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

Lei Ma*

March 19, 2025

(Job Market Paper)

Click here for the latest version.

Abstract

New construction has favored bigger, more expensive houses. This paper studies the causes of this pattern and evaluates 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.

*Boston University, Department of Economics. Email: leim@bu.edu. This paper has greatly benefited from invaluable feedback and support from my dissertation committee: Adam Guren, Marc Rysman, Kevin Lang, and Jihye Jeon. I am also grateful for helpful discussions with Rubaiyat Alam, Ari Anisfeld, Sophie Calder-Wang, Leonel Drukker, Santiago Franco, Peter Ganong, Troup Howard, David Lagakos, Jetson Leder-Luis, Anran Li, Yuhei Miyauchi, Pascal Noel, Abby Ostriker, Christopher Palmer, Natalia Ramondo, Jordan Rosenthal-Kay, Zhongyi Tang, and Vickie Wang. Participants in the Boston University empirical micro workshop, the Urban Economics Association conference, Dallas Fed Women in Central Banking workshop, and GNLab conference provided many helpful comments. I am grateful to Katherine Levine Einstein and Maxwell Palmer for CoreLogic data access and to the Institute for Economic Development at Boston University for its financial support.

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, Larson, Oliner and Shui (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, Ferreira and McMillan (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 house-

holds' 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, Levinsohn and Pakes (1995) and Bayer, Ferreira and McMillan (2007). To estimate zoning regulatory costs, I apply the residual approach from Glaeser and 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, Mast and Reed (2023), and Damiano and 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, Ferreira and Rexer (2023); Song (2021)), public housing demolitions (Almagro, Chyn and Stuart (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 and 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, McQuade and Huang (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 and Gyourko (2003) and Glaeser, Gyourko and Saks (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 and 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, Turner and Zwick (2020) find that the temporary tax credit for new homebuyers from 2008 to 2010 boosted housing demand and increased prices. Eriksen and 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 and Waldfogel (2005), Eriksen and 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, Schneider and Stroebel (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 and 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 São Paulo, Brazil (Anagol, Ferreira and Rexer (2023)), affordability housing mandates in Seattle (Krimmel and Wang (2023)), the permission to build accessory dwelling units in Vancouver (Davidoff et al. (2022)), and upzoning in Auckland (Greenaway-McGrevy and Phillips (2023)) and Zurich (Büchler and 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 2 describes the data used in the analysis. Section 3 presents the empirical facts that motivate the research question and the design of the structural model. Section 4 outlines the setup of the structural model, while Section 5 explains the estimation procedure and presents the estimation results. Section 6 presents the findings of counterfactual exercises. Section 7 concludes.

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 Ap-

pendix 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.

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 C.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 C.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 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 hous-

[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.

[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.

ing 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 C.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, Larson, Oliner and Shui (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).

Zoning code and development standards I obtain parcel-level zoning jurisdiction and zoning code from CoreLogic's assessor records in 2019. In Georgia, both cities and counties have zoning authority. Typically, cities govern zoning within their city limits, while counties oversee zoning in unincorporated areas. To determine the zoning jurisdiction, I geocode each parcel's location and check if it falls within an incorporated area as of 2020, using the shape files of incorporated places from the Census Bureau.

Although CoreLogic assessor records include zoning codes at the parcel level, they do not provide definitions for these codes. Each zoning code corresponds to specific rules regarding permissible

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

use and development standards. Development standards mandate specific dimensional requirements and other criteria regarding the structure. These rules include detailed specifications such as minimum lot size, maximum density, maximum lot coverage, and minimum setbacks. I manually collect the development standards for each zoning code from MuniCode.com and municipal documentation. More details on the data collection are in Appendix C.4.

Final sample I merge property-level housing characteristics, housing values, household demographics, housing costs, zoning regulations, and neighborhood characteristics to build a dataset covering both housing demand and supply information 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 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 1996.

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 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 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:

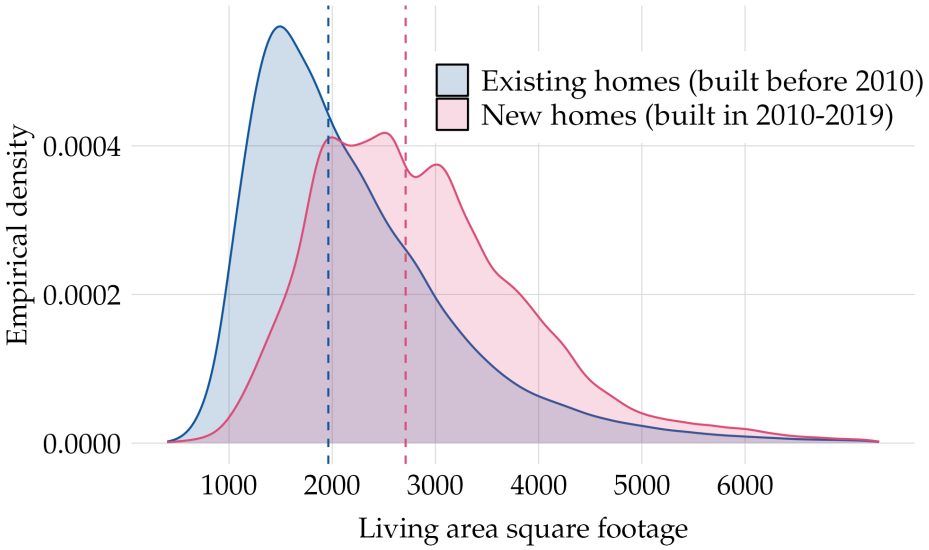
$$PropertyCharacteristics_j = \beta \cdot \mathbf{1}(BuiltAft2010_j) + \alpha_{zcta(j)} + \epsilon_j. \tag{1}$$

Table 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, Larson, Oliner and Shui (2021), with all values winsorized at the 1st and 99th percentiles.

Figure 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.

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

[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 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). 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.

ward 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

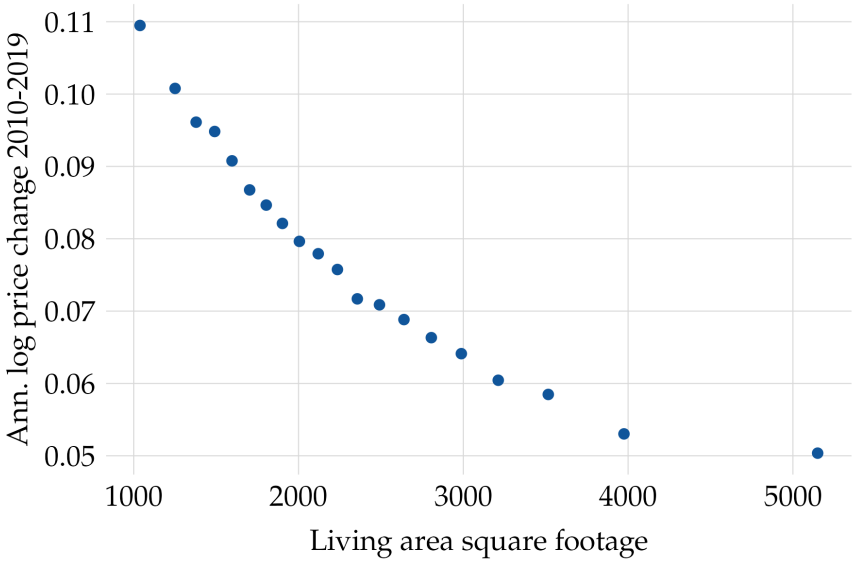
Next, I explore the geographic distribution of new homes and the demand factors associated with new construction activities. Panel (a) of Figure 3 presents a map of the number of new homes at the Census tract level. While new construction occurs in many neighborhoods, development is concentrated in suburban 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 characteristics separately:

$$\frac{NewHome_n}{TotalHomes_n} = \beta \cdot X_n^{2010} + \epsilon_n. \tag{2}$$

Panel (b) of Figure 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

[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 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.

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.

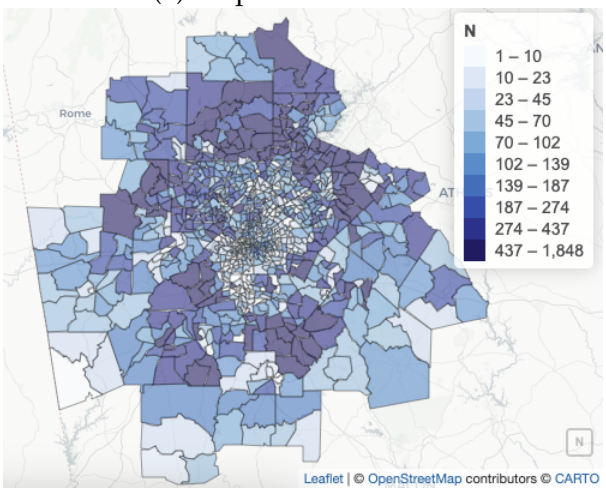
Fact 4: Larger homes are associated with more stringent zoning development standards

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, setbacks, and densities.[6] Therefore, I explore how various zoning development standards shape house sizes. I focus on three common standards

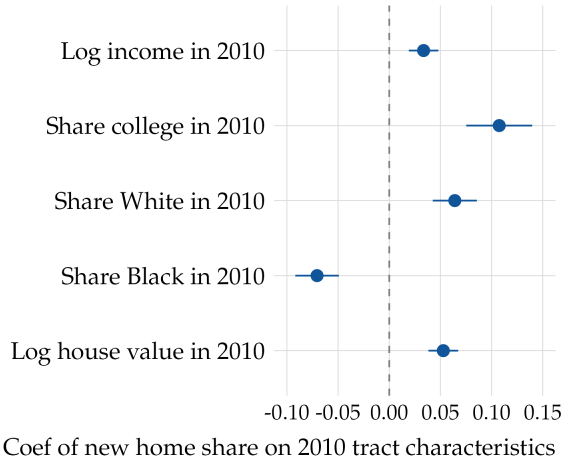
[6]Bartik, Gupta and Milo (2023) apply Large Language Models (LLM) to zoning documents to analyze the key features of zoning regulations. They find that density restrictions, particularly minimum lot size requirements, are prevalent and strongly correlate with exclusionary zoning.

Figure 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 (2). The horizontal bars represent the 95% confidence intervals calculated with robust standard errors.

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:

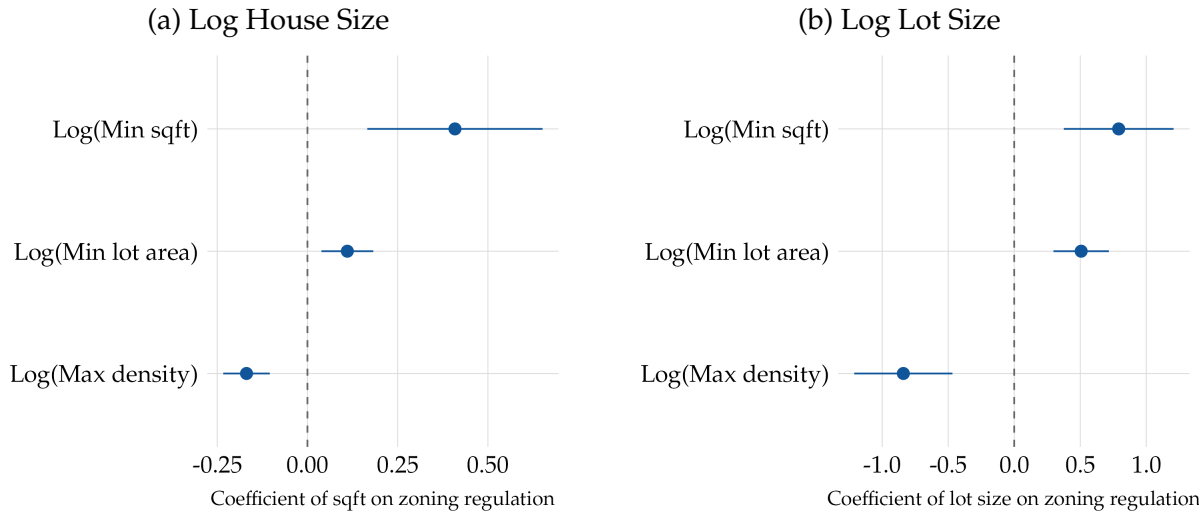
$$NewHomeCharacteristic_j = \beta \cdot Zoning_j + \alpha_m + \epsilon_j. \tag{3}$$

Panel (a) of Figure 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

[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 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 (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.

these heterogeneous zoning costs and assess their impact on housing supply.

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, Ferreira and McMillan (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.

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

$$u_{i,j,h} = \alpha_i p_h + \mathbf{X}_h \boldsymbol{\beta_i} + \xi_h + \epsilon_{i,j}. \tag{4}$$

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 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

[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.

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

$$\delta_h = \alpha^0 p_h + \mathbf{X}_h \boldsymbol{\beta^0} + \xi_h \tag{5}$$

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

$$\mu_{i,h} = \left(\sum_r \alpha^r z_i^r \right) p_h + \left(\sum_r \boldsymbol{\beta^r} z_i^r \right) \mathbf{X}_h \tag{6}$$

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:

$$\Pr(i \text{ chooses } h) \equiv d_{i,h}(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, \mathbf{Z_i}; \alpha^0, \boldsymbol{\beta}^0, \alpha^r, \boldsymbol{\beta}^r) = \frac{N_h \exp(\delta_h + \mu_{i,h})}{\sum_{h'} N_{h'} \exp(\delta_{h'} + \mu_{i,h'})}, \tag{7}$$

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

$$\mathcal{D}_h = \sum_i d_{i,h}(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, \mathbf{Z_i}; \alpha^0, \boldsymbol{\beta}^0, \alpha^r, \boldsymbol{\beta}^r), \tag{8}$$

which depends on the empirical distribution of demographics.

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

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 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 and 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

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.

[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 B.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).

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 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:

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

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

[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.

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]

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

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.

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 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

[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.

[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.

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 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.

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

$$LL(\delta, (\alpha^r, \boldsymbol{\beta}^r)_r) = \sum_i \sum_h y_{i,h} \ln d_{i,h} = \sum_i \sum_h y_{i,h} \ln \frac{N_h \exp(\delta_h + \mu_{i,h})}{\sum_{h'} N_{h'} \exp(\delta_{h'} + \mu_{i,h'})}, \tag{9}$$

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

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

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

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

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.

