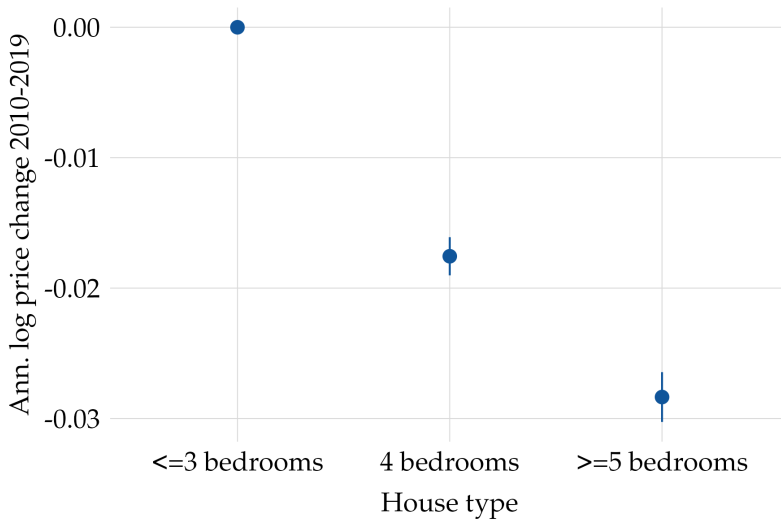
Notes: This figure plots the distribution of home sizes in square footage by house type for existing homes (built before 2010) and new homes (built after 2010) separately. Small homes are defined as homes with fewer than three bedrooms, medium homes are defined as homes with four bedrooms, and large homes are defined as homes with more than five bedrooms.

Figure A.5: Demolitions and upgrades of existing housing stock



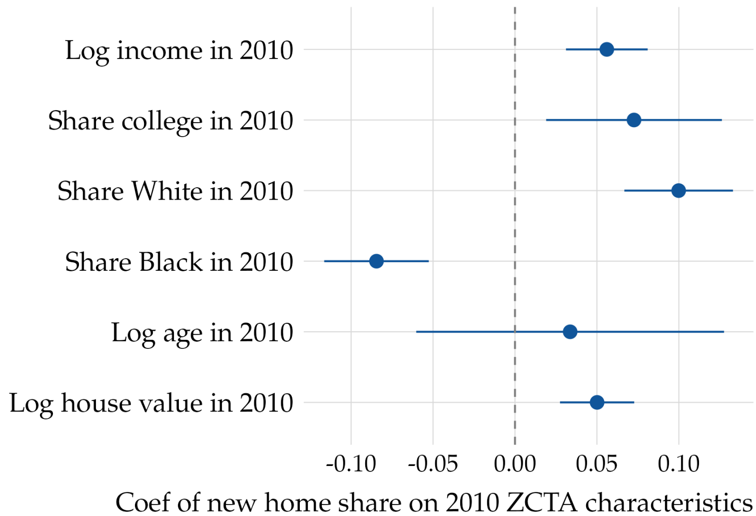
Notes: This graph plots the demolitions and upgrades of single-family homes built before 2010. It compares the number of single-family units recorded in the 2010 and 2019 Assessor records. The change in the number of units between these two years reflects alterations in the existing housing stock. The graph shows a decrease in the number of smaller homes in 2019 compared to 2010, suggesting that smaller homes are more likely to be demolished or redeveloped. In contrast, there is an increase in the number of larger homes in 2019, indicating that some of the existing housing stock has likely undergone upgrades or renovations to increase home size. Overall, the aggregate change in redevelopment is small, indicating that most new constructions are likely occurring on vacant land rather than through the redevelopment of existing properties.

Figure A.6: House price change of repeat-sale properties by house type in 2010-2019



Notes: The figure shows the annualized log price change of repeat-sale properties between 2010 and 2019, using data from CoreLogic transaction records. I partition all repeat-sale properties into three housing types based on their bedroom count, then I regress house price changes between transactions on their house type controlling for ZCTA and sale year fixed effects, taking three-bedroom homes as the omitted group.

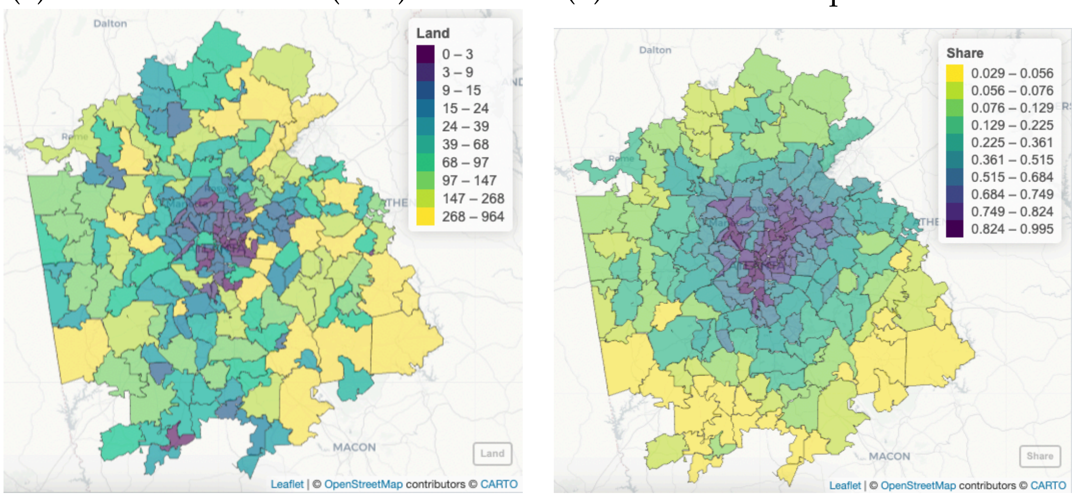
Figure A.7: Correlation between new home share and 2010 ZCTA characteristics



Notes: The figure shows the annualized log price change of repeat-sale properties between 2010 and 2019, using data from CoreLogic transaction records. I partition all repeat-sale properties into three housing types based on their bedroom count, then I regress house price changes between transactions on their house type controlling for ZCTA and sale year fixed effects, taking three-bedroom homes as the omitted group.

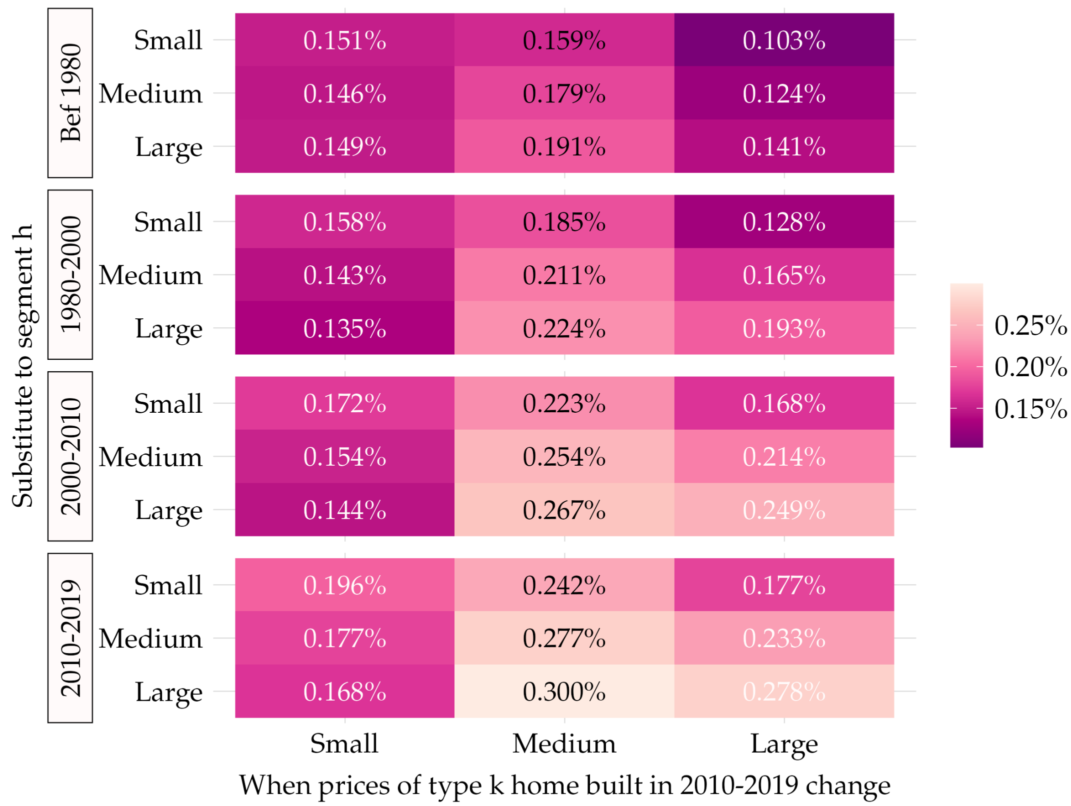
Figure A.8: Map of land availability

(a) Vacant barren land (acre) in 2016 (b) Share of developed land in 2016



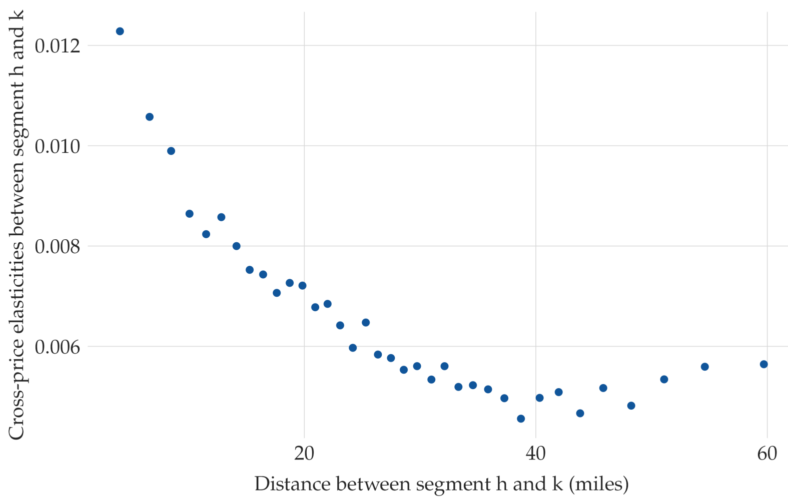
Notes: The two maps show land availability and land cover in the Atlanta MSA as of 2016, using data from the National Land Cover Database processed by the National Neighborhood Data Archive (NaNDA). The left panel shows the amount of vacant barren land – defined as land with less than 15% vegetation cover – measured in acres for each ZCTA. The right panel shows the share of already developed land relative to the total land area in each ZCTA.

Figure A.9: Estimated median cross-price elasticities by decade built and house type



Notes: This figure presents the median cross-price elasticity, aggregated by house type and decade built. The numbers within each rectangle gives the percentage change in market shares of segment h (row) when the prices of new homes of type k (column) changes by 1%. The colors indicate the magnitudes of the elasticities, with lighter shades indicating larger elasticities.

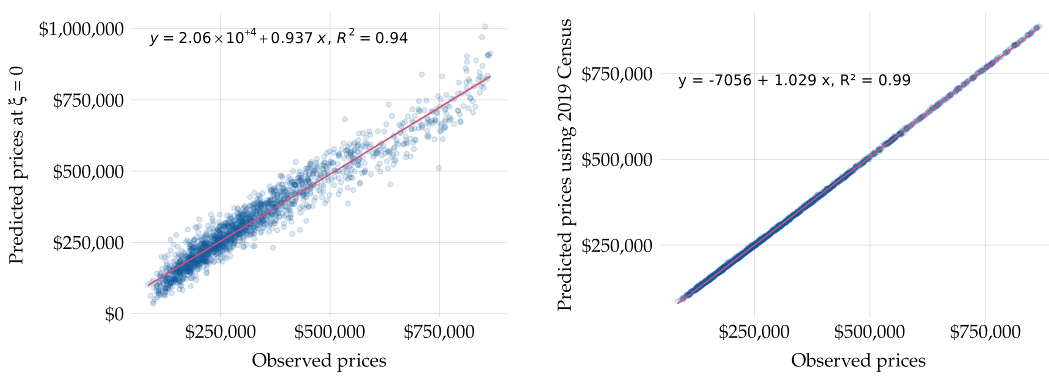
Figure A.10: Estimated cross-price elasticities by geographic distance



Notes: This figure displays a bin scatter plot illustrating the relationship between cross-price elasticities and the geographic distance between segments. The distances are measured using the centroids of the ZCTAs corresponding to each segment. Segments within the same ZCTAs are dropped.

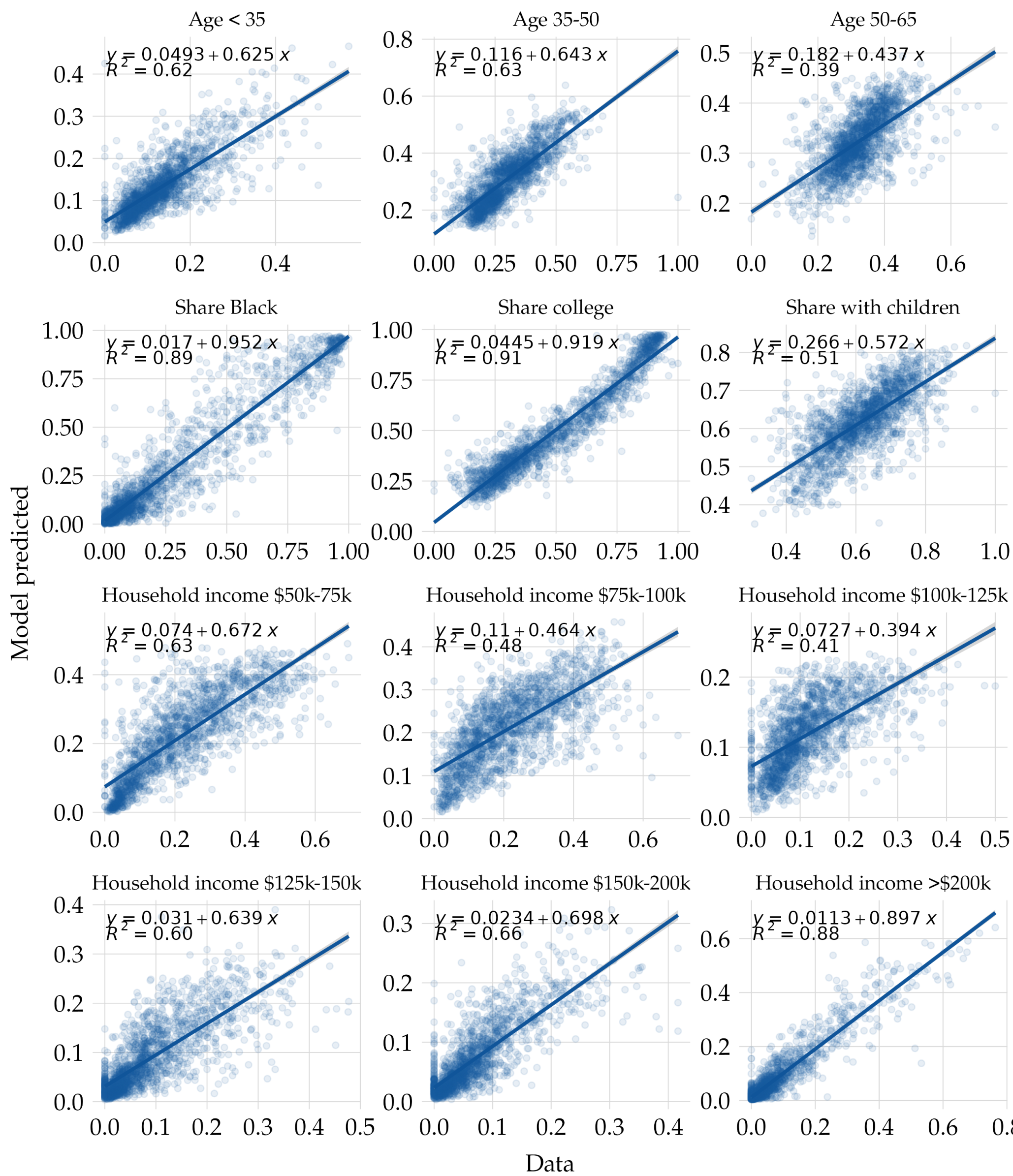
Figure A.11: Demand model validation

(a) $\xi_h = 0, \forall h$ (b) 2019 ACS population



Notes: The two graphs compare the actual housing prices with simulated prices under two scenarios. The simulated prices are calculated using the equilibrium condition and estimated demand parameters at the observed supply. The left panel recalculates housing prices by setting all unobserved segment quality to zero, $\xi_h = 0, \forall h$. The right panel recalculates housing prices using the population from the 2019 ACS.

