Job Market Paper

The Effects of Residential Zoning in U.S. Housing Markets

Jaehee Song*

November 28, 2021

Click here for most recent version.

Abstract

This paper examines the effects of minimum lot area restrictions on housing prices, construction, and residential sorting in the United States. First, I develop a structural break detection algorithm to estimate neighborhood-level minimum lot areas nationwide and provide new evidence on the prevalence and restrictiveness of residential zoning. Second, I use a spatial discontinuity design to evaluate the impact of minimum lot area restrictions. I find that doubling the minimum lot area increases sales prices by 14 percent and rents by 6 percent and intensifies residential segregation. Third, I develop a model of housing demand and supply to estimate households' preferences for neighborhood zoning stringency and regulatory costs in housing construction. I find that white households have strong preferences for strict zoning in their neighborhoods. I use the estimated model to evaluate a counterfactual zoning reform that halves minimum lot areas in Connecticut. The reform would substantially increase the supply of small and cheap homes and benefit racial minorities, while minimally affecting existing home values.

*Yale University, Department of Economics (email: jaehee.song@yale.edu). I thank Steven Berry, John Eric Humphries, and Costas Meghir for continuous support and helpful comments throughout the project. Jason Abaluck, Joseph Altonji, Paul Goldsmith-Pinkham, and Phil Haile provided valuable feedback that improved many parts of the paper. I also thank Julian Aramburu, Charles Cai, Paula Calvo, Deniz Dutz, Lucas Finamor, Robert Finlay, Hortense Fong, Disa Hynsjö, Yujung Hwang, Soonwoo Kwon, Jaewon Lee, Cormac O'Dea, Ahyan Panjwani, Luca Perdoni, Vitor Possebom, Suk Joon Son, Stephanie Weber, Seth Zimmerman, and participants of the Labor/Public Finance and Industrial Organization workshops at Yale University for helpful discussions. I gratefully acknowledge financial support from the Tobin Center for Economic Policy.

1 Introduction

Local governments in the United States regulate housing supply through residential zoning, which restricts the quantity and type of housing constructed. In response to recent shortages of affordable housing, policymakers and advocates have pushed for relaxing zoning regulations. Notably, they demand statewide and federal-level zoning reforms, arguing that local governments do not have sufficient incentives to relax regulations.[1] Proponents of relaxing zoning regulations argue that allowing higher-density developments will lower housing costs and increase the access of low-income households and racial minorities to higher-opportunity neighborhoods. In contrast, opponents argue that the construction of denser housing will decrease neighborhood quality and that it will lower existing property values.

The implications of such zoning reforms are not straightforward, as zoning regulations affect multiple dimensions of housing markets. First, strict zoning restricts density, which implies that each housing unit is required to occupy more land. Accordingly, housing units are more likely to be expensive. Second, zoning regulations affect neighborhood environments. Strict zoning may develop better neighborhood amenities through increased property tax bases, and relaxed density may itself be a valued amenity. Lastly, large-scale reforms are especially difficult to evaluate because they affect multiple municipalities simultaneously. Since households may move across locations, city-level or other local analyses that are common in the economics literature (Anagol et al., 2021; Kulka, 2019) do not directly extend to this setting.

This paper studies the impact of minimum lot area restrictions on housing prices, construction, and residential sorting across the United States. Minimum lot area restriction is the most common dimensional requirement in zoning, requiring each residential lot to occupy a certain minimum amount of land.[2] It is a crucial component in residential zoning, especially in subur-

1. Such zoning reforms have already been passed in Oregon in 2019 (House Bill 2001) and California in 2021 (Senate Bill 9) to allow duplexes in previously single-family zones. Similar “upzoning" reforms to relax zoning regulations have been proposed in Connecticut, Maryland, and many other states. In 2021, the federal government included a proposal in the infrastructure plan to provide incentives to local communities to eliminate local exclusionary zoning practices.

2. For instance, 91% of jurisdictions in the Wharton Land Use Regulatory Index survey and 96% in the Terner Center California Land Use Survey reply that they impose minimum lot area restrictions

ban communities. In this paper, I first construct a unique nationwide data set on minimum lot area restrictions. I then analyze the effects of the restrictions by using a boundary discontinuity design. Finally, I develop and estimate a model of housing demand and supply, and use this model to evaluate the effects of a statewide zoning reform on housing construction, prices, and welfare of different socioeconomic groups.

Zoning data is maintained by local jurisdictions, and no comprehensive nationwide database exists. Hence, existing zoning data sets do not cover broad geographic regions or do not provide information on how specific zoning regulations vary across municipalities. This lack of data has been a major challenge to studying zoning at the national and state levels. Faced with this challenge, I provide a new source of data on minimum lot area restrictions which covers over 29,000 municipalities across 48 contiguous states and Washington, D.C. To do so, I develop an algorithm that detects minimum lot areas from observed building characteristics in property tax records. The key idea is that minimum lot area restrictions make it substantially more difficult to build a house in a lot smaller than the cutoff. If the restriction is a meaningful constraint, then lot areas of new constructions will likely bunch right above the cutoff. I thus detect the point of bunching in the distribution of newly constructed lot areas using a structural break detection algorithm and define it as the minimum lot area.[3] I implement this approach at the neighborhood level, which I define by clustering Census block groups to proxy zoning districts. I validate the minimum lot area estimates by comparing them with a land use survey and hand-collected zoning data sets.

I use this constructed data set to provide an empirical description of the current state and stringency of zoning regulations at the national level. I document three facts. First, strict zoning is commonplace in the United States, with 40% of residential land (16% of single-family homes) subject to minimum lot area requirements of 1 acre (43,560 square feet) or greater. Second, zoning regulations likely distort housing characteristics. In particular, about 16% of single-family homes built since 1970 bunch within 5% of minimum lot areas. This suggests that the minimum lot area restriction is a binding constraint in lot area decisions. Third, homeowner demographics, especially race and ethnicity, are strongly associated with the stringency of zoning. In particular,

3. The algorithm takes a similar approach to OLS-based structural break detection in Andrews (1993) and Zeileis (2005).

white households are substantially more likely to live in neighborhoods with large minimum lot area requirements.

In the second part of this paper, I analyze the effects of minimum lot area restrictions on sales prices, rents, housing density, and residential sorting by using a boundary discontinuity design at municipal borders. I construct housing market outcomes from CoreLogic property tax history, deed records, Multiple Listings Service (MLS) data, and Home Mortgage Disclosure Act (HMDA) data. This rich data allows me to study the near universe of existing properties. Baseline results from the municipal border analysis indicate that doubling the minimum lot area increases sales prices by 14% and rents by 6%, while reducing housing density by 37%. In addition, strictly zoned neighborhoods disproportionately attract white households and wealthy households. These estimates are robust to municipality and school district controls.

I further decompose the price effects into building-level and neighborhood-level effects to investigate the differential impact of minimum lot areas on existing homes and new homes. The results imply that roughly two thirds of the 14% sales price effect is attributed to changes in building characteristics, and the remaining third is attributed to changes in neighborhood environments. Although existing properties are exempt from zoning changes, these results imply that their property values would be negatively affected by relaxation of zoning regulations via neighborhood-level effects. Hence, existing homeowners have an economic incentive to keep (or increase) the zoning stringency in their neighborhoods.

In the last part of this paper, I develop and estimate a model of housing demand and supply to evaluate a counterfactual statewide zoning reform. This framework allows me to account for migration across municipalities and preexisting housing stocks, which are crucial for evaluating large-scale zoning reforms. I first analyze the responses of construction and price to this reform. I then analyze this reform's welfare implications for different socioeconomic groups. On the demand side, households have preferences for zoning stringency and choose the housing location and building type in their core-based statistical area that maximize their utility. On the supply side, landowners make development decisions given their legacy housing stocks and the additional construction costs imposed by zoning and non-zoning supply restrictions. These model features allow me to characterize the locations, types, and amount of new construction under different zoning regimes.

I estimate a static discrete-choice model of housing demand embedded with a boundary discontinuity design (Bayer et al., 2007) using the household-level sales transactions data. In my model, households have preferences for neighborhood and building characteristics, and the preferences are allowed to vary by income and race. My demand estimates suggest that white households have *strong* preferences for strict zoning in their neighborhoods, while Asian, Hispanic, and Black households (ordered by their average preference for strict zoning) have *weak to neutral* preferences for strict zoning. This heterogeneity in housing preferences indicates differential welfare implications of zoning reforms by race and ethnicity. Within racial and ethnic groups, there is relatively little heterogeneity in preference for zoning stringency by income.

The goal of the supply-side framework is to estimate regulatory costs, which capture the costs of building new homes due to supply regulations such as zoning laws. I develop a static model of construction decisions where landowners choose the housing type that maximizes net profits (price minus financial costs minus regulatory costs). In the model, landowners face regulatory costs composed of (1) municipality-level costs applied to all types of housing and (2) municipality-level costs only applied to housing types violating existing zoning codes. I estimate these regulatory costs at the municipality level by using the observed shares of newly constructed lots and the observed compliance rates to minimum lot area restrictions.

I use the estimated model to evaluate a counterfactual zoning reform in Connecticut that reduces minimum lot areas by half. The zoning reform would increase construction by 25% and shift the supply towards smaller (by 7% on average) and cheaper (by $26,900 per housing unit on average) homes. The zoning reform would have little to no effects on existing home values because it changes the zoning stringency in all neighborhoods in the local housing markets simultaneously.

This zoning reform substantially benefits Black and Hispanic households and is approximately equivalent to a subsidy of $4,500 per household. Asian households also benefit from the reform by $1,500 to $4,600, depending on the extent to which neighborhood amenities change in response to the reform. In contrast, the sign of the effect of the zoning reform on white households depends on the response of neighborhood amenities to the reform. If neighborhood amenities fully respond to the reform, white households suffer a welfare loss of $6,800. At the other extreme, if neighborhood amenities do not respond to the reform, they benefit by a value of $4,400. Taken together, the counterfactual zoning reform that halve minimum lot areas in Connecticut would either benefit all race and income groups or act as a redistributive policy that benefits racial minorities. These policy implications broadly apply to other zoning reforms that double allowed housing density, such as allowing duplexes in single-family zoning and doubling maximum floor-to-area ratios.

This paper is related to several bodies of literature. First, this paper contributes to the literature on measuring the stringency of land use regulations. One popular data collection method is to conduct surveys. Works that use survey data either focus on individual states such as California (Quigley et al., 2008; Jackson, 2018; Mawhorter and Reid, 2018), or use data that surveys small numbers of jurisdictions across the United States, such as the Wharton Land Use Survey (Gyourko et al., 2008; Gyourko et al., 2021) or the Brookings Survey (Pendall et al., 2006), which respectively surveyed 2,450 and 1,844 jurisdictions. Another way to measure the stringency of land use regulations is to manually collect and aggregate local zoning ordinances. Notably, MAPC Zoning Atlas (2020), Menendian et al. (2020), and Bronin (2021) each built zoning data sets respectively covering 101 jurisdictions in Greater Boston, 100 jurisdictions in the Bay Area, and 180 jurisdictions in Connecticut. This paper contributes to this literature by proposing a new approach to estimate the stringency of local zoning. In particular, I estimate minimum lot area requirements for over 29,000 jurisdictions across the United States, which cover the near universe of urban and suburban communities.

Second, this paper contributes to the literature studying the impact of land use regulations on housing markets. Glaeser and Gyourko (2002), Glaeser and Ward (2009), and Kok et al. (2014) show the positive association between stricter regulations and higher housing prices. Other papers study the positive association of stricter regulations with less land development (Wu and Cho, 2007) and more racial segregation (Trounstine, 2020). This paper is especially related to works that use a boundary discontinuity design to address the endogeneity of land use regulations (Kahn et al., 2010; Turner et al., 2014; Kulka, 2019; Anagol et al., 2021). I contribute to this literature by decomposing the mechanisms of the price effects of zoning regulations in the border analysis. Moreover, Kulka (2019) and Anagol et al. (2021) adopt a discrete-choice model of neighborhood choices to study minimum lot area restrictions in Wake County, North Carolina, and to study maximum built-area-ratios in Sao Paulo, Brazil, respectively. I extend this framework by allowing households to have preferences for neighborhood zoning stringency and building types and by developing a new supply-side model incorporating the direct effect of minimum lot area on housing construction. The supply-side framework extends and builds on Glaeser and Gyourko (2018), who estimate regulatory costs as the gap between market prices of homes and financial costs of construction under the free entry assumption. Gyourko and Krimmel (2021) take a similar approach to estimate the so-called “zoning tax" in single-family housing markets. I develop a microeconomic model of housing supply that incorporates the regulatory costs imposed by both zoning and non-zoning supply regulations.

Third, this paper is part of a literature studying households' willingness to pay for neighborhood amenities (Black, 1999; Bayer et al., 2007; Handbury, 2019). Bayer et al. (2007) and Guerrieri et al. (2013) find that households have higher willingness to pay for wealthy and high-income neighbors. Similarly, Diamond and McQuade (2019) find that Low Income Housing Tax Credit construction has positive price spillovers in low-income neighborhoods, while having negative price spillovers in high-income neighborhoods. This paper contributes to this literature by estimating households' willingness to pay for neighborhood racial composition and income level in relation to how zoning affects these neighborhood demographics.

The rest of the paper is organized as follows. In Section 2, I describe the data used in the analysis. Section 3 describes the algorithm that estimates minimum lot areas from property characteristics data and characterizes the current state of zoning from the constructed data set. In Section 4, I conduct municipal border analyses on housing density, sales prices, rents, and residential sorting. Section 5 introduces a model of housing demand and supply and uses the estimated model to evaluate a statewide zoning reform. Section 6 concludes.

2 Data

2.1 Property Tax History, Deed, Listings, and Mortgage Data

The primary data sets I use are CoreLogic Tax Assessor data, CoreLogic deed data, and CoreLogic Multiple Listings Services data. I link these data sets using assessors' parcel numbers to assemble rich parcel-level information on property characteristics and sale and rental transactions. Throughout this paper, my focus of analysis is single-family homes. Therefore data description and analyses in this paper are restricted to single-family homes unless noted otherwise.

CoreLogic tax data includes property tax records from 2009 to 2019 for the near-universe of single-family homes. CoreLogic collects this data from local government entities responsible for levying property taxes, which are typically county governments. Each tax record includes property characteristics, such as building type, lot area, number of rooms, and construction year, tax assessed values and tax amounts in each tax year. I use this data to estimate minimum lot area requirements (see section 3.1 for more detail) and construct new construction indicators.

CoreLogic deed data, available since the 1980s, includes deed and mortgage transactions. I use this data in two ways. First, I collect sale transactions of single-family homes from the deed data. The data includes the date and price of sales. Second, I link mortgage deeds with Home Mortgage Disclosure Act (HMDA) loan application register data, which provides information on the race/ethnicity and income of mortgage applicants. Linking this data to Corelogic deed data allows me to characterize households' home purchase decisions by these demographics. I merge these two data sets following the approach of Bayer et al. (2007), which involves matching Census tract, mortgage year, and lender name.[4] The match rate is 58% for CoreLogic deed records of residential properties with mortgage information.[5]

CoreLogic listings data includes sale and rental listings collected from 154 multiple listings

4. Other papers that have implemented this merging procedure include Avenancio-Leon and Howard (2019) and Diamond and McQuade (2019).

5. This match rate is comparable to other papers that use the same approach. For example, Bayer et al. (2007) have a match rate of 60% in the Bay Area.

services (MLS).[6] Each MLS maintains its own database of property listings, and often covers specific local markets. CoreLogic aggregates these databases into a single data set. The data set includes list date and price, and, if the listing is closed, closing date and price. Since I already have sale information from CoreLogic deed data, I focus on rental listings in the MLS data and use closing prices of single-family home rental listings in the analysis. Although rental listings count for only small part of the data, I obtain more than 460,000 single-family home-level rent prices in the analysis sample.

2.2 Geographic Attributes

I geocode HMDA-linked CoreLogic data and merge it with several Census maps both to define municipalities and border regions and to augment neighborhood characteristics. Getting precise locations is particularly important for border analyses. Therefore, I geocode street addresses in CoreLogic data using both Smartystreets and ArcGIS StreetMap Premium. I use this data to complement CoreLogic's longitudes and latitudes data.

I define municipalities by (1) Census places or (2) Census county subdivisions that are not part of Census places. I use both definitions in 20 states where Census county subdivisions function as general-purpose governments, and only use definition (1) in the other states.[7] I identify 29,720 municipalities with more than 100 single-family homes (17,974 places and 11,746 county subdivisions) in CoreLogic data across 48 contiguous states and D.C. that belong to Core-Based Statistical Areas. Among these, I estimate minimum lot areas in 29,164 municipalities where I have sufficient data on lot areas of single-family homes. In the border analysis and the estimation of housing demand and supply, I restrict my attention to the 20 states where county subdivisions are general-purpose governments to ensure the complete coverage of states.[8] I illustrate the geographic coverage of the analysis sample in Appendix Figure C1.

I define border housing units as single-family homes located within 0.5km (0.3 miles) of

6. There are 597 multiple listing services as of 2020, according to the Real Estate Standards Organization.

7. These 20 states are CT, IL, IN, KS, ME, MA, MI, MN, MO, NE, NH, NJ, NY, ND, OH, PA, RI, SD, VT, and WI.

8. Census places are sparsely located. Because municipal border analysis compares adjacent municipalities, using Census places alone is insufficient.

municipal borders. Standard choices of the distance in spatial boundary discontinuity design range from 0.24km (0.15 miles) to 0.56km (0.35 miles) (Bayer et al., 2007; Turner et al., 2014). As for robustness, I also compare results using two other border region definitions: within 0.25km, 0.75km, and 1km of municipal borders. In my analysis, I drop corner parcels that are close to multiple municipal borders.

Lastly, I merge single-family home locations in CoreLogic data into 2019 elementary, secondary, and unified school district maps from Census. School districts are one of the most important factors in housing decisions. At the same time, because school district boundaries may coincide with municipal boundaries, they could be a major threat to the boundary discontinuity design. Therefore, I merge the geocoded data with school district boundaries and select municipal borders where school districts do not change. In the border analysis, I present the results using this subsample, in addition to the baseline results using all within-state municipal borders.

2.3 Summary Statistics

Table 1 reports summary statistics of CoreLogic tax, HMDA-linked CoreLogic deed, and CoreLogic MLS data. I define four samples: full sample (all single-family homes that appear on CoreLogic tax in contiguous United States); 20-state sample (all single-family homes that appear on CoreLogic tax in the 20 states where Census county subdivisions function as general-purpose local governments); analysis sample (single-family homes with robust minimum lot area estimates in the 20 states where Census county subdivisions function as general-purpose local governments); and border sample (within 0.5km of municipal borders among the analysis sample).

The full sample includes 87 million single-family homes. According to the Census, there are 94 million single-family homes in the United States as of 2019. Considering that my sample does not include rural single-family homes not located in Core-Based Statistical Areas, CoreLogic tax data covers nearly all single-family homes in the United States.

My analysis sample includes about 65% of the single-family homes in the 20-state sample.