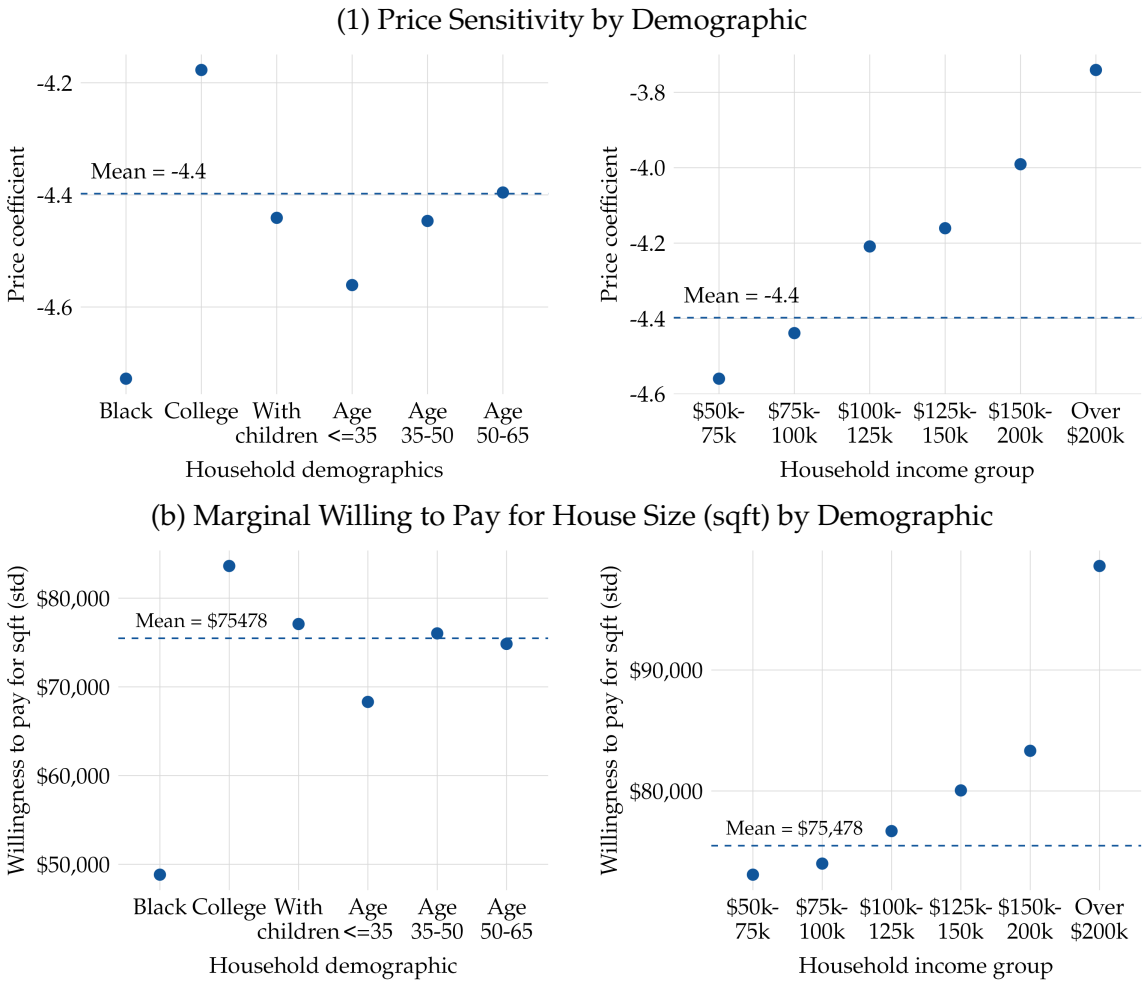
To address the potential endogeneity, I construct an instrument based on the methods of Berry, Levinsohn and Pakes (1995) and Bayer, Ferreira and McMillan (2007), using a set of immutable property and neighborhood characteristics from segments located at least five miles from the segment. Specifically, 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 (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 (5).

Estimates of demand parameters

Figure 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

Figure 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.

pay for one standard deviation increase in house size[19] is $75,478. In terms of neighborhood characteristics, households 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.

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

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 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, $Z_h^r = \left(\sum_i d_{ih} \times z_i^r\right) / \sum_i d_{ih}, \forall h$, 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.

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 likelihood 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.

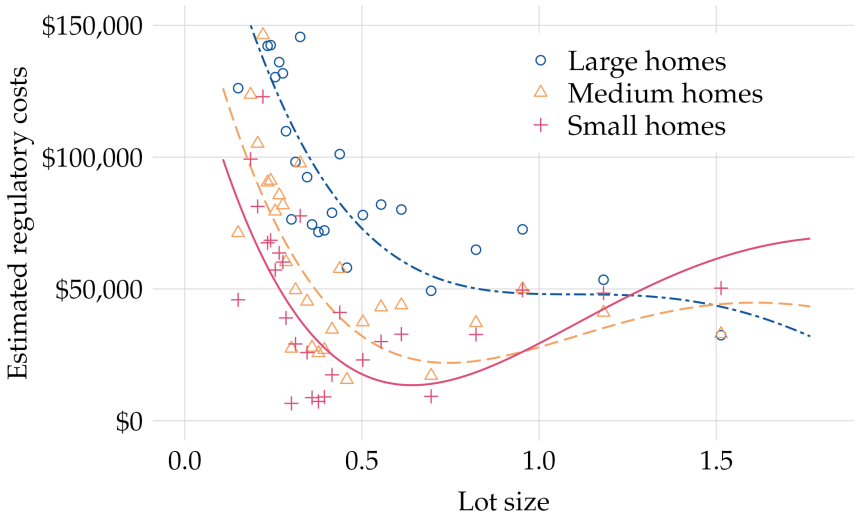
Supply estimation results

Figure 6 presents the estimated regulatory costs for each potential house type by lot size.[21] Zon-

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

[21] Appendix Table A.7 reports the underlying estimated parameters for regulatory costs. I standardize all variables

Figure 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.

ing 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 minorities, 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 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

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.

[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.

Table 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.

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.

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 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 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

[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.

Baum-Snow and 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.

6 Counterfactuals

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 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.

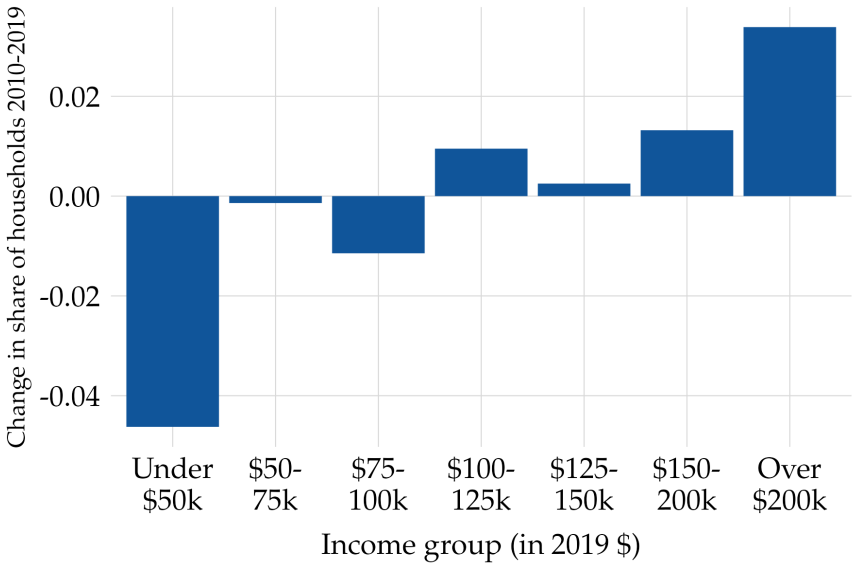
To assess the extent to which rising top income groups drive increases in house sizes, I simulate the model using the 2010 income distribution while keeping the total population and other demographic variables constant.[26] By holding the total population fixed, I can isolate the impact of

[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 and 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 C.5.

Figure 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.

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 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 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 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 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.

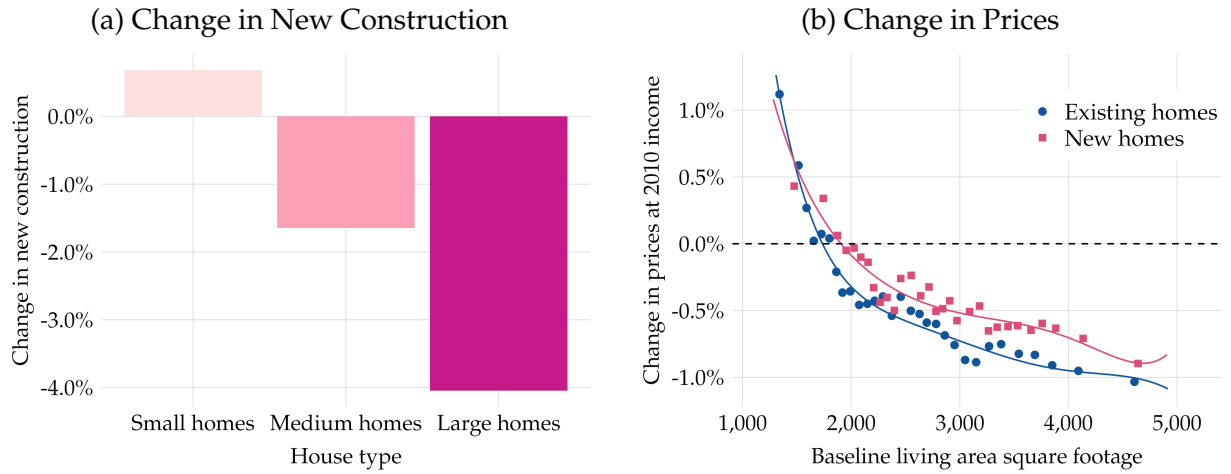
Role of zoning costs Next, I explore the effect of zoning costs on supply and prices. The supply estimation shows that zoning costs are substantial and differ by house type and parcel characteristics. Many studies have shown that zoning regulations primarily focus on density restriction, and minimum lot size requirement is one of the most common and restrictive rules for single-family homes (for example, Bartik, Gupta and Milo (2023), Gyourko and 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

[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.

[28] In Appendix B.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.

Figure 8: Simulated Equilibrium Under 2010 Income Distribution



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.

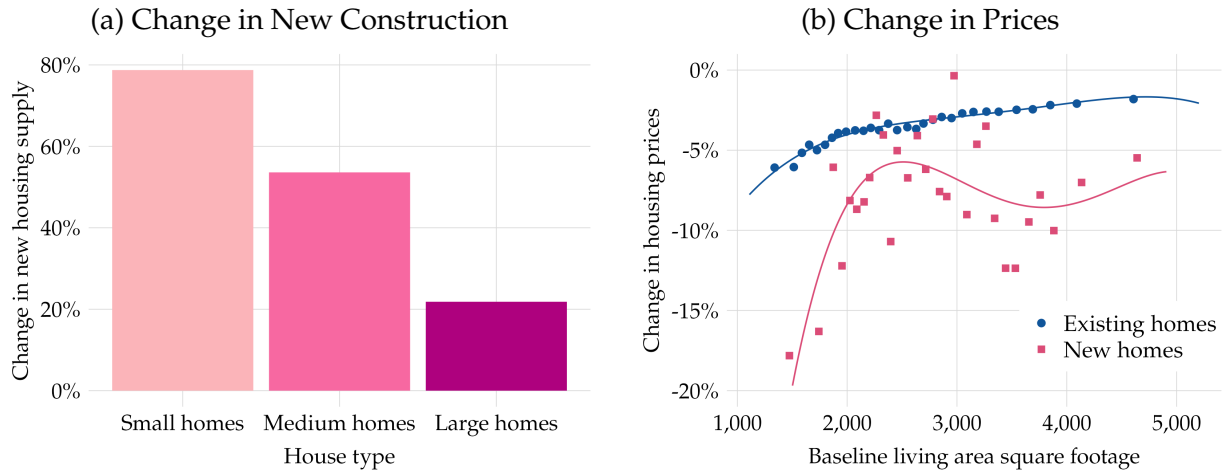
and 2019.[29] Mathematically, I split vacant parcels with sizes above the 25th percentile into two equal-sized parcels. This change nearly doubles the number of vacant parcels eligible for development, significantly increasing the potential housing supply at the extensive margin. I assume the total population 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 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 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

[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.

Figure 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.

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 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 restrictions, 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.

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 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

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

[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.

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 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 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.

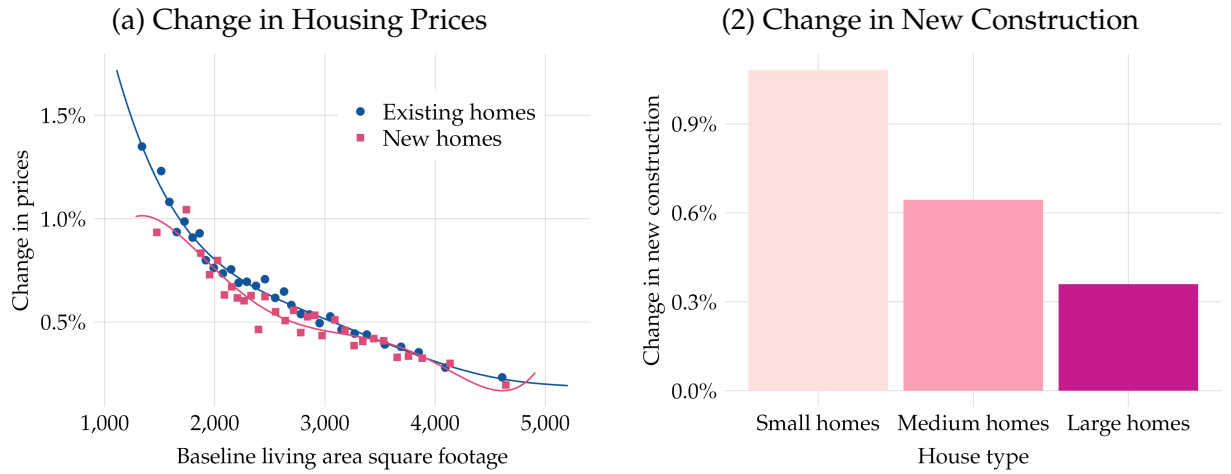
Why does the subsidy trigger such small equilibrium responses in housing markets? Under the counterfactual scenario, only 7.2% (96,962) of households are eligible for the subsidy, representing a small share of the total market. Moreover, these subsidized households mainly compete with older, unsubsidized 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]

[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.

[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 scenario 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 B.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 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.

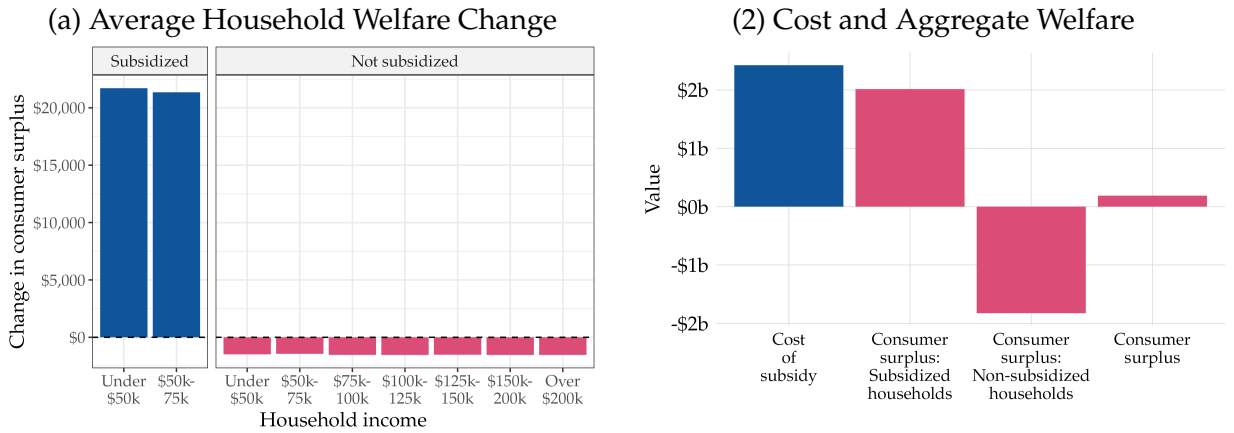
Next, I evaluate the distributional consequences of the subsidy on household welfare. For each household i, I measure the change in consumer surplus, which is the change in inclusive value normalized by the price coefficient,

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

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.

Figure 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 households, 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.

Figure 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.

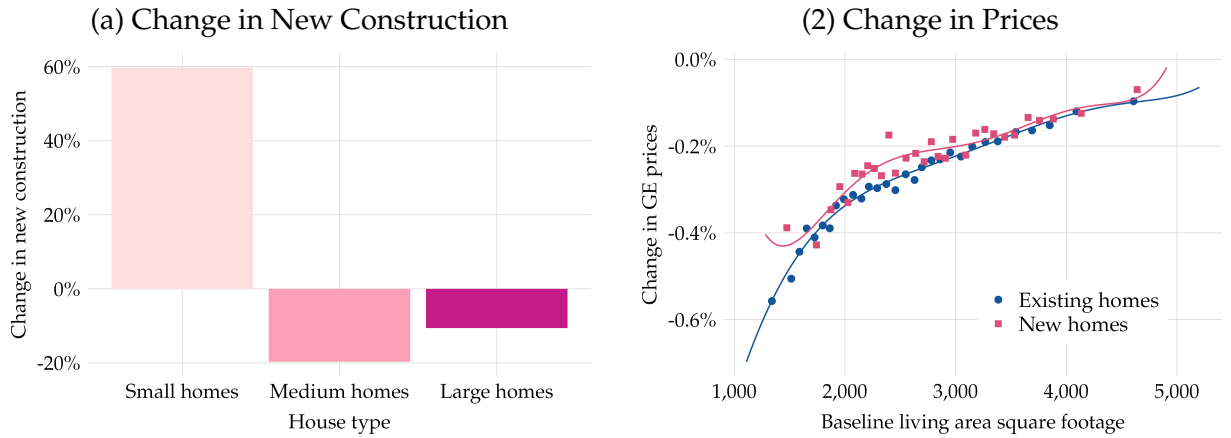
Subsidizing developers for small development In addition to demand-side subsidies, many housing advocates and policymakers are pushing for supply-side subsidies. While most existing government supply-side programs focus on housing for households at the lowest end of the income distribution, there is 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 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, representing a 9.31% growth in new construction

[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.