Figure A.12: Predicted and observed sociodemographic composition in each segment



Notes: These graphs compare the actual and predicted sociodemographic composition for each segment.

Figure A.13: Construction cost per square foot

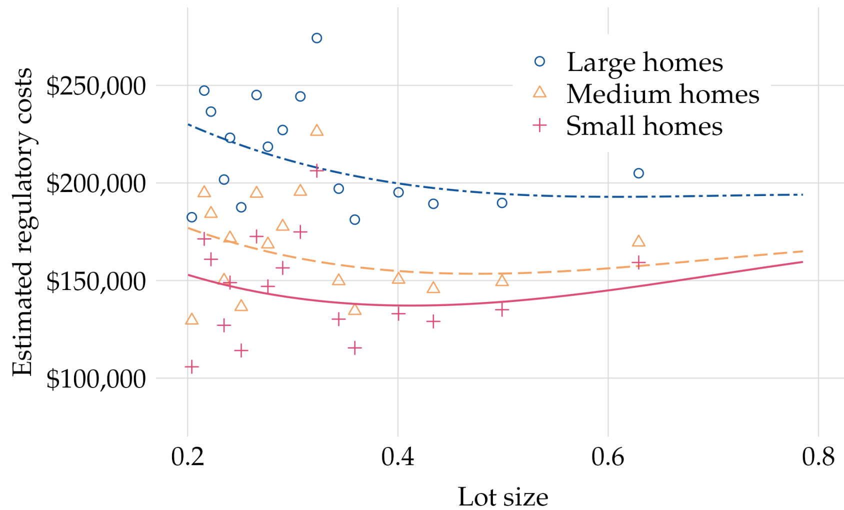
(a) By house size

(b) By neighborhood income

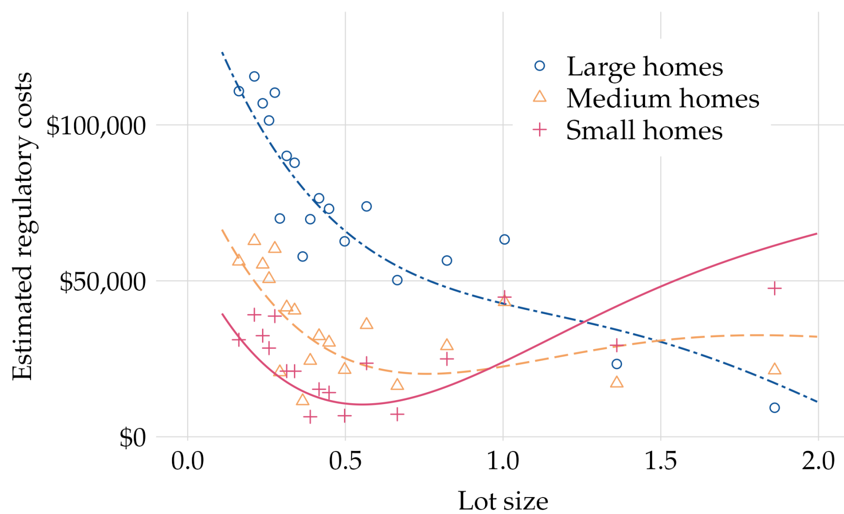
Notes: These graphs present the variation of construction cost per square foot by house size and location. Panel (a) presents a binned scatter plot showing how costs change with house size in square footage. Panel (b) highlights the variation in construction costs across neighborhoods with different income levels. The data are web-scrappd building permit data at the property level from the City of Atlanta between 2010 and 2019.

Figure A.14: Estimated regulatory costs by distance to city center

(a) Within 10 miles of city center



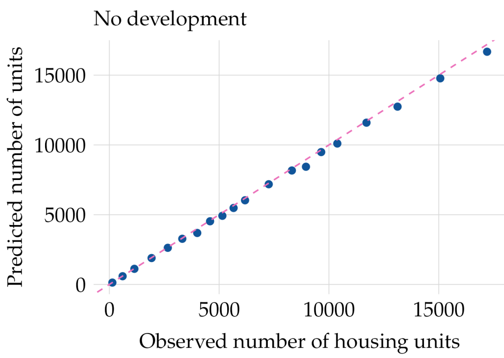
(b) Outside 10 miles of city center



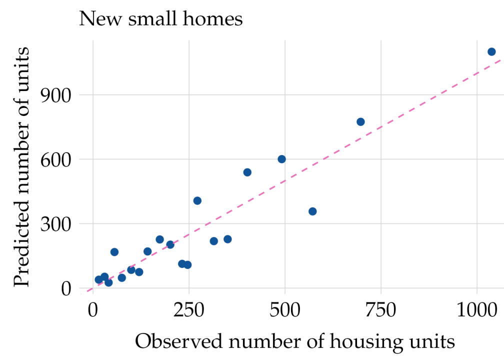
Notes: This figure plots the estimated zoning regulatory costs by distance to city center from the supply estimation. It is a binned scatter plot showing how these costs vary by lot size, potential house type, and location. Panel (a) shows the estimated costs for parcels located within 10 miles of the city center. Panel (b) shows the costs for parcels located further than 10 miles away from the city center. The fitted lines are cubic global polynomials applied to the data for each house type. Blue circles with dot-dashed lines represent large homes, yellow triangles with dashed lines represent medium homes, and '+' symbols with solid lines correspond to small homes. Small homes are defined as homes with ≤ 3 bedrooms, medium homes have four bedrooms, and large homes have ≥ 5 bedrooms.

Figure A.15: Predicted and observed lot development decision within neighborhood

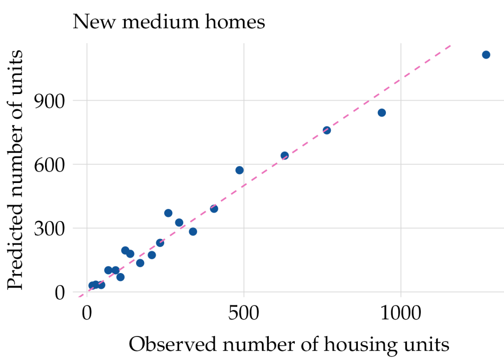
(a) No development (Keep existing structure)



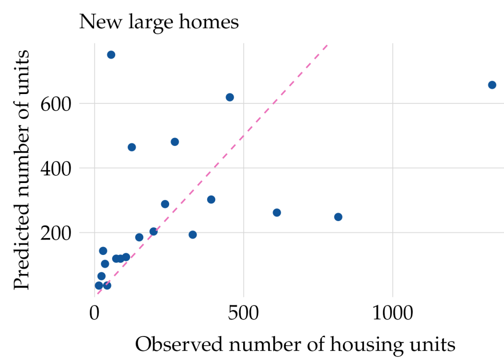
(b) New small homes



(c) New medium homes



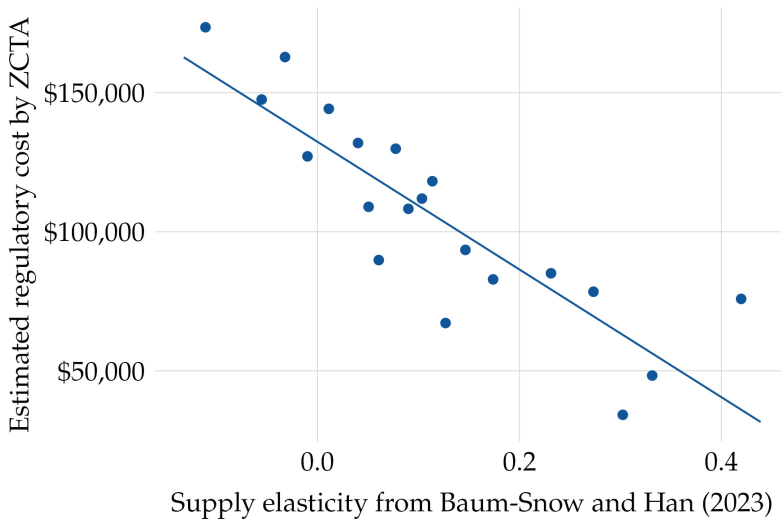
(d) New large homes



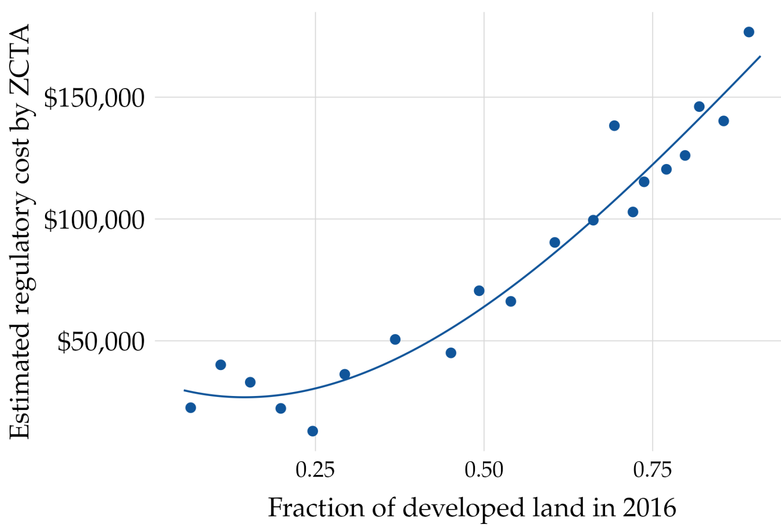
Notes: These binned scattered graphs compare the observed number of housing units to the model-predicted housing supply in each neighborhood. Each panel corresponds to a house type choice.The red dashed line is the 45-degree line.

Figure A.16: Estimated mean regulatory costs by ZCTA

(a) Correlation with supply elasticity from Baum-Snow and Han (2023)

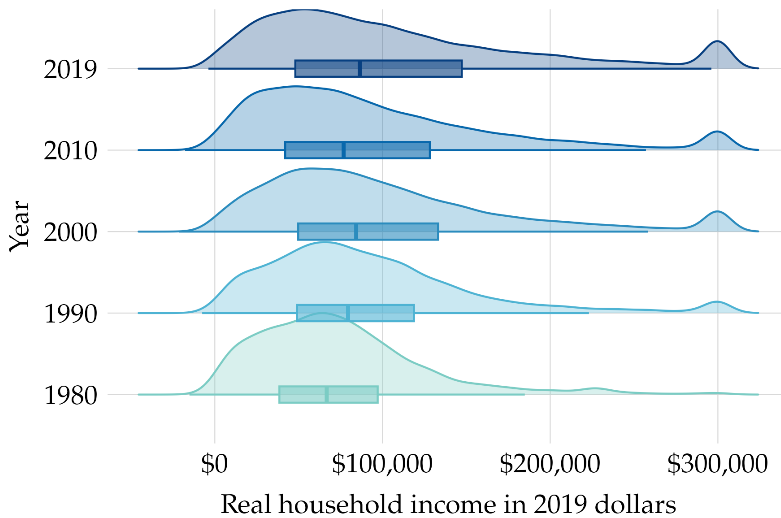


(b) Correlation with share of developed land in 2016 from NCLD



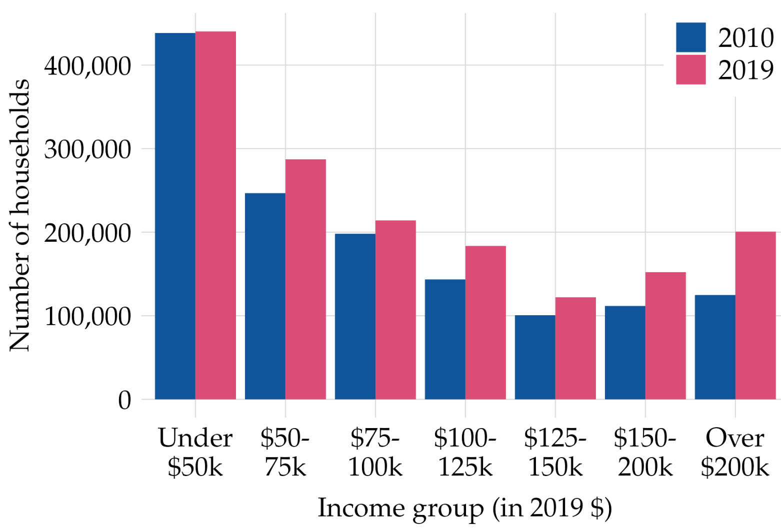
Notes: These two figures compare estimated regulatory costs aggregated at the ZCTA level with two other measures. Panel (a) uses the supply elasticities estimated by Baum-Snow & Han (2024). Panel (b) uses the fraction of developed land in 2016 from the National Land Cover Database.

