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

Jaehee Song*

January 15, 2025

Abstract

I construct a new nationwide dataset to measure the stringency of residential zoning in the U.S. and analyze its effects on housing prices and demographic sorting. First, I develop and implement a structural break detection algorithm to infer minimum lot size regulations from property tax records. The dataset covers over 16,000 local jurisdictions within Core-Based Statistical Areas, capturing both cross-jurisdictional and within-jurisdictional variations in zoning stringency. I find that 18.5 percent of single-family constructions bunch at the regulated lot size and that density regulations are strongly associated with a higher share of white homeowners, higher incomes, and higher housing prices. Second, using this dataset and a spatial discontinuity design at municipality borders, I estimate the impacts of minimum lot size regulations on housing markets. The results indicate that doubling the minimum lot size increases home sales prices by 14 percent and rents by 9 percent. Moreover, restrictive zoning disproportionately attracts high-income white homeowners, exacerbating residential segregation.

JEL Classification: R0, H7, C8, K2

Keywords: Zoning laws, minimum lot size, housing affordability, segregation

---

*University of Colorado, Leeds School of Business (email: jaehee.song@colorado.edu). I gratefully acknowledge financial support from the Tobin Center for Economic Policy at Yale University. I thank Steven Berry, John Eric Humphries, and Costas Meghir for their continuous support and helpful comments throughout the project. I thank Labor/Public Finance and Industrial Organization groups at Yale University. I also thank seminar and conference participants at AREUEA, BFI Women in Empirical Microeconomics, NBER Summer Institute, UEA, Arizona State University, Federal Housing Finance Agency, Federal Reserve Bank of Cleveland, Federal Reserve Bank of Minneapolis, Harvard Kennedy School, MIT Center for Real Estate, NYU Furman Center, Seoul National University, Syracuse University, University of British Columbia, University of Colorado Boulder, University of Toronto, Wharton Real Estate, and Yonsei University.

1 Introduction

Local governments in the U.S. regulate housing supply through residential zoning, restricting the quantity and type of housing that can be constructed. A growing body of empirical literature finds that local zoning laws increase housing prices and intensify segregation (Glaeser and Ward, 2009; Kok et al., 2014; Trounstine, 2020). Accordingly, recent shortages of affordable housing have sparked nationwide debates on relaxing zoning laws.[1] Despite this, most research relies on case studies or surveys with limited geographic coverage, lacking comprehensive evidence on the stringency of zoning laws and their effects on housing markets.

Zoning data is typically maintained by local authorities, and no nationwide datasets exist. Consequently, empirical studies have relied on land use surveys (Turner et al., 2014; Trounstine, 2020; Gyourko and Krimmel, 2021) or manually collected local zoning data (Bronin, 2021; Sahn, 2021), which are costly and nearly infeasible at a national scale given that over 30,000 local zoning authorities exist in the U.S.[2] The most widely used dataset, the Wharton Land Use Survey, covers only 2,450 municipalities (Gyourko et al., 2021), leaving significant gaps in geographic coverage. As such, empirical studies on zoning laws often face challenges stemming from sparse geographic coverage of available zoning data.

This paper addresses these gaps by providing the first nationwide examination of minimum lot size regulations, one of the most prevalent types of density restriction in residential zoning. I construct a new dataset covering 52 million single-family homes in 16,217 municipalities across 825 Core-Based Statistical Areas (CBSAs), encompassing over 60 percent of U.S. households. Using this dataset and a spatial discontinuity design, I analyze zoning stringency and its effects on housing prices and demographic sorting.

In the first part of the paper, I develop and implement an algorithm to estimate minimum lot sizes using observed building characteristics from CoreLogic property tax records, which cover

1. For example, Oregon enacted House Bill 2001 in 2019, and California passed Senate Bill 9 in 2021 for statewide zoning reform. Since 2017, California has introduced over 100 new housing-related laws (Fulton et al., 2023). The Zoning Reform Tracker by the Othering & Belonging Institute at UC Berkeley reports 143 municipal zoning reforms as of November 2023.