Figure 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.

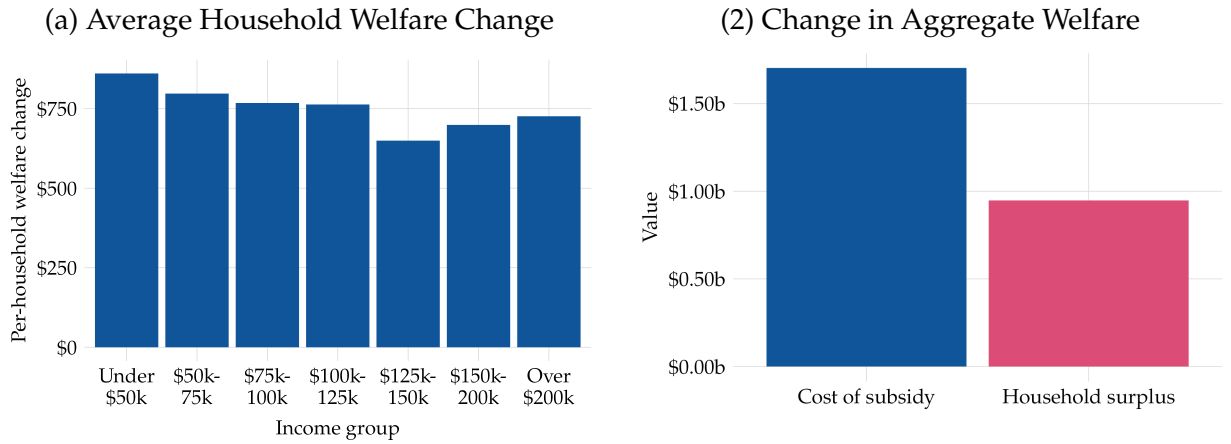
relative to the baseline. The results show that developers are responsive to profit incentives. As small home construction becomes more profitable, they adjust their development decision at the intensive margin and expand development at the extensive margin to capture the gains. However, in terms of 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.

Figure 13 presents the household welfare change with the supply-side subsidy. Panel (a) shows

[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 13: Impact of Supply-Side Subsidy on 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.

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.

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.

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.

References

Almagro, Milena, Eric Chyn, and Bryan A Stuart, "The Welfare and Distributional Consequences of Neighborhood Change: Evidence from Chicago's Public Housing Demolitions," *Working paper*, 2023.

Anagol, Santosh, Fernando Ferreira, and Jonah Rexer, "Estimating the Economic Value of Zoning Reform," *Working paper*, 2023.

Anenberg, Elliot and Edward Kung, "Can More Housing Supply Solve the Affordability Crisis? Evidence from a Neighborhood Choice Model," *Regional Science and Urban Economics*, January 2020, *80*, 103363.

Asquith, Brian J., Evan Mast, and Davin Reed, "Local Effects of Large New Apartment Buildings in Low-Income Areas," *The Review of Economics and Statistics*, March 2023, *105* (2), 359–375.

Babalievsky, Fil, Kyle F Herkenhoff, Lee E Ohanian, and Edward Prescott, "The Impact of Commercial Real Estate Regulations on U.S. Output," *Working paper*, 2023.

Bartik, Alexander, Arpit Gupta, and Daniel Milo, "The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement," *SSRN Electronic Journal*, 2023.

Baum-Snow, Nathaniel and Lu Han, "The Microgeography of Housing Supply," *Journal of Political Economy*, April 2024, pp. 000–000.

Bayer, Patrick and Christopher Timmins, "On the Equilibrium Properties of Locational Sorting Models," *Journal of Urban Economics*, May 2005, *57* (3), 462–477.

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

Berger, David, Nicholas Turner, and Eric Zwick, "Stimulating Housing Markets," *The Journal of Finance*, February 2020, *75* (1), 277–321.

Berry, Stephen, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995.

Bratu, Cristina, Oskari Harjunen, and Tuukka Saarimaa, "JUE Insight: City-wide Effects of New Housing Supply: Evidence from Moving Chains," *Journal of Urban Economics*, January 2023, *133*, 103528.

Büchler, Simon and Elena Lutz, "Making Housing Affordable? The Local Effects of Relaxing Land-Use Regulation," *Journal of Urban Economics*, September 2024, *143*, 103689.

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

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "The Production Function for Housing: Evidence from France," *Journal of Political Economy*, 2021, *129* (10).

Cook, Cody, Pearl Z Li, and Ariel J Binder, "Where to Build Affordable Housing? Evaluating the Tradeoffs of Location," 2024.

Damiano, Anthony and Chris Frenier, "Build Baby Build?: Housing Submarkets and the Effects of New Construction on Existing Rents," *Working paper*, 2020.

Davidoff, Thomas, Andrey Pavlov, and Tsur Somerville, "Not in My Neighbour's Back Yard? Laneway Homes and Neighbours' Property Values," *Journal of Urban Economics*, March 2022, *128*, 103405.

Davis, Morris A., William D. Larson, Stephen D. Oliner, and Jessica Shui, "The Price of Residential Land for Counties, ZIP Codes, and Census Tracts in the United States," *Journal of Monetary Economics*, March 2021, *118*, 413–431.

Diamond, Rebecca, Timothy McQuade, and Zong Huang, "The Unequal Effects of Up-Zoning: Evidence from Cook County," *Working paper*, 2024.

Dobbels, Gregory and Suren Tavakalov, "Not in My Back Yard: The Local Political Economy of Residential Land-Use Regulations," *Unpublished*, 2023.

Epple, Dennis, Brett Gordon, and Holger Sieg, "A New Approach to Estimating the Production Function for Housing," *American Economic Review*, June 2010, *100* (3), 905–924.

Eriksen, Michael D. and Amanda Ross, "Housing Vouchers and the Price of Rental Housing," *American Economic Journal: Economic Policy*, August 2015, *7* (3), 154–176.

_ **and Stuart S. Rosenthal**, "Crowd out Effects of Place-Based Subsidized Rental Housing: New Evidence from the LIHTC Program," *Journal of Public Economics*, December 2010, *94* (11-12), 953–966.

Favilukis, Jack, Pierre Mabille, and Stijn Van Nieuwerburgh, "Affordable Housing and City Welfare," *The Review of Economic Studies*, January 2023, *90* (1), 293–330.

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

Glaeser, Edward L and Joseph Gyourko, "The Impact of Building Restrictions on Housing Affordability," *FRBNY Economic Policy Review*, 2003.

Glaeser, Edward L., Joseph Gyourko, and Raven Saks, "Why Is Manhattan So Expensive? Regulation and the Rise in Housing Prices," *The Journal of Law and Economics*, October 2005, *48* (2), 331–369.

Greenaway-McGrevy, Ryan and Peter C.B. Phillips, "The Impact of Upzoning on Housing Construction in Auckland," *Journal of Urban Economics*, July 2023, *136*, 103555.

Greenwald, Daniel L and Adam Guren, "Do Credit Conditions Move House Prices?," *Working paper*, 2021.

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

_ **and** _ , "The Distaste for Housing Density," *NBER Working Paper*, October 2024.

Howard, Greg and Jack Liebersohn, "Why Is the Rent so Darn High? The Role of Growing Demand to Live in Housing-Supply-Inelastic Cities," *Journal of Urban Economics*, July 2021, *124*, 103369.

Krimmel, Jacob and Betty Wang, "Upzoning with Strings Attached: Evidence from Seattle's Affordable Housing Mandate," *NYU Furman Center*, 2023.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider, "The Housing Market (s) of San Diego," *American Economic Review*, April 2015, *105* (4), 1371–1407.

Li, Xiaodi, "Do New Housing Units in Your Backyard Raise Your Rents?," *Journal of Economic Geography*, September 2021, *22*.

Liu, Liyi, Doug McManus, and Elias Yannopoulos, "Geographic and Temporal Variation in Housing Filtering Rates," *Regional Science and Urban Economics*, March 2022, *93*, 103758.

Mast, Evan, "JUE Insight: The Effect of New Market-Rate Housing Construction on the Low-Income Housing Market," *Journal of Urban Economics*, January 2023, *133*, 103383.

McFadden, Daniel, "Modeling the Choice of Residential Location," 1978.

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

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

Nathanson, Charles G, "Trickle-down Housing Economics," *Working paper*, 2023.

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

Piazzesi, Monika, Martin Schneider, and Johannes Stroebel, "Segmented Housing Search," *American Economic Review*, March 2020, *110* (3), 720–759.

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

Schubert, Gregor, "House Price Contagion and U.S. City Migration Networks," *Working paper*, 2016.

Sinai, Todd and Joel Waldfogel, "Do Low-Income Housing Subsidies Increase the Occupied Housing Stock?," *Journal of Public Economics*, December 2005, *89* (11-12), 2137–2164.

Soltas, Evan, "Tax Incentives and the Supply of Low-Income Housing," 2024.

Song, Jaehee, "The Effects of Residential Zoning in U.S. Housing Markets," *SSRN Electronic Journal*, 2021.

Steven Manson, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles, "IPUMS National Historical Geographic Information System: Version 19.0 [dataset]," 2024. Minneapolis, MN: IPUMS.

Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler, "IPUMS USA: Version 15.0 [dataset]," 2024. Minneapolis, MN: IPUMS.

Susin, Scott, "Rent Vouchers and the Price of Low-Income Housing," *Journal of Public Economics*, January 2002, *83* (1), 109–152.

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

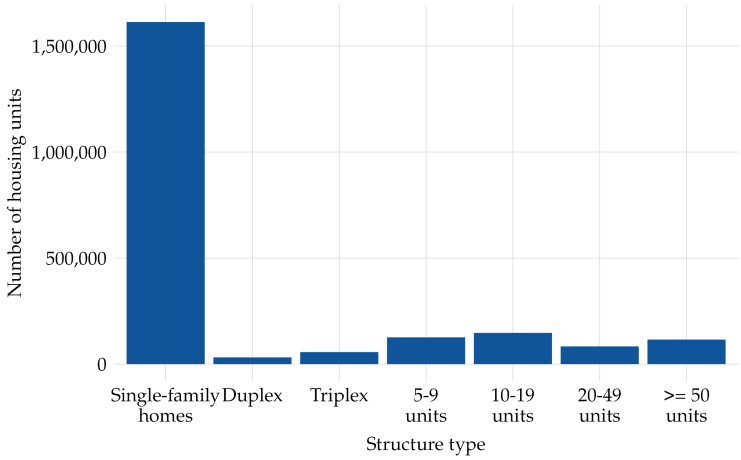
Tokman, Anthony E, "Density Restrictions and Housing Inequality," *Unpublished*, 2023.

Appendix

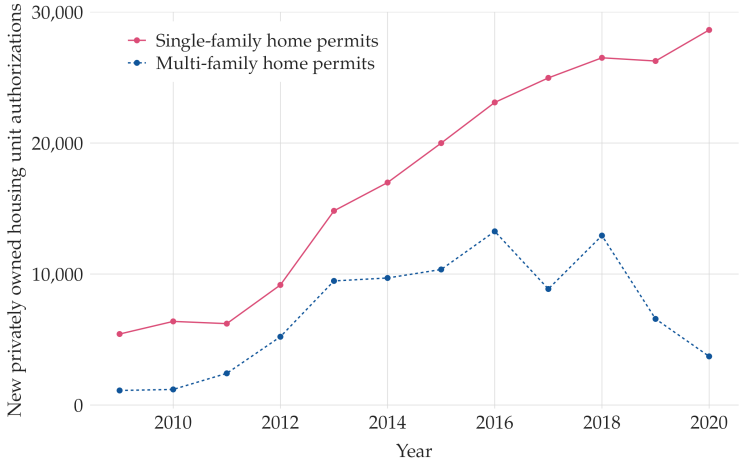
A Additional figures and tables

Figure A.1: Number of Housing Units by Structure Type

(a) Total number of housing units by structure type

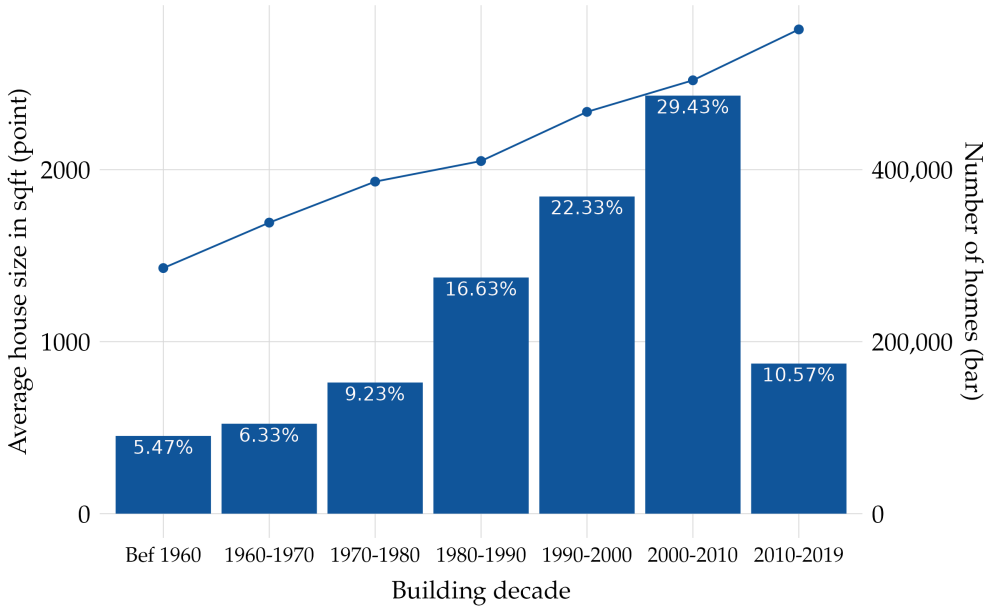


(b) Number of permits for new homes by property type



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.