Table 1 — Summary Statistics (CoreLogic and HMDA)

A. Sample Comparison in the Tax Data

	Full sample (48 states and D.C.)	20-state sample (20 states)	Analysis sample (20 states)	Boundary sample (20 states)
# tax records	753,841,762	331,420,333	208,073,746	63,990,296
# single-family homes	87,199,099	41,479,354	27,172,518	8,141,032
# municipalities	29,313	18,833	9,781	9,610
# counties	1,792	672	564	559
Construction year				
before 1930	8,410,091	6,075,724	3,512,320	874,759
1930 to 1959	16,748,360	8,639,697	6,288,528	1,911,961
1960 to 1989	25,170,460	10,634,640	8,733,284	2,631,668
1990 to 2019	22,031,462	7,208,386	5,920,448	1,919,828
Lot size (sq. ft)				
25th percentile	6,540	7,000	7,405	7,501
50th percentile	9,973	11,863	12,102	11,988
75th percentile	19,166	28,750	27,188	23,958
mean	30,940	43,679	37,593	33,038
std. deviation	87,826	111,198	93,530	81,969

B. HMDA-Linked Deed Data and Multiple Listing Service Data

	HMDA-CoreLogic deed ('90-'19)		MLS rental listings	
	Analysis sample	Boundary sample	Analysis sample	Boundary sample
# transactions	7,699,349	2,355,483	467,312	153,401
# single-family homes	5,860,373	1,793,932	276,322	90,337
# municipalities	8,665	8,292	4,248	3,523
# counties	455	449	273	226
Price ($)				
25th percentile	133,299	134,252	1,167	1,173
50th percentile	204,354	202,124	1,550	1,537
75th percentile	320,720	314,077	2,126	2,054
mean	263,352	258,995	1,824	1,775
std. deviation	219,901	211,823	1,353	1,258
HMDA applicants info.				
% white	75.4	77.0	.	.
% Black	6.1	5.7	.	.
% Hispanic	5.4	4.9	.	.
% Asian/Pacific islanders	4.4	4.1	.	.
mean income ($)	103,182	102,383	.	.
std. deviation	139,042	140,508	.	.
% owner-occupied	93.8	94.5	.	.

Note. This table presents summary statistics for the full sample of single-family homes (SFHs), the analysis sample of SFHs, and the boundary sample of SFHs. The full sample includes SFHs in 48 contiguous states and D.C., excluding those not located in any core-based statistical areas. The analysis sample is restricted to SFHs with estimated minimum lot area and located in the 20 states where county subdivisions are general-purpose governments. The boundary sample is further restricted to those within 0.5 km of municipal borders. Multiple listing service (MLS) rental listings sample only includes closed transactions. Panel A presents summary statistics from tax data. Panel B presents summary statistics for HMDA-linked deed data and multiple listing service data. Prices are adjusted to 2010 dollar value.

The missing 35% of the single-family homes is due to either (i) not enough single-family homes in the municipality (less than 100 units); (ii) missing lot area information in tax data to estimate minimum lot areas; or (iii) estimated minimum lot areas not robust. More details about cases (ii) and (iii) are discussed in Section 3.1. Still, my analysis sample covers a substantial share of single-family homes in the regions of interest.

The boundary sample includes 31% of single-family homes in the analysis sample. Single-family homes in the boundary sample are qualitatively similar to the the analysis sample on observable characteristics. For example, lot sizes and home prices are very similar. Moreover, homeowners in the boundary and analysis samples have similar income distribution and racial composition.

3 Restrictiveness of Residential Zoning

Zoning is a primary tool for local governments to regulate the land development of communities in the United States. It defines allowed land uses, building types, and building dimensions by neighborhood. Zoning was initially introduced in the early 1900s to separate incompatible uses, such as factories and residences, and to promote public health and safety (Silver, 2016). In the mid- to late-1900s, however, it became popular to adopt restrictive zoning policies in order to protect property values and exclude blue-collar workers and racial minorities (Fischel, 2004; Nicolaides and Wiese, 2017; Sahn, 2021).

Such strict zoning laws remain today and have been accused of limiting the supply of affordable housing. The American Jobs Plan fact sheets released by the White House in 2021 mention the challenge to providing affordable housing created by local zoning laws:

> *"For decades, exclusionary zoning laws — like minimum lot sizes, mandatory parking requirements, and prohibitions on multifamily housing — have inflated housing construction costs and locked families out of areas with more opportunities."*

As a growing number of households are housing cost-burdened nationwide, policymakers and advocates have increasingly demanded relaxing zoning regulations.[9] Proponents of zoning reforms expect relaxing zoning regulations will induce more housing supply, especially the supply of affordable housing. Suggested reforms primarily focus on allowing denser housing developments through multi-family housing, decreasing minimum lot areas, and increasing maximum floor-to-area ratios and maximum heights.

Notably, policymakers and advocates have pushed for statewide zoning reforms arguing that local governments may not have enough motivation to relax existing regulations.[10] Local governments have been considered to reflect homeowners' financial and exclusionary motives to keep (or increase) the zoning stringency (Fischel, 2005). Accordingly, state-level or even federal-level intervention in local zoning is proposed due to a lack of motivation for local governments to relax zoning regulations, especially in suburbs.

Zoning reforms often face opposition from existing homeowners, who are concerned that such reforms will change neighborhoods in undesirable ways. Opponents believe dense housing development generates negative externalities, such as noise, pollution, and crime. Furthermore, they are worried that more housing supply in their neighborhood will result in the devaluation of their properties.

This paper evaluates counterfactual statewide zoning reforms on minimum lot-area restrictions, which are the most common type of zoning control. Restrictive zoning takes various forms of local ordinances to mandate low-density housing development, such as banning multiplexes and imposing strict minimum lot areas or other dimensional requirements. These different regulations have the same goals: to require the minimum consumption of land and cap housing density. For example, banning duplexes is conceptually equivalent to doubling the minimum lot area or halving the allowed floor-to-area ratio. Therefore, although the main focus of this paper is minimum lot-area regulations applied to single-family homes, the implications apply to zoning reforms on single-family zoning and other density restrictions.

9. Cost-burdened households are defined by spending more than 30% of their income on housing. According to 2019 ACS five-year estimates, 19.3% of all homeowners and 45.9% of all renters are housing cost-burdened. 49.3% of homeowners and 75.2% of renters who earn less than $50,000 are housing cost-burdened.

10. Oregon House Bill 2001 (passed in 2019), California Senate Bill 9 (passed in 2021), and Connecticut House Bill 6107 (passed in 2021) are examples of recent statewide reforms.

3.1 New Nationwide Dataset of Minimum Lot Area Restrictions

To understand the impact of zoning regulations across the United States, research on zoning regulations with broad geographic coverage is essential. However, existing zoning data sets have limitations. First, most data sets have narrow and/or granular coverage of jurisdictions.[11] Second, land use surveys omit details of zoning ordinances, making them inadequate for studying specific zoning reforms.[12]

To overcome these issues, I develop a scalable algorithm to detect zoning regulations from observed property characteristics and construct a nationwide data set on neighborhood-level stringency of residential zoning. In particular, I estimate minimum lot areas applied to single-family homes by detecting structural breaks in newly constructed lots by neighborhoods. The constructed data set covers 75 million single-family homes in 909 CBSAs across the United States. To do this, I first create proxy zoning district maps by municipality. This step ensures that the constructed data reflect neighborhood-level zoning stringency, as dimensional requirements may vary within a municipality across zoning districts. Then, I apply a structural break detection algorithm at each proxy zoning district to estimate minimum lot areas. I describe these steps in greater detail in what follows.

Proxying Zoning Districts

To construct proxy zoning districts, I partly use zoning district information in CoreLogic tax data. This zoning district data is not always clean and has limited coverage. Therefore, I use the data only in the municipalities where (1) the missing rate is less than 30%, or (2) the missing rate is less than 70% and the data is missing at random geographically.[13] This zoning district data

11. Data scarcity is especially relevant when data is manually collected; manual compilation of local zoning laws is very expensive and implausible at the national level, as the United States has over 30,000 zoning jurisdictions. For instance, Bronin (2021) assembles zoning data in 180 jurisdictions in Connecticut, taking more than four months with 20 research assistants.

12. For example, the Density Restriction Index (DRI) in the Wharton Residential Land Use Regulation Index indicates whether a community has any minimum lot size requirement and, if it does, whether the largest minimum lot size is smaller than 0.5 acres, between 0.5 and 1 acre, 1 to 2 acres, or larger than 2 acres. As such, it is common to only report the "strictest" or the "most typical" regulation levels in intervals.

13. I test the randomness by comparing the mean of longitudes and latitudes of single-family homes by the data missing status. I approximate the distribution of the test statistic by resampling data within a municipality fixing the data missing rate. See Appendix A.1 for more details.

is valid in 8,203 municipalities under these criteria (28% of all municipalities). In these 8,203 municipalities, I impute missing data using a random forest algorithm based on the locations of single-family homes. In the other 20,691 municipalities, I cluster Census block groups by land use mixes and building characteristics to construct proxy zoning districts. See Appendix A for more detail on this construction.

I choose the Census block group for the spatial unit of interest because its boundary partly reflects local legal boundaries and its map is available nationwide in a shapefile format. Although constructed proxy zoning districts may not exactly map onto actual zoning districts, they aim to capture the category of zoning districts, such as high-density mixed districts or low-density residential districts. They help illustrate the variation in zoning stringency within a municipality, especially in large cities.

Detecting Minimum Lot Areas

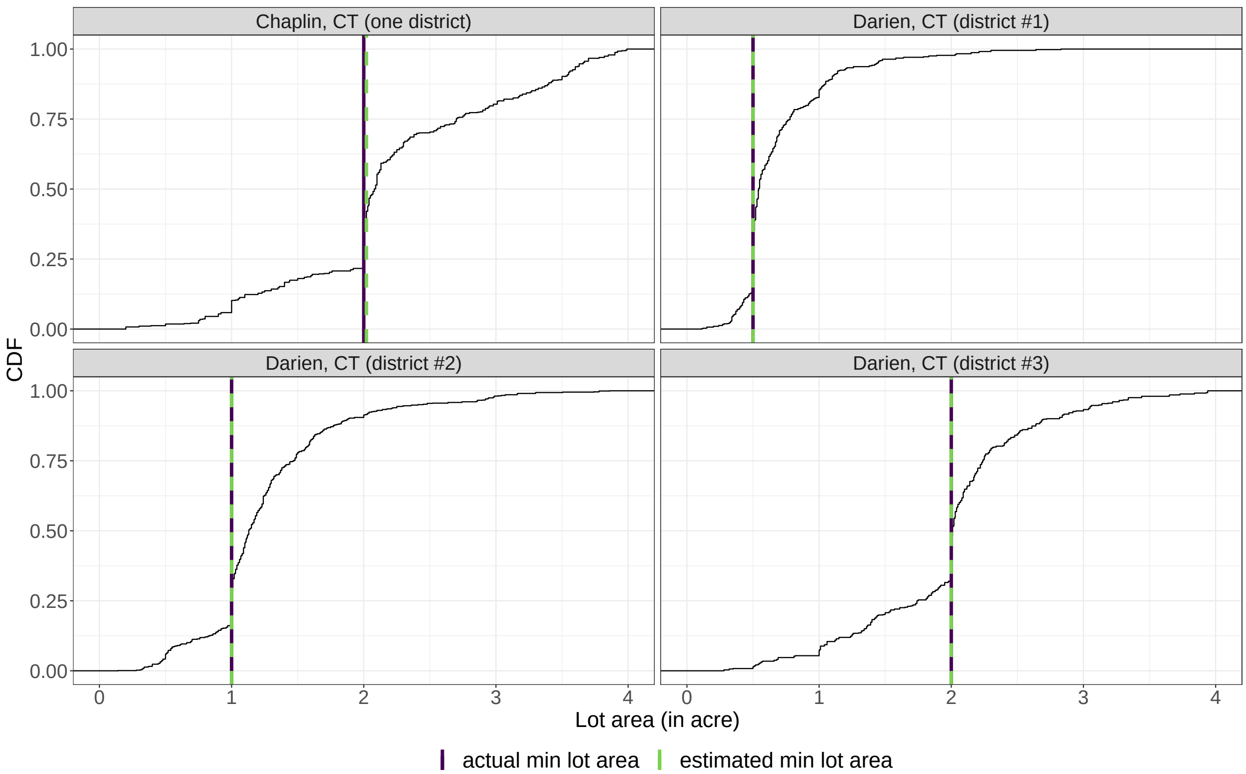
New constructions must comply with minimum lot area regulations unless they meet exception criteria or apply for zoning variances or rezoning. Such non-compliance is expensive and uncommon. Therefore, new constructions that aim for small lot area are forced to construct at the minimum lot area or not construct at all. This creates a discontinuity or kink in the distribution of residential lots at the minimum lot area. Figure 1 illustrates such structural breaks in four example zoning districts. In these districts, structural breaks in the distribution functions are evident at actual minimum lot area requirements. I confirm that, in the 119 municipalities where I manually collect local zoning data, such kinks in lot area distributions consistently appear in many cases and coincide with their minimum lot areas.

I detect kinks in the distribution of lot areas motivated by the structural break detection literature (Andrews, 1993; Zeileis, 2005) and define these break points as the minimum lot areas. Specifically, I solve the following minimization problem:

$$\min_{\bar{x}_h}\left[\min_{F_{h,1},F_{h,2}\in\mathcal{F}}\frac{1}{N_h}\sum_{x_{i,h}}\left(\hat{F}_h(x_{i,h})-F_{h,1}(x_{i,h})\mathbb{1}(x_{i,h}<\bar{x}_h)-F_{h,2}(x_{i,h})\mathbb{1}(x_{i,h}\geq\bar{x}_h)\right)^2\right]$$

where h is a proxy zoning district, i is a single-family home, $x_{i,h}$'s are the observed lot areas, N_h is the # of single-family homes in the zoning district, and $\hat{F}_h(\cdot)$ is the empirical distribution of

Figure 1 — Minimum Lot Area Detection from the Distribution of Constructed Lot Areas



Note. The figure depicts the distribution of lot areas of single-family homes built after 1940 in four example zoning districts. For each district, actual minimum lot areas are denoted by purple solid lines and estimated minimum lot areas are denoted by dashed green lines. In Chaplin, the actual minimum lot area is 2 acres municipality-wide. In Darien, which is comprised of three districts, districts #1–#3 respectively have 1/2-acre, 1-acre, and 2-acre minimum lot areas. The estimated minimum lot areas are constructed by detecting structural breaks in the lot area distribution, as described in Section 3.1.

single-family home lot areas in the zoning district. $F_{h,1}$ and $F_{h,2}$ are the (estimated) cumulative distribution functions for single-family homes smaller than the minimum lot area and for single-family homes bigger than or equal to the minimum lot area, respectively. $F_{h,1}$ and $F_{h,2}$ are chosen from a family of smooth functions ($\mathcal{F}$), which I set to be the family of seventh-order polynomials. For each zoning district h, I identify the value $\bar{x}_h$ that minimizes the sum of squared residuals with a single breakpoint at each zoning district. I use this estimate as my minimum lot area estimate.

I focus on minimum lot area regulations for single-family homes in urban and suburban places. I exclude rural areas by only considering single-family parcels in CBSAs. Since most local communities adopted zoning regulations in the mid-1900s, I further restrict parcels to those built after 1940. I assume that minimum lot areas have remained constant since the mid-1900s, and that each proxy zoning district is subject to at most one minimum lot area cutoff.

Selecting Robust Minimum Lot Area Estimates

One threat to this minimum lot area detection is development of houses with similar dimensions. Such standardized development is common to take advantage of economies of scale, especially in suburbs. To address this issue, I only select minimum lot areas robust to the choice of construction period. Specifically, I estimate structural breaks among single-family homes built after 1940 and among those built after 1970.[14] I drop the minimum lot area estimates that substantially vary by different choice of construction period (26% of all estimates) from the analyses.

Validating the Minimum Lot Area Estimates

I compare my minimum lot area estimates with minimum lot area measures from two other sources: the Terner Center California Land Use Survey of 2017-2018 (TCLUS) in 256 jurisdictions, and a hand-collected zoning data set in 119 municipalities across 34 states.[15] Comparing minimum lot area measures across these sources is not straightforward because these three data

14. I perform this step only for municipalities where enough new construction took place in the late 1900s and 2000s.

15. The entire TCLUS sample includes 271 jurisdictions. 256 of them report their municipality-level minimum lot area.

sets have different geographic units. In particular, my zoning data is at the proxy zoning district level, TCLUS is at the municipality level, and the hand-collected zoning data is at the zoning district level and not georeferenced. An additional complication arises from the fact that TCLUS data, which is collected by surveying jurisdictions asking for "typical" minimum lot area, may be biased.

Notwithstanding the difficulty in such a comparison exercise, the quality of my zoning data can be assessed by first confirming that the TCLUS data is consistent with the hand-collected data, and subsequently verifying that my zoning data is consistent with the TCLUS data (and thus the hand-collected data). In the first step, I find that, for 88% of the 65 municipalities that commonly appear in both TCLUS and the hand-collected data, the TCLUS-reported municipality minimum lot area is equal to (up to 10% measurement error) one of the hand-collected zoning district minimum lot areas within that municipality. This indicates that TCLUS-reported municipality-level minimum lot areas are fairly accurate. The median error rate is 0.0% and the mean error rate is 7.4%. In the second step, I find that, for 78% of the 233 municipalities that commonly appear in both TCLUS and my zoning data, the TCLUS-reported municipality minimum lot area is equal to (up to 10% measurement error) one of the estimated minimum lot areas within that municipality. The median error rate is 2.3% and the mean error rate is 8.9%. Together, these two steps indicate that my estimates successfully capture the "typical" minimum lot area in municipalities, and provides evidence in favor of the quality of my data.

3.2 The State of Local Zoning Laws in the United States

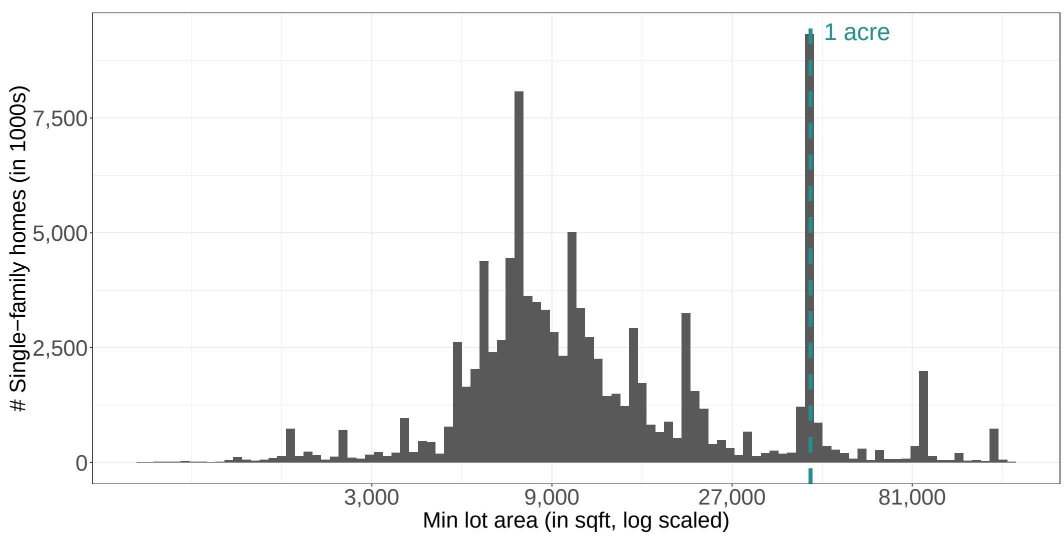
I provide new descriptive evidence of the prevalence and restrictiveness of zoning laws across the United States from the constructed data set of minimum lot area restrictions. Understanding the stringency of zoning and its variation across municipalities is especially important in evaluating statewide and federal-level zoning reforms. Moreover, I document that there is a strong relationship between zoning stringency and homeowner demographics.