Figure A.17: Change in income distribution 1980-2019



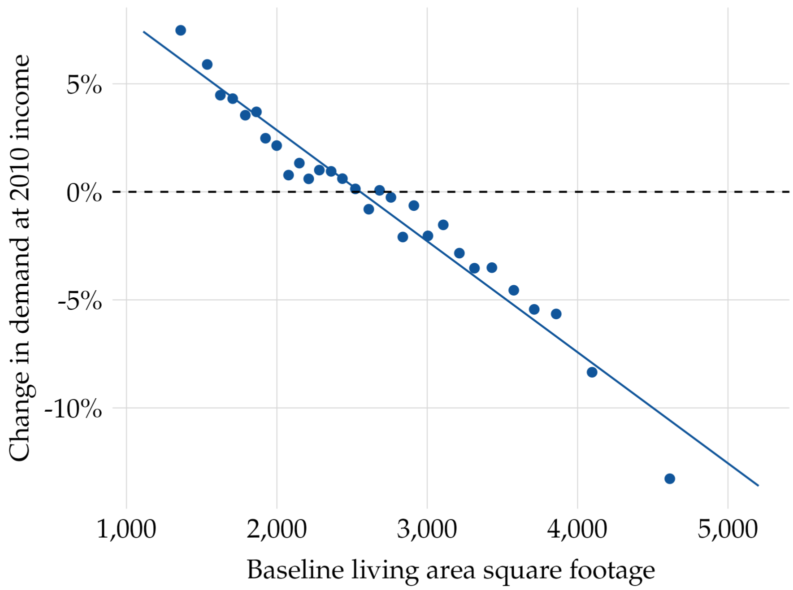
Notes: This figure show the change in income distribution from 1980 to 2019. The boxplot within the density plots shows the 25th percentile, median, and 75th percentile values. All incomes are adjusted to 2019 real dollars using the urban CPI for the Atlanta MSA. Incomes above $300,000 are grouped into a single category. Data for 1980, 1990, and 2000 are from the Decennial Census, and data for 2010 and 2019 come from the American Community Survey.

Figure A.18: Number of households by income in 2010 and 2019



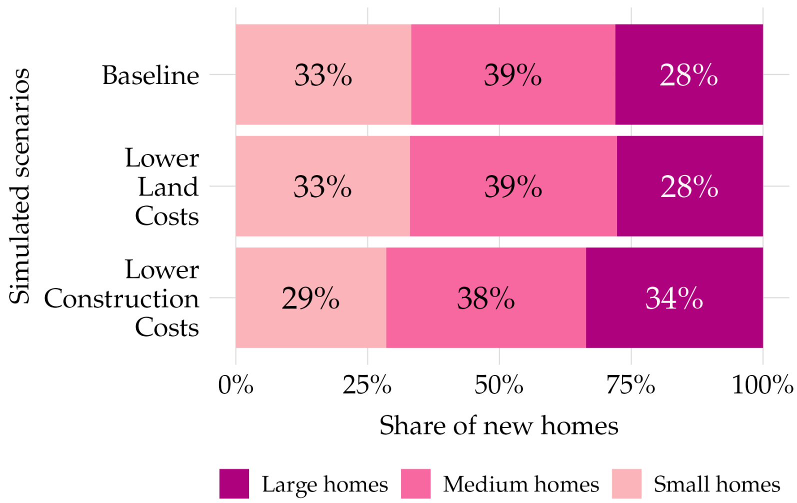
Notes: This figure show the number of households in each income group in 2010 to 2019. The data are from the American Community Surveys in 2010 and 2019. I convert the 2010 income to real 2019 dollars using the urban CPI in the Atlanta MSA.

Figure A.19: Partial equilibrium change in housing demand under 2010 income distribution



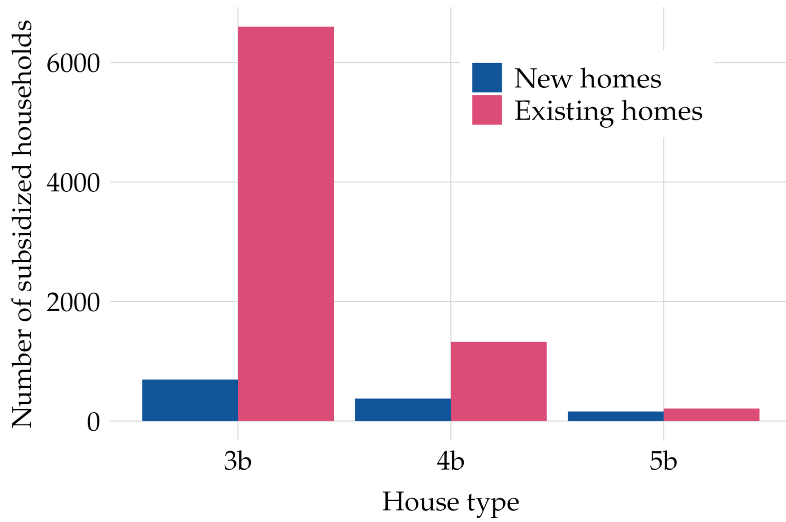
Notes: This figure plots the change in housing demand under the 2010 income distribution at observed prices compared to demand at the baseline scenario. Housing demand is calculated under partial equilibrium without adjusting supply and prices.

Figure A.20: Simulated distribution of new construction under lower production Costs



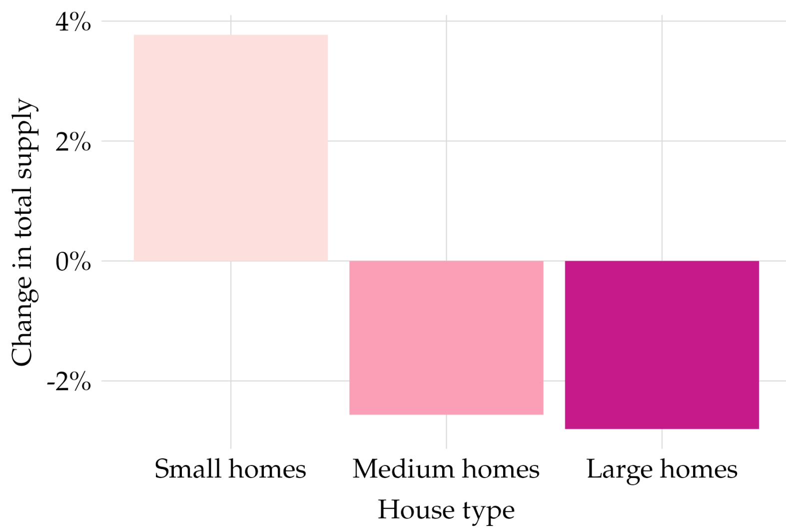
Notes: This figure presents the composition of new construction under the baseline scenario and the two simulated scenarios with the 2010 land and construction costs. To calculate the two simulated equilibria, I first calculate developers' decisions under the counterfactual costs. Given the updated supply, I calculate the new price vectors that clear the market. Then given these updated prices, I simulate developers' choices again. I keep iterating these steps until prices converge and supply equals to demand in each segment.

Figure A.21: Change in housing demand of subsidized households



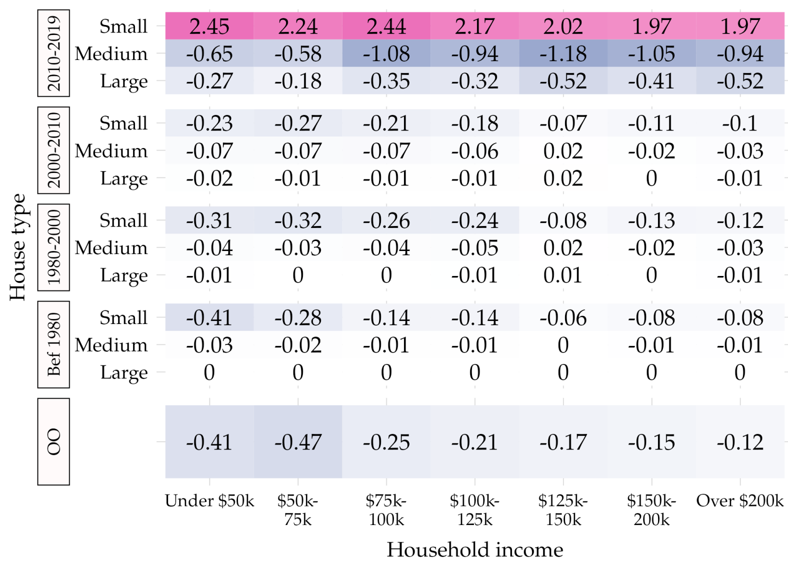
Notes: This figure shows the differences in housing choices of subsidized households between the baseline and subsidy scenarios. The numbers indicate the count of subsidized households switching to a different house type with the subsidy relative to baseline.

Figure A.22: Change in total housing supply with supply subsidy



Notes: This figure shows the percentage change in the total housing supply by house type, including both new and existing homes, under the supply subsidy scenario relative to the baseline.

Figure A.23: Change in housing demand with supply subsidy



Notes: This figure shows the percentage point change in the share of households choosing each house type and decade built (rows) for each income group (columns). The bottom row shows the percentage point change in households choosing the outside option (OO). For example, the top-left number indicates that for households earning less than $50,000, there is a 2.45 percentage point increase in the share of those households choosing small homes built between 2010 and 2019.

Table A.1: Examples of housing segments

Segment	Neighborhood (ZCTA)	Year built	Bedroom count
1	30305	2010-2019	≤ 3 (Small homes)
2	30305	2010-2019	4 (Medium homes)
3	30305	2010-2019	≥ 5 (Large homes)
4	30305	2000-2010	≤ 3 (Small homes)
5	30305	2000-2010	4 (Medium homes)
6	30305	2000-2010	≥ 5 (Large homes)
7	30305	1980-2000	≤ 3 (Small homes)
8	30305	1980-2000	4 (Medium homes)
9	30305	1980-2000	≥ 5 (Large homes)
10	30305	Bef 1980	≤ 3 (Small homes)
11	30305	Bef 1980	4 (Medium homes)
12	30305	Bef 1980	≥ 5 (Large homes)
...	...	...	...

Notes: This table illustrates an example of how housing segments are categorized. Houses in each neighborhood are grouped by their decade of construction and bedroom count. Houses in different segments are vertically differentiated in property and neighborhood characteristics.

Table A.2: The relationship between housing price changes and new homes

	Dependent variable: $\Delta \log P_j^{2010-2019}$			
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS w. FE	IV w. FE
$\Delta \log N_{h(j),c(j)}^{2010-2019}$	-0.0140***	-0.0131***	-0.0145***	-0.0279***
	(0.0001)	(0.0001)	(0.0031)	(0.0049)
Nbhd Characteristics		✓		
House Type FE			✓	✓
ZipCode FE			✓	✓
N	1,343,728	1,343,728	1,343,728	1,343,728
R^2	0.012	0.098	0.267	0.265

Notes: The table shows the regression results of property-level housing price change between 2010 and 2019 on the number of new homes that are of the same size type in the same zip code. The sample includes repeat-sale properties that were transacted at least twice between 2010 and 2019. j indexes property, $h(j)$ indexes one of house types each property j belongs to, and $c(j)$ indexes the zip code the property j is located at. Column (1) shows the OLS regression results without any controls. Column (2) adds neighborhood characteristics as controls, which includes the share of black population, share of college residents, median log household income, number of restaurants, school quality, and distance to city center. Column (3) uses zip code and house type fixed effects. Column (4) uses a shift-share instrument, which is the interaction between the 2010 initial share of houses in each housing type in each zip code and the growth rate of new housing supply in each housing type at the CBSA level.

Table A.3: Demographics of households

Sample	(1) Single-Family Homes	(2) Multi-Family Bldgs
Share Black	0.29	0.45
Share College	0.45	0.36
Share With Children	0.47	0.30
Age	52	44
Share Age ≤ 35	0.15	0.41
Share Age 35-50	0.33	0.28
Share Age 50-65	0.32	0.19
Household Income ($)	106,065	54,719
Share Income $\leq \$50$k	0.30	0.60
Share Income $\$50 - 75$k	0.18	0.18
Share Income $\$75 - 100$k	0.14	0.10
Share Income $\$100 - 125$k	0.11	0.05
Share Income $\$125 - 150$k	0.07	0.02
Share Income $\$150 - 200$k	0.09	0.02
Share Income $\geq \$200$k	0.11	0.02

Notes: This table presents summary statistics of the demographics of households living in single-family homes and multi-family buildings using the 2019 American Community Survey.

Table A.4: Estimates of mean housing preference parameters

	(1) OLS	(2) IV
	A. Housing characteristics	
Price ($100k)	-0.2261*** (0.0269)	-4.254*** (0.4763)
Sqft (std)	0.0810 (0.0472)	3.211*** (0.3948)
Built in 2010-2019	-1.420*** (0.0483)	1.672*** (0.4330)
Built in 2000-2010	-1.531*** (0.0445)	0.2251 (0.2459)
Built in 1980-2000	-0.8194*** (0.0380)	0.2749 (0.1822)
Sqft×Built in 2010-2019	-0.3236*** (0.0405)	0.1409 (0.1498)
Sqft×Built in 2000-2010	-0.2059*** (0.0407)	0.3927** (0.1489)
Sqft×Built in 1980-2000	-0.0689. (0.0415)	0.2282 (0.1383)
Lot size (std)	0.1423*** (0.0187)	0.5971*** (0.0934)
	B. Neighborhood characteristics	
Log pop density (std)	-0.0175 (0.0479)	0.2754 (0.2005)
Share Black (std)	-0.6398*** (0.0268)	-2.441*** (0.2188)
Share college (std)	-0.2510*** (0.0471)	2.706*** (0.3699)
Log household income (std)	-0.3213*** (0.0412)	-1.163*** (0.1806)
Share incorporated (std)	0.0517** (0.0159)	0.2736*** (0.0711)
Log dist to CBD (std)	0.1994*** (0.0455)	-1.243*** (0.2220)
School index (std)	0.0753*** (0.0228)	0.0810 (0.1046)
Job index (std)	-0.1501*** (0.0271)	0.4218** (0.1292)
Share developed low intensity (std)	-0.1383** (0.0484)	-1.020*** (0.2285)
Share developed medium intensity (std)	-0.0098 (0.0434)	-0.7749*** (0.1975)
Share developed high intensity (std)	0.1133** (0.0388)	-0.0132 (0.1772)
Share primary road (std)	0.0828*** (0.0129)	0.3372*** (0.0600)
Num public transport stops per sqmi (std)	0.2156*** (0.0388)	0.7792*** (0.1351)
Share park area (std)	-0.0757*** (0.0096)	-0.3428*** (0.0564)
Share woods (std)	-0.1922** (0.0738)	-0.8470** (0.3220)
Num restaurants per sqmi (std)	-0.2827*** (0.0447)	-1.012*** (0.1999)
Arts & entertainment estabs per sqmi (std)	-0.0873. (0.0499)	1.230*** (0.2253)

Notes: This table presents the mean preference coefficients for the housing demand model. It shows both the OLS and IV results from the second step of the demand estimation using $\hat{\delta}_h = \alpha p_h + \beta X_h + \xi_h$. Column (2) uses the optimal BLP instrument to address the endogeneity due to the correlation between unobserved quality ξ and prices p.