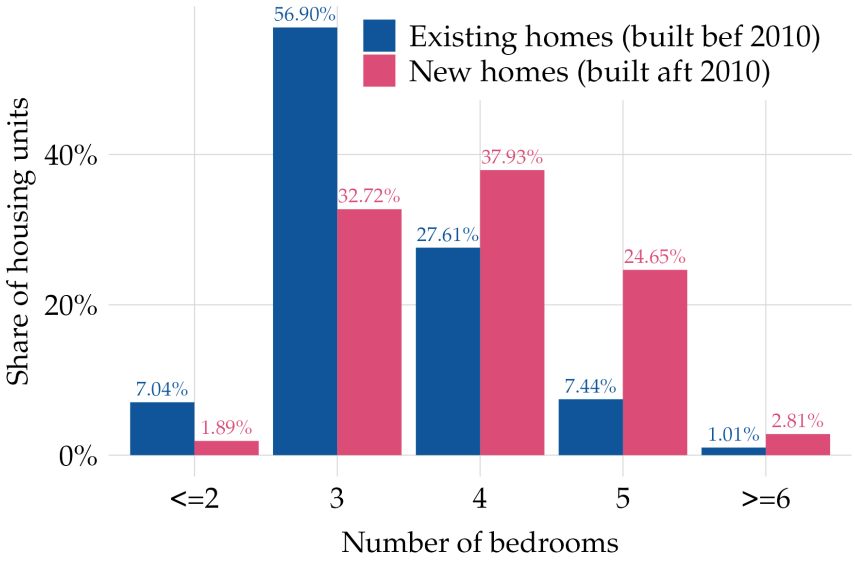
Figure A.2: Housing Stock Composition as of 2019



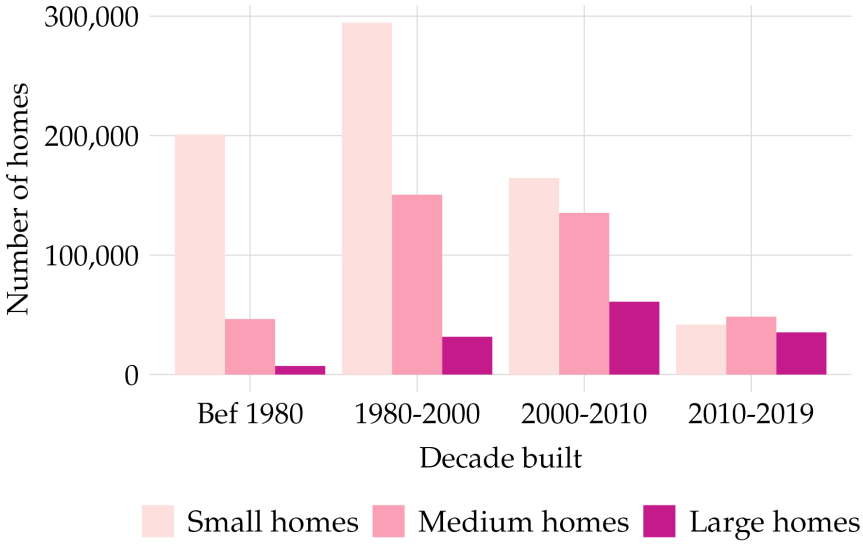
Notes: This graph shows housing stock composition by year built in the 2019 CoreLogic Assessor data. The points and line correspond to the left axis and show the mean house size for houses built in different years. The bars correspond to the right axis and show the number of houses built in different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

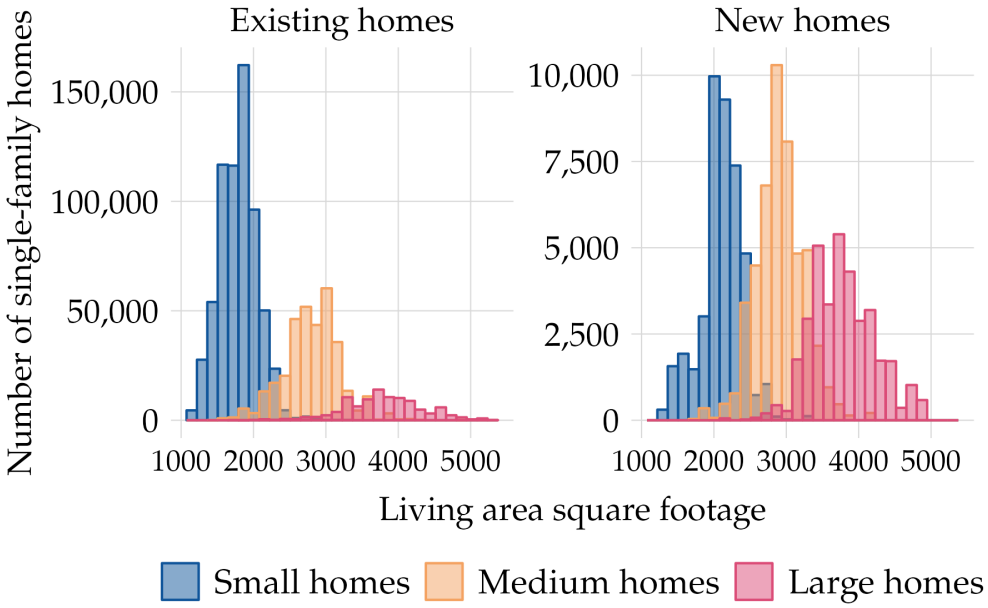


(b) Number of homes by decade built and house type



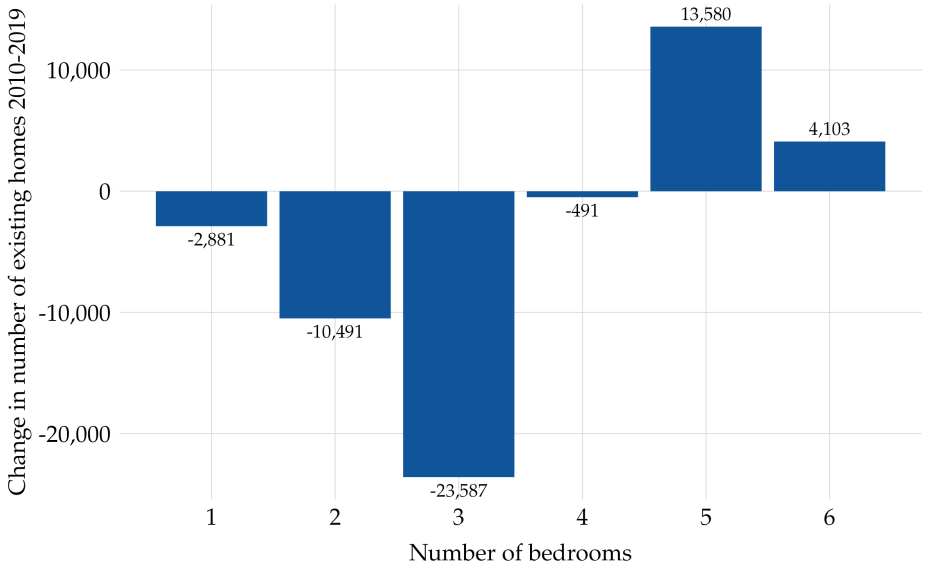
Notes: These graphs show housing stock composition by decade built and house type in the 2019 CoreLogic Assessor data. Small homes are defined as homes with fewer than three bedrooms, medium homes are defined as homes with four bedrooms, and large homes are defined as homes with more than five bedrooms.

Figure A.4: Single-Family Home Size Distribution by House Type



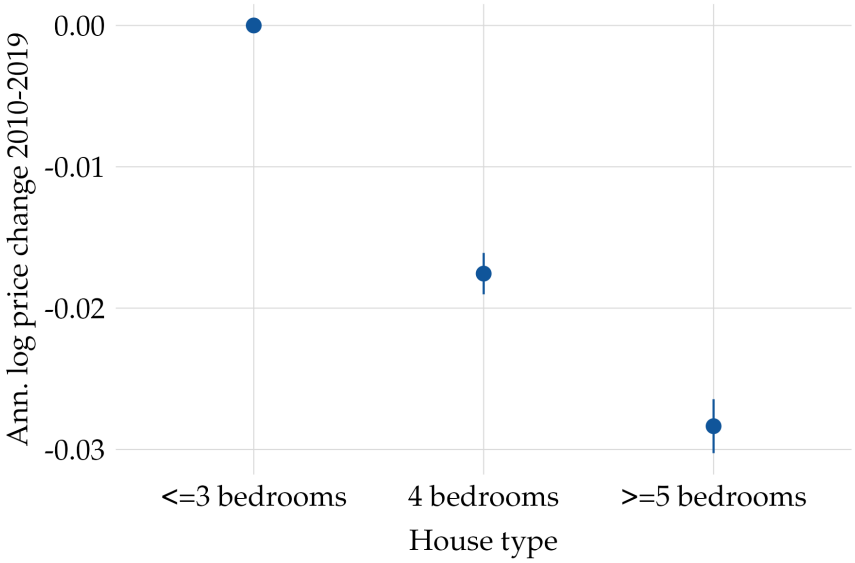
Notes: This figure plots the distribution of home sizes in square footage by house type for existing homes (built before 2010) and new homes (built after 2010) separately. Small homes are defined as homes with fewer than three bedrooms, medium homes are defined as homes with four bedrooms, and large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

Figure A.5: Demolitions and Upgrades of Existing Housing Stock



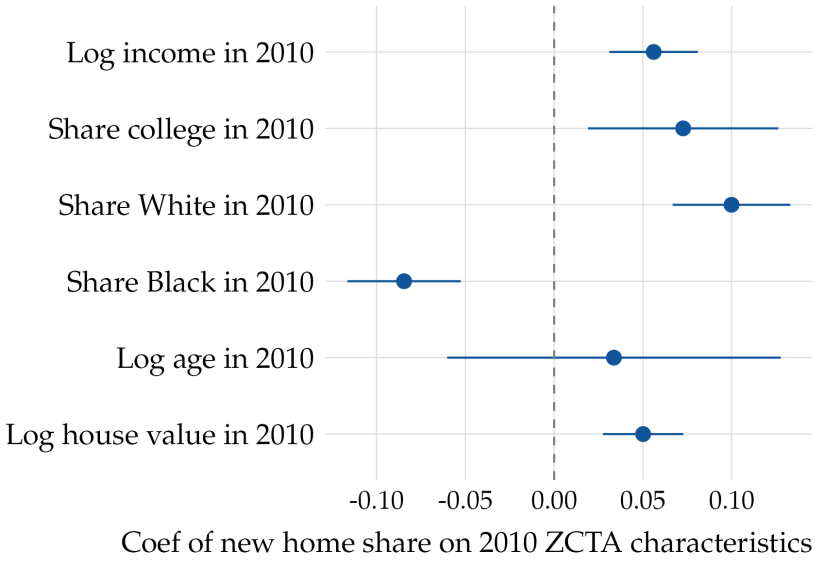
Notes: This graph plots the demolitions and upgrades of single-family homes built before 2010. It compares the number of single-family units recorded in the 2010 and 2019 Assessor records. The change in the number of units between these two years reflects alterations in the existing housing stock. The 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.

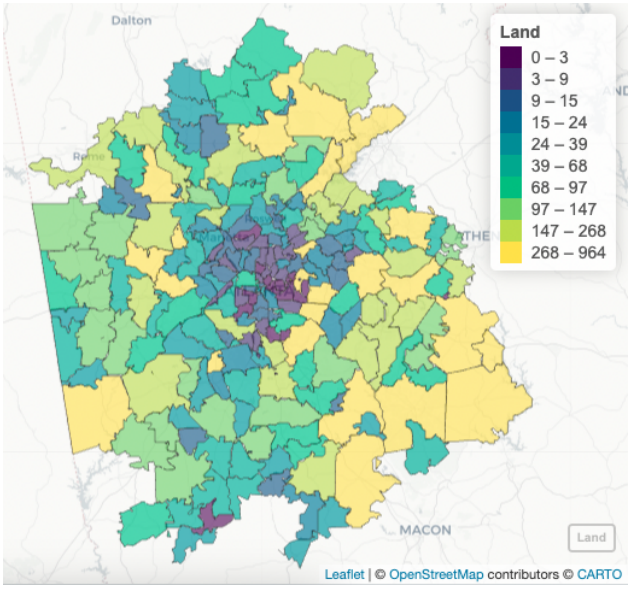
Figure A.7: Correlation Between New home Share and 2010 ZCTA Characteristics



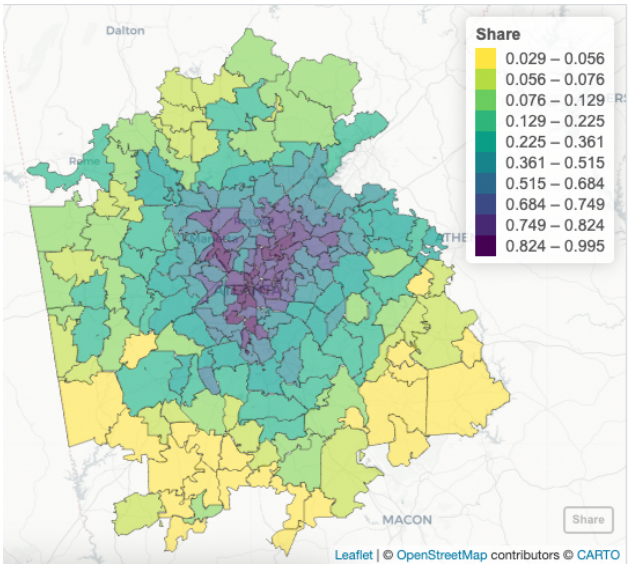
Notes: The figure shows the annualized log price change of repeat-sale properties between 2010 and 2019, using data from CoreLogic transaction records. I partition all repeat-sale properties into three housing types based on their bedroom count, then I regress house price changes between transactions 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

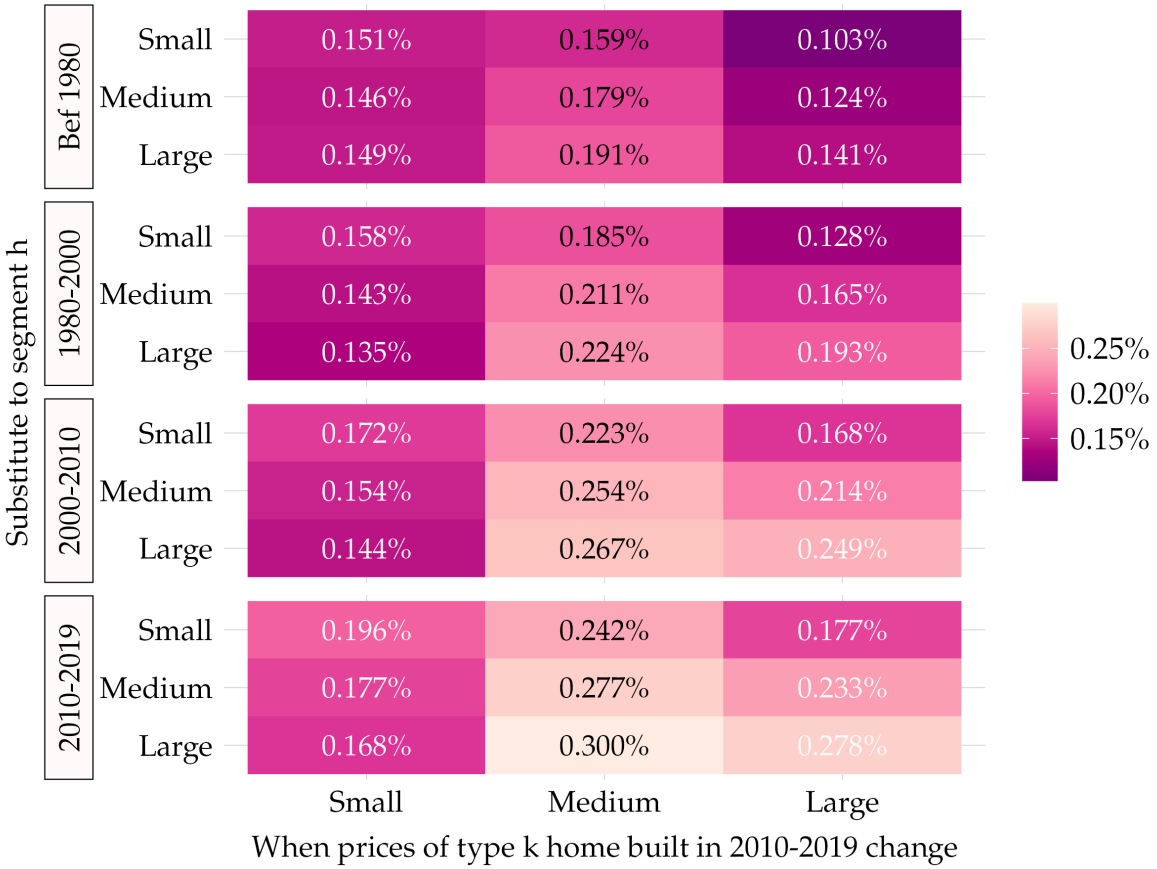


(b) Share of developed land in 2016



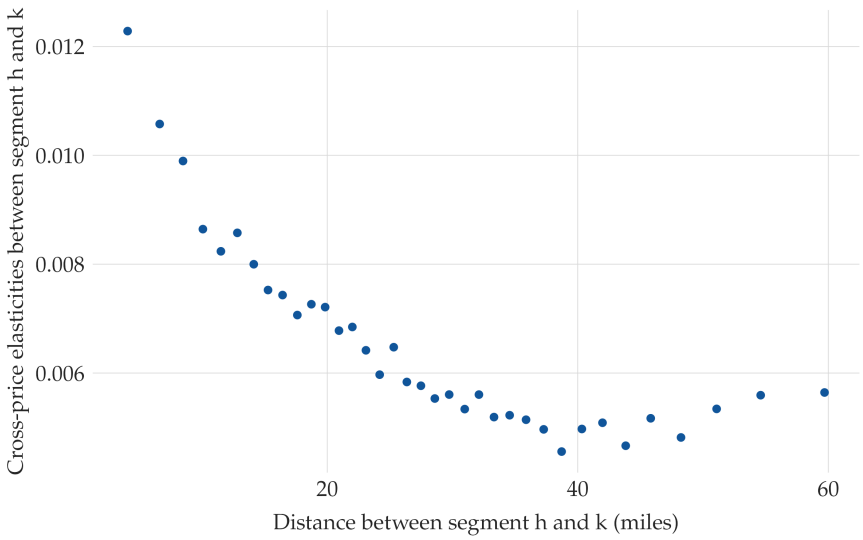
Notes: The two maps show land availability and land cover in the Atlanta MSA as of 2016, using data from the National Land Cover Database processed by the National Neighborhood Data Archive (NaNDA). The left panel shows the amount of vacant barren land – defined as land with less than 15% vegetation 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