Finding 1. Strict zoning is prevalent, especially in smaller jurisdictions.

Figure 2 illustrates the distribution of minimum lot areas over the United States. 39% of municipalities have extremely high minimum lot areas (≥ 1 acre) in at least one district. These strictly-zoned neighborhoods account for 40% of the residential land and 16% of existing single-family homes.

Municipalities with strict zoning laws are often smaller communities. The average municipality with an minimum lot area of one acre or more has 1,189 single-family homes. In contrast, places with relatively lax zoning laws are often highly populated areas. For example, the average municipality with a minimum lot area of 0.5 acres or less in all districts has 5,129 single-family homes. This relationship, between the number of single-family homes and the stringency of zoning, conforms to our understanding about the history of zoning: following the end of World War II, many small suburban communities were created, and these communities enacted protective and restrictive zoning policies (Nicolaides and Wiese, 2017).

Figure 2 — Distribution of Minimum Lot Areas



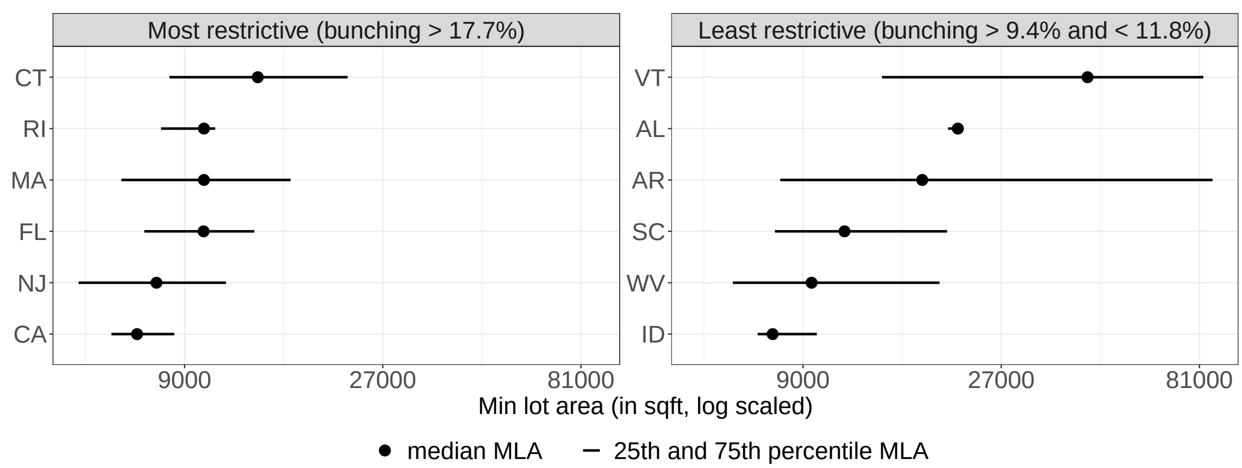
Note. The figure depicts a histogram of minimum lot areas in the constructed data set using the approach described in 3.1. The data is restricted robust minimum lot area estimates of neighborhoods located in municipalities with more than 100 single-family homes. The x-axis is minimum lot areas (in square feet) on a logarithmic scale. The plot depicts the frequency of single-family homes.

Finding 2. Zoning laws distort housing characteristics.

The distribution of constructed single-family lot area exhibits substantial bunching around the minimum lot areas. It indicates that zoning laws are binding constraints in determining building structures. 16% of all single-family homes built after 1970 have lot areas within 5% of the minimum lot area. This suggests that many of these single-family homes would have chosen to be smaller under more relaxed zoning regulations.

Such bunching is common in all states in the United States. Figure 3 illustrates the distribution of minimum lot area by the bunching rate around minimum lot area in each state. Of all states, CA, CT, FL, MA, NJ, and RI have the highest bunching rates, which range from 17% to 23%. In these states, minimum lot area likely substantially distorts the choice of residential lot areas. On the other side of the distribution, AL, AR, ID, SC, VT, and WV have the lowest bunching rates, which range from 9% to 12%. In these latter states, the bunching rates, which can be interpreted as the restrictiveness of local zoning, are relatively low.

Figure 3 — State with the Most and Least Stringent Zoning Restrictions



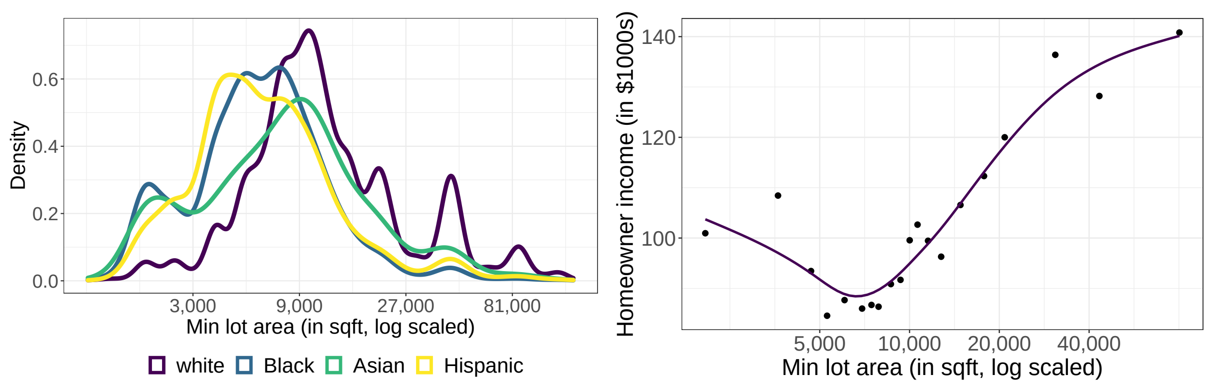
Note. The figure depicts the median (dots) and 25th percentile to 75th percentile ranges (lines) of minimum lot area restrictions (MLA) by state. The percentiles are calculated in terms of the number of single-family homes. States are grouped by the restrictiveness of zoning. The left panel includes six states with the most restrictive zoning, and the right panel includes six states with the least restrictive zoning. See Appendix Figure C2 for all states. The restrictiveness of zoning is defined by the amount of bunching around the MLA for single-family homes built after 1970, which is a proxy for how binding the MLA restriction is. A single-family home is counted as bunching at the MLA if its lot is within 5% of the MLA. In each panel, the states are ordered by their median MLA, a proxy for the stringency of zoning.

Finding 3. Residential sorting is strongly related to zoning stringency.

Minimum lot areas are strongly correlated with homeowner demographics. Figure 4 illustrates the relationship between homeowners' demographics and zoning stringency in their neighborhoods. I find strong racial segregation with respect to neighborhood zoning: white households are far more likely to buy homes in neighborhoods with large minimum lot areas. This racial sorting pattern persists after conditioning on income (Appendix Figure C3). Indeed, white households in the lowest income quartile are more likely to live in strictly-zoned areas than non-white households in the highest income quartile.

Plotting homeowner income by neighborhood zoning stringency in Figure 4 reveals a U-shaped sorting pattern. Homeowners in neighborhoods with the smallest minimum lot areas have high mean incomes, which reflects the above-average mean earnings of highly-urbanized cities such as New York City. Beyond these cities, I find that homeowners in neighborhoods with larger minimum lot areas have higher incomes.

Figure 4 — Patterns of Residential Sorting by Zoning Stringency



Note. In the left panel, the figure depicts the conditional density functions of neighborhood minimum lot areas by race and ethnicity of homeowners (white: purple, Black: blue, Asian: green, Hispanic: yellow). In the right panel, the figure depicts the binscatter plot (with 20 bins) and its fitted line for homeowner incomes. The data is restricted to the 20-state sample as described in Section 2.2 with robust minimum lot area estimates, including single-family homes in both city centers and border regions. Homeowner demographics are taken from HMDA data.

4 Municipal Border Analysis

A major challenge to understanding the causal effects of zoning on housing supply and housing prices is that the stringency of zoning regulations may be correlated with unobserved location amenities. For example, suburban places with higher quality residential environments may set stricter zoning to prevent low-income households from entering. In contrast, populated cities with well-developed infrastructures may allow denser development with relaxed zoning. As such, minimum lot areas may be positively (in the former case) or negatively (in the latter case) correlated with location amenities. As location amenities affect housing market outcomes, such as prices and construction, the OLS coefficients of minimum lot area on these outcomes will be biased.

To address this endogeneity concern, I adopt a boundary discontinuity design developed by Black (1999). In my setting, I compare housing market outcomes at municipal borders where the stringency of zoning changes. The key identification assumption is that unobserved amenities are as good as random within some prespecified small region around the border. This assumption likely holds if location amenities are continuous in geography. Proximity to physical locations and natural landscapes are examples of such amenities. In contrast, discontinuous amenities that change at municipal borders, such as municipal public services or school districts, may be threats to this identification strategy. To validate my empirical approach, I verify that my results are robust to the inclusion of municipality and school district controls.

As discussed in Section 2.2, I define a home as being near a border if it is located within 0.5 km of a municipal border. Results under different choices of distance are presented in Appendix B. The estimates are robust to different choices of border distance. For any pair of municipalities, I define their border region as the set of homes within 0.5 km of their joint border. To illustrate this idea, in Appendix Figure C4 I show a map of an example city, New Haven, CT, where I depict border homes colored by border regions.

My baseline transaction-level regression takes the form

$$y_{ihbmt} = \beta_{MLA} \log MLA_{hm} + \lambda_b + \lambda_t + \varepsilon_{ihbmt}, \tag{1}$$

where i denotes a single-family home transaction, h is the zoning district, b is the border region, m is the municipality, t is the market (CBSA when using the tax data, or CBSA × transaction year when using deed or MLS data), y is the outcome variable (such as sales prices and density), and MLA_{hm} is minimum lot area (MLA). The border region (b) is defined by an unordered pair of municipalities. Therefore, the regressions include thousands of fixed effects to absorb any location amenities that do not vary across borders. In this specification, I assume the treatment effects of minimum lot area restrictions are linear in log of MLA. In Appendix Figure B3–B5, I present alternative binscatter regressions. The conclusions are largely the same.

4.1 Baseline Results

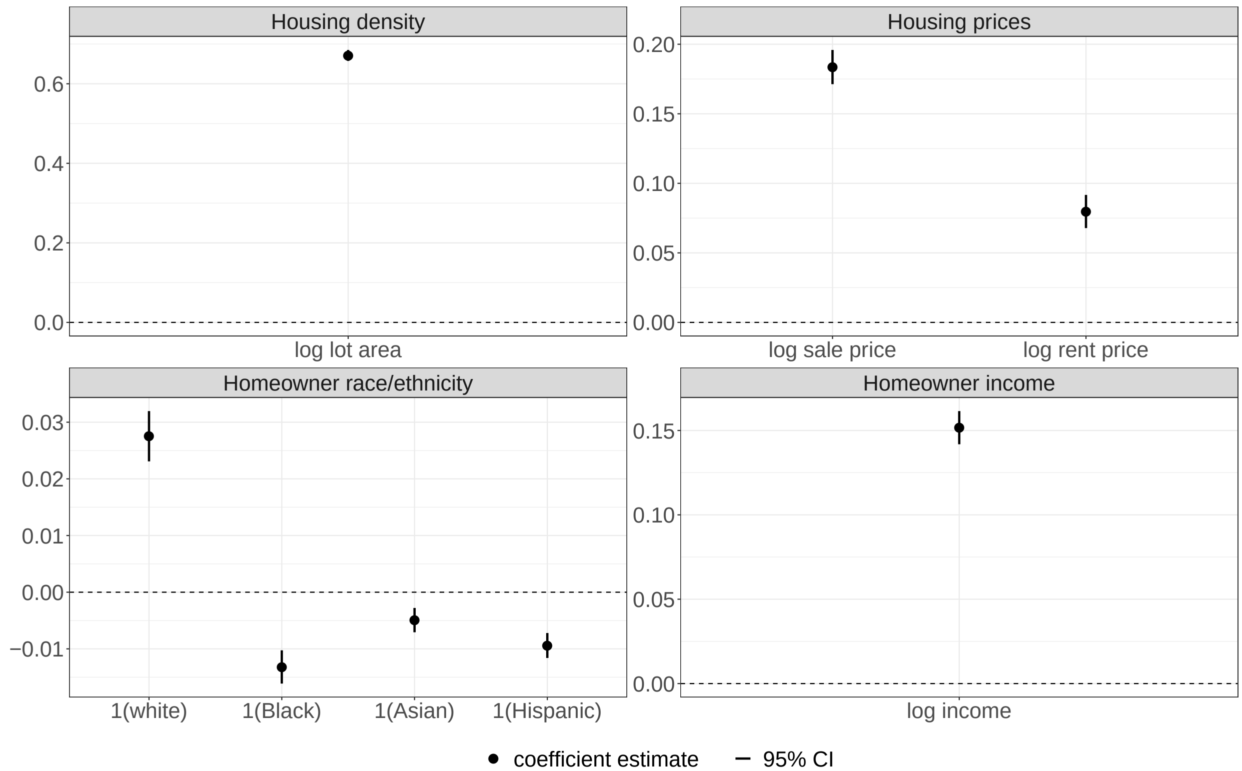
My boundary discontinuity design analysis yields three broad conclusions. First, as expected, stricter zoning (i.e, higher minimum lot areas) increases lot areas and decreases housing density. Second, stricter zoning increases housing prices. Third, stricter zoning disproportionately attracts white and wealthier households.

Stricter zoning increases lot areas and decreases housing density. The first panel of Figure 5 reports $\hat{\beta}_{MLA}$ for lot areas. Minimum lot area restrictions (MLA) increase lot areas of single-family homes (SFHs), with $\hat{\beta}_{MLA} = 0.67$ (p-value < 0.01) for log lot area. Translating this effect in terms of housing density (# of housing units per acre), the estimate indicates that doubling MLA decreases housing density by 37%. This substantial decrease emphasizes that minimum lot area requirements are binding constraints in lot area decisions.

Strict zoning increases housing prices. The second panel of Figure 5 reports $\hat{\beta}_{MLA}$ for sales prices and rents. Both sale and rents increase in response to higher MLA: $\hat{\beta}_{MLA}$ is 0.18 (p-value < 0.01) for log sales price, and $\hat{\beta}_{MLA}$ is 0.08 (p-value < 0.01) for log rents. These results indicate that doubling MLA will increase sales prices by 13.6% and rents by 5.7%. These price premiums capture the overall long-term effects of a higher MLA on several aspects of the housing market, including changes in housing supply, building characteristics, and neighborhood amenities. In Section 4.2, I decompose the different mechanisms that may drive these price effects.

Stricter zoning attracts white households and wealthy households. The bottom two panels

Figure 5 — Estimated Effects of Minimum Lot Areas from Municipal Border Analysis



Note. The figure depicts the estimates (dots) and 95% confidence intervals (lines) of the coefficients of minimum lot area restrictions (MLA) on various housing market outcomes. They correspond to β_{MLA} coefficients of $\log MLA$ in specification 1 with different outcome variables y. The first panel shows the estimated effects on lot areas. The second panel shows the estimated effects on sales prices and rents. The third panel shows the estimated effects on homeowners being white, Black, Asian, and Hispanic. The last panel shows the estimated effects on homeowner income. Robust standard errors are clustered at the municipality level.

of Figure 5 report $\hat{\beta}_{MLA}$ when the outcomes are homeowners' demographics. Examining the effect of zoning stringency on racial compositions, I find that higher MLA induces racial sorting, with homeowners more likely to be white and less likely to be Asian, Black, or Hispanic. The corresponding $\hat{\beta}_{MLA}$ are 0.028, -0.013, -0.005, and -0.009 (all p-values < 0.01).[16] The estimates indicate that when MLA is doubled, homeowners are 1.93pp more likely to be white, 0.91pp less likely to be Black, 0.34pp less likely to be Asian, and 0.65pp less likely to be Hispanic. The reduction in the share of minorities is especially noteworthy given that Blacks, Asians, and Hispanics respectively only make up 6.3%, 4.1%, and 5.0% of homeowners.

16. These results are robust to using Probit specifications.

I also find that higher MLA neighborhoods draw wealthier homeowners. The estimate $\hat{\beta}_{MLA}$ for log income is 0.15 (p-value < 0.01). Thus, doubling MLA will induce homeowners who have, on average, 11% higher income to move in. Importantly, this effect is about the income level of homeowners at the time of housing purchase, and not a statement about whether living in higher MLA neighborhoods causes homeowners to earn more income. Therefore, the results indicate that wealthy households select into high MLA neighborhoods, not that high MLA neighborhoods provide high-paid job opportunities.

These results on homeowners highlight the sorting patterns of demographics with respect to zoning stringency. These patterns potentially arise from heterogeneous preferences for housing and neighborhood characteristics. The demographic group who selects into strictly zoned neighborhoods (white and wealthy households) likely has stronger preference for strict zoning and/or for big lots. In contrast, racial minorities and lower-income households likely have weaker (or even negative) preference for strict zoning and/or for big lots.

4.2 Decomposing Building Size Effects and Neighborhood Effects

In this subsection, I decompose the price effects of minimum lot area restrictions (MLA) by their underlying mechanisms. In particular, I consider two primary mechanisms: (1) changes to minimum lot areas change building characteristics and (2) changes to minimum lot areas change neighborhood environments. The first mechanism reflects the price increase due to changes in building structure in response to changes in the MLA. This building-level effect is driven by the fact that a higher MLA shifts the distribution of lot areas to the right, and thus a purchased home will likely be larger than if the MLA were lower.

The second mechanism governs how zoning regulations shape the neighborhood environments and shift the demographics of your neighbors. If this neighborhood-level effect is strong, houses in more strictly zoned neighborhoods will be substantially more expensive than observably similar houses in less strictly zoned neighborhoods.

I decompose these channels by comparing estimates of $\hat{\beta}_{MLA}$ under the inclusion of building and neighborhood controls. The regression equation is

$$y_{ihbmt} = \beta_{MLA} \log MLA_{hm} + X_i \beta_X + W_h \beta_W + \lambda_b + \lambda_t + \varepsilon_{ihbmt} \tag{2}$$

where X_i are building characteristics, such as lot area, building footprint, and age, and W_h are neighborhood characteristics, such as demographic compositions. The rest of the specification is the same as in Equation 1. Table 2 reports the estimation results for sales price and rent. In what follows, I focus on the sales price effect for demonstration purposes.

In Section 4.1, we showed that the baseline $\hat{\beta}_{MLA}$ for sales price is 0.18. This estimate captures the overall effect of MLA on sales price. I begin by first only including lot area in X_i and omitting W_h in column (2) of Table 2. Under this specification, $\hat{\beta}_{MLA}$ decreases to 0.10 and is still significantly different from zero (p-value < 0.01). This decrease in $\hat{\beta}_{MLA}$ captures the building size effect—houses in higher MLA neighborhoods are big, and therefore more expensive. After controlling for this effect, I find that houses with the same lot area are still 7.5% more expensive.