2. The National Zoning Atlas is a recent initiative aiming to build a nationwide dataset.

nearly all single-family homes in the U.S. The central idea of the algorithm is that minimum lot size restrictions make it significantly difficult to construct a house on a lot smaller than the regulated cutoff.[3] As a result, these restrictions result in construction bunching at the cutoff, creating identifiable structural breaks in the distribution of constructed lot sizes. I apply a structural break detection method (Zeileis, 2005; Andrews, 1993) to single-family homes in the CoreLogic property tax data. The structural break detection is performed at the zoning district level, utilizing zoning district names directly available in the CoreLogic data or proxied with Census block groups when zoning district names are unavailable. To validate these minimum lot size estimates, I compare them with zoning data from 228 single-family zoning districts across 55 municipalities, compiled using the MAPC Zoning Atlas and other local zoning codebooks. The validation results demonstrate the accuracy of the estimated minimum lot sizes, with a median error of 11 percent and a correlation of 0.96 with the actual minimum lot sizes.

The constructed minimum lot size dataset captures housing density restrictions at the zoning district level for 16,217 municipalities, representing nearly 80 percent of local governments in Core-Based Statistical Areas (CBSAs).[4] This dataset offers significantly better geographic coverage than existing zoning datasets, capturing within-municipality variations and providing broader coverage of across-municipality differences.[5] Using this dataset, I first characterize the stringency of residential zoning across the U.S. The findings suggest that minimum lot size restrictions are stringent nationwide, with a median of 10,000 square feet. Approximately 18.5 percent of single-family homes are constructed right at the cutoff, indicating that these regulations impose binding constraints on lot size decisions. Moreover, zoning stringency varies considerably across and within CBSAs and is strongly correlated with neighborhood and demographic characteristics. Stricter zoning is associated with lower residential density, higher home values and rents, and demographic patterns characterized by a higher share of higher-income, non-Hispanic white

3. Under perfect compliance, minimum lot size restrictions would prohibit construction on lots smaller than the regulated size. However, some municipalities allow zoning variances to circumvent existing zoning laws, though these are often a complex and lengthy process.

4. To identify municipalities with zoning authority, I mapped 32,000 functioning local governments using the 2010 Census Guide to State and Local Census Geography.

5. Most land use surveys and zoning code compilations are limited to the municipality level, covering only a few thousand areas. For instance, the largest manual compilation, the Connecticut Zoning Atlas, includes 2,260 zoning districts. Broader initiatives, like the National Zoning Atlas, remain works in progress.

populations.

In the second part of this paper, I estimate the effects of minimum lot size regulations on home prices, rents, and homeowner demographics using a boundary discontinuity design at municipality borders. For this analysis, I link CoreLogic deed records and Home Mortgage Disclosure Act (HMDA) data, which cover most single-family homes in the U.S. and provide detailed information on sales prices, property characteristics, and homeowner demographics, along with CoreLogic MLS data, which includes rental listings from 154 MLS regions. By focusing on housing transactions within small geographic areas across municipal borders, the boundary discontinuity design addresses endogeneity concerns that locational characteristics, such as land availability and proximity to the Central Business District, may influence both zoning stringency and housing market outcomes. I further account for historical municipality characteristics derived from the 1940 Full-Count Census, including population, demographic composition, homeownership rates, home values, and rents, along with an indicator for whether the municipality was developed during the postwar suburbanization period, a time when zoning laws were most actively adopted. Additionally, I incorporate municipality-by-school district fixed effects to control for variations in municipal public goods and school district quality, which may change discontinuously at borders and correlate with zoning stringency, thereby providing the robustness of the results.