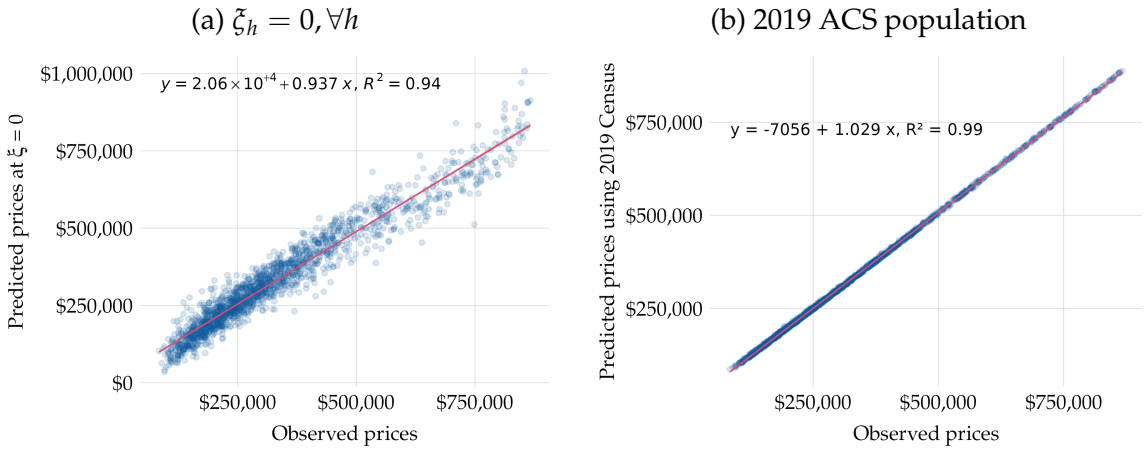
Notes: This figure presents the median cross-price elasticity, aggregated by house type and decade built. The numbers within each rectangle gives the percentage change in market shares of segment h (row) when the prices of new homes of type k (column) changes by 1%. The colors indicate the magnitudes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

Figure A.10: Estimated Cross-Price Elasticities By Geographic Distance



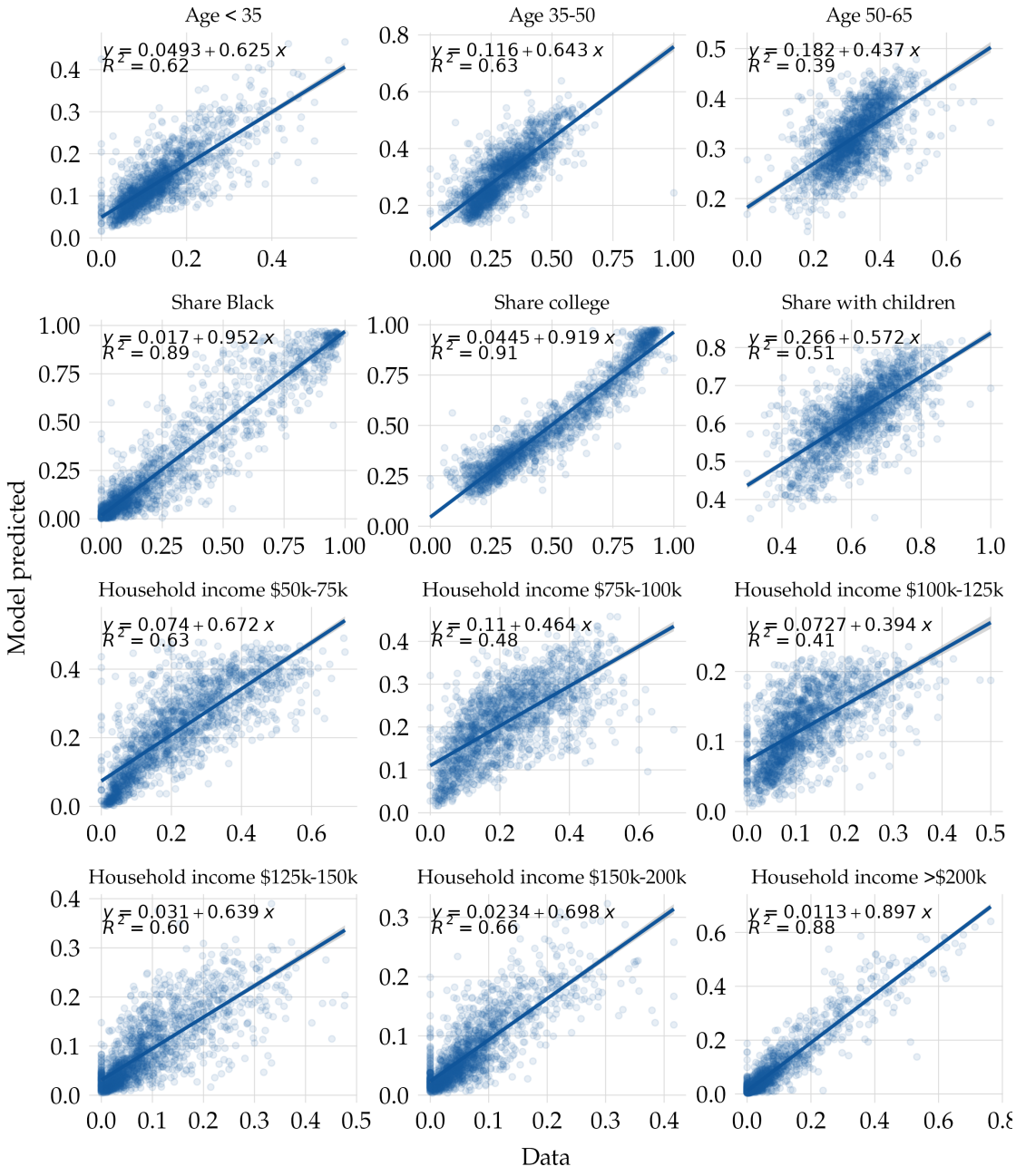
Notes: This figure displays a bin scatter plot illustrating the relationship between cross-price elasticities and the geographic distance between segments. The distances are measured using the centroids of the ZCTAs corresponding to each segment. Segments within the same ZCTAs are dropped.

Figure A.11: Demand Model Validation



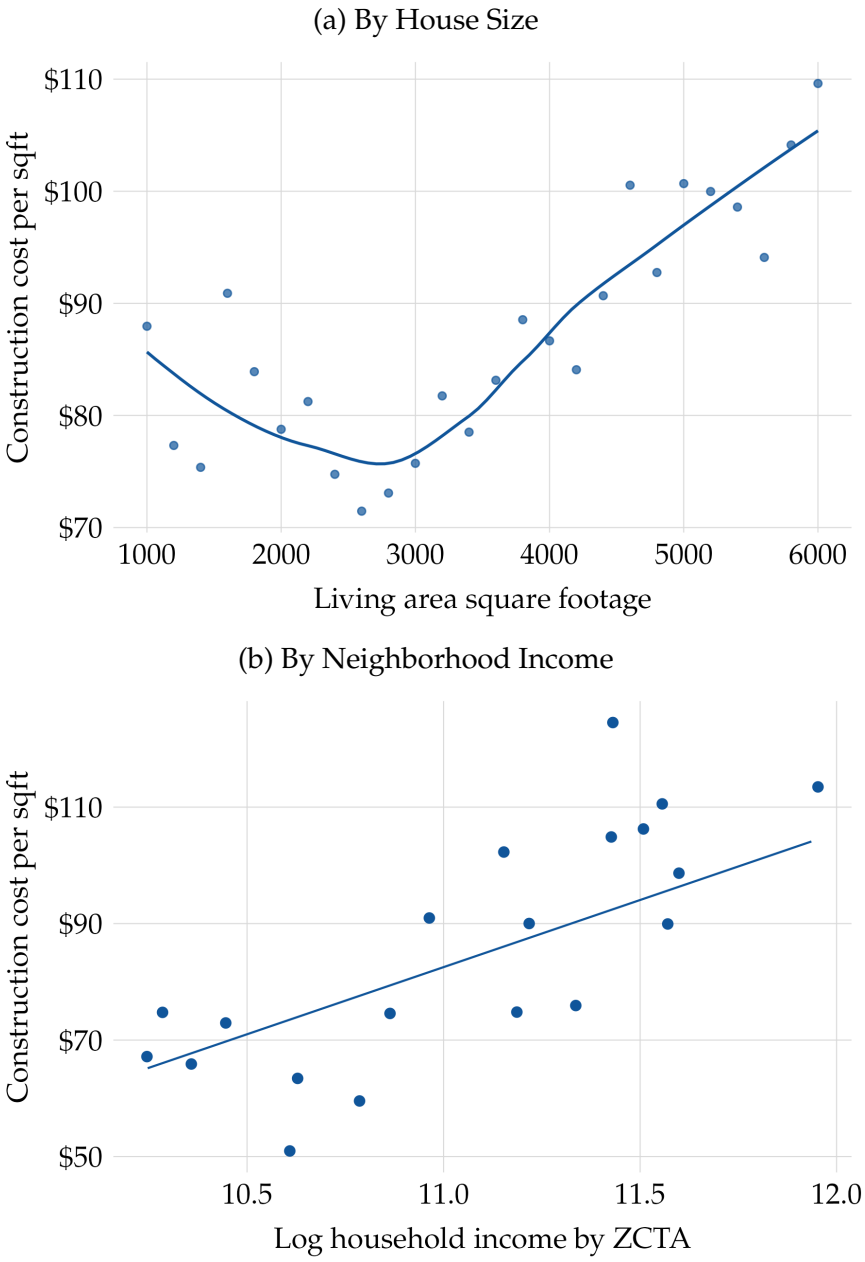
Notes: The two graphs compare the actual housing prices with simulated prices under two scenarios. The simulated prices are calculated using the equilibrium condition and estimated demand parameters at the observed supply. The left panel recalculates housing prices by setting all unobserved segment quality 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



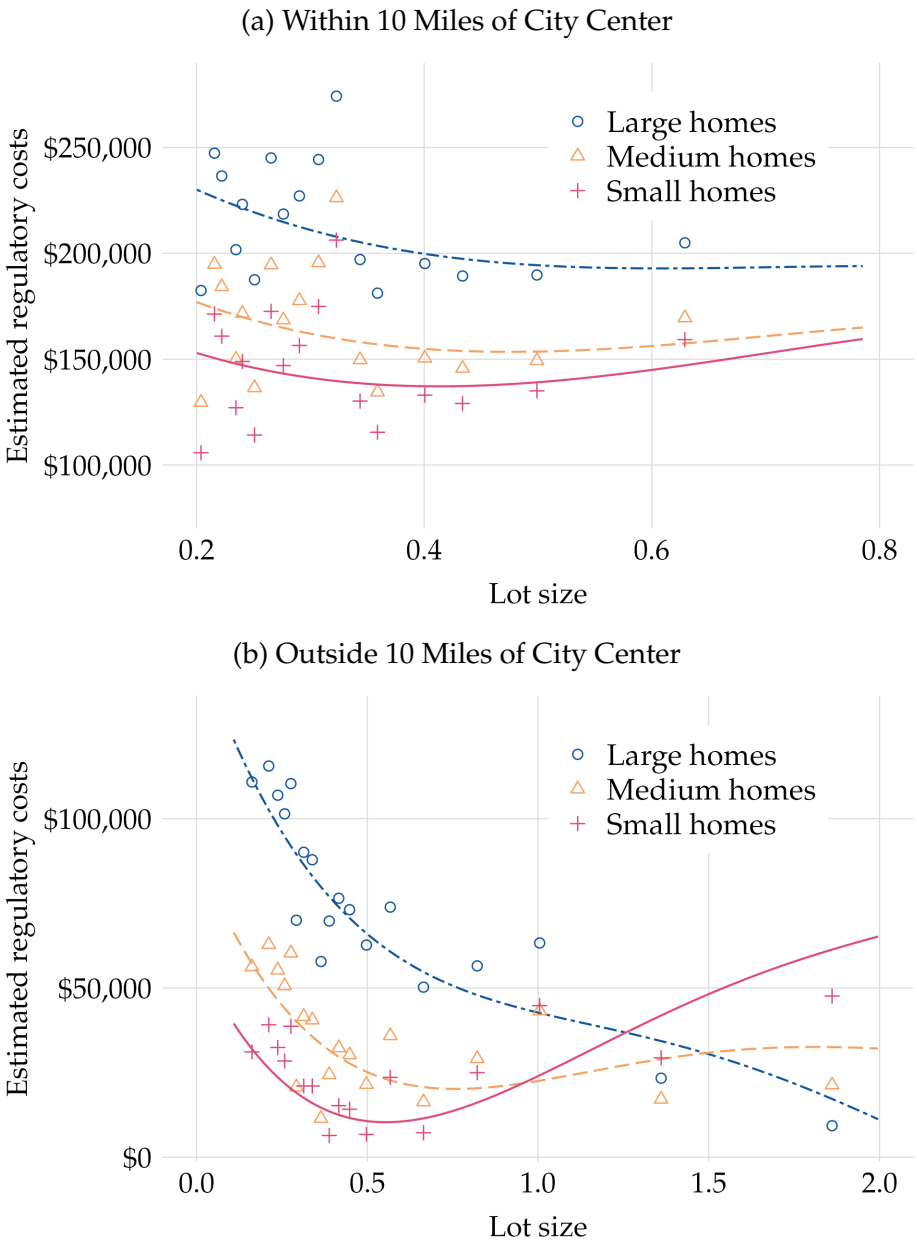
Notes: These graphs compare the actual and predicted sociodemographic composition for each segment.