Table A.5: Estimates of heterogeneous housing preference parameters

	Black	College	With Children	Age ≤ 35	Age 35-50	Age 50-65
Price ($100k)	-0.4080	0.2628	-0.0005	-0.2061	-0.1479	-0.1176
Sqft (std)	-0.9345	0.1269	0.0688	-0.1644	0.1359	0.0608
Built in 2010-2019	-0.4766	-0.0354	-0.1340	0.0419	-0.0364	0.0581
Built in 2000-2010	2.0846	0.2355	-0.1657	2.0497	1.5956	0.7935
Built in 1980-2000	2.1180	0.2512	-0.0330	1.1215	1.1015	0.7104
Sqft×Built in 2010-2019	1.4206	0.1844	-0.0156	0.4220	0.3696	0.4376
Sqft×Built in 2000-2010	1.0354	-0.0216	0.2164	-0.0858	0.0870	0.1650
Sqft×Built in 1980-2000	0.8984	-0.0365	0.1081	-0.1826	-0.1640	0.0641
Lot size (std)	0.8171	0.0217	0.0728	-0.2181	-0.1994	-0.0366
Log pop density (std)	-0.1439	0.0025	0.1307	0.1485	0.0552	0.0799
Share Black (std)	1.3411	-0.0367	0.0644	-0.3191	-0.3204	-0.1867
Share college (std)	0.0307	0.6167	-0.2683	0.0945	0.0855	-0.0311
Log household income (std)	0.0284	-0.2154	0.1399	-0.0188	0.0133	0.0761
Share incorporated (std)	-0.0080	-0.0147	0.0055	0.0216	0.0215	0.0227
Log dist to CBD (std)	-0.3205	-0.0883	0.2449	-0.4237	-0.2960	-0.0834
School index (std)	0.0390	-0.0934	0.0681	-0.0873	-0.0798	-0.0885
Job index (std)	0.1367	0.1131	0.0375	0.0357	-0.0177	-0.0490
Share developed low intensity (std)	0.2383	0.0223	0.2169	-0.0611	0.0286	0.1216
Share developed medium intensity (std)	-0.1812	-0.0671	0.0217	0.2591	0.1312	0.1338
Share developed high intensity (std)	-0.0634	-0.0018	0.1261	-0.0861	0.0299	0.0840
Share primary road (std)	-0.0259	-0.0130	-0.0314	-0.0049	-0.0275	-0.0311
Num public transport stops per sqmi (std)	-0.1898	-0.1118	0.0561	-0.0507	0.0888	0.0248
Share park area (std)	-0.0111	0.0210	-0.0023	0.0324	0.0135	0.0191
Share woods (std)	0.2547	-0.0705	0.4081	-0.0284	0.0718	0.1625
Num restaurants per sqmi (std)	0.1854	0.1044	-0.0794	0.0263	0.0053	0.0064
Arts & entertainment estabs per sqmi (std)	0.1514	-0.0095	0.0606	-0.0017	-0.0224	0.0152

Notes: This table presents the heterogenous preference coefficients by demographic from the first step of estimation of the housing demand model.

Table A.6: Estimates of heterogeneous housing preference parameters

Income group	\$50-75k	\$75-100k	\$100-125k	\$125-150k	\$150-200k	≥\$200k
Price (\$100k)	-0.1569	-0.0664	0.0453	-0.0207	0.1490	0.3990
Sqft (std)	0.0127	-0.0139	0.0166	0.2110	0.2095	0.5903
Built in 2010-2019	0.0669	-0.0543	-0.0542	-0.1545	-0.1368	-0.0249
Built in 2000-2010	-0.0782	0.3364	0.1891	0.5486	0.1297	-0.0264
Built in 1980-2000	0.0864	0.6050	0.5647	0.7679	0.4507	0.3650
Sqft×Built in 2010-2019	-0.1208	0.3665	0.5061	0.5788	0.4368	0.3199
Sqft×Built in 2000-2010	-0.2420	-0.1414	-0.0911	-0.1436	-0.2716	-0.2562
Sqft×Built in 1980-2000	-0.3073	-0.0279	0.0678	0.1080	-0.0969	-0.0237
Lot size (std)	-0.3694	-0.1488	-0.0261	0.0495	-0.0724	-0.0314
Log pop density (std)	-0.0487	-0.0279	-0.0147	0.2087	0.0881	0.2330
Share Black (std)	0.0117	0.2423	-0.1293	-0.0338	0.0948	0.1556
Share college (std)	0.0005	-0.0915	0.4284	0.6317	0.2946	0.2221
Log household income (std)	0.1409	0.5426	0.3607	0.4878	0.5460	0.5267
Share incorporated (std)	-0.0522	-0.0506	-0.0276	-0.0712	-0.0089	0.0996
Log dist to CBD (std)	-0.1581	-0.1448	-0.2429	0.0080	-0.0049	0.0178
School index (std)	-0.0346	0.2560	-0.0060	0.0015	0.0936	0.0098
Job index (std)	-0.1776	-0.2423	-0.2450	0.0137	0.1134	0.1002
Share developed low intensity (std)	-0.2830	-0.6244	-0.1887	-0.2281	-0.1291	-0.1028
Share developed medium intensity (std)	0.2383	0.4441	0.1462	-0.0365	-0.0761	-0.0796
Share developed high intensity (std)	-0.2834	-0.3303	0.0364	0.2589	0.2092	0.2061
Share primary road (std)	-0.0453	-0.0123	0.0144	-0.1337	-0.1175	-0.0662
Num public transport stops per sqmi (std)	-0.3537	-0.2760	-0.3959	-0.0325	-0.2055	-0.1144
Share park area (std)	0.0581	0.0340	0.0198	0.1068	0.0543	0.0625
Share woods (std)	-0.2685	-0.5017	-0.0881	0.0712	0.2326	0.2340
Num restaurants per sqmi (std)	0.4718	0.3097	-0.0022	-0.0441	-0.0175	-0.2281
Arts & entertainment estabs per sqmi (std)	-0.0931	-0.2388	-0.1007	-0.0602	0.1924	0.1560

Notes: This table presents the heterogenous preference coefficients by income from the first step of estimation of the housing demand model.

Table A.7: Estimates of regulatory cost parameters

($)	Coef.	S.E.
A. Observed lot characteristics		
Lot Size (std)$\times$Medium homes	-26,644	(5,536)
Lot Size (std)$\times$Large Homes	-61,188	(5,998)
Lot Size (std)	43,504	(9,690)
B. Observed neighborhood characteristics		
Vacant Land (std)	-13,818	(1,861)
Incorporated (std)	2,277	(2,460)
Log(House Price) (std)	4,844	(3,696)
Share Black (std)	-3,634	(3,181)
Log (Dist to CBD) (std)	-57,735	(3,514)
Scale parameter σ_ν	30,334	

Notes: This table reports the estimated supply parameters for regulatory costs for building new homes between 2010 and 2019. All variables are standardized by subtracting the mean and standard deviation. The coefficients represent the additional regulatory costs in dollar terms from one standard deviation increase in each variable. Standard errors are calculated using nonparametric bootstrap with 200 replications.

Table A.8: Estimated regulatory costs by homes type ($)

($)	Median	Mean	25th percentile	75th percentile
New Small Homes	49507	56860	23069	84538
New Medium Homes	59262	67215	29359	96022
New Large Homes	98755	102655	60389	140377

Notes: This table presents the estimated regulatory costs for building new homes of each house type between 2010 and 2019 in the Atlanta MSA.

A.2 SUPPLEMENTAL ANALYSES

A.2.1 Developers' decisions on parcel size

When undertaking new developments, developers make two decisions: First, they choose how to divide vacant land into parcels with various sizes, and second, they choose which type of new homes to build on each parcel. In the main text, I focus on the second-stage choice of house characteristics given parcel attributes. Here, I present a simple model to illustrate the first-stage choice on parcel size. The choice of parcel size is modeled as a reduced-form function of local zoning regulations, land availability, and land costs.

Table A.9 reports the regression results using different specifications. Columns (1) to (4) uses parcels with new homes for which I can match the corresponding zoning standards. The results show that a 10% increase in minimum lot size is associated with 4.4% larger parcels. Additionally, the availability of vacant land is positively correlated with parcel sizes, indicating that neighborhoods with more vacant land generally have larger parcels. Conversely, land value appears to have minor impact on parcel sizes. Given that I cannot match zoning standards for many newly developed parcels, I use neighborhood characteristics as proxies for zoning stringency. Columns (5) and (6) report that lot sizes are bigger in neighborhoods with a lower share of Black residents, higher income, and lower population density. These neighborhoods with higher socioeconomic status tend to implement stricter zoning laws that limits denser development.

The results are consistent with a broad literature documenting the role of minimum lot size as the crucial factors determining lot sizes, such as Song (2021) and Mei (2022). This suggests that municipalities effectively use minimum lot size requirements as an exclusionary zoning strategy. Given the zoning environment, it is less financially feasible for developers to make profits from building small homes on large parcels with high land costs, and thus it encourages the construction of larger homes catering to wealthier households.

Table A.9: Determinants of parcel size

	Dep.Var: Log(Parcel Size)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Min Lot Size)	0.4326***	0.4497***	0.4429***	0.4424***		
	(0.0047)	(0.1030)	(0.0049)	(0.1025)		
Log(Vacant Land Area)	0.0809***	0.1483**	0.0867***	0.1510**		
	(0.0030)	(0.0483)	(0.0031)	(0.0508)		
Log(Land Value Per Acre)			0.0315***	-0.0621		
			(0.0034)	(0.0422)		
Share Black					-0.1327**	-0.1711
					(0.0143)	(0.1322)
Log(Household Income)					0.5038***	0.4253***
					(0.0122)	(0.1171)
Log(Pop Density)					-0.4990***	-0.3149***
					(0.0042)	(0.674)
Zoning jurisdiction FE	No	Yes	No	Yes	No	Yes
N	47,395	47,395	47,395	47,395	124,449	124,449
R^2	0.205	0.289	0.206	0.290	0.132	0.324

Notes: This table regresses observed parcel size on zoning regulations, land value, and land availability. Columns (1) and (2) uses the minimum lot size to measure the stringency of local zoning regulations. Minimum lot size is the government-mandated minimum lot size in the zoning area each parcel belongs to. Vacant land area is the sum of land size of all vacant land in the Census tract the parcel belongs to. Land value per acre is the average land value at the Census tract level. Since minimum lot size is not always observed in the data.

A.2.2 Role of heterogeneous demand

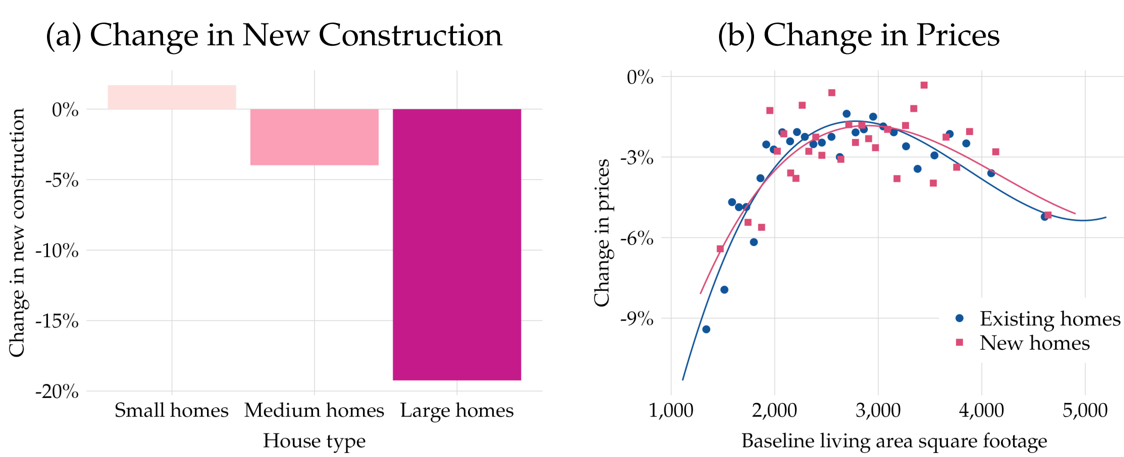
In this subsection, I evaluate the impact of household heterogeneity on housing supply and prices. To quantify the role of heterogeneous demand, I consider an extreme scenario in which all households are assumed to have homogenous preferences, and then I compare the counterfactual equilibrium outcomes to the baseline equilibrium with heterogeneous preferences. To simulate the counterfactual, I set all preference parameters (α_i, β_i) at the average value across all households and recalculate market-clearing prices based on the updated demand. Developers then re-optimize their development decisions when facing these new prices, and households re-optimize given the updated supply. I keep iterating this process until prices converge. This homogenous preference scenario eliminates heterogeneous sorting. All households derive the same level of utility from a given segment, and their choice probabilities are independent of their demographics.

Panel (a) of Figure A.24 shows that with homogeneous preferences, the number of newly built large homes decreases by almost 20%, while medium-sized homes see a 4% reduction. This result highlights the role of the interaction between heterogeneous demand and endogenous supply in driving the construction of homes of different sizes. When eliminating preference heterogeneity, the preferences of higher-income households shift toward smaller homes, lowering the demand for larger homes and leading developers to reduce the construction of large homes.

However, this shift in demand results in only a 2% increase in small home construction. Small homes are the house type that best aligns with the preferences of an average household. As shown in Table 1.1, the average household income is $90,000, while the average home has 3.5 bedrooms and 2,300 square feet. Figure 1.5 shows that the average preferences closely match the preferences of households earning between $75,000 and $125,000. Among these households, 53% reside in small homes. The small increase in small home construction indicates that there is a net decrease in total new construction. This suggests that developers are more responsive to the preferences of high-income households with a higher willingness to pay for newer, larger homes, rather than the average household.

Panel (b) shows the general equilibrium price changes for homes of different sizes under the counterfactual, compared to the baseline. When no households have strong preferences for large homes and low price sensitivity, prices decline across all home sizes. This indicates that demand from high-income households contributes to overall price growth, as they have greater financial capacity to bid up prices. Moreover, converging household preferences toward the average leads to substantial price declines at the two ends of the size spectrum. Prices for large homes drop by 3 to 6% (around $20,000), driven by the reduced demand from high-income households for large homes, despite a drop in the supply of newly built large homes. Similarly, prices for small homes drop by as much as 9% (around $10,000). This is because the preferences of low-income households shift to the middle, which leads to lower demand for the smallest homes and lower prices.