Baseline price effect estimates from the border analysis suggest that doubling the minimum lot size increases home sales prices by 14 percent and rents by 9 percent. The price premium associated with stringent zoning may operate through two primary mechanisms. First, larger minimum lot sizes mandate the construction of bigger homes, which directly increases housing costs—referred to here as the direct effect. Second, zoning regulations influence neighborhood characteristics, which in turn affect housing prices. For instance, stringent zoning may reduce neighborhood congestion, boost local tax revenues, and alter neighborhood demographics—referred to here as the neighborhood effect. My analysis attributes 78 percent of the price effect and 72 percent of the rent effect to the direct effect, with the remainder attributed to the neighborhood effect. The statistically significant and positive neighborhood effect suggests that restrictive zoning contributes to higher housing values by enhancing certain neighborhood attributes.

Additionally, I examine the effects of local zoning regulations on demographic sorting by repeating the border analysis with homeowner race, ethnicity, and income from CoreLogic-HMDA data as outcome variables. The results reveal that doubling the minimum lot size increases the likelihood of homeowners being non-Hispanic white by two percentage points and raises homeowners' income at the time of mortgage applications by 11 percent. These findings underscore the significant role zoning laws play in shaping the demographic composition of neighborhood homeowners.

This paper contributes to several bodies of literature. First, this paper contributes to the literature on measuring the stringency of zoning laws. A common method of data collection is through surveys, as seen in studies focusing on individual states like California (Quigley et al., 2008; Jackson, 2018; Mawhorter and Reid, 2018), or nationwide efforts such as the Wharton Land Use Survey (Gyourko et al., 2008, 2021) and the Brookings Survey (Pendall et al., 2006), which respectively survey 2,450 and 1,844 jurisdictions. Another approach involves manually collecting and aggregating local zoning ordinances, exemplified by the MAPC Zoning Atlas (2020), Menendian et al. (2020), and Bronin (2021), which document zoning regulations for 101 jurisdictions in Greater Boston, 100 jurisdictions in the Bay Area, and 180 jurisdictions in Connecticut, respectively. This paper builds on these efforts by developing a data-driven method to infer minimum lot sizes at the zoning district level, constructing a nationwide dataset with unique geographic coverage and detail. It is closely related to studies employing structural break detection (Zabel and Dalton, 2011; Cui, 2023), machine learning (Nechamkin and MacDonald, 2019; Mleczko and Desmond, 2023; Bartik et al., 2023), and judicial case analysis (Ganong and Shoag, 2017) to measure zoning stringency. However, this study uniquely focuses on inferring minimum lot sizes nationwide at the zoning district level, providing unparalleled geographic coverage and granularity.

Second, the paper contributes to research on the impact of land use regulations on housing markets. Prior work demonstrates that stricter regulations are associated with higher housing prices (Glaeser and Gyourko, 2002; Glaeser and Ward, 2009; Kok et al., 2014), reduced land development (Wu and Cho, 2007), higher land prices (Gyourko and Krimmel, 2021), losses in economic output (Hsieh and Moretti, 2019), and welfare loss (Brueckner and Sridhar, 2012; Albouy and Ehrlich, 2018). This paper is particularly connected to studies employing boundary discontinuity designs to address endogeneity concerns in zoning regulations (Kahn et al., 2010; Turner et al., 2014; Kulka, 2019; Anagol et al., 2021; Resseger, 2022; Kulka et al., 2023). By compiling a comprehensive nationwide map of local governments, this study applies a boundary discontinuity design at municipality borders to estimate the impacts of minimum lot size regulations.

Lastly, this paper contributes to the literature on the relationship between density restrictions in residential zoning and neighborhood demographics (Rothwell and Massey, 2009; Trounstine, 2020; Freemark, 2023). By using transaction-level homeowner demographic data from HMDA, this study estimates the extent to which zoning laws shift neighborhood demographics, including race and income. The spatial discontinuity design embedded in the analysis addresses the endogeneity of zoning laws, further enriching the understanding of zoning's influence on demographic composition.