When I include other observable building characteristics, including the building area and age of the building, while continuing to omit neighborhood characteristics in column (3) of Table 2, the estimate further decreases to 0.063, and this effect is still significant (p-value $<$ 0.01).[17] This additional decrease indicates that minimum lot area requirements affect sales price through changes to building characteristics other than lot area. This result has two potential interpretations. First, minimum lot area regulations may be correlated with other zoning controls. Higher MLA districts typically have stricter zoning controls on other building dimensions as well (Sahn, 2021). In this case, we can interpret MLA as a summary statistic of stringency of zoning, and the baseline $\hat{\beta}_{MLA} = 0.18$ is the effects of overall zoning stringency proxied by MLA. Second, this result may reflect a bundled choice of building characteristics. A typical house with a larger lot area has a larger building area and more rooms, not because of other restrictions but because of the preference of developers or homeowners. As such, MLA changes not only lot area but also other characteristics of houses.

17. Although CoreLogic tax records include many building characteristics such as the number of rooms and roof types, the data missing rate of most variables are considerably high. Therefore, I only use the lot area, building area, and construction year in this exercise.

Table 2 — Decomposing the Effects of Minimum Lot Areas on Sales Price and Rent

	Outcome variables									
	log sales price					log rent				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Zoning stringency										
log min lot area	0.184	0.103	0.063	0.108	0.024	0.080	0.040	0.032	0.049	0.015
	(0.006)	(0.007)	(0.004)	(0.005)	(0.003)	(0.006)	(0.007)	(0.004)	(0.005)	(0.004)
Building characteristics										
log lot area		0.119	0.055		0.053		0.060	0.028		0.027
		(0.006)	(0.004)		(0.004)		(0.005)	(0.003)		(0.003)
log building area			0.613		0.593			0.369		0.360
			(0.007)		(0.007)			(0.009)		(0.009)
age (50 years)			−0.151		−0.145			−0.110		−0.109
			(0.003)		(0.003)			(0.005)		(0.005)
Neighborhood characteristics										
% white				−0.005	0.178				0.044	0.056
				(0.047)	(0.053)				(0.066)	(0.065)
% Black				−0.281	−0.308				0.003	−0.001
				(0.065)	(0.066)				(0.102)	(0.100)
% Asian				0.266	0.153				0.321	0.417
				(0.064)	(0.068)				(0.090)	(0.092)
log median income				0.942	0.492				0.474	0.442
				(0.012)	(0.011)				(0.020)	(0.020)

Note. This table reports the coefficient estimates and their standard errors from border analysis regressions of zoning stringency, building characteristics, and neighborhood characteristics on sales price and rent. See Appendix Table B1 for details about the regression samples. For each outcome, the table presents the results under five different specifications. All specifications include transaction year fixed effects and border fixed effects (for each pair of adjacent municipalities). Columns (1) and (6) regress outcomes on log minimum lot area as a measure of neighborhood zoning stringency. Columns (2) and (7) additionally control for building-specific log lot areas. Columns (3) and (8) additionally include log building area and building age controls. Columns (4) and (9) regression outcomes on log minimum lot area and neighborhood demographics. Columns (5) and (10) regress outcomes on log minimum lot area and all considered building-level and neighborhood-level characteristics. All specification additionally include transaction year fixed effects and border fixed effects. Robust standard errors are clustered at the municipality level and are reported in parentheses.

The remaining $\hat{\beta}_{MLA} = 0.063$ captures the price effects that do not work through building characteristics; hence I define this residual effect as capturing neighborhood-level effects. This result implies that, even holding building characteristics fixed, doubling MLA will lead to a 4.5% increase in sales prices. These neighborhood-level effects apply to all properties, regardless of their structural characteristics.

I further investigate one crucial channel of the neighborhood effects: MLA shifts neighborhood demographics. To do so, I include neighborhood race/ethnicity compositions and median income controls, in addition to building characteristic controls, in the regression specification. The estimation result under this specification is presented in column (5) of Table 2. $\hat{\beta}_{MLA}$ reduces to 0.024 (p-value < 0.01). The reduction of $\hat{\beta}_{MLA}$ from 0.068 to 0.024 indicates that a substantial part of the neighborhood-level effects works through neighbor compositions: higher MLA attracts demographic groups that are preferred by average households in the housing markets.

These neighbor effects depend on two components: (1) the sales price premium for neighborhood demographics and (2) the extent to which MLA shifts neighborhood demographics. First, the price premium for neighborhood demographics is captured in the coefficient estimates of neighborhood demographics in column (5) of Table 2. These coefficients reflect the price premium arising from neighborhood demographics without being confounded by sorting on unobservable amenities, under the assumption of boundary discontinuity design.[18]

The positive coefficient of % white indicates that average households in the housing markets are willing to pay more for having white neighbors. Similarly, average households prefer neighborhoods with a higher share of Asian households and with higher median income, while they disprefer neighborhoods with a higher share of Black households.[19] Furthermore, the findings in Section 4.1 suggest that higher MLA attracts more white households and wealthier households. Taken together, we conclude that a stricter MLA draws more white households and higher-income households who are more preferred as neighbors relative to minorities and lower-income households by average households in the housing markets. It accordingly increases sales prices.

The remaining $\hat{\beta}_{MLA}$ of 0.024 in column (5) of Table 2 reflects the effects of minimum lot area on unobservable neighborhood factors. For example, these factors may include the value of feeling safer in low-density residential neighborhoods. Another example is the "insurance" value of strict zoning regulations against future neighborhood changes. For example, areas with lax residential zoning may be more susceptible to potential unfavorable changes in neighborhood environments. As such, this $\hat{\beta}_{MLA}$ reflects any effects of MLA through other channels than

18. Bayer et al. (2007) conduct a similar exercise with school attendance zone boundaries.

19. Note that this statement does not take a stand on the basis of the sales price premium. In particular, this analysis does not identify whether the price premium arises from racial animus or homophily.

building characteristics and neighborhood demographics.

This decomposition exercise is critical in understanding the local politics of zoning. Local zoning regulations reflect homeowners' interests (Fischel, 2005). Although existing properties would be grandfathered and thus immune to new zoning reforms, their property values will still be affected by zoning reforms through the neighborhood-level effect. Hence, the positive and significant estimate of neighborhood-level effect partly explains homeowners' opposition to zoning reforms. Homeowners have an economic motivation to keep or increase the zoning stringency in their neighborhoods.

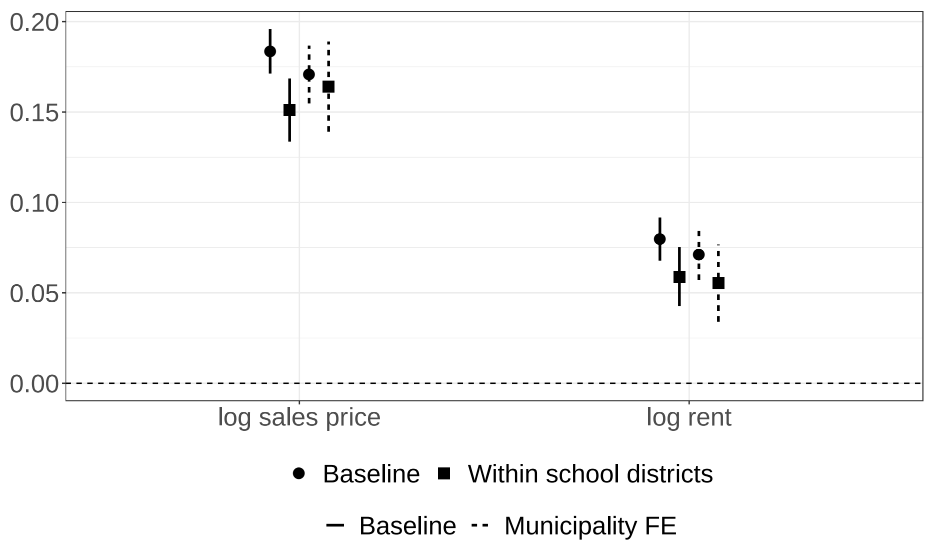
4.3 Incorporating Municipality and School District Controls

Causally interpreting the above border estimates relies on the key assumption that unobserved amenities of neighborhoods within the same border region are as good as random, except for those created by residential zoning. If public services and school district quality discontinuously change at municipal borders and are not driven by zoning, for example, then this assumption will be violated.

To address this concern, I first estimate β_{MLA} when including municipality fixed effects. Municipalities that are adjacent to multiple neighboring municipalities and have variation in zoning stringency across their border regions allow me to include the fixed effects in the analysis. Including municipality fixed effects will control for municipal services such as refuse removal or utility services. Second, I restrict my analysis to municipal borders where school districts do not change and estimate β_{MLA}. This specification will control for the quality of school districts. The results for all outcome variables are presented in Appendix Figure B2. The conclusion stay the same in both specifications: higher minimum lot area results in lower housing density, higher prices, and residential sorting with respect to race/ethnicity and income.

In particular, Figure 6 depicts the estimation results for sales price and rent. The coefficients of minimum lot area on these outcomes decrease slightly with the inclusion of municipality fixed effects and restricting municipal borders to those within school districts. These decreases indicate that neighborhoods with larger minimum lot areas tend to have better municipality-level amenities and better school districts. Despite this decrease, the results show that the price effects of minimum lot area are still substantial and positive under these specifications. While this analysis highlights the robustness of the main border analyses results, these estimates should be interpreted with a caveat. Namely, municipal services and school district quality may be part of the price effects that we would like to include in the estimation rather than controlling for them. For example, imposing stricter minimum lot area restrictions increases the property tax collection. Increased local tax base may result in better public services and/or increased local funding for school districts, therefore resulting in higher sales prices and rents.

Figure 6 — Robustness to Municipality and School District Controls



Note. This figure depicts the estimated effects of minimum lot area $\hat{\beta}_{MLA}$ (dots and squares) and their 95% confidence intervals (solid and dashed lines) on sales price and rent under four different specifications. Results depicted with a dot correspond to specifications using all within-state municipal borders, and results depicted with a square correspond to specifications restricted to within-school district municipal borders. Results with solid 95% confidence intervals correspond to specifications without municipality fixed effects, while results with dashed 95% confidence intervals correspond to specifications with municipality fixed effects. The leftmost specifications depicted with a dot and solid line correspond to the baseline specification (as in the second panel of Figure 5). Robust standard errors are clustered at the municipality level.

5 Evaluating Zoning Reforms

In the previous section, I presented empirical evidence highlighting the significant impact of minimum lot area restrictions on housing prices, density, and residential sorting. I obtain these estimates using the cross-sectional variation in minimum lot areas under the current housing market equilibrium. These estimates can be used to evaluate local changes in zoning regulations. However, these results are not suitable to evaluate the effects of large-scale zoning reforms that affect broad geographic regions due to general equilibrium concerns. In particular, households migrate between municipalities, and therefore zoning changes in one municipality affect other municipalities. Therefore, local zoning changes and broader zoning changes likely have very different effects on the housing market.

This section develops a framework of housing demand and supply. This model allows us to understand the underlying housing preferences and supply factors affected by zoning regulations and characterize household location decisions and housing construction in response to zoning reforms. Importantly, the framework allows us to evaluate how zoning reforms affect multiple areas simultaneously. In addition, it accounts for existing housing stocks and remaining vacant land, which allows us to estimate the supply response more precisely. I apply this framework to estimate the effects of a statewide zoning reform in Connecticut that halves minimum lot areas.

On the demand side of the model, households choose the optimal location and housing type to purchase in their Core-Based Statistical Area. This choice depends in part on their preferences for neighborhoods and building characteristics. In particular, the model allows households to have preferences for amenities generated by the imposition of zoning regulations in neighborhoods. I estimate the demand-side model embedded with a municipal border discontinuity design.

On the supply side, landowners (who make development decisions in the model) choose which housing types to develop, and also have the option of keeping preexisting properties. Therefore, I characterize the supply response to zoning reforms, accounting for existing housing stocks. Landowners factor in the additional construction costs imposed by zoning and non-zoning regulations. These costs reflect the local regulatory environments of housing construction, and thus vary by municipality. Moreover, landlords can build small homes violating existing zoning codes by paying extra costs, which allows for potential zoning variances and exemptions.

5.1 Housing Demand

5.1.1 Housing Choice Model

I model households' housing decisions using a standard discrete choice framework developed by Bayer et al. (2007). Household i in Core-Based Statistical Area (CBSA) t chooses a pair of housing type j and neighborhood h to purchase among available alternatives $\mathcal{J}_t$. Housing type j is defined by observable building characteristics. Neighborhood h is defined by the combination of municipality m, subregion b, and neighborhood minimum lot area requirement (MLA). Subregion b includes municipality centers (for each municipality, a collection of houses with no municipal borders within their 0.5km radius) in addition to border regions (for each unordered pair of municipalities, a collection of houses within 0.5km of their joint border).

The utility for household i from choosing (j,h) is

$$u^i_{jh} = \beta^i_{MLA}\log MLA_h + X_j\beta^i_X - \beta^i_p \log p_{jh} + \lambda_{b(h)} + \varepsilon_{ij} \tag{3}$$

where MLA is the neighborhood minimum lot area, p is the sales price, X is a vector of building characteristics including the lot area and age. $b(h)$ indicates the subregion where neighborhood h belongs to and λ_b is the subregion fixed effect. CBSA t is omitted for notational simplicity.

I assume that subregion fixed effects capture inherent location amenities, and that the $\log MLA$ term captures neighborhood amenities produced by zoning regulations. The intuition follows from the municipal border analysis. Consider a border region b between Town A and Town B. Houses in this region provide similar location amenities, such as landscape and access to physical locations, regardless of on which side of the border they are located. λ_b captures such inherent location amenities, irrelevant to zoning regulations. The two neighborhoods (the border region on city A and the border region on city B) also have developed different neighborhood amenities due to zoning regulations, such as demographics of neighbors and local tax base. I do not specify neighborhood amenities developed by zoning and capture them collectively in $\log MLA$ term in Equation 3.

The preferences for housing characteristics, $(\beta^i_{MLA}, \beta^i_X, \beta^i_p)$, depend on race/ethnicity (z^i_1) and income (z^i_2) of household i.[20] In particular, they take the functional form of

$$\begin{pmatrix} \beta^i_{MLA} \\ \beta^i_X \\ \beta^i_p \end{pmatrix} = \begin{pmatrix} \beta^{z^i_1}_{MLA,0} \\ \beta^{z^i_1}_{X,0} \\ \beta^{z^i_1}_{p,0} \end{pmatrix} + z^i_2 \cdot \begin{pmatrix} \beta^{z^i_1}_{MLA,1} \\ \beta^{z^i_1}_{X,1} \\ \beta^{z^i_1}_{p,1} \end{pmatrix}.$$

$\beta^{z_1}_0 \equiv (\beta^{z_1}_{MLA,0}, \beta^{z_1}_{X,0}, \beta^{z_1}_{p,0})$ captures the mean preference of race/ethnicity and $\beta^{z_1}_1 \equiv (\beta^{z_1}_{MLA,1}, \beta^{z_1}_{X,1}, \beta^{z_1}_{p,1})$ captures the slope of the preference parameters with respect to income. Assuming that ε_{ij} is idiosyncratic shocks following the type-1 extreme value distribution, household i chooses (j, h) in market t with probability

$$Pr_t(j,h|z^i) = \frac{\exp(\beta^i_{MLA} \log MLA_h + X_j \beta^i_X - \beta^i_p \log p_{j,h} + \lambda_{b(h)})}{\sum_{(j',h')\in\mathcal{J}_t} N_{j',h'} \cdot \exp(\beta^i_{MLA} \log MLA_{h'} + X_{j'} \beta^i_X - \beta^i_p \log p_{j',h'} + \lambda_{b(h')})}$$

where $N_{j,h}$ is the number of type-j housing units in neighborhood h.[21] $\mathcal{J}_t$ is the set of single-family homes that appear in HMDA-linked CoreLogic deed data. I assume that all households in CBSA t have the same choice set of $\mathcal{J}_t$ and do not consider single-family homes that were not transacted during the time period.

20. Race and ethnic group z^i_1 is obtained from HMDA data and takes a value among white, Black, Asian, and Hispanic. Income level z^i_2 is normalized at CBSA levels by subtracting the mean income and dividing by the standard deviation.

21. The implicit assumption here is that a "product" in this market is defined as each house instead of each housing type, although several houses in the same neighborhood may have the same set of observable characteristics (and thus the same price).

5.1.2 Demand Estimation

I estimate $(\beta_0^{white}, \beta_0^{Black}, \beta_0^{Asian}, \beta_0^{Hispanic})$, $(\beta_1^{white}, \beta_1^{Black}, \beta_1^{Asian}, \beta_1^{Hispanic})$, and $\{\lambda_b\}$ in a one-step procedure maximizing the following log likelihood.

$$LL(\beta_0, \beta_1, \lambda) = \sum_t \sum_i \log \left(\frac{\exp(\beta^i_{MLA} \log MLA_{h(i)} + X_{j(i)}\beta^i_X - \beta^i_p \log p_{j(i),h(i)} + \lambda_{b(h(i))})}{\sum_{(j',h') \in \mathcal{J}_t} N_{j',h'} \cdot \exp(\beta^i_{MLA} \log MLA_{h'} + X_{j'}\beta^i_X - \beta^i_p \log p_{j',h'} + \lambda_{b(h')})} \right) \tag{4}$$

where $j(i)$ and $h(i)$ indicate the housing type and neighborhood chosen by i.

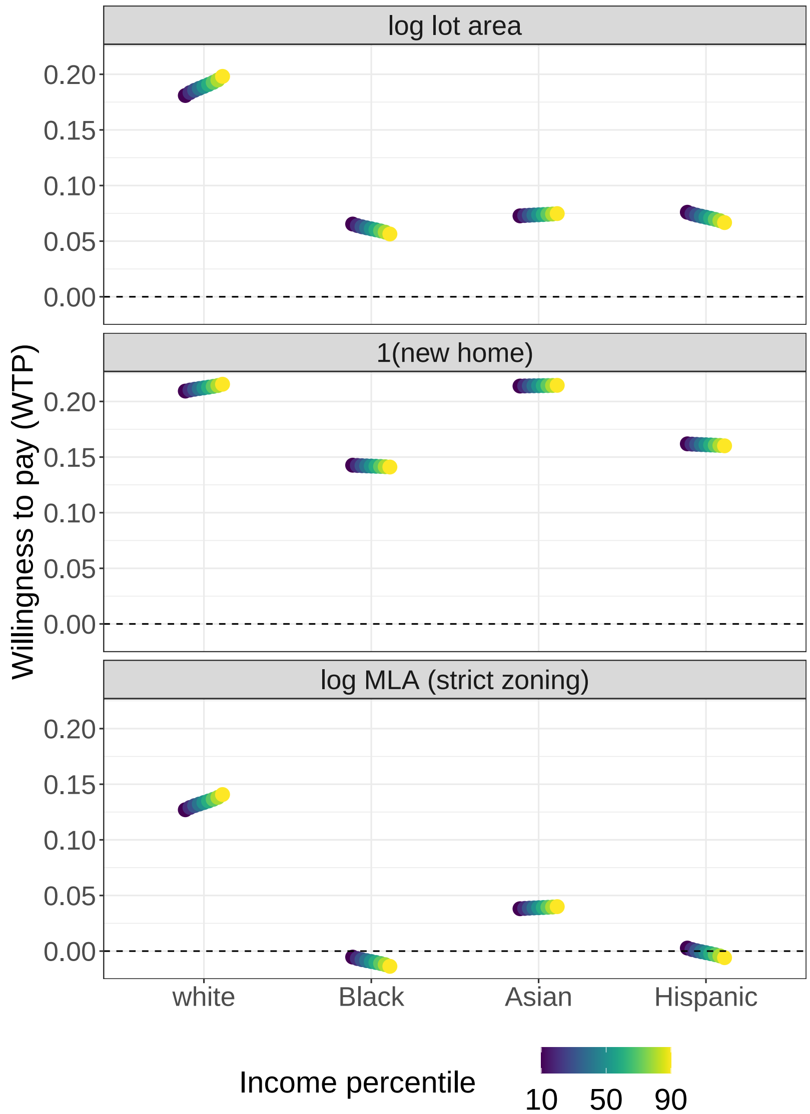
I estimate this static demand model in 6 CBSAs in Connecticut from 1989 to 2018. Solving this numerical optimization is computationally challenging due to its large set of fixed effects (1,327 subregions) and alternatives (395,429 housing types in total). To ease the computational burden, I discretize income (z_2^i) into 100 bins in each CBSA by race/ethnicity. This discretization reduces the number of demographic types (z_1^i, z_2^i) from 72,429 to 2,399. Then, I solve the numerical optimization using the Newton-Raphson method, taking advantage of the known analytical form of the first derivative of Equation 4.