Figure A.13: Construction Cost Per Square Foot



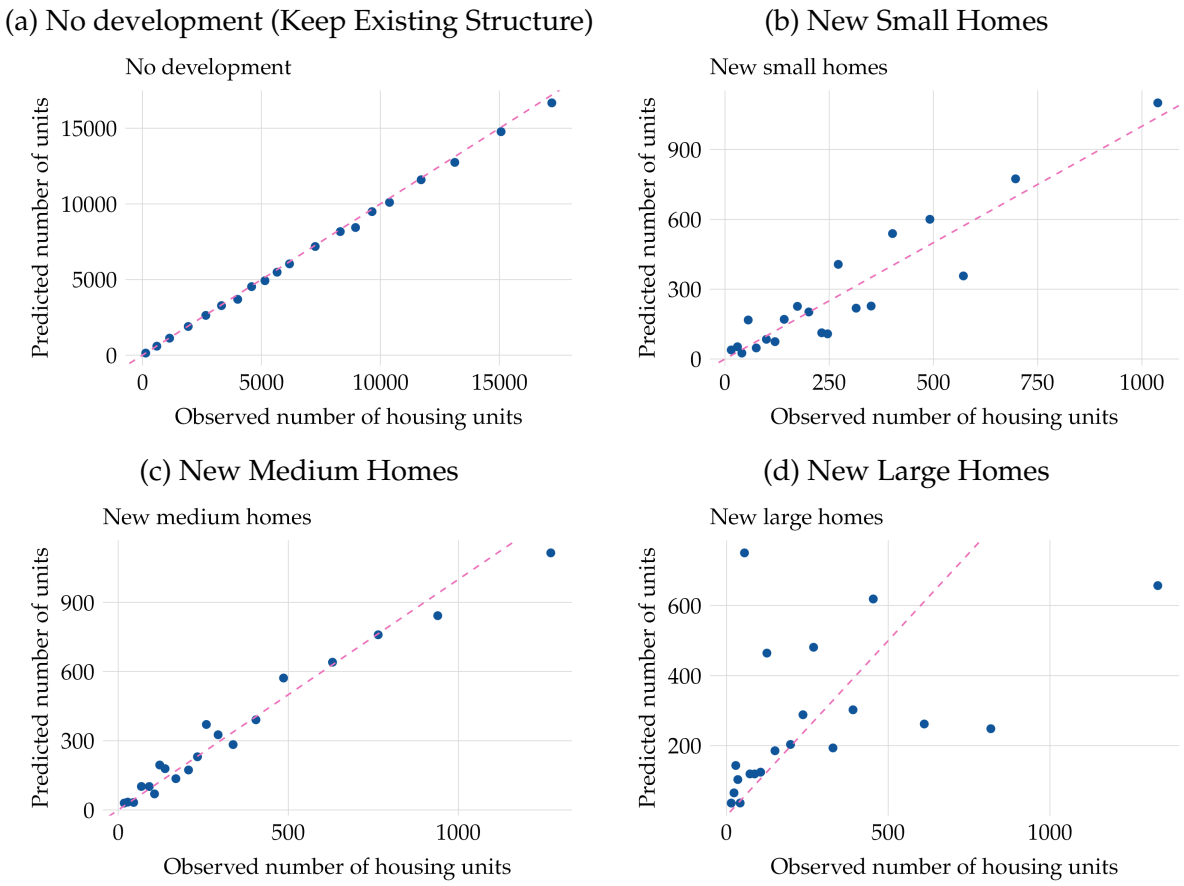
Notes: These graphs present the variation of construction cost per square foot by house size and location. Panel (a) presents a binned scatter plot showing how costs change with house size in square footage. Panel (b) highlights the variation in construction costs across neighborhoods with different 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



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 $\leq$ 3 bedrooms, medium homes have four bedrooms, and large homes have $\geq$ 5 bedrooms.

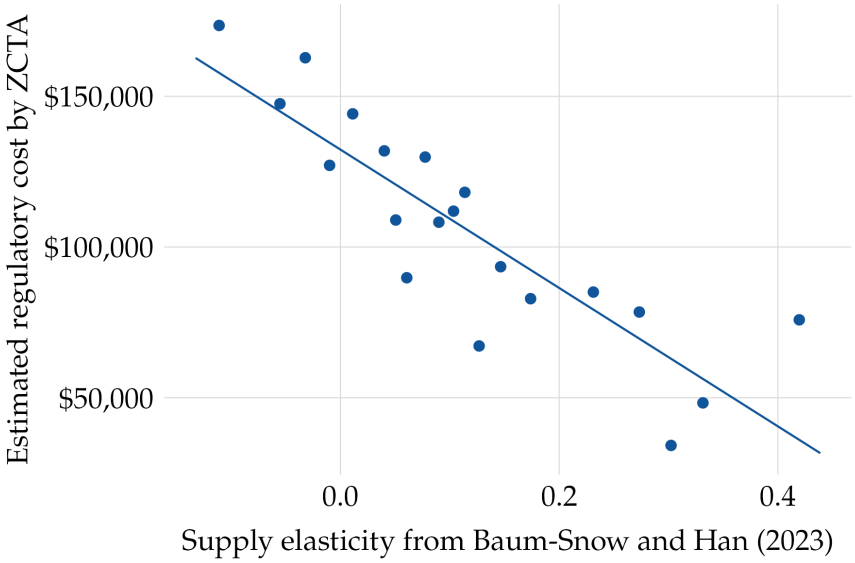
Figure A.15: Predicted and Observed Lot Development Decision Within Neighborhood



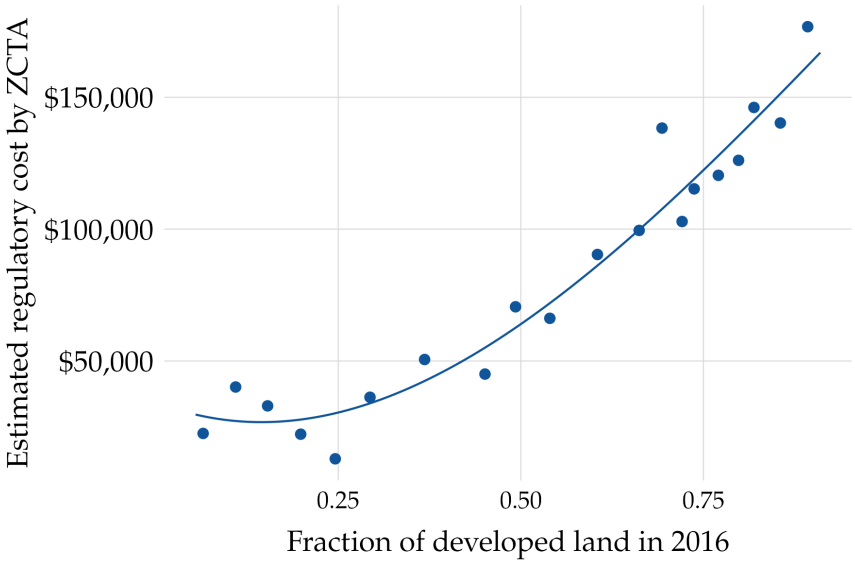
Notes: These binned scattered graphs compare the observed number of housing units to the model-predicted housing supply in each neighborhood. Each panel corresponds to a house type choice.The red dashed line is the 45-degree line.

Figure A.16: Estimated Mean Regulatory Costs By ZCTA

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

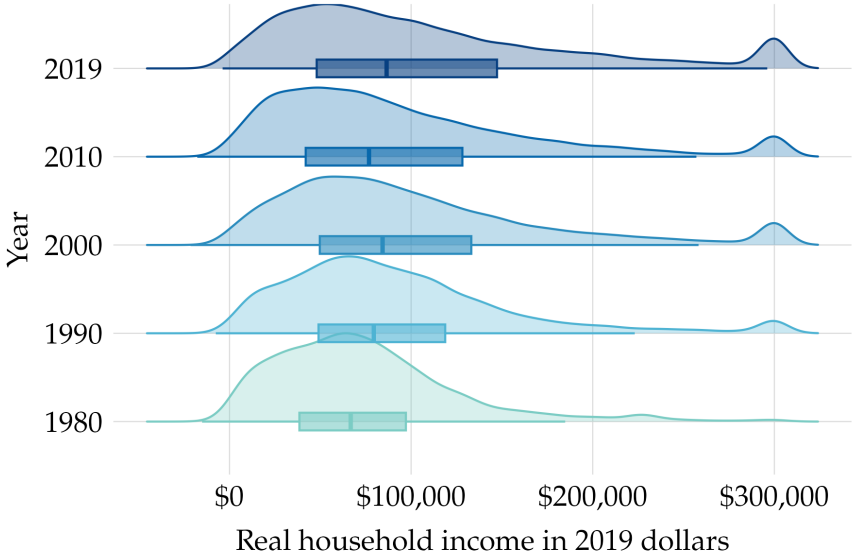


(b) Correlation with share of developed land in 2016 from NCLD



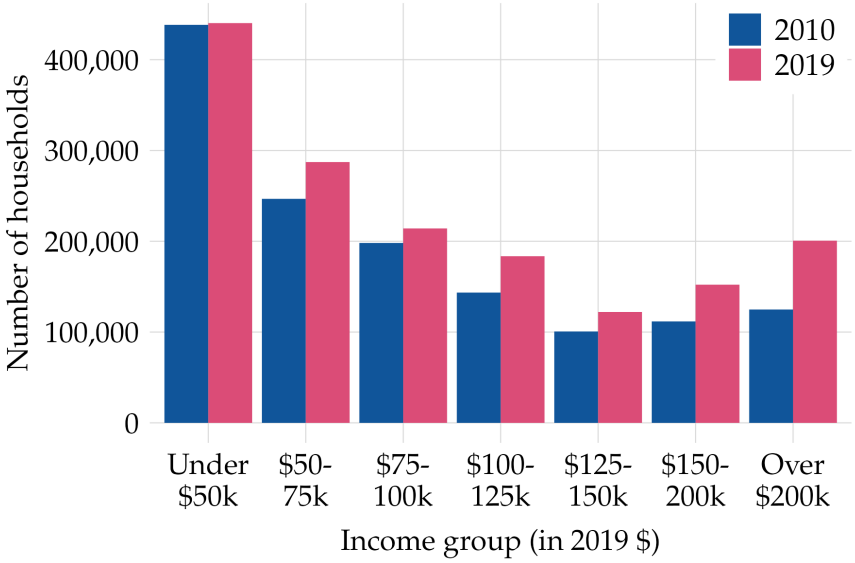
Notes: These two figures compare estimated regulatory costs aggregated at the ZCTA level with two other measures. Panel (a) uses the supply elasticities estimated by Baum-Snow and Han (2024). Panel (b) uses the fraction of developed land in 2016 from the National Land Cover Database.

Figure A.17: Change in Income Distribution 1980-2019



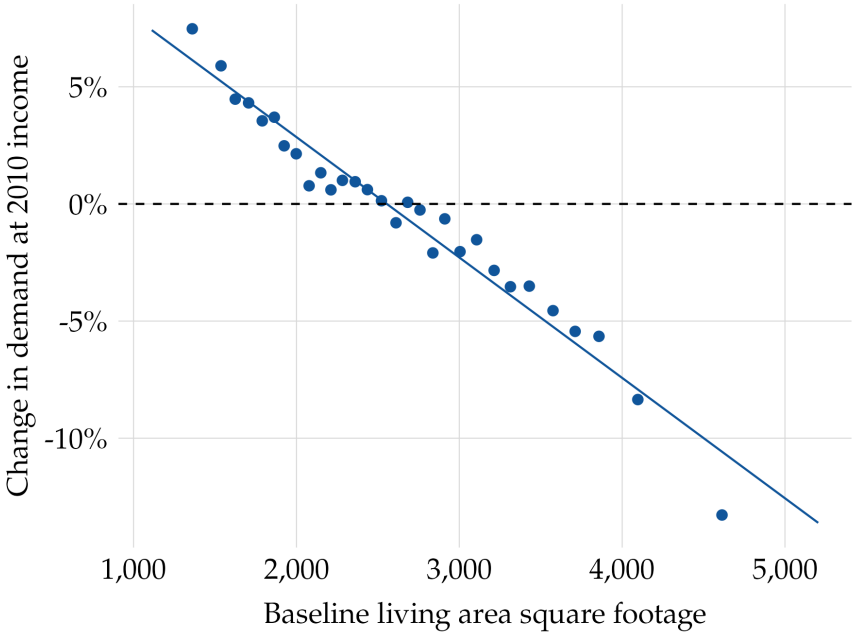
Notes: This figure show the change in income distribution from 1980 to 2019. The boxplot within the density plots shows the 25th percentile, median, and 75th percentile values. All incomes are adjusted to 2019 real dollars using the urban CPI for the Atlanta MSA. Incomes above $300,000 are grouped into 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



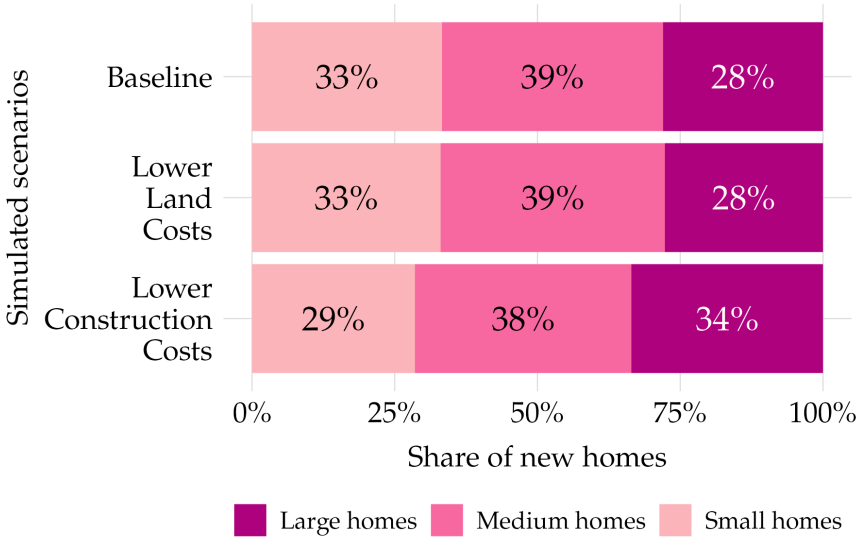
Notes: This figure show the number of households in each income group in 2010 to 2019. The data are from the American Community Surveys in 2010 and 2019. I convert the 2010 income to real 2019 dollars using the urban CPI in the Atlanta MSA.

Figure A.19: Partial Equilibrium Change in Housing Demand Under 2010 Income Distribution



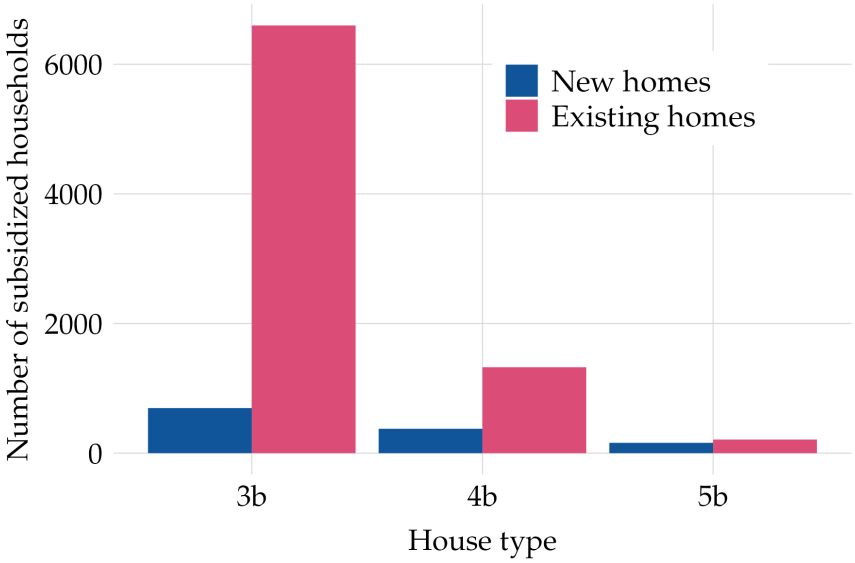
Notes: This figure plots the change in housing demand under the 2010 income distribution at observed prices compared to demand at the baseline scenario. Housing demand is calculated under partial equilibrium without adjusting supply and prices.