The remainder of the paper is organized as follows. Section 2 details the datasets used in the analyses. Section 3 introduces the algorithm for estimating minimum lot sizes based on property characteristics and provides an overview of the current state of zoning using the constructed dataset. Section 4 presents the municipality border analysis, examining the effects of zoning on housing prices and demographic composition. Finally, Section 5 concludes the paper.

2 Data

CoreLogic Tax Assessor Data

The primary dataset used for constructing zoning data is the CoreLogic Tax Assessor data, which includes property tax records from 2009 to 2019 for both residential and non-residential properties. CoreLogic collects this data from local government entities responsible for levying property taxes, typically county governments. Each record provides detailed property characteristics, including building type, lot area, number of rooms, construction year, and zoning district name, as well as tax-assessed values and tax amounts for each tax year. The analysis throughout the paper is restricted to single-family homes, identified using property type and land use category information.

CoreLogic tax records are merged with Census County Subdivision and Census Place maps to assign parcels to municipalities with functioning local governments as defined in the 2010 Census Guide to State and Local Census Geography (see Appendix Table A1 for details on municipal definitions). To focus on urban and suburban contexts, observations outside of Core-Based Statistical Areas (CBSAs) are excluded, thereby removing rural areas from the analysis. This dataset is utilized for structural break detection to infer minimum lot size restrictions in Section 3 and serves as a source of control variables in Section 4, where it is merged with CoreLogic Deed-HMDA and CoreLogic MLS data.

Corelogic Deed-Home Mortgage Disclosure Act Data

CoreLogic deed data, available since the 1980s, contains information on deed and mortgage transactions. I focus on sale transactions of single-family homes from 2000 to 2018 and merge them with property characteristics from CoreLogic Tax Assessor data, including lot size, building square footage, (effective) construction year, and the number of bedrooms and bathrooms. These transactions are then linked to Home Mortgage Disclosure Act (HMDA) loan application register data, following the approach of Bayer et al. (2007), which matches records by Census tract, mortgage year, and lender name.[6] The match rate is 65.7% for CoreLogic deed records of residential properties with mortgage information, resulting in a dataset with transaction-level prices and demographics (race, ethnicity, and income) of mortgage applicants for single-family home sales.[7]

To identify single-family homes near municipality borders, I overlay the CoreLogic-HMDA data on Census County Subdivision and Census Place maps.[8] Border regions are defined as areas within 0.1 km (0.06 miles) to 1.0 km (0.6 miles) of a border, with 0.5 km serving as the baseline distance. This baseline aligns with standard choices in spatial boundary discontinuity

6. Other studies implementing this merging procedure include Avenancio-Leon and Howard (2019) and Diamond and McQuade (2019).

7. This match rate is consistent with previous studies using the same method. For instance, Bayer et al. (2007) achieved a 60% match rate in the Bay Area and demonstrated the representativeness of the matched sample.

8. Appendix Table A1 provides more details on the definition of municipality boundaries.

designs, which typically range from 0.24 km (0.15 miles) to 0.56 km (0.35 miles) (Bayer et al., 2007; Turner et al., 2014). For price analyses, I use CoreLogic deed data prior to merging with HMDA to include cash sales and unmatched transactions, resulting in a sample of 9.2 million transactions (see Panel A of Table 1 for summary statistics). For demographic analyses, I rely on the merged CoreLogic-HMDA dataset, which includes information on borrowers' race and income, resulting in a sample of 4.0 million transactions (see Panel B of Table 1 for summary statistics).

Corelogic Multiple Listings Service Data

CoreLogic Multiple Listings Service (MLS) data comprises sale and rental listings collected from 154 multiple listing services (MLS).[9] Each MLS operates its own database of property listings, typically covering specific local markets. CoreLogic aggregates these individual databases into a single dataset, which includes information such as the list date and price, and, for closed listings, the closing date and price. Since sales information is already captured in the CoreLogic deed data, I focus on rental listings in the MLS data, using closing prices of single-family home rental listings in the analysis. Although rental listings represent only a small portion of the data, I identify approximately 185,000 geocoded single-family home rental listings with closing prices within 0.5 km of municipality borders. Summary statistics for these rental listings are provided in Panel C of Table 1.