5.1.3 Estimation Results: Preferences for Strict Zoning and Lot Areas

The key demand-side factors determining the impact of zoning regulations on housing markets are the willingness to pay (WTP) for zoning stringency and lot area of houses. The WTP of household i for characteristic x is defined as β^i_x/β^i_p. It reflects how much price premium (in %) the household is willing to pay for an additional unit of characteristic x. Figure 7 presents the estimated WTP of lot area of houses, age of houses, and the neighborhood minimum lot area by demographics. The complete set of estimated demand coefficients (β_0, β_1) are shown in Appendix Table C1.

Preference for neighborhood zoning stringency varies substantially across race and ethnic groups. White households at the 50th income percentile have a WTP for minimum lot area (MLA) equal to 0.134. The estimate implies that they are willing to pay *9.73% more* to live in neighborhoods with twice as strict zoning. In contrast, Black households at the 50th income percentile have a WTP for MLA equal to -0.007. It implies that they are willing to pay *0.51% less* for neighborhoods with twice as strict zoning. Asian households with the median income and

Figure 7 — Estimated Willingness to Pay by Demographics (in Connecticut)



Note. The figure depicts the estimated willingness to pay (WTP) for three housing characteristics: (1) log lot area of the building (top panel), (2) being a new home, defined by construction within 30 years (middle panel), and (3) minimum lot area restriction of the neighborhood (bottom panel). For each characteristic x, the WTP is defined by the ratio between the demand coefficient of the characteristics and the coefficient of log price ($\beta_x^i / \beta_{\log price}^i$ in equation 3). For each panel, nine WTPs are presented for each race/ethnicity (white, Asian, Hispanic, and Black). These nine WTPs correspond to the 1st to 9th income deciles in the CBSA. Race/ethnicity and income data are obtained from Home Mortgage Disclosure Act data.

Hispanic households with the median income are willing to pay *2.75% more* and *0.04% more* for twice as strict zoning, respectively.

Higher-income white households have stronger preferences for minimum lot area compared to lower-income white households, but the heterogeneity is not as prominent as the race/ethnicity dimension. For instance, white households in the 95th income percentile have a WTP for MLA equal to 0.143. In comparison, white households in the 5th income percentile have a WTP for MLA equal to 0.126. Even the low-income white households are willing to pay *9.10% more* for twice as strict zoning.

These WTPs for minimum lot area is a reduced-form estimate of the preferences for neighborhood amenities produced by zoning regulations. For instance, the 9.75% price premium white households are willing to pay for stricter zoning reflects their preference for neighborhood characteristics induced by stricter zoning. Such characteristics include changes in demographics of neighbors due to stricter zoning, local services provided from increased property taxes, and the insurance value on home prices of zoning. As such, we can interpret the price premium as indicating that white households are willing to pay 9.75% more to enjoy the neighborhood characteristics that come with doubling the zoning stringency.

Another relevant housing characteristic to understand the effects of minimum lot area restrictions is the lot area of the house. Minimum lot area restrictions increase the supply of bigger homes by design. Therefore, we expect households with stronger preferences for big homes to benefit more from higher minimum lot areas. The results indicate that households prefer bigger lot areas overall, but white households have far stronger preferences. White households at the 50th income percentile have a WTP for lot size equal to 0.190. It implies that they are willing to pay *14.0% more* for houses with twice big lots. In comparison, Black households, Asian households, and Hispanic households at their 50th income percentiles are willing to pay *4.48% more*, *5.26% more*, and *5.24% more* for houses with twice big lots, respectively.

5.2 Housing Supply

5.2.1 A Model of Housing Construction

Landlord ℓ, endowed with land in neighborhood h in municipality m in Core-Based Statistical Area (CBSA) t, makes a static decision to develop their land at a density level of τ over a 30-year time period (1989–2019). Neighborhood h is defined by municipalities, subregions (city centers and border regions), and zoning districts. They have an option to construct no new development (denoted as $\tau = 0$) and keep the value of preexisting properties with per-acre value $\pi_{\ell,0}$. If they choose to develop, they decide on the development density $\tau > 0$, measured in # single-family homes per acre. I assume they have rational expectation on the existing property value $\pi_{\ell,0}$ and on the prices of single-family homes by density type τ in the neighborhood, $\bar{p}_{hmt}(\tau)$, at the time of making the construction decision. Given the information on the prices and construction costs they face, they make a per-acre-profit-maximizing decision.

Landlords face total construction costs to build a new single-family home of type-τ,

$$CC_{\ell hmt}(\tau) = FC_{hmt}(\tau) + RC_{\ell hmt}(\tau)$$

where $FC_{hmt}(\cdot)$ represents the financial costs of construction and $RC_{\ell hmt}(\tau)$ represents the regulatory costs of construction, reflecting the monetary value of the difficulty of home construction due to government regulations. The regulatory costs depend on whether the landlord complies with existing zoning code ($\bar{\tau}_h$):

$$RC_{\ell hmt}(\tau) = \begin{cases} \tilde{\omega}_{\ell mt,1} = \omega_{mt,1} + \nu_{\ell hmt,1} & \text{if } \tau \le \bar{\tau}_h, \\ \tilde{\omega}_{\ell mt,2} = \omega_{mt,2} + \nu_{\ell hmt,2} & \text{if } \tau > \bar{\tau}_h. \end{cases}$$

Here, $\bar{\tau}_h$ represents the density restriction level, measured by the reciprocal of minimum lot area in acre, in the neighborhood. $\omega_{mt,1}$ is the regulatory costs to be paid if the housing type complies with existing zoning code while $\omega_{mt,2}$ is the costs to be paid if the housing type violates existing zoning code. Intuitively, $\omega_{mt,1}$ reflects the overall difficulty in developing new housing in municipality m due to administrative process or supply regulations other than minimum lot area. $\omega_{mt,2}$ includes an additional regulatory cost component due to zoning, specifically costs and difficulty of zoning variance or rezoning. $\nu_{\ell hmt,1}$ and $\nu_{\ell hmt,2}$ are idiosyncratic cost shocks.

The maximum per-acre profit from developing new properties in compliance of $\bar{\tau}_h$ is

$$\pi^*_{hmt,1}(\tilde{\omega}_{\ell mt,1}) = \max_{\tau \leq \bar{\tau}_h} \left(\tau \cdot (\bar{p}_{hmt}(\tau) - FC_{hmt}(\tau) - \tilde{\omega}_{\ell mt,1})\right).$$

Similarly, the maximum per-acre profit from developing new properties in violation of $\bar{\tau}_h$ is

$$\pi^*_{hmt,2}(\tilde{\omega}_{\ell mt,2}) = \max_{\tau > \bar{\tau}_h} \left(\tau \cdot (\bar{p}_{hmt}(\tau) - FC_{hmt}(\tau) - \tilde{\omega}_{\ell mt,2})\right).$$

After the shocks $(\nu_{\ell hmt,1}, \nu_{\ell hmt,2})$ are realized, landlord ℓ chooses the development option, solving

$$\max \Big[\underbrace{\pi_{\ell,0}}_{\text{No development } (\tau=0)}, \underbrace{\pi^*_{hmt,1}(\tilde{\omega}_{\ell mt,1})}_{\text{Building in compliance of } \bar{\tau}_h}, \underbrace{\pi^*_{hmt,2}(\tilde{\omega}_{\ell mt,2})}_{\text{Building in violation of } \bar{\tau}_h} \Big].$$

5.2.2 Supply Estimation

In this model, the density level τ is the only building characteristics landlords determine. Hence, I define τ-type housing as a typical building in the neighborhood with density τ. To illustrate, consider a bundle of all building characteristics, $\{x_1, x_2, ..., x_M\}$, to be chosen by developers, such as height, number of bedrooms, roof type, or interior design. Then, I define τ-type housing in neighborhood h by projecting the characteristics on τ, $\{x_{hmt,1}(\tau), x_{hmt,2}(\tau), ..., x_{hmt,M}(\tau)\}$. In my application, I consider two dimensions of building due to data limitation. Let x_1 be the lot area and x_2 the be building area. By definition, $x_{hmt,1}(\tau) = 1/\tau$. I obtain $x_{hmt,2}(\cdot)$ from penalized cubic splines at each neighborhood.

I estimate the cost parameters $\omega_{mt,1}$ and $\omega_{mt,2}$ at the municipality level by matching the share of lots that are newly constructed and the compliance rate of minimum lot area restrictions in each municipality. The parameters $\omega_{mt,1}$ and $\omega_{mt,2}$, together with $FC_{hmt}(\cdot)$, characterize the choice probabilities of construction actions,

$$D_{\ell hmt} = \begin{cases} 0 & \text{(no construction)}, \\ 1 & \text{(building in compliance of } \bar{\tau}_h), \\ 2 & \text{(building in violation of } \bar{\tau}_h). \end{cases}$$

If both regulatory costs $\tilde{\omega}_{\ell mt,1}$ and $\tilde{\omega}_{\ell mt,2}$ are so high that both building in compliance and in violation of zoning are not profitable ($\pi^*_{hmt,1}(\tilde{\omega}_{\ell mt,1}) < \pi_{\ell,0}$ and $\pi^*_{hmt,2}(\tilde{\omega}_{\ell mt,2}) < \pi_{\ell,0}$), they choose to keep the existing properties. Conditional on building, they comply with $\bar{\tau}_h$ if the cost of violation, $\tilde{\omega}_{\ell mt,2}$, is so high that $\pi^*_{hmt,1}(\tilde{\omega}_{\ell mt,1}) > \pi^*_{hmt,2}(\tilde{\omega}_{\ell mt,2})$. As such, $\omega_{mt,1}$ and $\omega_{mt,2}$ are identified by how much each action is taken given $\tau \cdot (\bar{p}_{hmt}(\tau) - FC_{hmt}(\tau))$. Note that we observe $\bar{p}_{hmt}(\cdot)$ in the data. I fix $FC(\cdot)$ at \$80 per building square foot, which is the median financial construction costs in the United States (Glaeser and Gyourko, 2018). Specifically, $FC_{hmt}(\tau) = 80 \cdot x_{hmt,2}(\tau)$, where $x_{hmt,2}(\tau)$ is the building square footage of type-τ housing in neighborhood h.

$\nu_{\ell hmt,1}$ and $\nu_{\ell hmt,2}$ are independently drawn from a Gumbel distribution with location 0 and scale $1,000$. Then, the probability of construction actions of landlord ℓ is

$$\begin{aligned}
Pr_{hmt}(D_\ell = 0; \pi_{\ell,0}, \Theta_t) &= Pr_{\nu_1,\nu_2}\big(\pi^*_{hmt,1}(\omega_{mt,1} + \nu_{\ell hmt,1}) < \pi_{\ell,0} \text{ and } \pi^*_{hmt,2}(\omega_{mt,2} + \nu_{\ell hmt,2}) < \pi_{\ell,0}\big), \\
Pr_{hmt}(D_\ell = 1; \pi_{\ell,0}, \Theta_t) &= Pr_{\nu_1,\nu_2}\big(\pi^*_{hmt,1}(\omega_{mt,1} + \nu_{\ell hmt,1}) > \pi_{\ell,0} \text{ and } \\
&\qquad \pi^*_{hmt,1}(\omega_{mt,1} + \nu_{\ell hmt,1}) > \pi^*_{hmt,2}(\omega_{mt,2} + \nu_{\ell hmt,2})\big), \\
Pr_{hmt}(D_\ell = 2; \pi_{\ell,0}, \Theta_t) &= Pr_{\nu_1,\nu_2}\big(\pi^*_{hmt,2}(\omega_{mt,2} + \nu_{\ell hmt,2}) > \pi_{\ell,0} \text{ and } \\
&\qquad \pi^*_{hmt,2}(\omega_{mt,2} + \nu_{\ell hmt,2}) > \pi^*_{hmt,1}(\omega_{mt,1} + \nu_{\ell hmt,1})\big).
\end{aligned}$$

where $\Theta_t = \{\omega_{mt,1}, \omega_{mt,2}\}_m$. The idiosyncratic errors enter the choice probabilities nonlinearly, and, therefore, the choice probabilities do not have simple analytical forms. Hence, I simulate $\nu_{\ell hmt,1}$ and $\nu_{\ell hmt,2}$ for a given set of parameters Θ_t to approximate the probabilities. I estimate

Θ_t by solving

$$\max_{\Theta_t} \sum_m \sum_h \sum_{d \in \{0,1,2\}} \left(\underbrace{\frac{\sum_\ell \left(x_{\ell hmt,1} \cdot \mathbb{1}(D_{\ell hmt} = d) \right)}{\sum_\ell x_{\ell hmt,1}}}_{\text{realized share of action } d} - \underbrace{\frac{\sum_\ell \left(x_{\ell hmt,1} \cdot Pr_{hmt}(D_\ell = d; \pi_{\ell,0}, \Theta_t) \right)}{\sum_\ell x_{\ell hmt,1}}}_{\text{predicted share of action } d} \right)^2 .$$

I use the per-acre tax assessed values for $\pi_{\ell,0}$. However, we do not observe the assessed value of previous buildings demolished in most new construction sites. Therefore, I assume that all new properties in the last 30 years were built on empty land and let $\pi_{\ell,0}$ for these landlords to be their per-acre land assessed value. This assumption may lead to overestimation of the cost parameters.

5.2.3 Estimation Results: Regulatory Costs of Single-Family Home Construction

Table 3 shows the distribution of regulatory cost estimates. The regulatory cost if the construction complies with the existing zoning code ($\hat{\omega}_1$) is $110,000 in the median, and the interquartile range is $57,000 to $195,000. These estimates are comparable to the $94,000 regulatory cost estimates from the survey of builders and developers conducted by the National Association of Home Builders in 2021 (Emrath, 2021). The regulatory cost if the construction does not comply with the existing zoning code ($\hat{\omega}_2$) is $164,000 in the median, and the interquartile ranges from $99,000 to $273,000. The results suggest that developers pay substantial costs to get zoning variances or rezoning, and the median of this extra cost is $28,000 per housing unit.

Table 3 — Estimated Regulatory Costs (in Connecticut)

	percentile					
Regulatory costs (in $)	10th	25th	50th	75th	90th	mean
Complying with existing zoning ($\hat{\omega}_1$)	6,221	56,748	109,841	194,909	301,022	162,853
With zoning variance/rezoning ($\hat{\omega}_2$)	65,358	99,361	164,229	272,915	471,993	224,515

Note. This table reports the distribution of regulatory cost estimates in Connecticut.

5.3 Housing Market Equilibrium

The equilibrium prices equalize the supply and demand of each housing type in each neighborhood. Recall that housing type j is defined by structure characteristics X_j and neighborhood h is defined by a combination of municipality (m), subregion (b), and neighborhood minimum lot area. Demand of housing type j in neighborhood h is determined by neighborhood-level housing price schedules $(\{p_{ht}(X_j)\}_h$, denoted by $P_t)$, demand parameters $(\beta_0, \beta_1, \lambda)$, and the distribution of demographics in Core-Based Statistical Area t (denoted by Z_t). Accordingly, I define the demand of housing type j in neighborhood h as $\mathcal{D}_{jht}(P_t; \beta_0, \beta_1, \lambda, Z_t)$. Similarly, the supply of housing type j in neighborhood h is determined by prices (P_t), supply parameters $(\{\omega_{mt,1}, \omega_{mt,2}\}_m$ denoted by $\Omega_t)$, and the distribution of the value of existing properties (denoted by $\Pi_{0,h}$). Therefore, I define the supply of housing type j in neighborhood h as $\mathcal{S}_{jht}(P_t; \Omega_t, \Pi_{0,ht})$. The market-clearing prices P_t satisfy

$$\mathcal{D}_{jht}(P_t; \beta_0, \beta_1, \lambda, Z_t) = \mathcal{S}_{jht}(P_t; \Omega_t, \Pi_{0,ht}), \quad \forall j, h.$$

5.4 Counterfactual Analysis: Connecticut Zoning Reform

The state of Connecticut is currently actively debating introducing sweeping zoning reforms. In 2021, the state senate passed a zoning reform bill which requires towns to allow accessory dwelling units in single-family homes. Advocates continue to push for further reforms, especially regarding the state's extremely restrictive minimum lot area restrictions, as described in Section 3.2. For example, DesegregateCT, a housing advocacy group in Connecticut, has their 2022 policy agenda explicitly set for reducing minimum lot areas. This policy debate makes Connecticut an ideal setting to apply my framework and evaluate a counterfactual zoning reform.

To understand the effects of relaxing minimum lot area restrictions across Connecticut, I evaluate a counterfactual zoning reform that halves minimum lot areas statewide. Conceptually, such a reform is equivalent to allowing duplexes in single-family zoning districts by doubling the minimum allowed density while halving the minimum consumption of land. I examine how such a reform would affect the next thirty years of housing demand and supply. I compare the housing market outcomes and welfare of different demographic groups under such a reform to those when maintaining the existing minimum lot area restrictions.

5.4.1 Equilibrium Assumptions

I consider two types of equilibrium in the counterfactual analysis: partial equilibrium and general equilibrium with market expansion. Partial equilibrium takes the observed price schedules of different housing types and neighborhoods from the data and evaluates the supply responses under different zoning regimes. To do so, I solve the construction decisions under the two alternative zoning regimes given the prices and the existing housing stocks and developable vacant land as of 2019.[22]

In general equilibrium with market expansion, I compute the housing market equilibrium by equalizing matching the market shares of housing types in all neighborhoods. To do so, I assume that the population of Core-Based Statistical Areas (CBSAs) grows as more housing units are constructed, while preserving the distribution of household demographics. The inflow of households may come from the local rental market (i.e., renters decide to buy houses), from other states, and/or from internal population growth. Therefore, if renters are more likely to be minorities and low-income households, for example, the assumption to keep the distribution of household demographics constant may underestimate the demand for small and cheap houses.

To find a general equilibrium, I first solve for the partial equilibrium with the observed sales prices. Then, I solve for market-clearing prices that equate housing demand and supply of all houses in CBSAs and that are the closest to the observed prices. Using these new prices, I solve the construction decisions again. I iterate these steps until I find equilibrium prices and corresponding construction outcomes. Note that the uniqueness of equilibrium is not guaranteed and this procedure picks an equilibrium close to the current equilibrium.