Figure A.20: Simulated Distribution of New Construction under Lower Production Costs



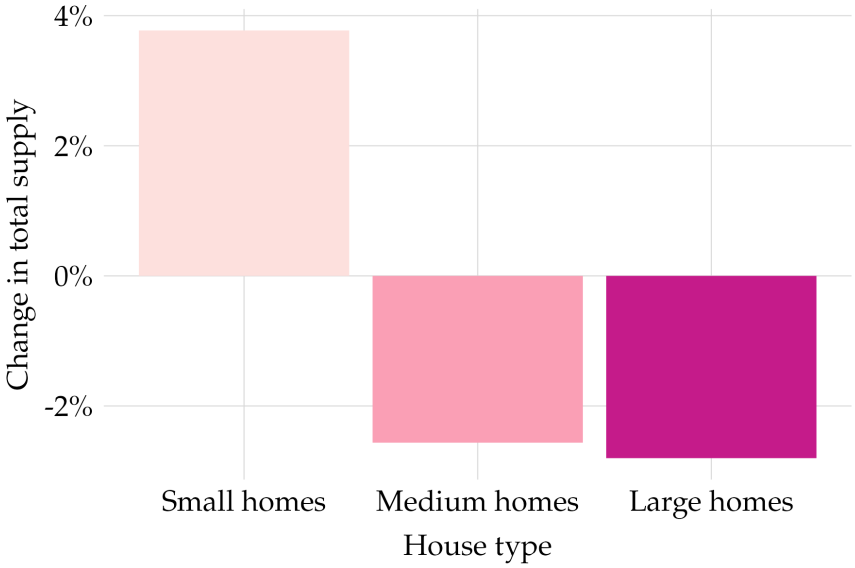
Notes: This figure presents the composition of new construction under the baseline scenario and the two simulated scenarios with the 2010 land and construction costs. To calculate the two simulated equilibria, I first calculate developers' decisions under the counterfactual costs. Given the updated supply, 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 $\leq$ 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 $\leq$ \$50k	0.30	0.60
Share Income \$50 $-$ 75k	0.18	0.18
Share Income \$75 $-$ 100k	0.14	0.10
Share Income \$100 $-$ 125k	0.11	0.05
Share Income \$125 $-$ 150k	0.07	0.02
Share Income \$150 $-$ 200k	0.09	0.02
Share Income $\geq$ \$200k	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 $\leq$ 35	Age 35-50	Age 50-65
Price ($100k)	-0.4080	0.2628	-0.0005	-0.2061	-0.1479	-0.1176
Sqft (std)	-0.9345	0.1269	0.0688	-0.1644	0.1359	0.0608
Built in 2010-2019	-0.4766	-0.0354	-0.1340	0.0419	-0.0364	0.0581
Built in 2000-2010	2.0846	0.2355	-0.1657	2.0497	1.5956	0.7935
Built in 1980-2000	2.1180	0.2512	-0.0330	1.1215	1.1015	0.7104
Sqft$\times$Built in 2010-2019	1.4206	0.1844	-0.0156	0.4220	0.3696	0.4376
Sqft$\times$Built in 2000-2010	1.0354	-0.0216	0.2164	-0.0858	0.0870	0.1650
Sqft$\times$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	$\geq$\$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$\times$Built in 2010-2019	-0.1208	0.3665	0.5061	0.5788	0.4368	0.3199
Sqft$\times$Built in 2000-2010	-0.2420	-0.1414	-0.0911	-0.1436	-0.2716	-0.2562
Sqft$\times$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)×Medium homes	-26,644	(5,536)
Lot Size (std)×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.

B Supplemental Analyses

B.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.

B.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.

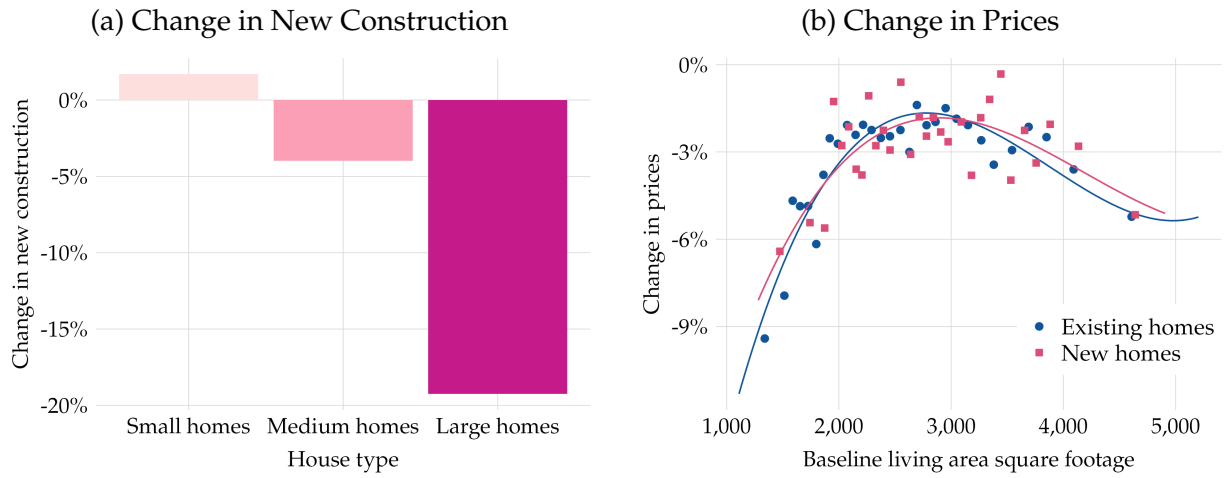
Panel (a) of Figure A.24 shows that with homogeneous preferences, the number of newly built

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, the average household income is $90,000, while the average home has 3.5 bedrooms and 2,300 square feet. Figure 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.

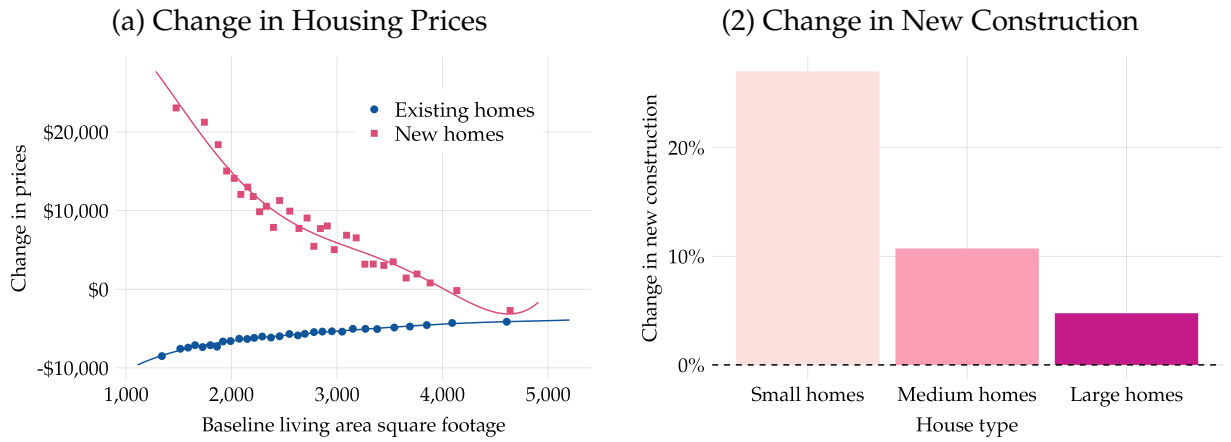
B.3 Restrict House Types for Subsidized Households

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 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 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.

C Data Construction

C.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

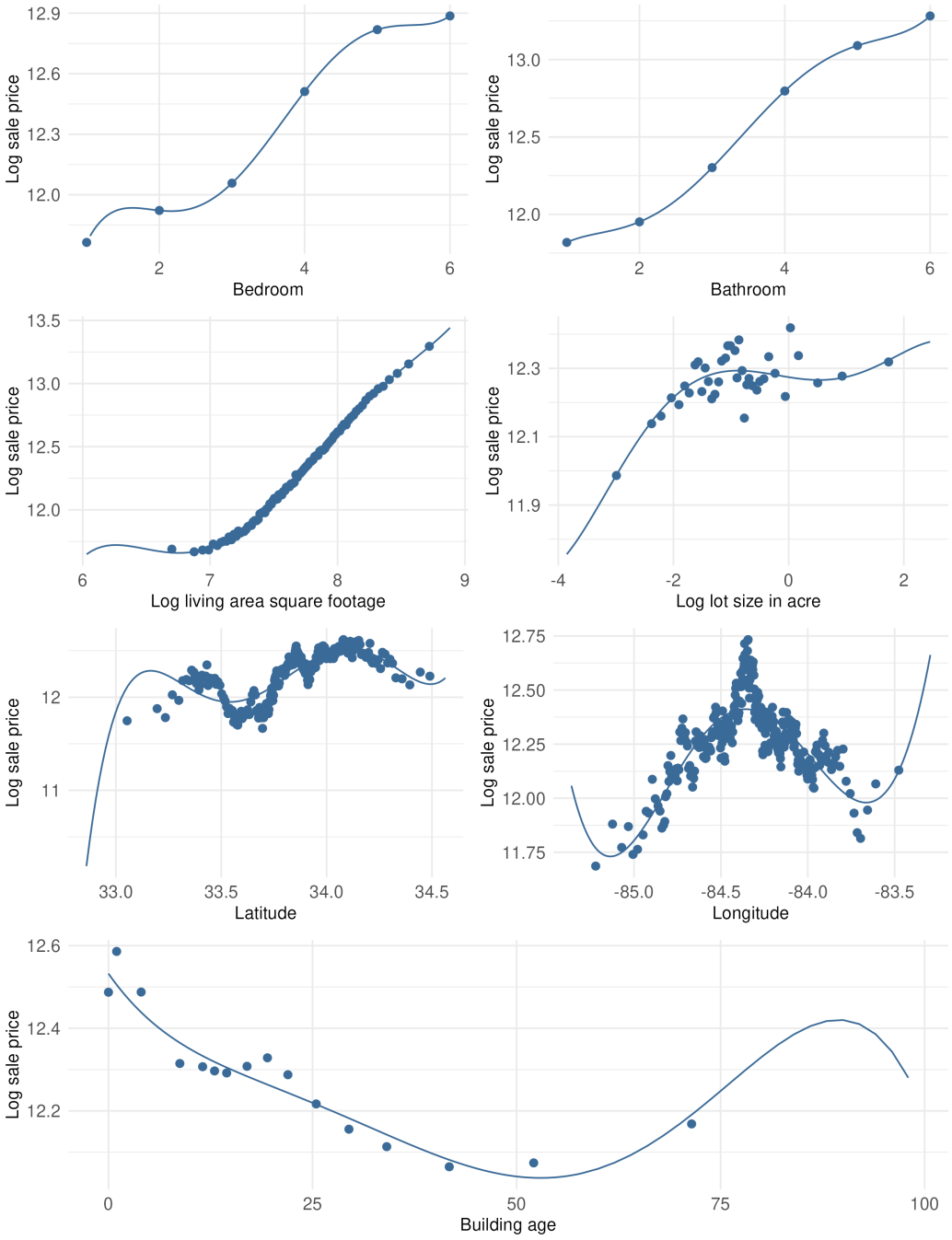
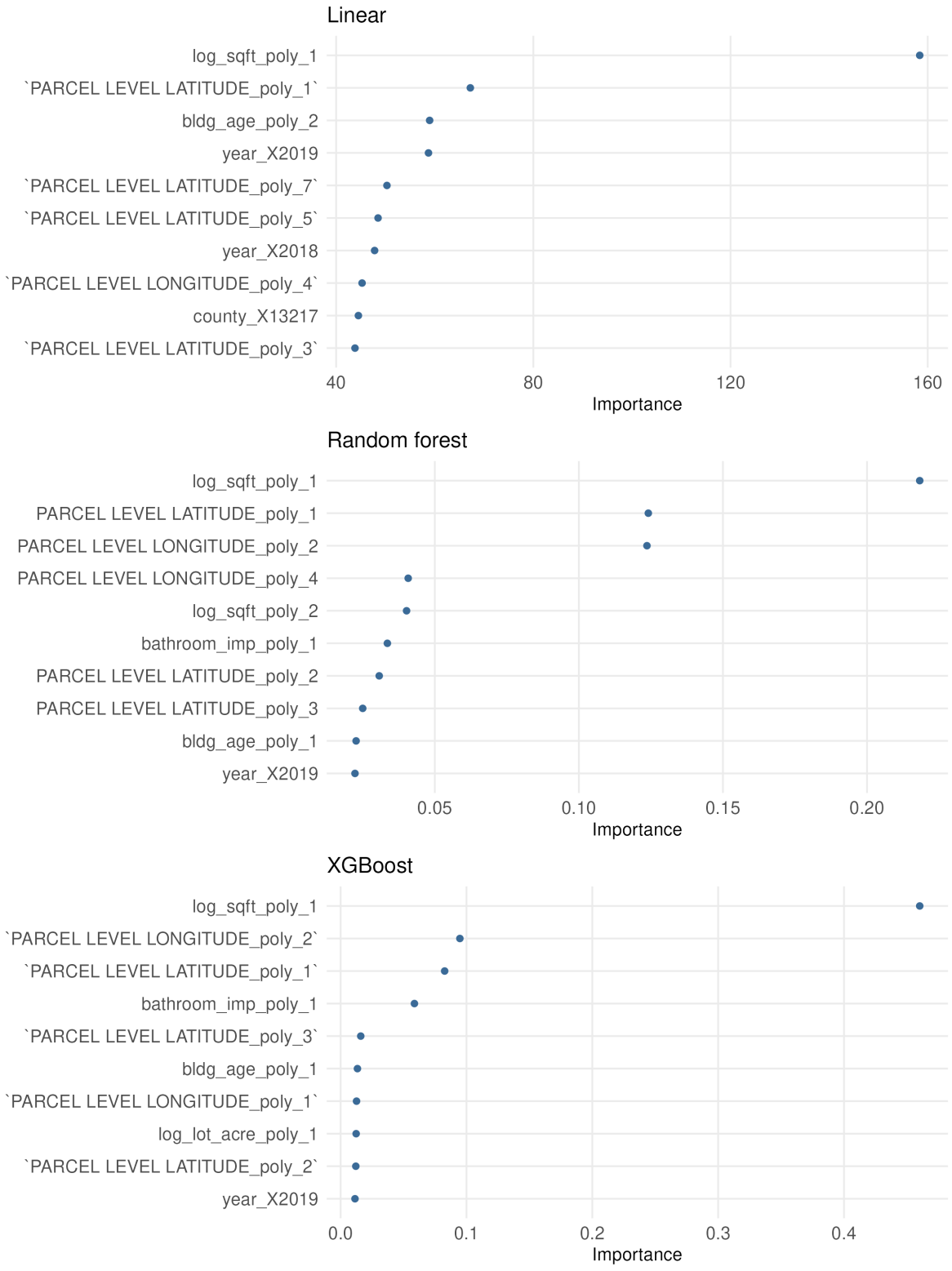


Table A.10: Compare performances of machine learning models

	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



C.2 Validate L2 Data

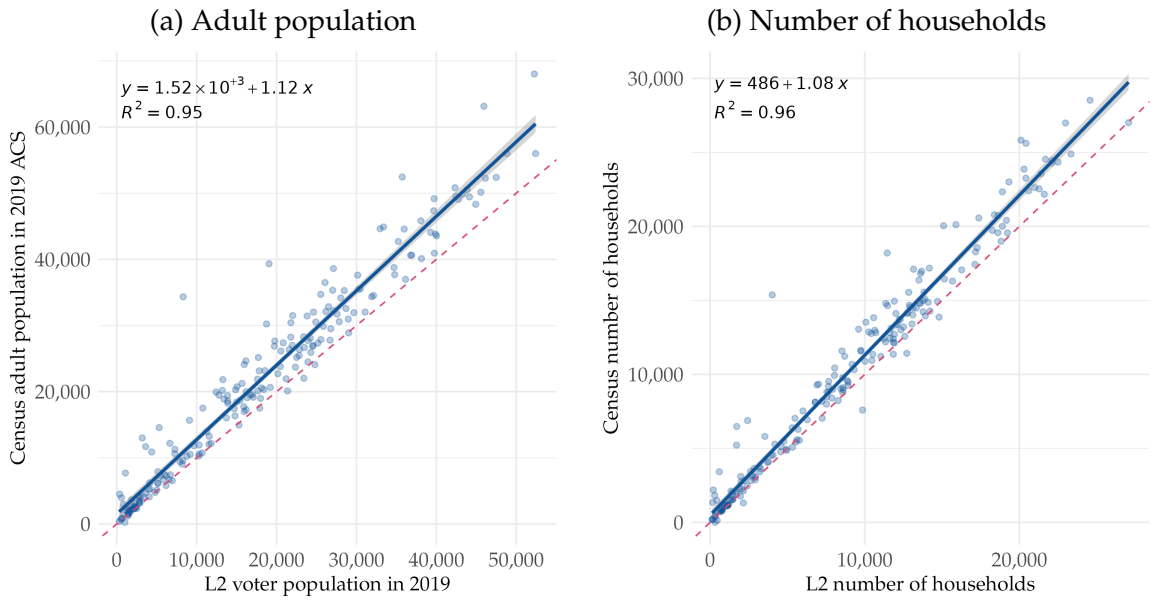
L2 compiles basic demographic information, including age, race, address, and gender, from government voter registration records. L2 further augments the base files with socioeconomic characteristics from other sources, such as credit bureaus, telecommunication providers, and other consumer data databases. 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

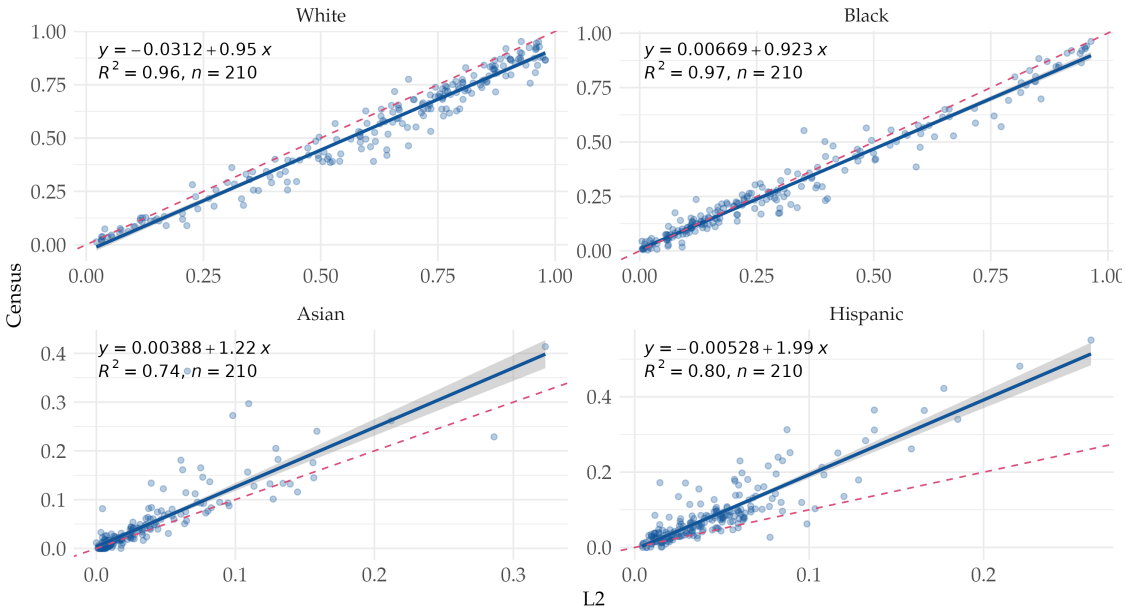
Figure A.29: L2 population vs. 2019 ACS



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.