Figure A.24: Simulated equilibrium with homogeneous preferences



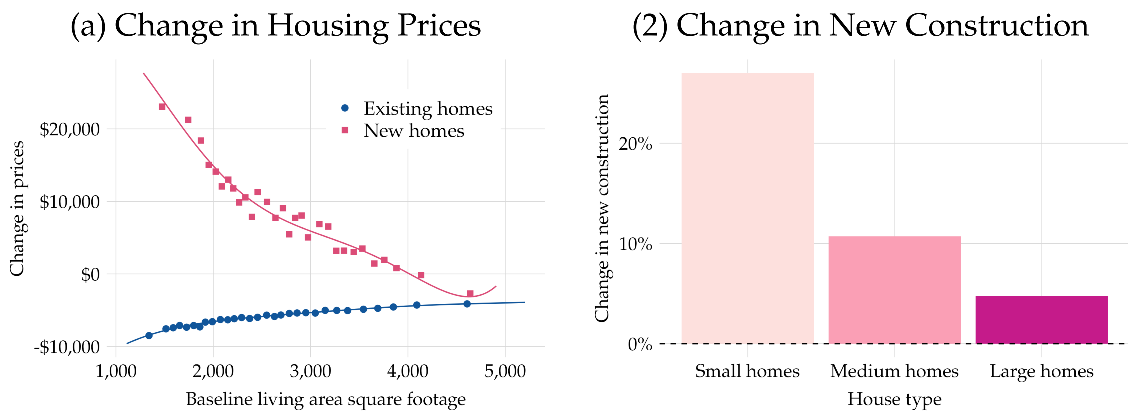
Notes: These figures present the simulated equilibrium under a counterfactual scenario where housing preferences are homogeneous across all households. In this scenario, the preference parameters of all households are set to the average values. The left panel shows the percent change in new construction by house type under the counterfactual scenario with homogeneous preferences, compared to the baseline scenario with heterogeneous preferences. The right panel is a binned scatter plot showing the average percent change in housing prices by house size for both existing and new homes under the counterfactual relative to the baseline.

A.2.3 Restrict house types for subsidized households

Many are concerned that subsidizing households will simply drive up housing prices by the same amount of the subsidy, especially in markets with inelastic supply, without increasing overall housing stock. The quantitative analysis in Section 1.6.2 shows that small home prices will rise by an average of $2,000. The modest price effect is largely due to the fact that only 7.2% of households receive the subsidy, and they only need to slightly outbid other households. Furthermore, my equilibrium calculation assumes that all households re-optimize their housing choices, reflecting a long-term equilibrium averaged over ten years. However, in reality, only a subset of homes are available on the market in the short term, which could result in more pronounced price effects. Therefore, I consider an extreme scenario in which subsidized households are restricted to purchasing new homes. This scenario gives an upper bound for the potential price impact.

Figure A.25 reports the changes in prices and new construction under this scenario. The distribution of price and supply changes of new homes is similar to the main results in Figure 1.10. However, the magnitudes here are much bigger. Prices of new homes around 1,500 square feet increase by $20,000, nearly 80% of the subsidy amount given to each eligible household. Developers also build 27% more new homes relative to the baseline new construction in response to the price growth. However, at the aggregate level, the total stock of small homes rises by only 1.7%. The large magnitude is mostly driven by the restricted choice set. The counterfactual excludes existing homes from the choice set of subsidized households, so it creates a larger demand shock for new homes. The large price response also reflects the slow adjustment of new housing supply to demand shocks, which leads to substantial price changes. The decline in prices of existing homes also reflects a mechanical effect of the counterfactual that reduced demand for existing homes lowers prices, especially smaller ones.

Figure A.25: Impact of demand-side subsidy on the housing market



Notes: These two figures plot the simulated equilibrium with a one-time $25,000 transfer to households under age 35 and income below $75,000, but restricts subsidized households to choose from new homes only. The left panel is a binned scatter plot showing the average percent change in housing prices by house size for both existing and new homes, relative to the baseline. The right panel shows the percent change in new construction by house type in the metro under this counterfactual scenario compared to the baseline levels.

A.3 DATA CONSTRUCTION

A.3.1 Predict housing values

I use a machine learning approach to predict the market value of non-transacted properties. I run three statistical models on the sample of transacted properties: linear regression, random forest, and extreme gradient boosting (XGBoost). The linear regression model is equivalent to a hedonic model. Then I compare the performances between the three models, and apply the best performing model on the set of untransacted properties to predict their prices.

The predictors include the number of bedrooms, number of bathrooms, living area square footage, lot size, building age, property longitude, property latitude, and county-year fixed effects. Figure A.26 shows the binscatter plots between each property characteristic and sale prices. Sale prices increase linearly with the size of the property, but change nonlinearly with property location. To address the potential nonlinear relationships, I use polynomials of the predictors to allow for more flexible functional forms.

I split the sample of transacted properties to a training dataset (75% of the sample) and a test dataset (25% of the sample). I tune the hyperparameters in the random forest and xgboost models to minimize the root mean squared errors. Specifically, I choose the number of features and number of trees for random forest, and choose the tree depth, number of trees, and learning rate for xgboost. The dataset includes over 800,000 sales records, and its large size imposes substantial computational costs for the tuning process. To reduce the computational cost, I perform the tuning procedure on several smaller random subsets of the data and then choose well-performed hyperparameters heuristically on the entire dataset.

Table A.10 compares the in-sample and out-of-sample predictive performances of the three models. The conventional hedonic linear model has the worst performance in terms of root mean squared error and R^2. Both random forest and XGBoost have much better predictive performances. The random forest has the best in-sample performance with 0.116 root mean squared error and 0.97 R^2, but it has slightly worse out-of-sample performances probably due to overfitting. XGBoost has the best out-of-sample performance with 0.249 root mean squared error and 0.829 R^2, and its in-sample performance is similar to random forest. So I choose XGBoost to predict housing prices for non-transacted properties.

To further understand the machine learning models, I use the permute-and-predict method to understand the contribution of each variable to the final prediction. The permute-and-predict method randomly permute values for each variable and then compare the predictive performance relative to the baseline values. The intuition is that the predictive performance will deteriorate significantly when permuting important variables. Figure A.27 plots the importance of each variable in these three machine learning models. All methods agree that the living square footage is the most important variable in predicting housing prices.

Figure A.26: Correlation between property characteristics and sale prices

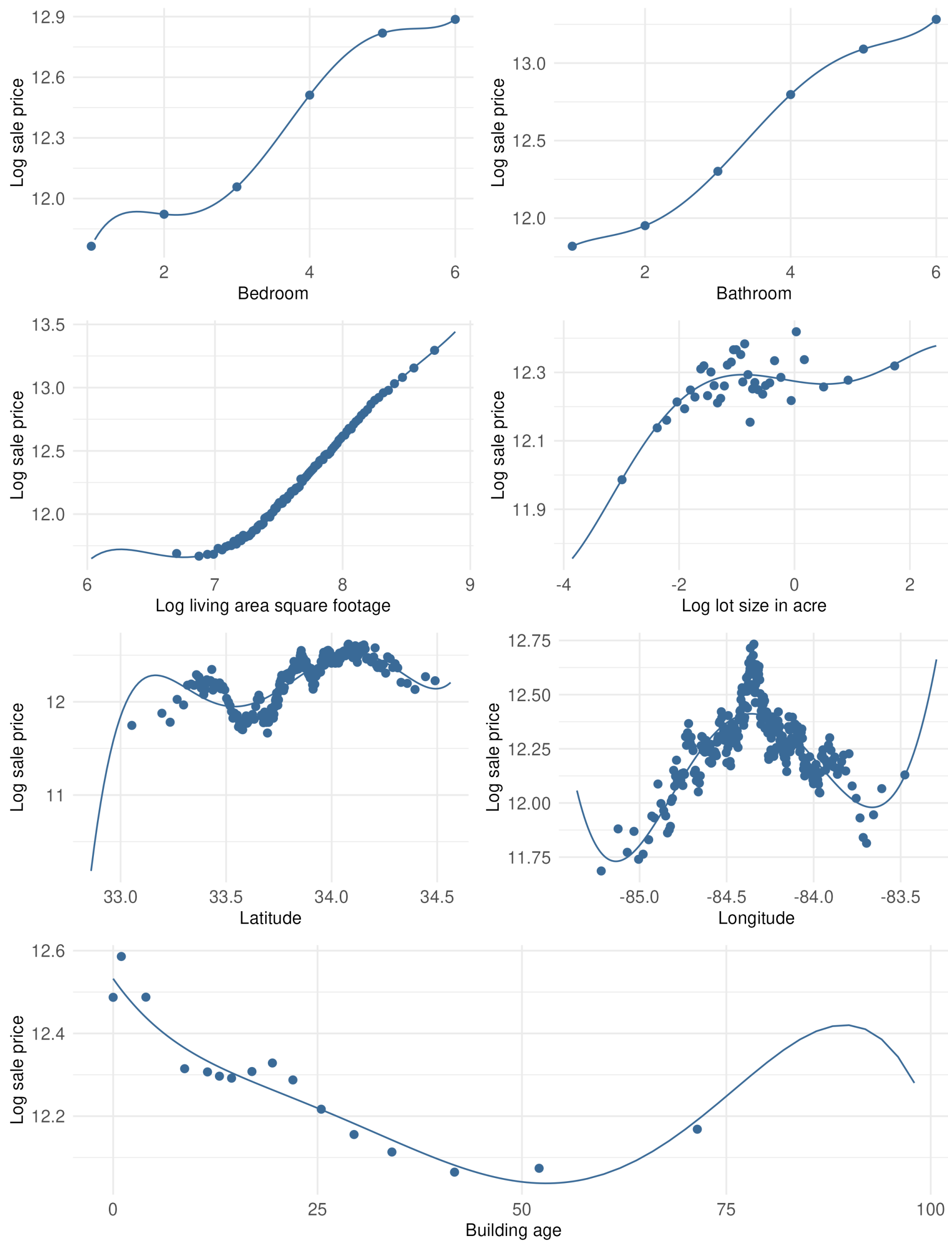
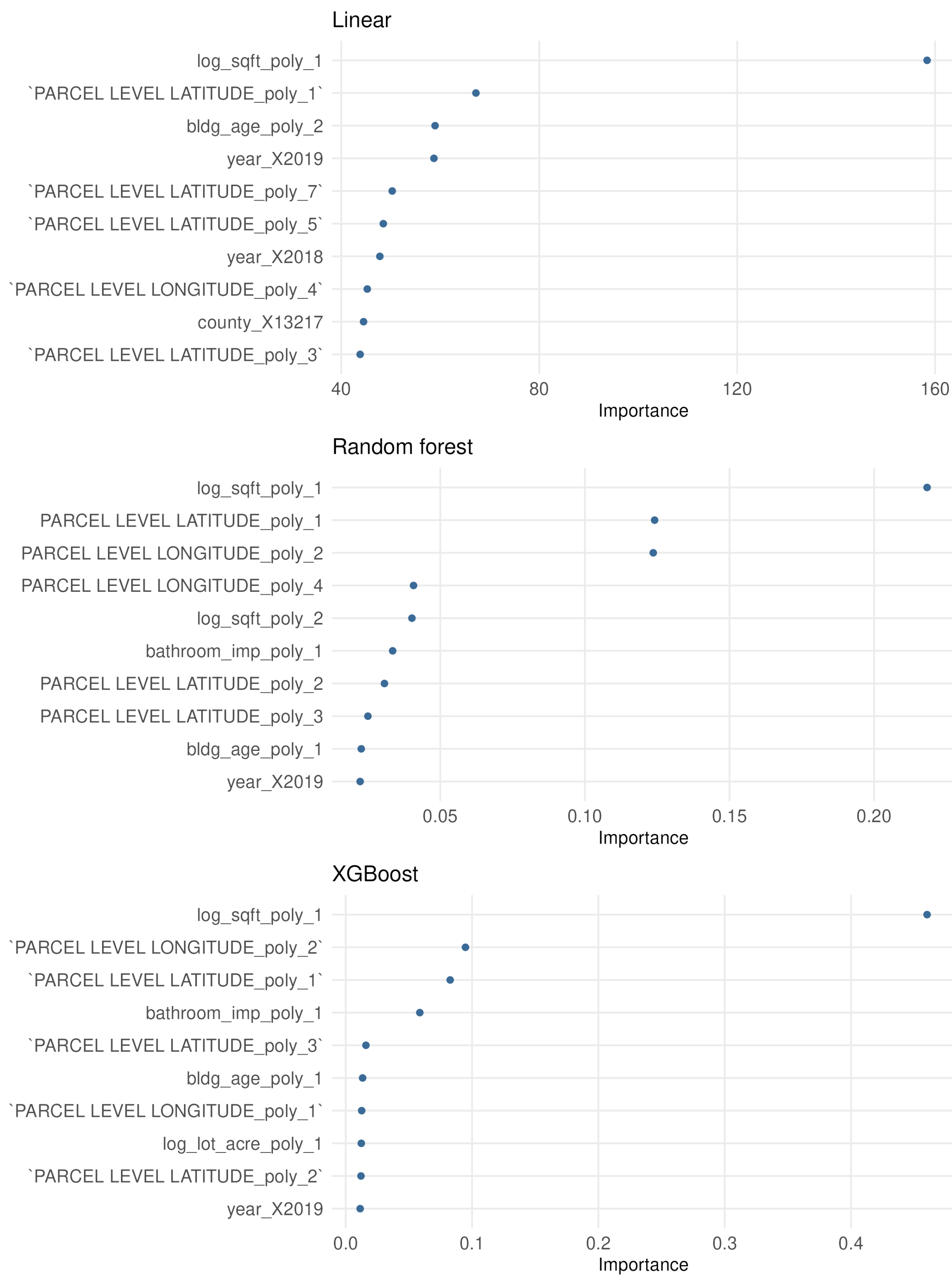


Table A.10: Compare performances of machine learning models

	In-sample		Out-of-sample	
	RMSE	R^2	RMSE	R^2
Linear	0.350	0.662	0.351	0.662
Random Forest	0.116	0.966	0.262	0.812
XGBoost	0.140	0.947	0.249	0.829

Figure A.27: Variable importance in machine learning models



A.3.2 Validate L2 data

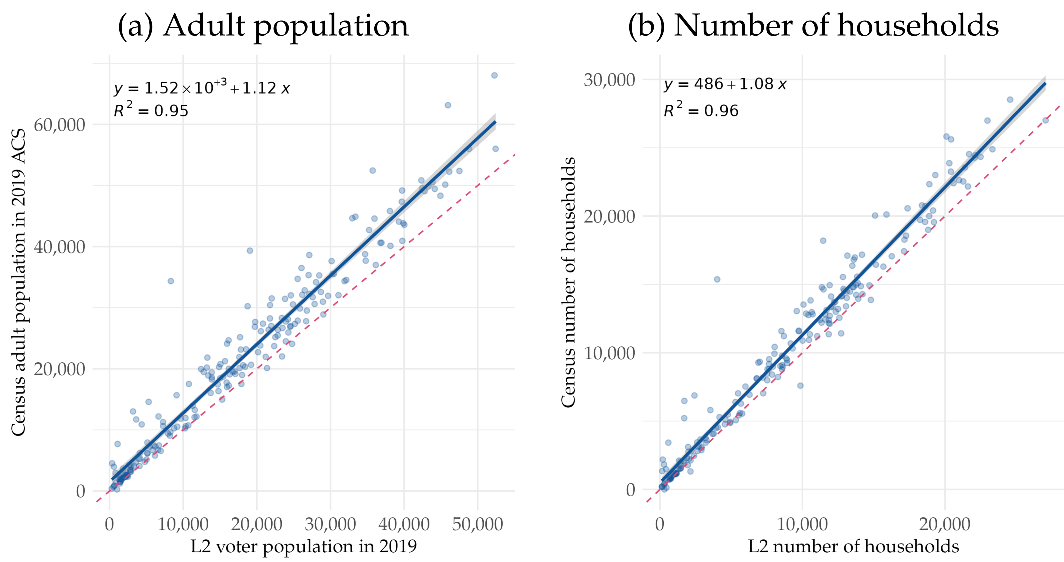
L2 compiles basic demographic information, including age, race, address, and gender, from government voter registration records. L2 further augments the base files with socioeconomic characteristics from other sources, such as credit bureaus, telecommunication providers, and other consumer data databases. Some of the demographic variables are imputed using proprietary models. To check the accuracy and reliability of L2 data, I compare the average socioeconomic characteristics at the Census tract level against the 2019 American Community Survey.