Census Block Group Characteristics

I collect Census Block Group (CBG) characteristics from the 2020 American Community Survey, including population, demographic composition, income, education, the number of housing units, median home value, and median rent. Additionally, I calculate the distance to the Central Business District (CBD) for each CBG using CBD coordinates from Fee et al. (2013). These CBG characteristics are utilized in the descriptive analysis of minimum lot sizes in Section 3.3.

9. According to the Real Estate Standards Organization, there were 597 multiple listing services in 2020.

1940 Census Full-Count Data

The pre-zoning municipal characteristics are derived from the 1940 Ancestry.com and IPUMS complete-count Census restricted data (Ruggles et al., 2017). I geocode the street addresses from individual-level 1940 Census data using Smarty geocoder and merge them with 2019 municipal boundaries to obtain historical municipality characteristics. Since many municipalities were established during the postwar suburbanization period, this dataset is relatively sparse. Therefore, I construct a variable indicating whether a municipality existed as of 1940 and include it as a control variable. For municipalities that existed in 1940, I incorporate their characteristics, such as total population, household size and structure, mobility, demographic composition (e.g., race and income), homeownership rates, average home values, and average rents. These control variables are applied in the border analysis in Section 4.

2019 School District Boundaries

I merge the geocoded CoreLogic deed, deed-HMDA, and MLS datasets with 2019 school district boundaries. This allows me to identify the unified school district, elementary school district, and secondary school district where each transacted home is located. I construct school district fixed effects for each combination of unified, elementary, and secondary school districts. These fixed effects are interacted with municipality fixed effects and included as controls in the border analysis in Section 4.

3 New Nationwide Dataset of Minimum Lot Size Restrictions

Existing zoning datasets primarily rely on survey methods or manual compilation of local zoning laws, but they face several limitations. First, they often have narrow and/or granular coverage of jurisdictions.[10] Second, land use surveys often lack specific details about zoning ordinances,

10. Data scarcity is especially pronounced when data is manually collected; compiling local zoning laws manually is both expensive and logistically challenging at the national level, given the existence of over 30,000 zoning jurisdictions in the United States. For example, Bronin (2021) assembled zoning data for 180 jurisdictions in Connecticut, a process that required over four months and the efforts of 20 research assistants. Recent works, such as Mleczko and Desmond (2023) and Bartik et al. (2023), have begun developing natural language processing methods to collect and clean zoning details more efficiently.

making them less suitable for studying particular zoning reforms.[11] Finally, land use surveys are prone to low response rates and measurement errors (Mleczko and Desmond, 2023). These limitations in zoning datasets have posed significant challenges to the expanding empirical literature on zoning, particularly for researchers aiming to study broad geographic regions.

To overcome these challenges, I develop a scalable algorithm to infer zoning regulations from observed property characteristics, constructing a nationwide dataset on neighborhood-level stringency of residential zoning. Specifically, I estimate minimum lot sizes applied to single-family homes by detecting structural breaks in constructed single-family lot sizes at the zoning district and Census Block Group levels using CoreLogic Tax Assessor data. This geographically detailed zoning dataset encompasses 52 million single-family homes across 16,217 municipalities within 825 Core-Based Statistical Areas (CBSAs) in the United States.[12] Finally, I validate these estimates against actual regulations using zoning data from 228 zoning districts across 55 municipalities, compiled from the MAPC Zoning Atlas and other local zoning codebooks.