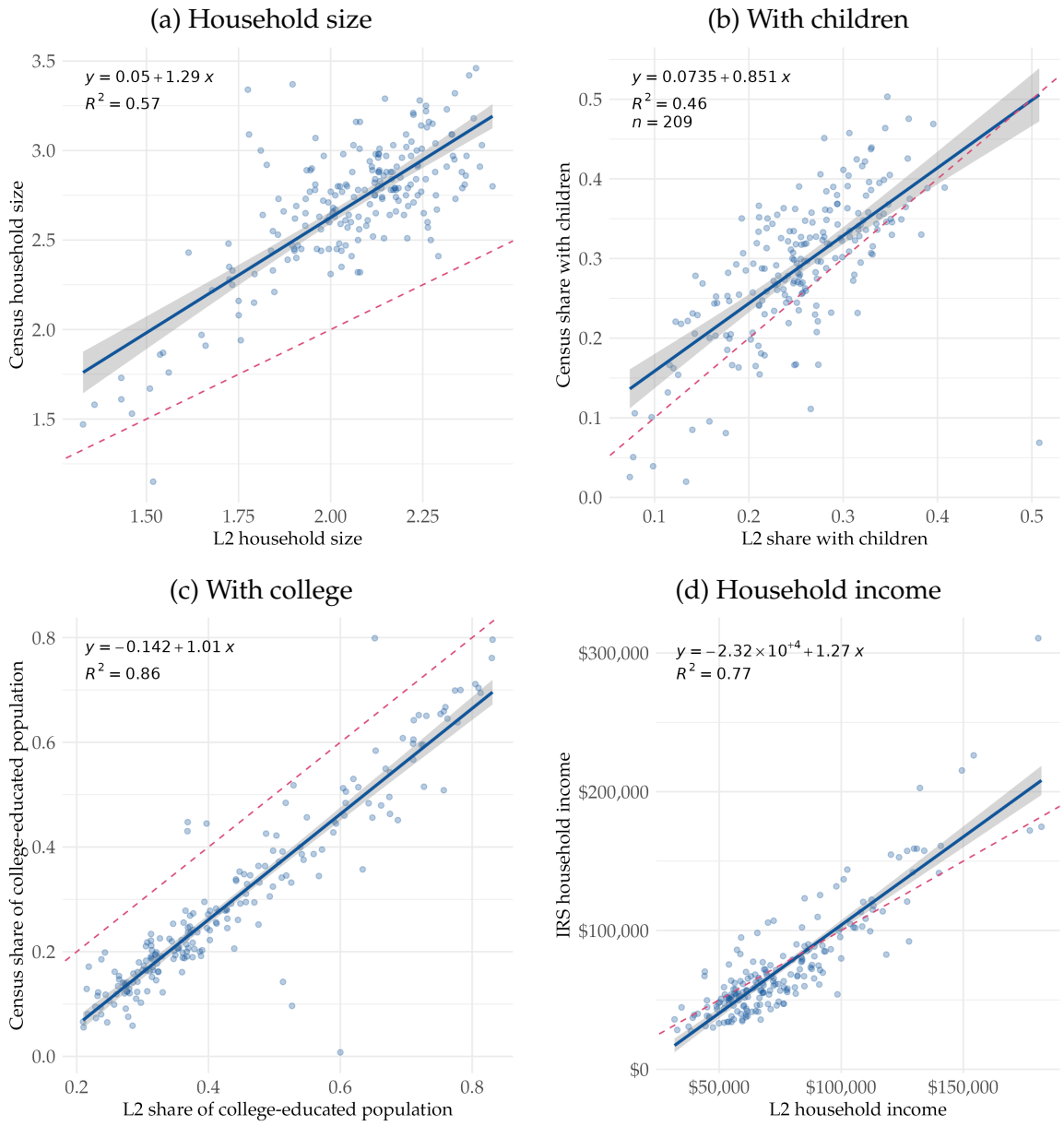
Figure A.30: L2 racial composition vs. 2019 ACS



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.

C.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.

C.4 Zoning Code

In Georgia, both cities and counties have zoning authority. Typically, cities govern zoning within their city limits, while counties oversee zoning in unincorporated areas. To determine the zoning jurisdiction, I geocode each parcel's location and check if it falls within incorporated areas as of 2020, using the shape files of incorporated places from the Census Bureau. Several new cities were established between 2010 and 2020, such as Brookhaven in 2012 and Stonecrest in 2016. One common goal for forming new cities is to gain greater local control over zoning and land use planning. I use the incorporation status as of 2020 to determine the level of zoning authrority. Given that these new cities occupy a small portion of the region's total land area, the impact of these changes can be reasonably excluded. My approach assumes that zoning districts align closely with pre-annexation zoning boundaries and that zoning stringency remains consistent. It would be interesting to examine how zoning regulations change after city formation in future research.

Zoning regulations generally consist of two main components: permitted uses and development standards. Zoning codes are typically formatted as R-1, R-2, C-1, NR, PUD, and AG, with the letter indicating the permitted uses and the number specifying the level of development standards. Permitted uses determine whether a parcel can be developed for residential, commercial, industrial, or mixed-use purposes. Development standards mandate specific dimensional requirements and other criteria regarding the structure. These rules include detailed specifications such as:

- Minimum lot area, minimum lot width, minimum lot depth, and minimum lot frontage
- Minimum front yard setback, minimum side yard setback, minimum rear yard setback
- Maximum lot coverage, maximum floor-to-area ratio, maximum density
- Minimum living area, minimum ground floor area
- Maximum structure height
- Minimum parking requirements
- Sewer and water requirements

While CoreLogic provides the zoning codes, it does not explain their meanings. To fill this gap, I hand-collect the development standards for each zoning code from MuniCode, a digital platform that compiles municipal codes and ordinances for local governments. For each zoning code, I focus on the development standards and standardize the requirements across zoning jurisdictions. Although zoning regulations are public information, researchers often lack data on them due to the localized nature of these regulations. My data digitalization efforts complement a recent initiative to collect local zoning regulations in a central database, notably the National Zoning Atlas. This ongoing project aims to compile zoning information nationwide. However, their most recent

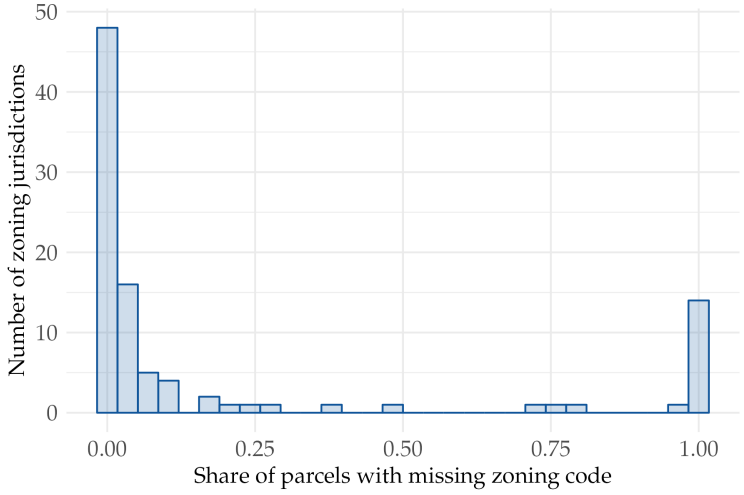
data release only includes basic information, such as the presence of single-family zoning and the number of zoning districts within each jurisdiction.

I compare the zoning codes recorded in CoreLogic with those listed in the government ordinances on MuniCode. Occasionally, zoning codes change when local governments amend their Codes of Ordinances. I use the version closest to 2019 and standardize older zoning codes to match this version, as CoreLogic sometimes still uses the older codes for some parcels.

Since my primary focus is understanding the development of single-family homes, I focus on residential zoning districts and planned residential development zones. About 7% of single-family homes in my sample are built in non-residential zones, such as commercial and industrial zones. Among newly constructed single-family homes, less than 1% of them were built in non-residential zones.

Although CoreLogic provides zoning codes at the parcel level, data quality varies. At the aggregate level, about 25% of the 1.2 million parcels with single-family homes have missing zoning codes. As shown in Figure A.32, these parcels are concentrated in a few zoning jurisdictions. For example, CoreLogic does not report zoning codes for Cobb County and the cities within it, such as Marietta, Powder Springs, and Kennesaw. For parcels with missing codes in jurisdictions that do report zoning codes, I fill out the missing values with the mode of zoning codes within their Census block group or Census tract in the same zoning jurisdiction. For each zoning code, I match the development standards I hand-collected.

Figure A.32: Share of missing zoning code by zoning jurisdiction



Notes: This graph shows the histogram of number of zoning jurisdictions with missing zoning code.

C.5 Counterfactual dataset

I describe how I create the counterfactual dataset with the 2010 income distribution used in 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	...
...	...	...	...	...	...	...	...	...

D Model Estimation

D.1 Demand Estimation Details

I estimate the heterogenous preference coefficients and mean indirect utilities by maximizing the log likelihood function,

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

My sample includes over 1.3 million households. Calculating choice probabilities for such a large number of households is computationally intensive. Given that the heterogeneity arises from demographics only, I group households based on their demographic characteristic. I discretize demographics and group households into 224 household types. Specifically, I use 4 age bins, 7 income bins, 2 racial groups, 2 education levels, and 2 children present status. Then I estimate the

following likelihood function,

$$LL(\delta, (\alpha^r, \beta^r)_r) = \sum_g \sum_h y_{g,h} \ln s_{g,h}$$

where g indexes the household type, $y_{g,h}$ is the number of households of type g choosing segment h, and $s_{g,h}$ is the model-predicted probability of households of type g choosing h.

To improve computational efficiency, I also standardize all continuous characteristics by subtracting the mean and dividing by the standard deviation at the CBSA level. When solving the numerical optimization problem, I estimate all parameters simultaneously by calculating the gradient using reverse mode auto differentiation. Although I have a large number of fixed effects to estimate, I find this method much faster than the fixed point iteration commonly used in the literature. I use the Newton-Raphson algorithm to optimize the function.

D.2 Price Impact of Supply Changes

The effect of an exogenous supply change on prices can be derived analytically using the equilibrium condition

$$D_h(\mathbf{p}, \mathcal{S}) = S_h, \forall h.$$

Taking the total derivative of the equilibrium condition gives

$$dS_h = \frac{\partial D_h}{\partial p_h} dp_h + \sum_{k \neq h} \frac{\partial D_h}{\partial p_k} dp_k + \frac{\partial D_h}{\partial S_h} dS_h + \sum_{k \neq h} \frac{\partial D_h}{\partial S_k} dS_k.$$

Dividing both sides by dS_h gives the change in housing prices for segment h resulting from a change in the supply of segment h,

$$\begin{aligned}
1 &= \frac{\partial D_h}{\partial p_h} \frac{dp_h}{dS_h} + \sum_{k \neq h} \frac{\partial D_h}{\partial p_k} \frac{dp_k}{dS_h} + \frac{\partial D_h}{\partial S_h} + \sum_{k \neq h} \frac{\partial D_h}{\partial S_k} \frac{dS_k}{dS_h} \\
\frac{\partial D_h}{\partial p_h} \frac{dp_h}{dS_h} &= 1 - \sum_{k \neq h} \frac{\partial D_h}{\partial p_k} \frac{dp_k}{dS_h} - \frac{\partial D_h}{\partial S_h} \\
\frac{dp_h}{dS_h} &= \left(\frac{\partial D_h}{\partial p_h} \right)^{-1} \left(1 - \sum_{k \neq h} \frac{\partial D_h}{\partial p_k} \frac{dp_k}{dS_h} - \frac{\partial D_h}{\partial S_h} \right).
\end{aligned}$$

Rearranging the terms differently, we can get the change in housing prices for segment $k \neq h$ resulting from a change in the supply of segment h,

$$\frac{dp_k}{dS_h} = \left(\frac{\partial D_h}{\partial p_k} \right)^{-1} \left(1 - \frac{\partial D_h}{\partial p_h} \frac{dp_h}{dS_h} - \sum_{h' \neq h,k} \frac{\partial D_h}{\partial p_{h'}} \frac{dp_{h'}}{dS_h} - \frac{\partial D_h}{\partial S_h} \right).$$

Dividing both sides by $dS_{h'}$ gives the change in housing prices for segment h resulting from a change in the supply of other segment $h' \neq h$,

$$\begin{aligned}
\frac{dS_h}{dS_{h'}} &= \frac{\partial D_h}{\partial p_h}\frac{dp_h}{dS_{h'}} + \sum_{k \neq h}\frac{\partial D_h}{\partial p_k}\frac{dp_k}{dS_{h'}} + \frac{\partial D_h}{\partial S_h}\frac{dS_h}{dS_{h'}} + \sum_{k \neq h}\frac{\partial D_h}{\partial S_k}\frac{dS_k}{dS_{h'}} \\
\frac{\partial D_h}{\partial p_h}\frac{dp_h}{dS_{h'}} &= -\sum_{k \neq h}\frac{\partial D_h}{\partial p_k}\frac{dp_k}{dS_{h'}} - \frac{\partial D_h}{\partial S_{h'}} \\
\frac{dp_h}{dS_{h'}} &= \left(\frac{\partial D_h}{\partial p_h}\right)^{-1}\left(-\sum_{k \neq h}\frac{\partial D_h}{\partial p_k}\frac{dp_k}{dS_{h'}} - \frac{\partial D_h}{\partial S_{h'}}\right).
\end{aligned}$$

The analytical expression shows that supply affects prices through direct supply effects and indirect price effects driven by substitution. The direct supply effect is straightforward: for instance, if developers build more homes in one segment, prices are expected to decrease in order to clear the market. The indirect price effect depends on the degree of market segmentation. For example, when supply increases in segment h, prices in other segments k that are close substitutes to h can decrease as demand shifts from k to h. The price declines in p_k can put additional downward pressure on prices in segment h. The indirect spillover effects are stronger if there are many segments similar to h.

D.3 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:

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

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 and 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.