5.4.2 Supply Responses to the Zoning Reform

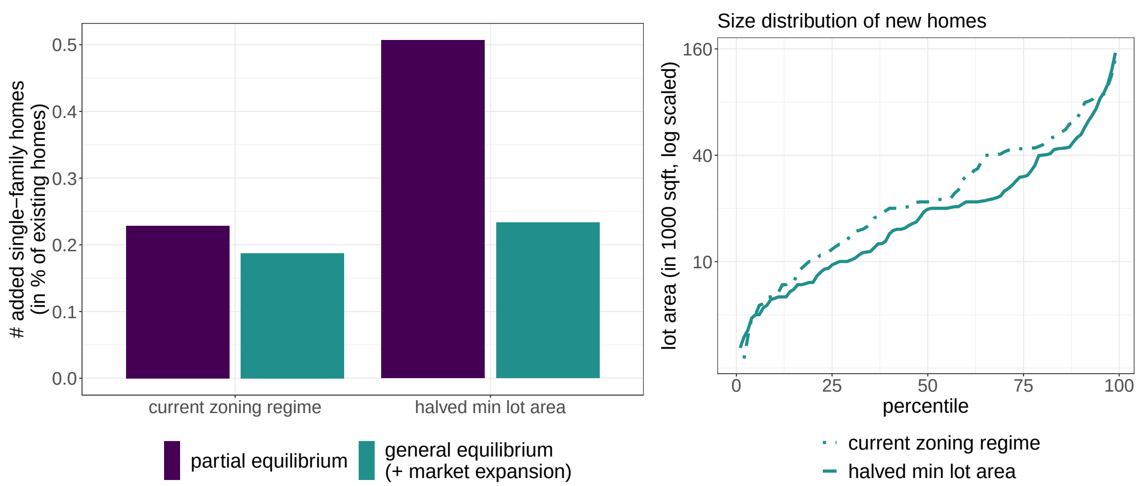
Figure 8 depicts the number of constructed single-family homes both under the current zoning regime and under the counterfactual zoning reform. In the partial equilibrium, zoning

22. Since HMDA-CoreLogic deed sample used in this analysis covers a portion of single-family homes in the tax data (about 25%), I accordingly scale down the amount of vacant land in the counterfactual analysis.

reforms would induce the development of 122% more single-family homes. This occurs because denser development is typically more profitable, and thus relaxing zoning regulations increases housing supply. This substantial increase in housing supply arises from both extensive margins and intensive margins: (1) 82% more vacant land would be developed under the zoning reform; (2) on average, houses would be 7% smaller.

In the general equilibrium, however, zoning reforms would increase the number of construction by only 25%. This is largely due to the fact that, as more housing is constructed, prices would adjust by decreasing, and hence denser housing development would no longer be as profitable. We observe this price adjustment in Figure 9. However, in practice this general equilibrium effect may not be fully internalized in development decisions, as housing markets are prone to overbuild (Grenadier, 1995). Therefore, we expect to find the supply increase in between 25% and 122%.

Figure 8 — Housing Construction in the Next 30 Year Under Different Zoning Regimes

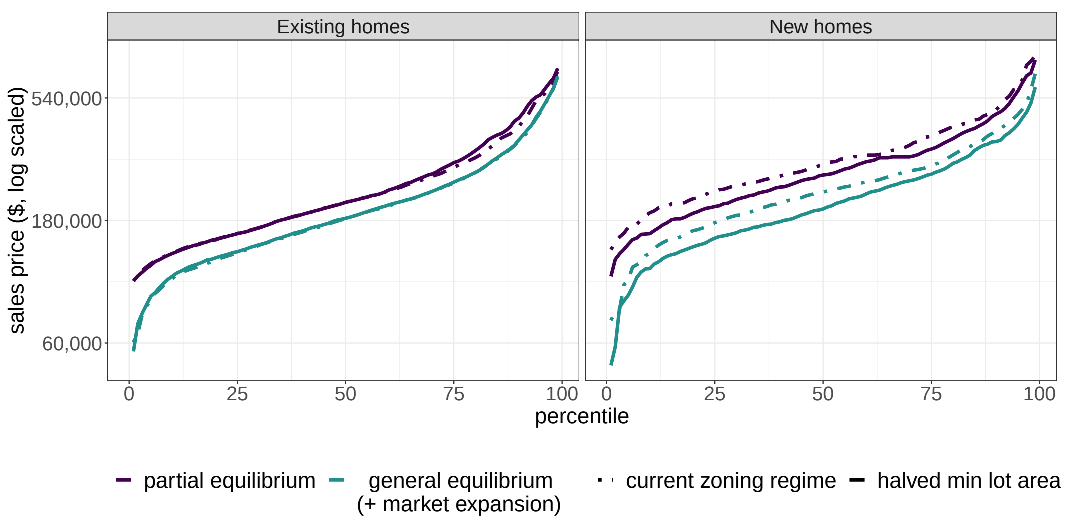


Note. The figures depict the numbers and characteristics of new single-family homes. The left panel depicts the number of new constructions in % of existing homes in the four different scenarios. The left two bars depict new construction in the next 30 years if the zoning regime stays the same in Connecticut, while the right two bars depict new construction under the zoning reform of halving minimum lot areas statewide. The partial equilibrium (purple) computes the housing supply given the observed price schedule of single-family homes by housing types and locations. The general equilibrium (green) equalizes the demand of each house to one, while assuming the markets expand, holding the distribution of preference parameters constant. The plot in the right panel depicts the distributions of lot area of new homes in general equilibrium under two alternative zoning regimes.

5.4.3 Price Responses to the Zoning Reform

Figure 9 shows the price distributions of existing homes and newly constructed homes under both counterfactuals. A new single-family home is $26,900 cheaper on average under the zoning reform relative to the current zoning regime. In contrast, the price effects of reducing minimum lot areas on existing homes are negligible. This finding may seem at odds with the neighborhood effect estimates in Section 4.2. However, note that the statewide zoning reform changes zoning stringency at the same time in all municipalities. The results indicate that a coordinated relaxation of zoning does not affect the prices of existing homes when within-CBSA migration is considered—all neighborhoods in the households' choice sets were subject to the minimum lot area reduction. This observation highlights the limitation of applying the hedonic regression results from the municipal border analysis when evaluating large-scale zoning reforms, as it fails to factor in the interaction between municipalities.

Figure 9 — Home Price Distribution Under Different Zoning Regimes



Note. The figure depicts the distribution of home prices in the two alternative zoning regimes. The left panel depicts the price distribution of existing homes, and the right panel depicts the price distribution of new homes (current zoning regimes in dashed lines and statewide zoning reform in solid lines). The color indicates the equilibrium scenarios. The partial equilibrium (purple) computes the housing supply given the observed price schedule of single-family homes by housing types and locations. The general equilibrium (green) equalizes the demand of each house to one, while assuming the markets expand, holding the distribution of preference parameters constant. The y-axis is on a logarithmic scale.

5.4.4 Welfare Effects of the Zoning Reform

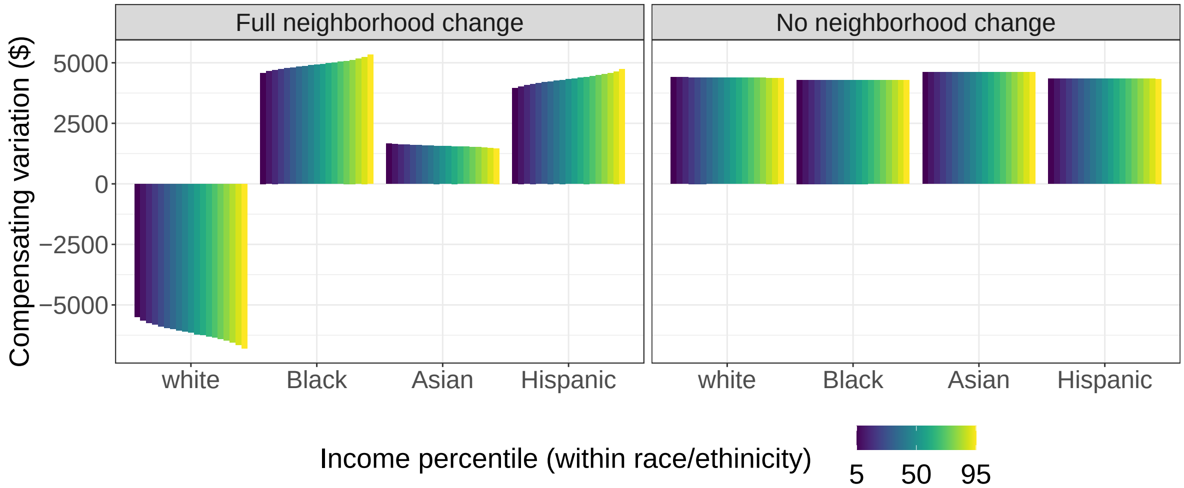
I present the welfare implications of the counterfactual zoning reform under two distinct assumptions on the extent to which neighborhood amenities respond to changes in zoning. In the first scenario, neighborhood amenities change to the level implied by halving minimum lot areas. For instance, neighborhoods with currently 1-acre minimum lot areas would have amenities typically offered in currently 1/2-acre minimum lot areas. I denote this scenario as a full neighborhood change.

This scenario may not be plausible, given the long history of zoning regulations since the early 1900s. A substantial number of buildings and infrastructures have been already constructed. Hence, neighborhood amenities may not respond as dramatically as in the full neighborhood change scenario. Therefore, I consider a second scenario in which neighborhood amenities do not change in response to the changes in zoning. In this case, I isolate the role of minimum lot areas on housing supply while ignoring their potential effects on neighborhood amenities. Note that, in both of these scenarios, the construction and price counterfactual results remain unchanged and are equal to the results discussed above.

To measure the welfare, I take the representative consumer approach (Herriges and Kling, 1999). I compute the compensating variation of 40 household types: 10-income quantile in four race/ethnic groups. The compensating variation equates the expected utility before and after the zoning reform.

Figure 10 shows the compensating variation by demographics, averaged across CBSAs, under the two scenarios. Under the full neighborhood change scenario, the counterfactual zoning reform benefits minorities while hurting white households. The welfare loss of white households arises from their strong preference for zoning stringency in their neighborhoods. They lose \$5,501–\$6,801 due to the reform, since the neighborhood changes induced by relaxing zoning are undesirable to them. In contrast, Black and Hispanic households gain \$3,952–\$5,340 from the supply of cheaper houses due to the reform. Although the new homes have smaller lots, Black and Hispanic households are more willing to make the price-size trade-off than white households. Asian households enjoy higher welfare for a similar reason, but the welfare gain is subdued because they moderately value strict zoning.

Figure 10 — Welfare Effects of Zoning Reform for Representative Existing Homeowners



Note. The figure depicts the welfare effects of the statewide zoning reform in Connecticut to halve minimum lot areas under two different scenarios of neighborhood changes. For each panel, 19 compensating variations of the zoning reform are presented for each race (white, Asian, Hispanic, and Black). The 19 compensating variations correspond to the 5th to 95th percentiles within-race/ethnicity income groups.

When we assume that neighborhood amenities do not change in response to the zoning reform, all households gain welfare by $4,289–$4,617. These welfare gains further illustrate how minimum lot area requirements restrict housing supply and how it is socially beneficial to increase the supply of smaller and cheaper homes. Even white households with strong preferences for big lots benefit from the reform because they strongly prefer buying new homes. As in the first scenario, Black and Hispanic households gain from the fact that new homes are smaller and cheaper. Asian households gain the most, as they strongly prefer new construction and only weakly prefer big lots.

The welfare effect of a statewide relaxation of zoning depends in part on how neighborhood amenities change in response to the zoning. Although the true extent to which these amenities respond to zoning is hard to quantify, analyzing the two ends of the spectrum of possible responses yields two important policy conclusions. First, regardless of the extent to which amenities respond to relaxed zoning, Black and Hispanic households benefit substantially from the statewide relaxation of zoning. The amount they benefit remains relatively constant and is approximately equivalent to a subsidy of $4,500 to Black and Hispanic households. Asian households also benefit positively irrespective of amenity response, but the amount depends on the response. Second, the amount white households benefit from the statewide relaxation of zoning varies in sign and magnitude depending on the amenity response to the change in zoning. Their change in welfare is between −$6,800 and $4,400, with lower values when the amenity response is higher. At the extreme where the amenity fully responds to the zoning change, the welfare analyses suggest that the statewide relaxation of zoning has clear redistribution effects by benefiting Black and Hispanic households and harming white households.

It should be noted that these welfare analyses do not account for the increase in home ownership. The counterfactual zoning regime includes 3.2% more homeowners than the current zoning regime. These added households become homeowners because cheaper homes are supplied in the housing market due to the reform. Although the current framework does not allow for a direct calculation of the welfare of these new homeowners, the welfare effects are likely positive and substantial.

6 Conclusion

In this paper, I study the effects of minimum lot area restrictions on housing prices, construction, and residential sorting and evaluate a statewide relaxation of the zoning regulation. I find that minimum lot area restrictions play significant roles in increasing housing prices and limiting housing supply. I show that halving minimum lot areas statewide in Connecticut, for example, would substantially increase the supply of small and cheap homes. It would benefit minority households irrespective of the response of amenities to zoning reforms and either benefit or harm white households depending on the response of neighborhood amenities to zoning reforms.

My analysis begins with constructing a nationwide data set of neighborhood-level minimum lot area estimates. I propose a new approach to estimate dimensional requirements by detecting structural breaks in observed property characteristics data. I apply a structural break detection algorithm to estimate minimum lot areas across the United States and build a data set on minimum lot area estimates with substantially broader coverage than existing zoning data sets. This data set allows me to characterize the state of zoning at the national level. Strict zoning is prevalent and substantially limits the supply of small houses.

I then leverage the spatial discontinuity in the stringency of minimum lot area restrictions at municipal borders to estimate their effects on housing market outcomes. I find that doubling minimum lot areas will increase sales prices by 14% and rents by 6%. Zoning changes will affect the prices of existing homes through the impact of minimum lot areas on the neighborhood environment—doubling minimum lot areas will increase sales prices by 4% and rents by 2% of existing homes. I also show that strictly zoned neighborhoods disproportionately attract white households and higher-income households.

Finally, I evaluate a statewide reform which halves minimum lot areas in Connecticut. To do so, I develop and estimate a model of housing demand and supply, acknowledging the limitation of the municipal border analysis in evaluating large-scale zoning reforms affecting multiple neighborhoods. On the demand side, households make location decisions within their Core-Based Statistical Area under their preference for building and neighborhood characteristics. The housing preference parameters are estimated using a boundary discontinuity design, similar to the municipal border analysis. On the supply side, landlords make construction decisions facing additional construction costs imposed by zoning and non-zoning supply restrictions. The regulatory costs are estimated by matching construction rates by housing type at the municipality level.

With the estimated demand and supply model in hand, I analyze the construction and price responses to zoning reforms in the next 30 years. The zoning reform would increase housing supply by 25% or up to 122%. The new single-family homes would be substantially smaller (by 7% on average) and cheaper (by $26,900 per unit on average). Hence, the reform substantially benefits Black and Hispanic households, as they have little to no preferences for strictly zoned neighborhoods and have weak preferences for large lots. They greatly benefit from the increased availability and affordability of housing. I find mixed evidence on whether the zoning reform is total welfare-improving. On the one hand, if neighborhood amenities change substantially according to the reform, then the reform is welfare-reducing by hurting white households who have strong preferences for strict zoning. On the other hand, if neighborhood amenities do not substantially change, then the reform is welfare-improving. Indeed, the reform increases the welfare of all demographic groups in this case.

References

Anagol, Santosh, Fernando Ferreira, and Jonah Rexer. 2021. "Estimating the Economic Value of Zoning Reform." *Working Paper.*

Andrews, Donald WK. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica: Journal of the Econometric Society* 821–856.

Avenancio-Leon, Carlos, and Troup Howard. 2019. "The Assessment Gap: Racial Inequalities in Property Taxation." *Available at SSRN 3465010.*

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of political economy* 115 (4): 588–638.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The quarterly journal of economics* 114 (2): 577–599.

Bronin, Sara C. 2021. "Exclusion, Control, and Consequence in 2,622 Zoning Districts." *Control, and Consequence in* 2.

Diamond, Rebecca, and Tim McQuade. 2019. "Who Wants Affordable Housing in their Backyard? An Equilibrium Analysis of Low-Income Property Development." *Journal of Political Economy* 127 (3): 1063–1117.

Emrath, Paul. 2021. "Government Regulation in the Price of a New Home: 2021." *Special Study for Housing Economics, National Association of Home Builders.*

Fischel, William A. 2004. "An Economic History of Zoning and a Cure for its Exclusionary Effects." *Urban Studies* 41 (2): 317–340.

Fischel, William A. 2005. *The Homevoter Hypothesis: How Home Values Influence Local Government Taxation, School Finance, and Land-Use Policies.* Harvard University Press.

Glaeser, Edward, and Joseph Gyourko. 2018. "The Economic Implications of Housing Supply." *Journal of Economic Perspectives* 32 (1): 3–30.

Glaeser, Edward L, and Joseph Gyourko. 2002. "The Impact of Zoning on Housing Affordability." *NBER Working Paper* 8835.

Glaeser, Edward L, and Bryce A Ward. 2009. "The Causes and Consequences of Land Use Regulation: Evidence from Greater Boston." *Journal of urban Economics* 65 (3): 265–278.

Grenadier, Steven R. 1995. "The Persistence of Real Estate Cycles." *The Journal of Real Estate Finance and Economics* 10 (2): 95–119.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. "Endogenous Gentrification and Housing Price Dynamics." *Journal of Public Economics* 100 45–60.

Gyourko, Joseph, Jonathan S Hartley, and Jacob Krimmel. 2021. "The Local Residential Land Use Regulatory Environment across US Housing Markets: Evidence from a New Wharton Index." *Journal of Urban Economics* 124 103337.

Gyourko, Joseph, and Jacob Krimmel. 2021. "The Impact of Local Residential Land Use Restrictions on Land Values across and within Single Family Housing Markets." *Journal of Urban Economics* 126 103374.

Gyourko, Joseph, Albert Saiz, and Anita Summers. 2008. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index." *Urban Studies* 45 (3): 693–729.

Handbury, Jessie. 2019. "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living across U.S. Cities." *NBER Working Paper* 26574.

Herriges, Joseph A, and Catherine L Kling. 1999. "Nonlinear Income Effects in Random Utility Models." *Review of Economics and Statistics* 81 (1): 62–72.

Jackson, Kristoffer Kip. 2018. "Regulation, Land Constraints, and California's Boom and Bust." *Regional science and urban economics* 68 130–147.

Kahn, Matthew E, Ryan Vaughn, and Jonathan Zasloff. 2010. "The Housing Market Effects of Discrete Land Use Regulations: Evidence from the California Coastal Boundary Zone." *Journal of Housing Economics* 19 (4): 269–279.

Kok, Nils, Paavo Monkkonen, and John M Quigley. 2014. “Land Use Regulations and the Value of Land and Housing: An Intra-Metropolitan Analysis.” *Journal of Urban Economics* 81 136–148.

Kulka, Amrita. 2019. “Sorting into Neighborhoods: The Role of Minimum Lot Sizes.”Technical report, Mimeo., University of Wisconsin.