Figure A.29 plots the number of adult population and households from L2 against the 2019 ACS for each Census tract. I also report the coefficients and R-squared when regressing the Census estimates on the L2 estimates for each plot. Figure A.29 shows that Census captures a larger population. The result aligns with expectations, as the L2 data only includes registered voters and therefore excludes immigrants and citizens who are not registered to vote. Nevertheless, the population figures between the two sources are fairly close.

Figure A.28: Summary statistics of ACS, L2, and matched CoreLogic-L2 in 2019

	ACS	L2	CL-L2-SFH
# adults	4,414,026	3,716,906	2,961,991
# households	2,104,360	1,891,276	1,400,222
White	0.471	0.542	0.596
Black	0.336	0.360	0.311
Asian	0.058	0.043	0.045
Hispanic	0.107	0.055	0.048
Median household income	$68316	$76854	$77833
College	0.339	0.485	0.492

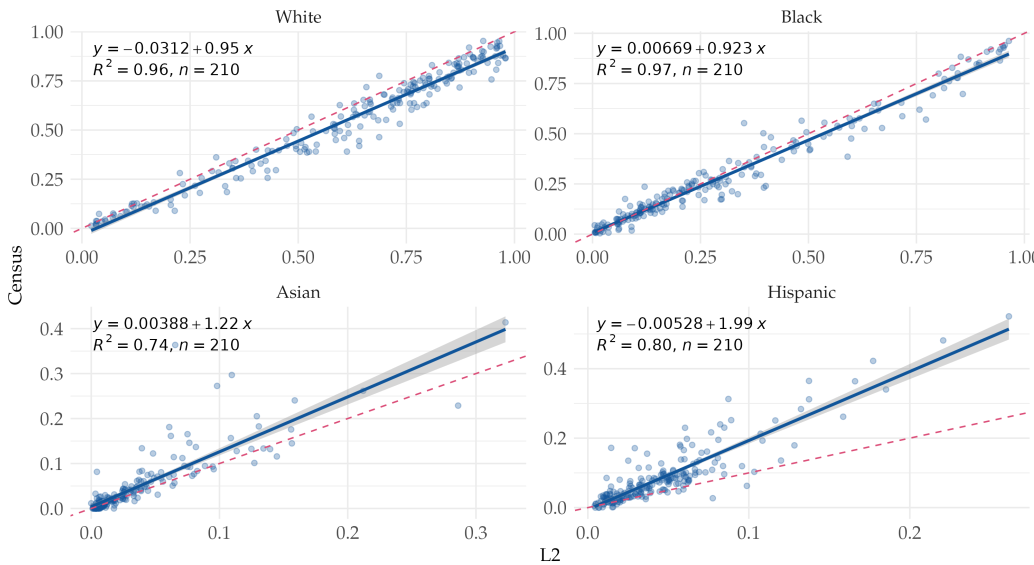
Figure A.29: L2 population vs. 2019 ACS



Notes. Each dot represents a Census tract. The blue line is the linear fitted line between the L2 and ACS estimates. The dashed pink line is the 45 degree line.

Figure A.30 compares the racial composition between the two datasets. L2 accurately represents the White and Black populations but underestimates the Asian and Hispanic populations. This is also expected, as Asian and Hispanic population are more likely to be immigrants and not eligible to vote. In my main analysis, I do not distinguish between the Asian and Hispanic populations because they constitute a small share of the total population in Atlanta, and the data cannot allow me to accurately capture their heterogenous preferences for housing.

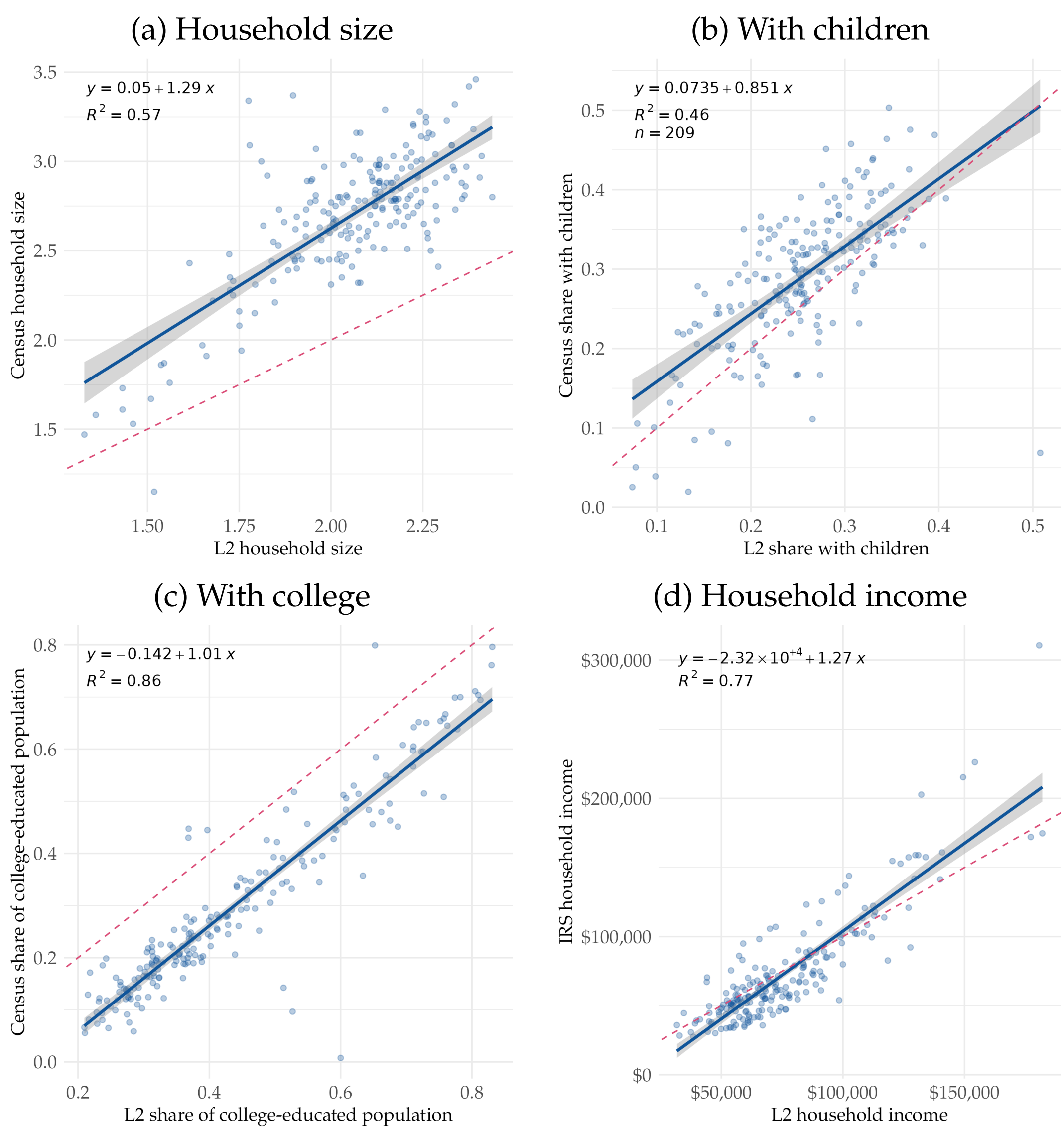
Figure A.30: L2 racial composition vs. 2019 ACS



Notes. Each dot represents a Census tract. The blue line is the linear fitted line between the L2 and ACS estimates. The dashed pink line is the 45 degree line.

Figure A.31 compares household composition, education, and income between the two datasets. L2 undercounts the number of individuals in households, likely because it only includes adults, thus excluding any children. Panel (b) compares the share of households with children. Although the numbers are noisier, this panel provides a better approximation of household size. Panel (c) shows the share of individuals with a college degree. L2 overestimates the level of college education in almost all neighborhoods, possibly because highly educated individuals are more likely to register to vote or because L2 systematically over-predicts education levels. However, the slope is roughly one, indicating that this discrepancy should not significantly affect demand estimation. Panel (d) presents household income, which is of fairly good quality.

Figure A.31: L2 socioeconomic characteristics vs. 2019 ACS



Notes. Each dot represents a Census tract. The blue line is the linear fitted line between the L2 and ACS estimates. The dashed pink line is the 45 degree line.

A.3.3 Link CoreLogic to L2

I build a panel dataset that tracks the residential location, property characteristics, and demographics of each household over time. The L2 dataset provides details information of each household and individual voters, including name, gender, age, race, household size, imputed education, imputed income, and residential address. To obtain the property characteristics of each household's housing choice, I merge L2 with detailed property features from CoreLogic's Assessor records using residential address.

The CoreLogic Assessor provides property-level records for single-family homes. However, it does not contain unit-level information for many multi-family properties, such as rental apartment and Co-op. Instead, the Assessor only contains one single record for the whole multi-family building or complex. In some other cases where unit-level records are available, such as condos and townhouses, over half of these properties have missing unit numbers. Therefore, I do not distinguish units within the same multi-family building.

To match L2 and CoreLogic using address, I first standardize the format of addresses in L2 and CoreLogic. Each address is parsed into house number, prefix direction, street name, designator, suffix direction, unit number, city, and five-digit zip code. I only keep unique addresses in both datasets to improve matching efficiency.

I use an iterative procedure to match addresses. I apply exact string match using the full address: house number, prefix direction, street name, designator, suffix direction, city, and five-digit zip code. Over 75% of unique addresses in L2 are matched to unique addresses in CoreLogic in the first step. For the rest of unmatched L2 addresses, I drop unit number and city name subsequently to do the merge.

Almost 90% of addresses in L2 are successfully matched to CoreLogic. The rest 10% of addresses cannot be matched, primarily due to misspelled or missing house number (for example, some L2 addresses are in the middle of a road), missing street names in L2, misspelled street suffix direction (for example, NW vs. SW), non-residential addresses in L2 (such as PO Box), and missing property records in CoreLogic. Theoretically, I can use a fuzzy string match for the unmatched addresses, but this method will likely lead to incorrect matches. Moreover, most of the unmatched addresses are in multi-family buildings. So I take the conservative approach to stop the matching process here.

After matching each address in L2 to address in CoreLogic, I find the property characteristics of each individual's residence from CoreLogic Assessor records. I also match each property with its predicted price and actual sale price if it was sold. Most addresses in CoreLogic are uniquely identified by the proprietary property identifier created by CoreLogic, which is equivalent to the combination of county and parcel number (APN). About 1% of addresses are matched to several property identifiers, or in other words, matched to multiple parcels. This happens when there exist parcels owned by local government or Homeowner Associations at the address, and when addresses are misspelled (For example, two properties share the same street addresses except for their street designators. One address misspells its street designator as "RD" but its true value should be "PL"). For the first case, I drop non-residential parcels at the address. For the second case, I randomly choose one parcel matched to the address.

A.3.4 Counterfactual dataset

I describe how I create the counterfactual dataset with the 2010 income distribution used in 1.6.1. Table A.11 provides an example of the calculation. First, I group households by their non-income demographics, including race, education, age, and whether they have children (column 2 to 5). For each group, I calculate the share of households within each income bracket based on the 2010 ACS data (column 7). The sum of income shares (column 7) equals to 1 for each household group. Next, I calculate the total number of households in each non-income demographic group using the 2019 baseline data (column 8). Multiplying columns 7 and 8 gives the simulated number of households within each demographic group fitted to the 2010 income distribution (column 9). This approach allows me to only update the income distribution while keeping other demographic distributions unchanged.

Table A.11: Counterfactual data example with 2010 income distribution

(1) Group	(2) Black	(3) College	(4) With Child	(5) Age	(6) Income	(7) Income Share in 2010	(8) Total num of households in 2019	(9) Simulated num with 2010 inc dist
1	1	1	1	35-50	<50k	0.2	4000	800
1	1	1	1	35-50	50-75k	0.3	4000	1200
1	1	1	1	35-50	75-100k	0.1	4000	400
1	1	1	1	35-50	100-125k	0.1	4000	400
1	1	1	1	35-50	...	...	4000	...
2	0	1	1	35-50	<50k	0.3	5000	1500
2	0	1	1	35-50	50-75k	0.3	5000	1500
2	0	1	1	35-50	75-100k	0.2	5000	1000
2	0	1	1	35-50	100-125k	0.1	5000	500
2	0	1	1	35-50	...	...	5000	...
...	...	...	...	...	...	...	...	...