Throughout this paper, I focus on minimum lot size regulation as it represents the most critical dimension of zoning laws in single-family home construction for several reasons. First, minimum lot size regulation is the most prevalent type of zoning law. For instance, 91% of jurisdictions surveyed in the Wharton Land Use Regulatory Index and 96% in the Terner Center California Land Use Survey report imposing minimum lot size restrictions. Second, minimum lot size regulation is among the most distortionary zoning laws affecting single-family home construction. To illustrate, I calculate the bunching rate of single-family home construction around minimum lot sizes and maximum floor-area ratios (FAR), another commonly studied zoning regulation. In CoreLogic data merged with the MAPC Zoning Atlas, 18% of single-family home constructions after 1940 bunch at minimum lot sizes, compared to only 5% that bunch at maximum FAR limits. Finally, minimum lot size regulation often proxies for the overall

11. For instance, the Density Restriction Index (DRI) in the Wharton Residential Land Use Regulation Index only indicates whether a community has any minimum lot size requirement and, if so, whether the largest minimum lot size falls into one of four broad categories: smaller than 0.5 acres, 0.5 to 1 acre, 1 to 2 acres, or larger than 2 acres. Such reporting is typically limited to the "strictest" or "most typical" regulation levels, often provided in intervals.

12. For comparison, the 2020 American Community Survey reports 86 million single-family homes within CBSAs. The minimum lot size data excludes single-family homes in areas without functioning local government entities. Consequently, the dataset covers at least 60% of single-family homes subject to local zoning laws.

stringency of zoning, as municipalities frequently pair large minimum lot size requirements in low-density districts with other restrictive zoning measures, such as low maximum FAR and stringent height limits. Thus, while residential zoning regulates multiple aspects of building characteristics, minimum lot size regulation is particularly impactful, playing a central role in shaping single-family home construction, and reflecting the broader stringency of zoning laws.

3.1 Constructing the Minimum Lot Size Dataset

Here, I describe the procedure for constructing the minimum lot size dataset in greater detail. I explain how geolocation, construction year, and lot size data of single-family home constructions are utilized to apply a structural break detection algorithm to infer minimum lot size regulations.

Construct Neighborhood × Construction Period Grids

Local governments divide their land into zoning districts and establish zoning laws specific to each district. For example, stricter zoning restrictions are often imposed in residential-only neighborhoods on the periphery, while more flexible regulations are applied in mixed-use neighborhoods closer to city centers.[13] As a result, zoning laws can vary considerably across zoning districts, even within the same municipality. This makes the zoning district level the most appropriate scale for measuring the stringency of zoning laws.

To proxy zoning districts and capture within-municipality variations in zoning, I partially rely on zoning district information available in the CoreLogic tax data. However, this zoning district data is incomplete and inconsistent.[14] Therefore, in municipalities where zoning district information is unavailable in the data, I use Census block groups to define neighborhoods. Census block groups are a suitable spatial unit because their boundaries partially align with local legal boundaries and are consistently available nationwide in shapefile format.

Since existing structures are typically grandfathered in when zoning laws are newly adopted or

13. For instance, many cities designate R1 and R2 districts, both of which are residential zones, but R2 permits denser housing development. Municipalities frequently have multiple residential districts with varying levels of restriction.

14. In the sample, 46% of single-family home parcels have a populated CoreLogic zoning data field. When populated, the field often contains unstandardized zoning district names, such as R1, R-1, or Residential low density.

amended, effective minimum lot size restrictions may vary by construction year. Unfortunately, the adoption years of zoning laws are not directly observable, and collecting this information is extremely challenging (Cui, 2023). As a result, for my baseline minimum lot size estimates, I focus on single-family homes constructed after 1940, as most municipalities implemented local zoning regulations during the mid-to-late 1900s (Fischel, 2004). When validating these estimates, I analyze how the choice of construction year restrictions influences the precision of the minimum lot size estimates, emphasizing the trade-off between shorter periods that capture only recent constructions and longer periods that include older constructions as well. Furthermore, in the border analysis in Section 4, I test the robustness of the results using various versions of minimum lot size estimates.