Mawhorter, Sarah, and Carolina Reid. 2018. “Terner California Residential Land Use Survey.” *University of California, Berkeley.*

Menendian, Stephen, Samir Gambhir, Karina French, and Arthur Gailes. 2020. “Single-Family Zoning in the San Francisco Bay Area.” *Berkeley, CA: Haas Institute for a Fair and Inclusive Society.*

Nicolaides, Becky, and Andrew Wiese. 2017. “Suburbanization in the United States after 1945.” In *Oxford research encyclopedia of American history.*

Pendall, Rolf, Robert Puentes, and Jonathan Martin. 2006. “From Traditional to Reformed: A Review of the Land Use Regulations in the Nation’s 50 Largest Metropolitan Areas.”

Quigley, John M, Steven Raphael, and Larry A Rosenthal. 2008. “Measuring Land-Use Regulations and their Effects in the Housing Market.”

Sahn, Alexander. 2021. “Racial Diversity and Exclusionary Zoning: Evidence from the Great Migration.”

Silver, Christopher. 2016. “Zoning in 20th-Century American Cities.” In *Oxford Research Encyclopedia of American History.*

Trounstine, Jessica. 2020. “The Geography of Inequality: How Land Use Regulation Produces Segregation.” *American Political Science Review* 114 (2): 443–455.

Turner, Matthew A, Andrew Haughwout, and Wilbert Van Der Klaauw. 2014. “Land Use Regulation and Welfare.” *Econometrica* 82 (4): 1341–1403.

Wu, JunJie, and Seong-Hoon Cho. 2007. "The Effect of Local Land Use Regulations on Urban Development in the Western United States." *Regional Science and Urban Economics* 37 (1): 69–86.

Zeileis, Achim. 2005. "A Unified Approach to Structural Change Tests based on ML Scores, F Statistics, and OLS Residuals." *Econometric Reviews* 24 (4): 445–466.

Appendix A Constructing Zoning Data Set: Additional Details

A.1 Testing Randomness in Missing in Zoning District Data Field

I examine the data quality of zoning district information in CoreLogic tax data. As the first step, I test the randomness in the missing by municipalities. Denote municipality by m. In each municipality, I observe $\{z_{1i,m}, z_{2i,m}, d_{i,m}\}_{i=1}^{N_m}$ where $z_{1i,m}$ and $z_{1i,m}$ are longitude and latitude of single-family home i, $d_{i,m}$ is whether zoning district is observed for i (where 0 indicates missing CoreLogic “zoning” field, and 1 indicates that the data field is filled), and N_m is the number of parcels in m. The test statistic is the distance between mean longitude and latitude conditional on $d_{i,m}$: the test statistic with observed $d_{i,m}$,

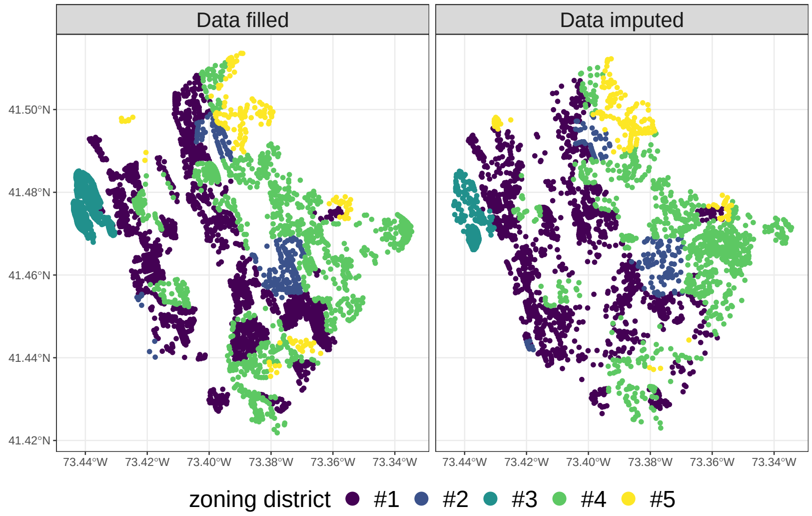
$$T_m = \underbrace{\left[\frac{z_{1i,m} d_{i,m}}{N_m \cdot (1-p_m)} - \frac{z_{1i,m}\left(1-d_{i,m}\right)}{N_m \cdot p_m}\right]^2}_{\substack{\text{difference in mean longitude} \\ \text{conditional on } d_{i,m}}} + \underbrace{\left[\frac{z_{2i,m} d_{i,m}}{N_m \cdot (1-p_m)} - \frac{z_{2i,m}\left(1-d_{i,m}\right)}{N_m \cdot p_m}\right]^2}_{\substack{\text{difference in mean latitude} \\ \text{conditional on } d_{i,m}}}$$

where observed data missing rate $p_m = Pr\left(d_{i,m} = 0\right)$. I simulate the distribution of T_m under the null hypothesis of random missing: I randomly draw $d_{i,m}$ while keeping the missing rate at p_m and compute T_m in each draw. I reject the null if and only if $Pr(T_m > T_{m,observed}) < 0.05$.

A.2 Zoning District Imputation

In the municipalities where (1) the missing rate is less than 30%, or (2) the missing rate is less than 70%, and it fails to reject the permutation test describe above, I impute missing zoning districts using locations of single-family homes. I train a random forest algorithm by municipality only using longitudes and latitudes of homes. To exclude any planned districts, I exclude any single-family homes if their district name includes "PD", "PUD" or "Planned". The prediction accuracy is over 90%. Figure A1 illustrates the imputed zoning district in a sample municipality.

Figure A1 — Location-Based Imputation of Zoning Using a Random Forest Algorithm

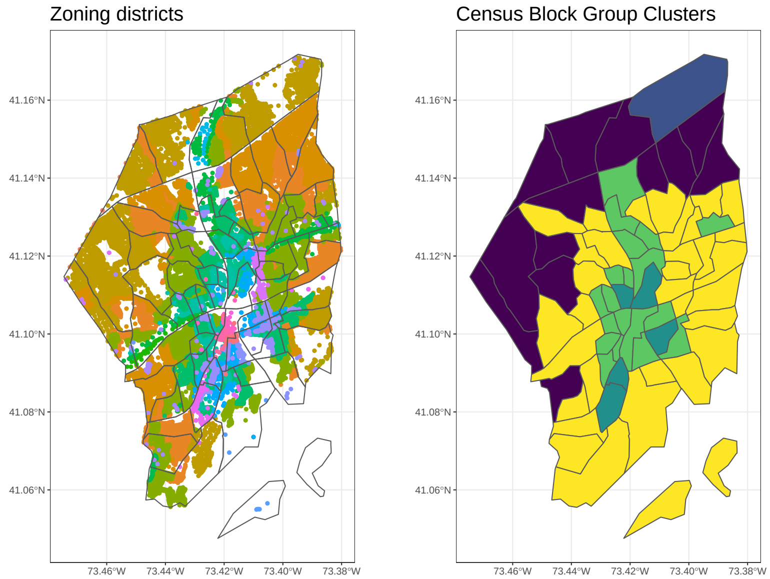


Note. The figure depicts single-family homes in an example municipality. The left panel includes the single-family homes whose zoning district data field in CoreLogic is filled, and the right panel includes those with missing data. Each dot is colored by zoning districts. In the left panel, color indicates the observed zoning districts while, in the right panel, color indicates the imputed zoning districts using the random forest algorithm.

A.3 K-means Clustering of Census Block Groups

In the rest of the municipalities, I cluster Census block groups using k-means clustering algorithms to construct proxy zoning districts. I include % single-family housing, % multi-family housing, % agricultural, % industrial, % commercial, and % vacant land share as predictors of zoning districts. I also include the mean, standard deviation, mode, median, 25th quantile, and 75th quantile of lot sizes of all properties in the algorithm. Figure A2 depicts the constructed Census Block Group clusters in comparison with actual zoning districts in a sample municipality. The clustering approach successfully distinguish higher-density districts in the city center from lower-density districts in the periphery in this case.

Figure A2 — Constructing Proxy Zoning Districts by Clustering Census Block Groups



Note. The left plot depicts actual zoning districts in an example municipality, and the right plot depicts proxy zoning districts constructed by k-means clustering of Census block groups using land use compositions and building characteristics. In the left panel, each dot depicts a single-family home colored by zoning district overlaid on the Census block group map (line). In the right panel, each polygon represents a Census block group colored by proxy zoning districts.

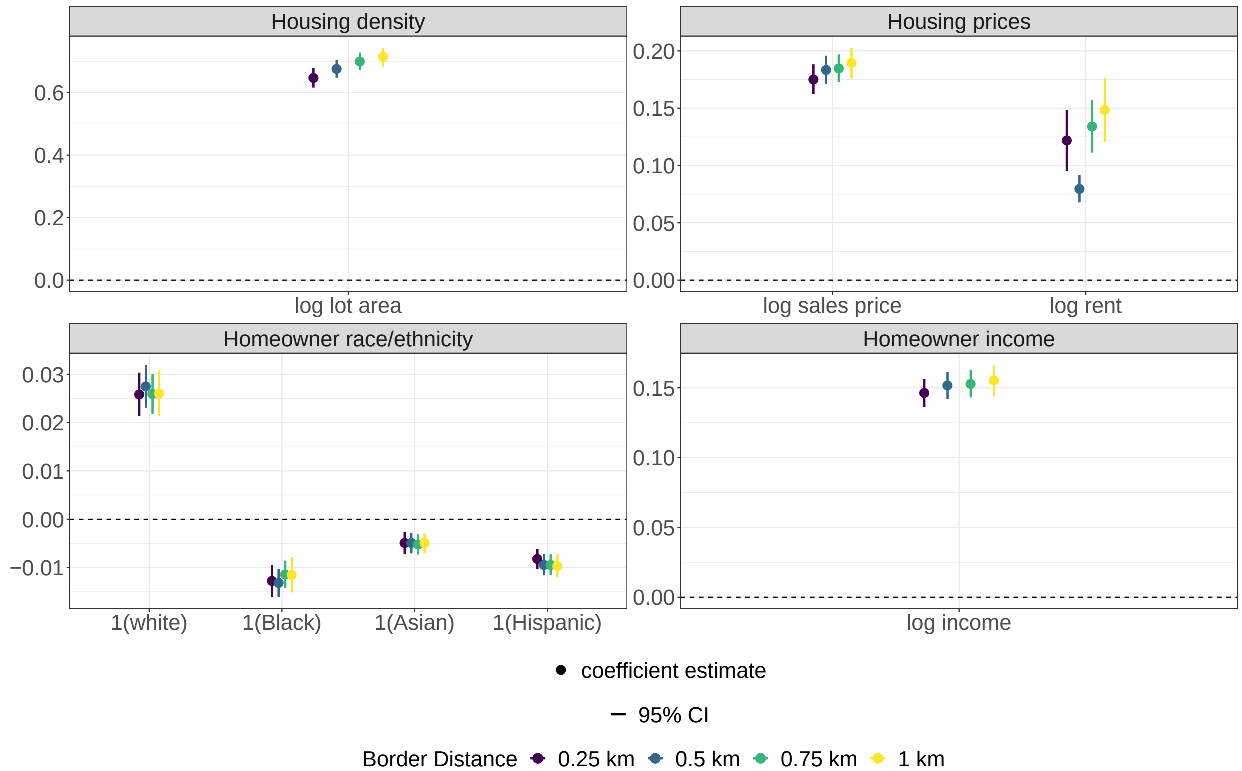
Appendix B Municipal Border Analysis: Additional Details and Results

Table B1 — Regression Sample Descriptions

	HMDA-CoreLogic deed		MLS rental listings	
	All borders	Within school districts	All borders	Within school districts
# transactions	2,122,884	625,735	142,775	39,087
# borders	6,884	2,309	5,828	1,685
# municipalities	5,557	3,068	3,510	1,739
25th percentile (in $)	136,116	142,413	1,229	1,294
50th percentile (in $)	203,929	207,587	1,571	1,602
75th percentile (in $)	318,900	315,795	2,065	2,038
mean (in $)	267,498	268,625	1,772	1,767
std. deviation (in $)	257,590	258,046	846	736

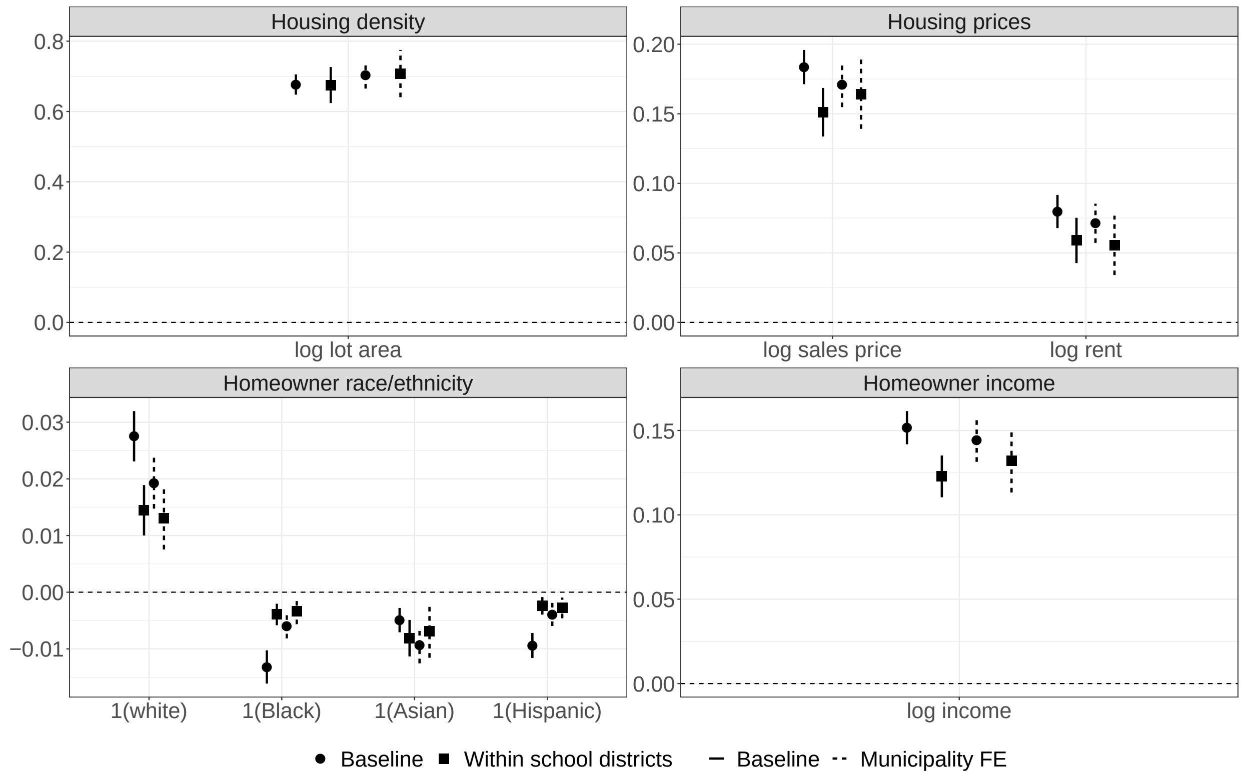
Note. The table reports the numbers of transactions, municipal borders, and municipalities in the regression samples. The table also reports the quartiles, mean, and standard deviation of sales prices and rents in each sample. The data is restricted to the boundary sample near municipal borders where the robust estimates of minimum lot areas exist on the both sides of the borders, and the minimum lot areas differ. In addition, the data is restricted to municipal borders where at least 10 transactions are observed on the both sides. Prices are adjusted to 2010 dollar value.

Figure B1 — Robustness to Different Choices of Border Distances: All Outcomes



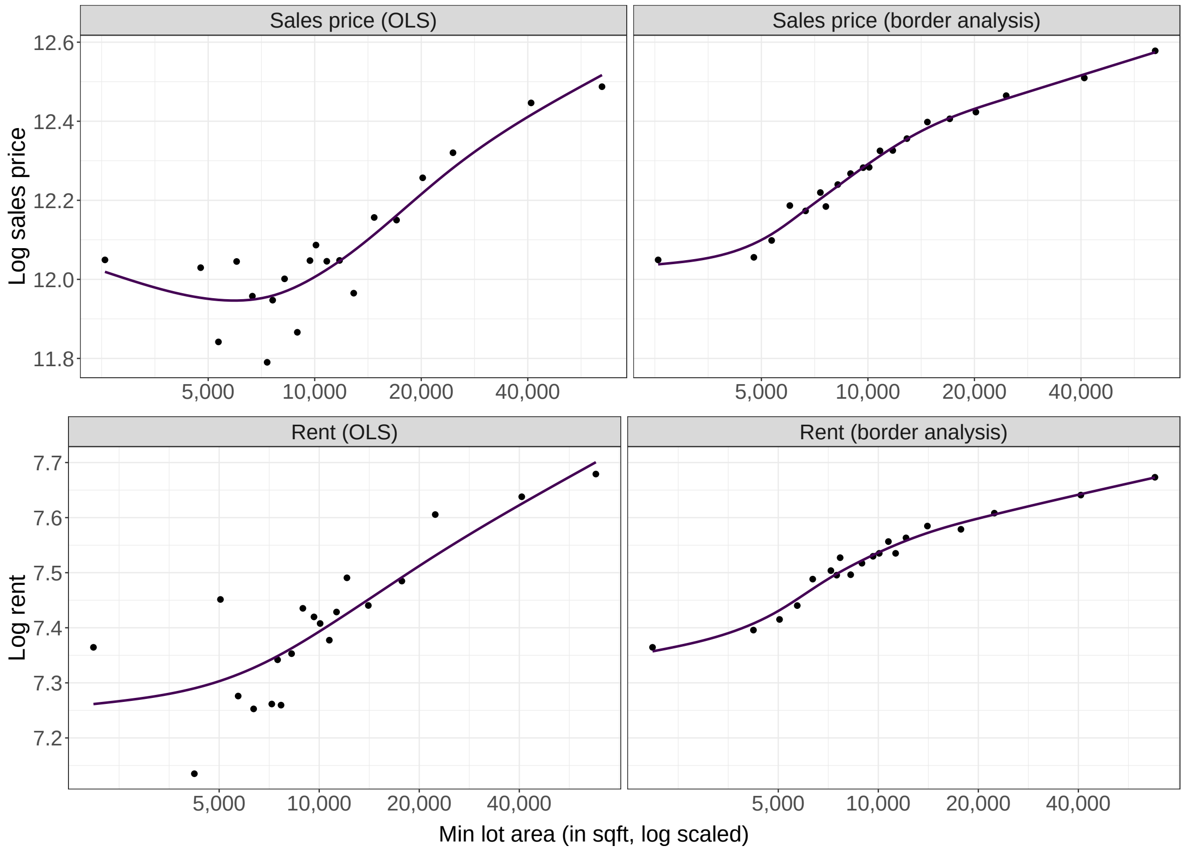
Note. The figure depicts the estimates (dots) and 95% confidence intervals (lines) of the coefficients of minimum lot area restrictions (MLA) on various housing market outcomes. They correspond to β_{MLA} coefficients of $\log MLA$ in specification 1 with different outcome variables y. The first panel shows the estimated effects on lot areas. The second panel shows the estimated effects on sales prices and rent prices. The third panel shows the estimated effects on home buyers being white, Black, Asian, and Hispanic. The last panel shows the estimated effects on home buyers income. Errors are clustered at municipality level. For each outcome, the figure presents the results from four different choices of distances, 0.25km (purple, leftmost), 0.5km (blue), 0.75km (green), and 1km (yellow, rightmost), to define border regions (See Section 2.2 for border region definition).