A.4 EQUILIBRIUM ALGORITHM

The equilibrium aims to find the price vector that equalizes demand and supply in every segment. Given preference parameters, cost parameters, and exogenous housing and parcel characteristics, the price vector solves the system of equations:

$$\mathcal{D}_{n,k}(\mathbf{p}; \mathbf{X}, \boldsymbol{\xi}, \mathbf{Z}_{\boldsymbol{i}}, \alpha^0, \boldsymbol{\beta}^0, \alpha^r, \boldsymbol{\beta}^r) = \mathcal{S}_{n,k}(\mathbf{p}; \mathbf{L}, \mathbf{W}, \gamma, \eta, \rho, \tau), \forall n \in \mathcal{N}, k \in \mathcal{K}.$$

One method is to directly solve the system of nonlinear equations by minimizing the difference between demand and supply,

$$\min_{\mathbf{p}} \|\mathcal{D}_{n,k}(\mathbf{p}) - \mathcal{S}_{n,k}(\mathbf{p})\|.$$

An alternative method is to find a fixed point using the equilibrium condition, and it is way much faster than the first method. The fixed point algorithm searches for the new price vector that matches predicted demand and supply in each segment. In each iteration, I calculate the predicted demand and supply at given prices. If the predicted demand are lower than the predicted supply, it indicates that prices need to be lower than the current guess in order the increase the predicted demand. The iteration keeps going until prices converge and there is no excess demand and supply. Mathematically,

$$\mathbf{p}' = \mathbf{p} - \sigma \cdot \left(\ln \hat{\mathcal{S}}_{n,k}(\mathbf{p}) - \ln \hat{\mathcal{D}}_{n,k}(\mathbf{p}) \right),$$

where σ is the tuning parameter that determines the step size of each price update. I use observed prices as the initial value when calculating the new equilibrium.

The equilibrium condition only requires that prices clear the market in every segment. However, one can define another set of equilibrium conditions by treating demographic composition in each neighborhood as endogenous and including them in the equilibrium. Under this alternative equilibrium definition, multiple equilibria may arise because there may exist different demographic distributions across neighborhoods satisfying the equilibrium. For example, if two types of households have strong sorting preferences for living near others of the same type and there are two neighborhoods with identical fixed attributes, it becomes uncertain how each type sort into the neighborhoods. Bayer & Timmins (2005) find that the uniqueness of equilibrium is more likely to be guaranteed if the agent faces a large number of choices, there is substantial variation among these choices, and households have heterogeneous preferences across choices. Even if I adopt this equilibrium definition, a unique equilibrium is likely to occur because my model incorporates all these three features.

While my model incorporates local interactions, where each household's decision is influenced by the residential choices of others and the resulting sociodemographic composition of the neighborhood, I do not update the demographic composition when calculating a new equilibrium under counterfactual scenarios. I assume that each household is too small to reasonably predict the final neighborhood composition after considering the aggregate decisions of all other households in response to a policy change, so I exclude demographics as an equilibrium condition.

APPENDIX B

Appendix to Chapter 2

B.1 ADDITIONAL FIGURES AND TABLES

Table B.1: Hedonic regressions

	(1)	(2)
Black seller	-0.0406***	
Black-only seller		-0.0462***
Observations	3,464,101	3,464,101
R^2	0.910	0.910

Notes: Regressions include control for building square footage and lot area, age, heating and cooling systems, pools, garages, and categorical dummies for number of bedrooms and bathrooms. Census tract $\times$ year-quarter of sale date fixed effects are included as well. We exclude transactions with sale values less than $10,000 and above the 99th percentile of sales values. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Table B.2: Seller and property characteristics: Any delinquency

	(1)	(2)	(3)	(4)	(5)	(6)
Black seller	-0.0032***	-0.0029***	-0.0029***			
	(0.0003)	(0.0003)	(0.0003)			
Black-only seller				-0.0032***	-0.0029***	-0.0029***
				(0.0003)	(0.0003)	(0.0003)
Delinquent 30d+		-0.0046***	-0.0046***		-0.0046***	-0.0046***
		(0.0002)	(0.0002)		(0.0002)	(0.0002)
Roof			0.0031***			0.0031***
			(0.0005)			(0.0005)
Replacement			0.0028***			0.0028***
			(0.0004)			(0.0004)
Remodel			0.0114***			0.0114***
			(0.0008)			(0.0008)
Observations	593,895	593,895	593,895	593,895	593,895	593,895
R^2	0.723	0.723	0.723	0.723	0.723	0.723

Notes: This table presents the estimates of the racial gap in annualized returns when we control for seller's mortgage delinquency status and property renovations. Sellers are identified as delinquent if they have any delinquency in the year prior to selling their home. All regressions follow the same specification in Equation 2.1. All regressions include the Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. The first three columns show the estimates for Black sellers, while the last three columns show the estimates for Black-only sellers. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Table B.3: Seller and property characteristics: Last month 90 day+ delinquency

	(1)	(2)	(3)	(4)	(5)	(6)
Black seller	-0.0032***	-0.0029***	-0.0029***			
	(0.0003)	(0.0003)	(0.0003)			
Black-only seller				-0.0032***	-0.0029***	-0.0029***
				(0.0003)	(0.0003)	(0.0003)
Delinquent 90d+		-0.0098***	-0.0098***		-0.0098***	-0.0098***
		(0.0003)	(0.0003)		(0.0003)	(0.0003)
Roof			0.0030***			0.0030***
			(0.0005)			(0.0005)
Replacement			0.0028***			0.0028***
			(0.0004)			(0.0004)
Remodel			0.0114***			0.0114***
			(0.0008)			(0.0008)
Observations	593,895	593,895	593,895	593,895	593,895	593,895
R^2	0.723	0.723	0.724	0.723	0.723	0.724

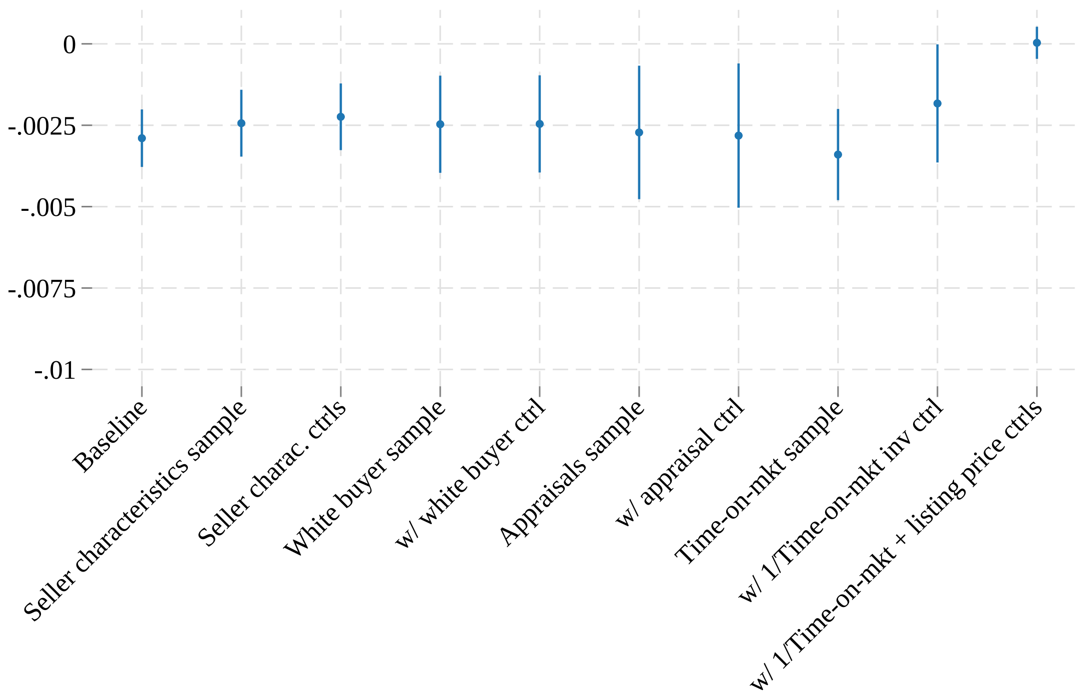
Notes: This table presents the estimates of the racial gap in annualized returns when we control for seller's mortgage delinquency status and property renovations. Sellers are identified as delinquent if they have a 90 day or longer delinquency in the month prior to selling their home. All regressions follow the same specification in Equation 2.1. All regressions include the Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. The first three columns show the estimates for Black sellers, while the last three columns show the estimates for Black-only sellers. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Table B.4: Seller and property characteristics: Last month any delinquency

	(1)	(2)	(3)	(4)	(5)	(6)
Black seller	-0.0032***	-0.0029***	-0.0029***			
	(0.0003)	(0.0003)	(0.0003)			
Black-only seller				-0.0032***	-0.0029***	-0.0029***
				(0.0003)	(0.0003)	(0.0003)
Delinquent last month		-0.0073***	-0.0073***		-0.0073***	-0.0073***
		(0.0003)	(0.0003)		(0.0003)	(0.0003)
Roof			0.0031***			0.0031***
			(0.0005)			(0.0005)
Replacement			0.0028***			0.0028***
			(0.0004)			(0.0004)
Remodel			0.0114***			0.0114***
			(0.0008)			(0.0008)
Observations	593,895	593,895	593,895	593,895	593,895	593,895
R^2	0.723	0.723	0.724	0.723	0.723	0.724

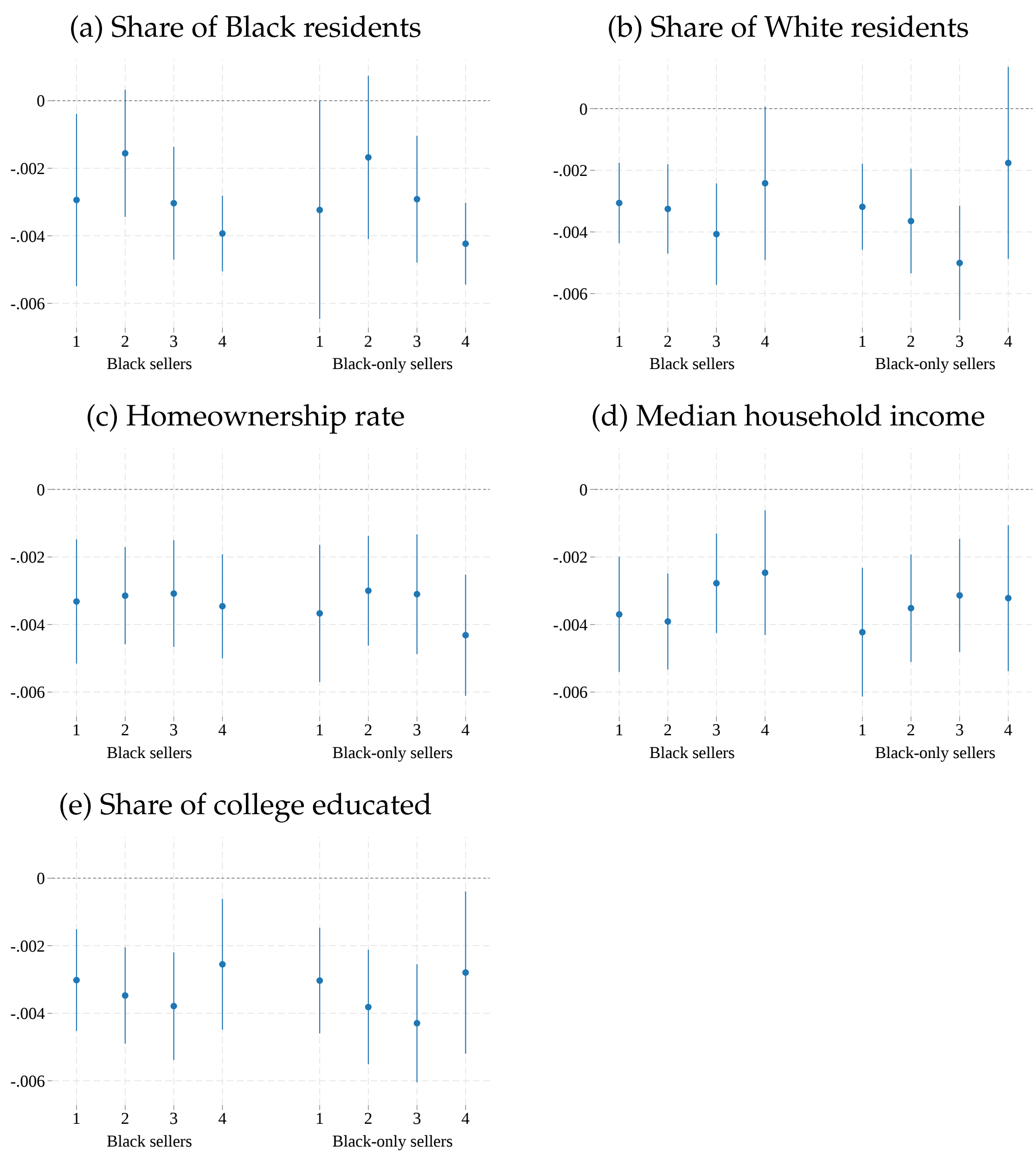
Notes: This table presents the estimates of the racial gap in annualized returns when we control for seller's mortgage delinquency status and property renovations. Sellers are identified as delinquent if they have any delinquency in the month prior to selling their home. All regressions follow the same specification in Equation 2.1. All regressions include the Census tract-level annualized mean house price growth rate between the purchase year and sale year, as well as the Census tract by sale year-quarter fixed effects. The first three columns show the estimates for Black sellers, while the last three columns show the estimates for Black-only sellers. Standard errors are clustered at the Census tract levels. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Figure B.1: Regressions with additional short-sale restriction



Notes: This figure presents the annualized housing returns gap with a variety of controls. The specification of the first plot corresponds to that in Table 2.1, the specification of the second and third plots correspond to those in Table 2.2, the specification of the fourth and fifth plots correspond to those in Table 2.3, the specification in the sixth and seventh plots correspond to those in Figure 2.1, and the specification in the eighth, ninth, and tenth plots correspond to those in Table 2.7. Standard errors are clustered at the tract level with 95% confidence intervals displayed.

Figure B.2: Annualized housing return gaps by socioeconomic characteristics



Notes: These figures present the annualized housing returns in Census tracts with different sociodemographic characteristics. Census tracts are grouped into quartiles based on their characteristics indicated by the title for each panel. All regressions include Census tract-level annualized mean house price index growth between purchase year and sale year and Census tract × year-quarter fixed effects. Standard errors are clustered at the tract level. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

BIBLIOGRAPHY

Allen, M. T., Rutherford, J., Rutherford, R., & Yavas, A. (2018). Impact of investors in distressed housing markets. *The Journal of Real Estate Finance and Economics*, *56*, 622–652.

Almagro, M., Chyn, E., & Stuart, B. A. (2023). The Welfare and Distributional Consequences of Neighborhood Change: Evidence from Chicago's Public Housing Demolitions. *Working paper*.

Anagol, S., Ferreira, F., & Rexer, J. (2023). Estimating the Economic Value of Zoning Reform. *Working paper*.

Anenberg, E., & Kung, E. (2014). Estimates of the size and source of price declines due to nearby foreclosures. *American Economic Review*, *104*(8), 2527–2551.

Anenberg, E., & Kung, E. (2020). Can more housing supply solve the affordability crisis? Evidence from a neighborhood choice model. *Regional Science and Urban Economics*, *80*, 103363.

Argyle, B., Indarte, S., Iverson, B., & Palmer, C. (2023). Explaining Racial Disparities in Personal Bankruptcy Outcomes.