Figure B2 — Robustness to Municipality and School District Controls: All Outcomes



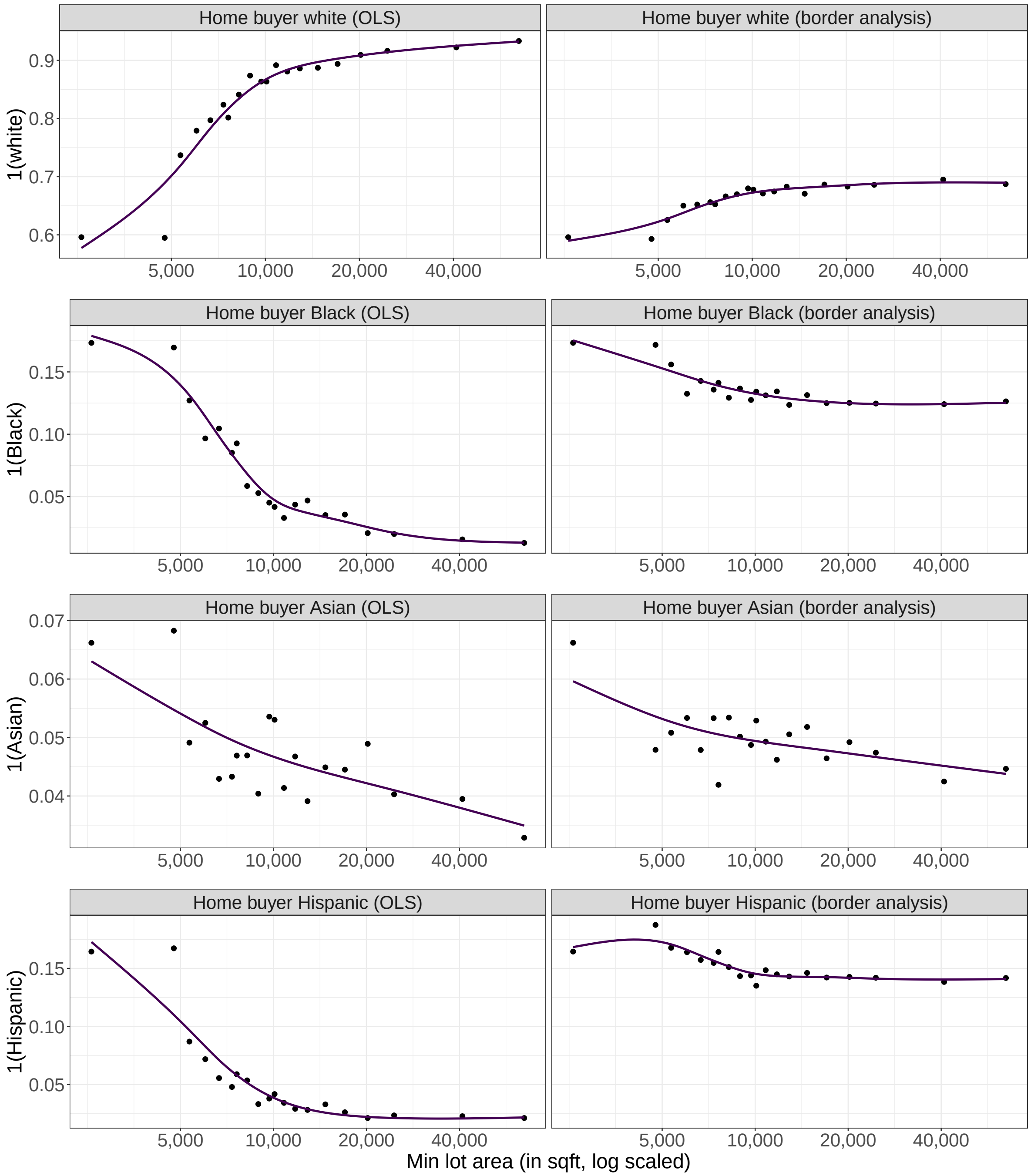
Note. This figure depicts the estimated effects of minimum lot area $\hat{\beta}_{MLA}$ (dots and squares) and their 95% confidence intervals (solid and dashed lines) on various housing market outcomes. The first panel shows the estimated effects on lot areas. The second panel shows the estimated effects on sales prices and rent prices. The third panel shows the estimated effects on home buyers being white, Black, Asian, and Hispanic. The last panel shows the estimated effects on home buyers income. For each outcome, the figure presents results from four different specifications. Results depicted with a dot correspond to specifications using all within-state municipal borders, and results depicted with a square correspond to specifications restricted to within-school district municipal borders. Results with solid 95% confidence intervals correspond to specifications without municipality fixed effects, while results with dashed 95% confidence intervals correspond to specifications with municipality fixed effects. The leftmost specifications depicted with a dot and solid line correspond to the baseline specification (as in Figure 5). Errors are clustered at the municipality level.

Figure B3 — Binscatter Plots of Housing Prices



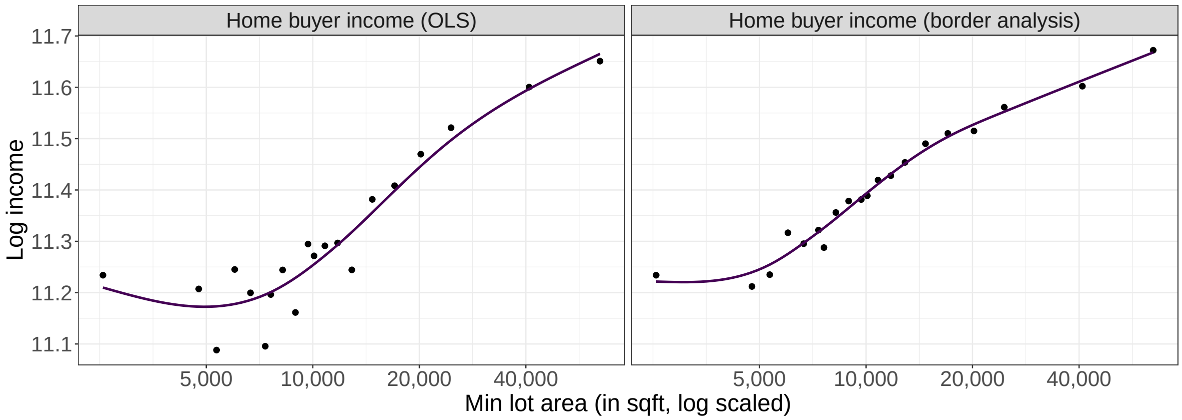
Note. The figure depicts the binscatter plots (with 20 bins) and their fitted lines for price outcomes: log sales price and log rent. The x-axis is minimum lot areas (in square feet) on a logarithmic scale. The left two plots illustrate the binscatter regressions without including border fixed effects. The right two plots illustrate the binscatter regressions with including border fixed effects.

Figure B4 — Binscatter Plots of Home Buyer Race/Ethnicity



Note. The figure depicts the binscatter plots (with 20 bins) and their fitted lines for race/ethnicity of home buyer: indicator of being white, Black, Asian, and Hispanic. The x-axis is minimum lot areas (in square feet) on a logarithmic scale. The left four plots illustrate the binscatter regressions without including border fixed effects. The right four plots illustrate the binscatter regressions with including border fixed effects.

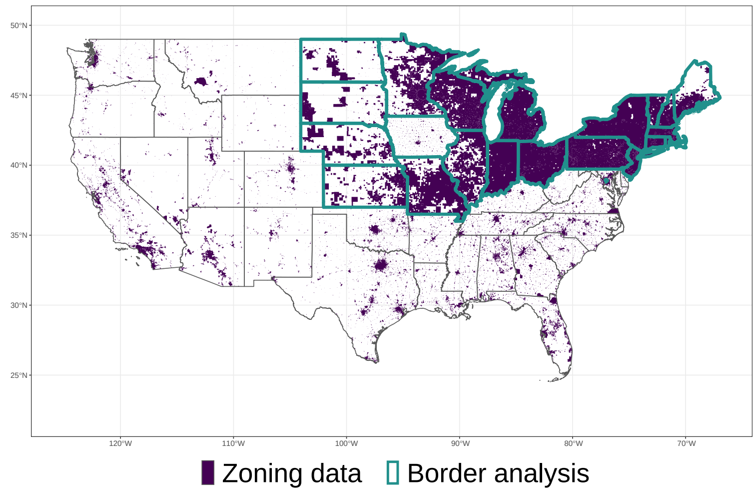
Figure B5 — Binscatter Plots of Home Buyer Income



Note. The figure depicts the binscatter plots (with 20 bins) and their fitted lines for income of home buyer. The x-axis is minimum lot areas (in square feet) on a logarithmic scale. The left plot illustrates the binscatter regressions without including border fixed effects. The right plot illustrates the binscatter regressions with including border fixed effects.

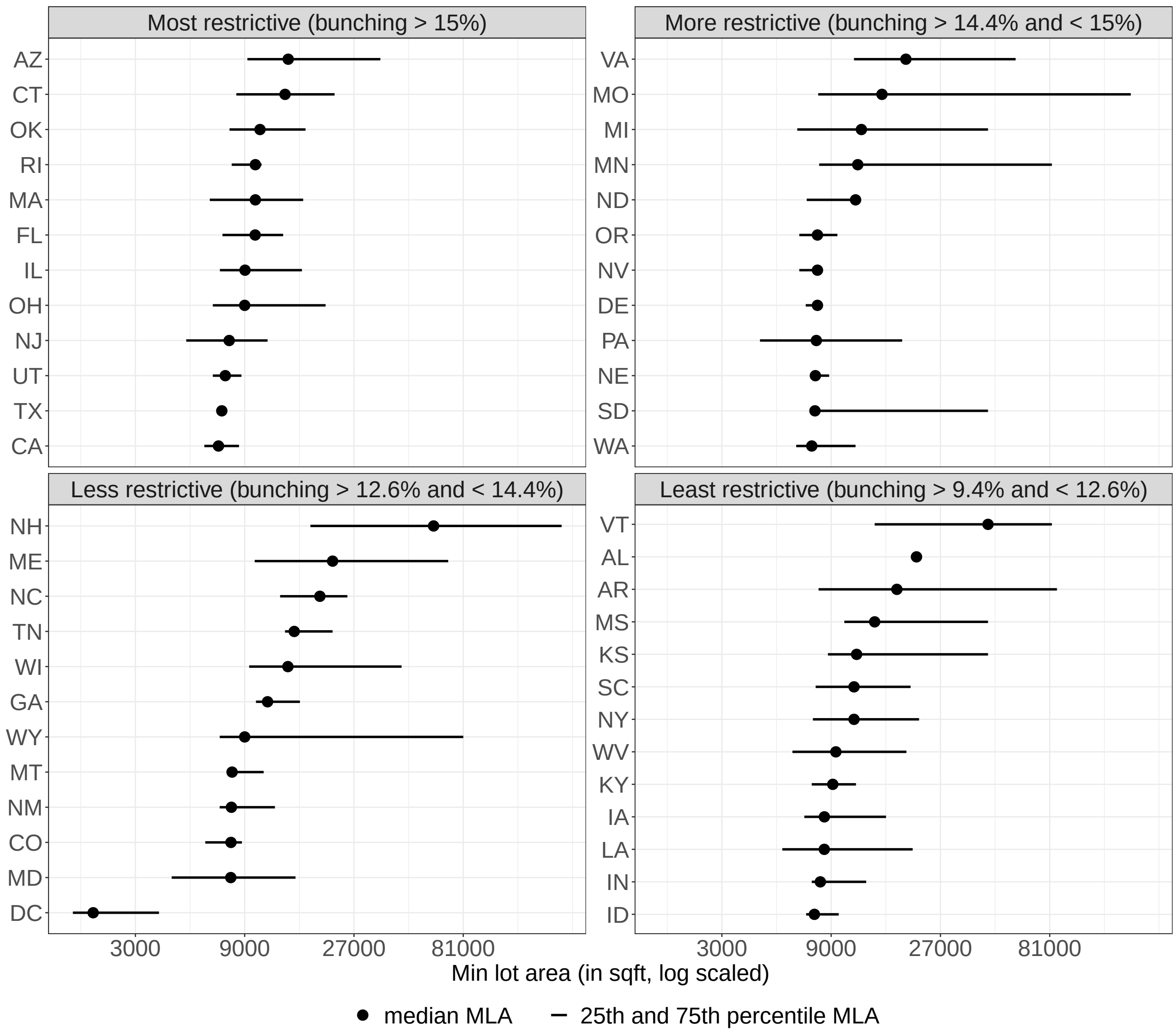
Appendix C Additional Figures and Tables

Figure C1 — Geographic coverage of zoning data and 20 states in the border analysis



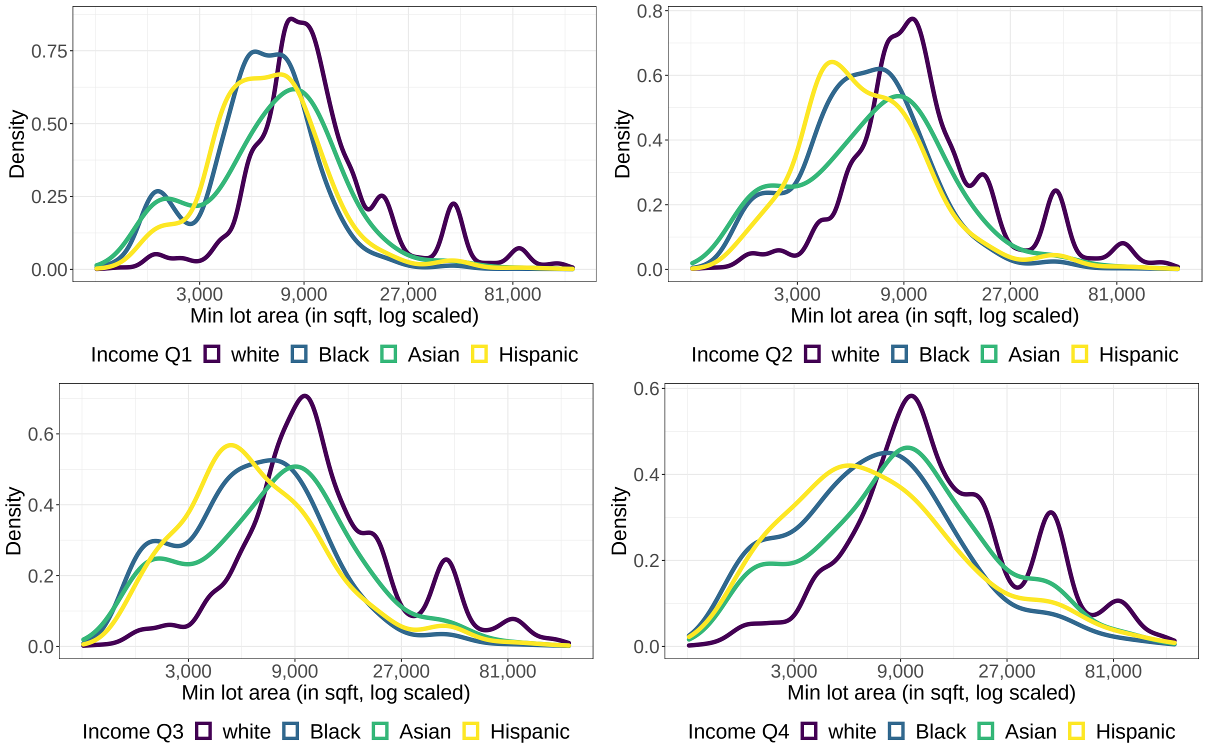
Note. This figure depicts the coverage of minimum lot area estimates and the 20 states included in the border analysis at the municipality level. Purple areas indicate Census places and Census county subdivisions where minimum lot area estimates exist. Green solid boundaries indicate the 20 states where county subdivisions are general-purpose governments, and hence included in the border analysis.

Figure C2 — Zoning Stringency by State



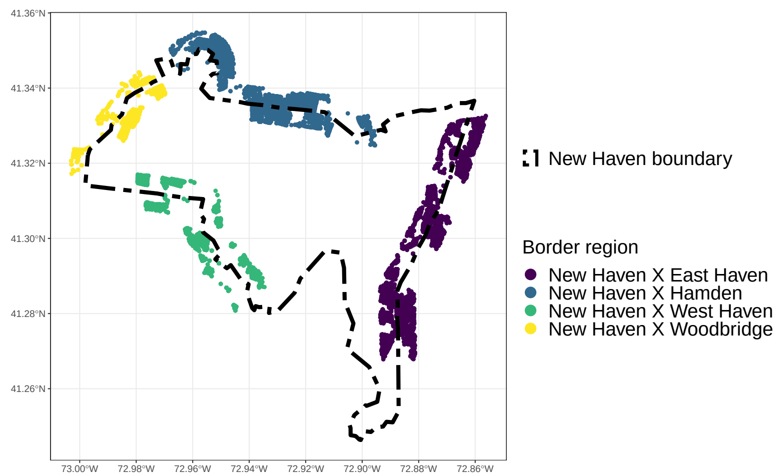
Note. The figure depicts the median (dots) and 25th percentile to 75th percentile ranges (lines) of minimum lot area restrictions (MLA) by state. The percentiles are calculated in terms of the number of single-family homes. States are grouped by the restrictiveness of zoning, and each panel depicts one such group. The restrictiveness of zoning is defined by the amount of bunching around the MLA for single-family homes built after 1970, which is a proxy for how binding the MLA restriction is. A single-family home is counted as bunching at the MLA if its lot is within 5% of the MLA. In each panel, the states are ordered by their median MLA, a proxy for the stringency of zoning.

Figure C3 — Racial Segregation by Income Quartile



Note. The figure depicts the conditional density functions of neighborhood minimum lot areas by race and ethnicity of homeowners by income groups. The data is restricted to the 20-state sample with robust minimum lot area estimates, including single-family homes in both city centers and border regions. The data is restricted to the 20-state sample as described in Section 2.2 with robust minimum lot area estimates, including single-family homes in both city centers and border regions. Homeowner demographics are taken from HMDA data. Each panel presents the conditional density functions of different income quartile (Q1: top left, Q2: top right, Q3: bottom left, Q4: bottom right). In each panel, four density functions conditional on race/ethnicity are depicted in different colors (white: purple, Black: blue, Asian: green, Hispanic: yellow).

Figure C4 — Example Border Regions



Note. This figure depicts the border regions near the boundary of New Haven, CT. The dashed line illustrates the municipal boundary of New Haven. Each point denotes a single-family home in border regions, located either in New Haven or in adjacent municipalities. The points are colored by the border regions where the single-family homes are situated.

Table C1 — Demand Estimates (in Connecticut)

	white		Black		Asian		Hispanic	
	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
log price	0.9059	−0.0035	0.9951	0.0038	0.9983	−0.0007	0.9937	0.0039
log lot area	0.1715	0.0054	0.0607	−0.0032	0.0736	0.0007	0.0710	−0.0033
$\mathbb{1}$(new home)	0.1924	0.0014	0.1412	−0.0001	0.2138	0.0001	0.1600	−0.0000
log MLA	0.1212	0.0044	−0.0094	−0.0033	0.0390	0.0007	−0.0015	−0.0034

Note. This table present the complete set of demand estimates except the subregion fixed effects. For each housing characteristics, β_0 represents the mean preference of each race/ethnicity and β_1 represents the race/ethnicity-specific slope of the preference with respect to income level. The income level is normalized by subtracting the CBSA-level mean and dividing by the CBSA-level standard deviation.