Asquith, B. J., Mast, E., & Reed, D. (2023). Local Effects of Large New Apartment Buildings in Low-Income Areas. *The Review of Economics and Statistics*, *105*(2), 359–375.

Austin, N. (2022). Keeping up with the blackstones: Institutional investors and gentrification. *Available at SSRN 4269561*.

Autor, D. H., Palmer, C. J., & Pathak, P. A. (2014). Housing market spillovers: Evidence from the end of rent control in cambridge, massachusetts. *Journal of Political Economy*, *122*(3), 661–717.

Avenancio-León, C. F., & Howard, T. (2022). The Assessment Gap: Racial Inequalities in Property Taxation. *The Quarterly Journal of Economics*, *137*(3), 1383–1434.

Babalievsky, F., Herkenhoff, K. F., Ohanian, L. E., & Prescott, E. (2023). The Impact of Commercial Real Estate Regulations on U.S. Output. *Working paper*.

Bartik, A., Gupta, A., & Milo, D. (2023). The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement. *SSRN Electronic Journal*.

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech Era. *Journal of Financial Economics*, *143*(1), 30–56.

Baum-Snow, N., & Han, L. (2024). The Microgeography of Housing Supply. *Journal of Political Economy*, (pp. 000–000).

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

Bayer, P., Ferreira, F., & McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, *115*(4), 588–638.

Bayer, P., Ferreira, F., & Ross, S. L. (2016). The Vulnerability of Minority Homeowners in the Housing Boom and Bust. *American Economic Journal: Economic Policy*, *8*(1), 1–27.

Bayer, P., & Timmins, C. (2005). On the equilibrium properties of locational sorting models. *Journal of Urban Economics*, *57*(3), 462–477.

Berger, D., Turner, N., & Zwick, E. (2020). Stimulating Housing Markets. *The Journal of Finance*, *75*(1), 277–321.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*.

Bhutta, N., Hizmo, A., & Ringo, D. (2022). How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions. *Finance and Economics Discussion Series*, (2022-067), 1–44.

Bratu, C., Harjunen, O., & Saarimaa, T. (2023). JUE Insight: City-wide effects of new housing supply: Evidence from moving chains. *Journal of Urban Economics*, *133*, 103528.

Broadnax, C., & Wylie, J. (2021). Reducing Valuation Bias by Addressing Appraiser and Property Valuation Commentary.

Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2022). Why is Intermediating Houses So Difficult? Evidence from iBuyers.

Büchler, S., & Lutz, E. (2024). Making housing affordable? The local effects of relaxing land-use regulation. *Journal of Urban Economics*, *143*, 103689.

Butler, A. W., Mayer, E. J., & Weston, J. P. (2022). Racial Disparities in the Auto Loan Market. *The Review of Financial Studies*, *36*(1), 1–41.

Calder-Wang, S. (2021). The Distributional Impact of the Sharing Economy: Evidence from New York City. *Working paper*.

Campbell, J. Y., Giglio, S., & Pathak, P. (2011). Forced sales and house prices. *American Economic Review*, *101*(5), 2108–2131.

Chilton, K., Silverman, R. M., Chaudhry, R., & Wang, C. (2018). The impact of single-family rental reits on regional housing markets: A case study of nashville, tn. *Societies*, *8*(4), 93.

Christensen, P., Sarmiento-Barbieri, I., & Timmins (2021). Racial Discrimination in the United States Rental Market. *NBER Working Paper*, (29516).

Christensen, P., & Timmins, C. (2022). Sorting or Steering: The Effects of Housing Discrimination on Neighborhood Choice. *Journal of Political Economy*, *130*(8), 2110–2163.

Christensen, P., & Timmins, C. (2023). The Damages and Distortions from Discrimination in the Rental Housing Market.

Christophers, B. (2023). How and why us single-family housing became an investor asset class. *Journal of Urban History*, *49*(2), 430–449.

Colburn, G., Walter, R. J., & Pfeiffer, D. (2021). Capitalizing on collapse: An analysis of institutional single-family rental investors. *Urban Affairs Review*, *57*(6), 1590–1625.

Combes, P.-P., Duranton, G., & Gobillon, L. (2021). The Production Function for Housing: Evidence from France. *Journal of Political Economy*, *129*(10).

Cook, C., Li, P. Z., & Binder, A. J. (2024). Where to Build Affordable Housing? Evaluating the Tradeoffs of Location.

Couillard, S., & Christensen, P. (2024). Racial Housing Price Differentials and Neighborhood Segregation. *NBER Working Paper*, (32815).

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

Currie, J., Davis, L., Greenstone, M., & Walker, R. (2015). Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *American Economic Review*, *105*(2), 678–709.

Damiano, A., & Frenier, C. (2020). Build Baby Build?: Housing Submarkets and the Effects of New Construction on Existing Rents. *Working paper*.

Davidoff, T., Pavlov, A., & Somerville, T. (2022). Not in my neighbour's back yard? Laneway homes and neighbours' property values. *Journal of Urban Economics*, *128*, 103405.

Davis, M. A., Larson, W. D., Oliner, S. D., & Shui, J. (2021). The price of residential land for counties, ZIP codes, and census tracts in the United States. *Journal of Monetary Economics*, *118*, 413–431.

Diamond, R., & Diamond, W. F. (2024). Racial Differences in the Total Rate of Return on Owner-Occupied Housing. *NBER Working Paper*, (32916).

Diamond, R., McQuade, T., & Huang, Z. (2024). The Unequal Effects of Up-Zoning: Evidence from Cook County. *Working paper*.

Dobbels, G., & Tavakalov, S. (2023). Not in My Back Yard: The Local Political Economy of Residential Land-Use Regulations. *Unpublished*.

DLima, W., & Schultz, P. (2022). Buy-to-rent investors and the market for single family homes. *The Journal of Real Estate Finance and Economics*, (pp. 1–37).

Epple, D., Gordon, B., & Sieg, H. (2010). A New Approach to Estimating the Production Function for Housing. *American Economic Review*, *100*(3), 905–924.

Eriksen, M. D., & Rosenthal, S. S. (2010). Crowd out effects of place-based subsidized rental housing: New evidence from the LIHTC program. *Journal of Public Economics*, *94*(11-12), 953–966.

Eriksen, M. D., & Ross, A. (2015). Housing Vouchers and the Price of Rental Housing. *American Economic Journal: Economic Policy*, *7*(3), 154–176.

Favilukis, J., Mabille, P., & Van Nieuwerburgh, S. (2023). Affordable Housing and City Welfare. *The Review of Economic Studies*, *90*(1), 293–330.

Ferreira, F., & Gyourko, J. (2015). A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012.

Fields, D., & Vergerio, M. (2022). Corporate landlords and market power: What does the single-family rental boom mean for our housing future?

Fisher, L. M., Lambie-Hanson, L., & Willen, P. (2015). The role of proximity in foreclosure externalities: Evidence from condominiums. *American Economic Journal: Economic Policy*, *7*(1), 119–140.

Frame, W. S., Huang, R., Mayer, E. J., & Sunderam, A. (2021). The Impact of Minority Representation at Mortgage Lenders.

Freemark, Y. (2020). Upzoning Chicago: Impacts of a Zoning Reform on Property Values and Housing Construction. *Urban Affairs Review*, *56*(3), 758–789.

Galster, G., & Godfrey, E. (2005). By Words and Deeds: Racial Steering by Real Estate Agents in the U.S. in 2000. *Journal of the American Planning Association*, *71*(3), 251–268.

Gerardi, K., Rosenblatt, E., Willen, P. S., & Yao, V. (2015). Foreclosure externalities: New evidence. *Journal of Urban Economics*, *87*, 42–56.

Gerardi, K., Willen, P., & Zhang, D. H. (2023). Mortgage prepayment, race, and monetary policy. *Journal of Financial Economics*, *147*(3), 498–524.

Gilbukh, S., & Goldsmith-Pinkham, P. (2023). Heterogeneous Real Estate Agents and the Housing Cycle.

Glaeser, E. L., & Gyourko, J. (2003). The Impact of Building Restrictions on Housing Affordability. *FRBNY Economic Policy Review*.

Glaeser, E. L., Gyourko, J., & Saks, R. (2005). Why Is Manhattan So Expensive? Regulation and the Rise in Housing Prices. *The Journal of Law and Economics*, *48*(2), 331–369.

Goldsmith-Pinkham, P., & Shue, K. (2023). The gender gap in housing returns. *The Journal of Finance*, *78*, 1097–1145.

Greenaway-McGrevy, R., & Phillips, P. C. (2023). The impact of upzoning on housing construction in Auckland. *Journal of Urban Economics*, *136*, 103555.

Greenwald, D. L., & Guren, A. (2021). Do Credit Conditions Move House Prices? *Working paper*.

Gurun, U. G., Wu, J., Xiao, S. C., & Xiao, S. W. (2023). Do wall street landlords undermine renters welfare? *The Review of Financial Studies*, *36*(1), 70–121.

Gyourko, J., & McCulloch, S. (2023). Minimum Lot Size Restrictions: Impacts on Urban Form and House Price at the Border. *NBER Working Paper*, (w31710).

Gyourko, J., & McCulloch, S. (2024). The Distaste for Housing Density. *NBER Working Paper*.

Howard, G., & Liebersohn, J. (2021). Why is the rent so darn high? The role of growing demand to live in housing-supply-inelastic cities. *Journal of Urban Economics*, *124*, 103369.

Ihlanfeldt, K., & Yang, C. F. (2021). Not in my neighborhood: The effects of single-family rentals on home values. *Journal of Housing Economics*, *54*, 101789.

Issler, P., Stanton, R., Vergara-Alert, C., & Wallace, N. (2023). Housing and Mortgage Markets with Climate-Change Risk: Evidence from Wildres in California. *Working paper*.

Kermani, A., & Wong, F. (2024). Racial Disparities in Housing Returns.

Krimmel, J., & Wang, B. (2023). Upzoning with Strings Attached: Evidence from Seattle's Affordable Housing Mandate. *NYU Furman Center*.

Lambie-Hanson, L., Li, W., & Slonkosky, M. (2019). Institutional investors and the us housing recovery.

Landvoigt, T., Piazzesi, M., & Schneider, M. (2015). The Housing Market (s) of San Diego. *American Economic Review*, *105*(4), 1371–1407.

Laouénan, M., & Rathelot, R. (2022). Can Information Reduce Ethnic Discrimination? Evidence from Airbnb. *American Economic Journal: Applied Economics*, *14*(1), 107–132.

Li, X. (2021). Do new housing units in your backyard raise your rents? *Journal of Economic Geography*, *22*.

Liu, L., McManus, D., & Yannopoulos, E. (2022). Geographic and temporal variation in housing filtering rates. *Regional Science and Urban Economics*, *93*, 103758.

Mac, F. (2021). Racial and Ethnic Valuation Gaps In Home Purchase Appraisals. *Freddie Mac Economics & Housing Research Note*.

Manson, S., Schroeder, J., Riper, D. V., Knowles, K., Kugler, T., Roberts, F., & Ruggles, S. (2023). IPUMS National Historical Geographic Information System: Version 18.0 Database. http://doi.org/10.18128/D050.V18.0.

Mast, E. (2023). JUE Insight: The effect of new market-rate housing construction on the low-income housing market. *Journal of Urban Economics*, *133*, 103383.

McFadden, D. (1978). Modeling the Choice of Residential Location.

Mei, M. (2022). House Size and Household Size: The Distributional Effects of the Minimum Lot Size Regulation. *Working paper*.

Mills, J., Molloy, R., & Zarutskie, R. (2019). Large-scale buy-to-rent investors in the single-family housing market: The emergence of a new asset class. *Real Estate Economics*, *47*(2), 399–430.

Murphy, A. (2018). A Dynamic Model of Housing Supply. *American Economic Journal: Economic Policy*, *10*(4), 243–267.

Nathanson, C. G. (2023). Trickle-down housing economics. *Working paper*.

Ondrich, J., Stricker, A., & Yinger, J. (1998). Do Real Estate Brokers Choose to Discriminate? Evidence from the 1989 Housing Discrimination Study. *Southern Economic Journal*, *64*(4), 880–901.

Oosthuizen, D. (2023). Institutional housing investors and the great recession.

OpenAI (2025). Chatgpt. https://chat.openai.com/. Assisted with proofreading. Not generating content or analysis.

Pennington, K. (2021). Does Building New Housing Cause Displacement?: The Supply and Demand Effects of Construction in San Francisco. *SSRN Electronic Journal*.

Piazzesi, M., Schneider, M., & Stroebel, J. (2020). Segmented Housing Search. *American Economic Review*, *110*(3), 720–759.

Romo, V. (2023). Black couple settles lawsuit claiming their home appraisal was lowballed due to bias. *NPR*.

Rosenthal, S. S. (2014). Are Private Markets and Filtering a Viable Source of Low-Income Housing? Estimates from a "Repeat Income" Model. *American Economic Review*, *104*(2), 687–706.

Schubert, G. (2016). House Price Contagion and U.S. City Migration Networks. *Working paper*.

Sinai, T., & Waldfogel, J. (2005). Do low-income housing subsidies increase the occupied housing stock? *Journal of Public Economics*, *89*(11-12), 2137–2164.

Soltas, E. (2024). Tax Incentives and the Supply of Low-Income Housing.

Song, J. (2021). The Effects of Residential Zoning in U.S. Housing Markets. *SSRN Electronic Journal*.

Susin, S. (2002). Rent vouchers and the price of low-income housing. *Journal of Public Economics*, *83*(1), 109–152.

Sweeney, J. L. (1974). Quality, Commodity Hierarchies, and Housing Markets. *Econometrica*, *42*(1), 147.

Tokman, A. E. (2023). Density Restrictions and Housing Inequality. *Unpublished*.

Turner, M. A., Santos, R., Levy, D. K., Wissoker, D., Aranda, C., Pitingolo, R., & Institute, T. U. (2013). Housing Discrimination Against Racial and Ethnic Minorities 2012. *U.S. Department of Housing and Urban Development, Office of Policy Development and Research*.

Vajja, A. (2023). Have Racial and Ethnic Valuation Gaps in Home Purchase Narrowed? *FHFA Insights*.

Yinger, J. (1998). Evidence on Discrimination in Consumer Markets. *Journal of Economic Perspectives*, *12*, 23–40.

Zhang, D., & Willen, P. (2020). Do Lenders Still Discriminate? A Robust Approach for Assessing Differences in Menus. Federal Reserve Bank of Boston Research Department Working Papers, Federal Reserve Bank of Boston.

Zhao, B., Ondrich, J., & Yinger, J. (2006). Why do real estate brokers continue to discriminate? Evidence from the 2000 Housing Discrimination Study. *Journal of Urban Economics*, *59*(3), 394–419.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]