Minimum Lot Size Detection

New construction must comply with minimum lot size regulations unless it qualifies for exception criteria, applies for zoning variances, or undergoes rezoning, all of which are less common and typically costly. Consequently, new construction on smaller lots is likely either to comply with the minimum lot size requirement or to forego development altogether. This creates a structural break in the distribution of observed residential lot sizes at the minimum lot size, characterized by a flattening of the distribution function to the left of the regulation threshold and a sharp increase at the threshold. Figure 1 illustrates this structural break using observed single-family lot sizes from CoreLogic tax data across four sample zoning districts. The figure clearly highlights the structural break at the actual minimum lot size requirement, depicted by the dashed line.

I detect a structural break in the distribution of lot sizes motivated by the structural break detection literature (Andrews, 1993; Zeileis, 2005) and define these breakpoints as the minimum lot sizes. 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 neighborhood (either zoning district or Census Block Group), i is a single-family home, $x_{i,h}$'s are the observed lot sizes, N_h is the # of single-family homes in the neighborhood, and $\hat{F}_h(\cdot)$ is the empirical distribution of single-family home lot sizes in the neighborhood. $F_{h,1}$ and $F_{h,2}$ are the cumulative distribution functions for single-family homes smaller than the minimum lot size and for single-family homes bigger than or equal to the minimum lot size, 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. I identify the value $\bar{x}_h$ that minimizes the sum of squared residuals with a single breakpoint at each neighborhood with more than 50 observed single-family home lot sizes. I use this estimate as my minimum lot size estimate. To illustrate, Figure 1 depicts the actual minimum lot sizes (in dashed lines) and detected structural breaks (in solid lines) in four example zoning districts. I apply the structural break detection algorithm nationwide.

I acknowledge a couple of limitations in this minimum lot size estimation approach using structural break detection. First, the method may inadvertently capture preference-driven bunching or mass development. For instance, if lots were uniformly divided into one-acre parcels during an initial development phase, the algorithm might falsely detect one acre as a structural break and infer it as the minimum lot size. Second, the algorithm identifies only the most prominent structural break, potentially overlooking multiple breaks. Multiple structural breaks may arise due to temporal changes in minimum lot sizes or inaccuracies in geographic proxies for zoning districts. To mitigate these concerns, I demonstrate the robustness of the subsequent analyses in Section 4 by using multiple versions of minimum lot sizes based on different construction periods. Additionally, when CoreLogic zoning district names are unavailable, I use the most granular proxy available—Census Block Groups—to minimize the risk of a single proxy encompassing multiple zoning districts with varying minimum lot sizes.

3.2 Validation of Minimum Lot Size Estimates

I validate the minimum lot size estimates using the MAPC Zoning Atlas and manually compiled local zoning codebooks. The MAPC Zoning Atlas aggregates zoning information for the Greater Boston area and provides minimum lot size data for each zoning district. As one of the few publicly available datasets offering actual minimum lot sizes across multiple municipalities, it provides a valuable resource for validating district-level minimum lot size estimates.[15] Because the MAPC Zoning Atlas is restricted to the Greater Boston area, I supplement it with a manually compiled zoning dataset that includes 21 municipalities spanning all four U.S. regions (Midwest, Northeast, South, and West).[16] The final validation dataset includes 228 single-family residential zoning districts across 55 municipalities matched to the district-level minimum lot size estimates.

Figure 2 illustrates the relationship between actual minimum lot size (MLS) and the estimates generated using the structural break detection method ($\widehat{MLS}$). Table 2 additionally provides detailed error statistics. The baseline $\widehat{MLS}$, derived from single-family homes constructed since 1940, exhibit a strong positive relationship with actual minimum lot sizes, with the correlation between $\widehat{MLS}$ and MLS of 0.96 and the correlation between $\log \widehat{MLS}$ and $\log MLS$ of 0.92.[17] Additionally, the median absolute error rate, defined as $\left|\frac{MLS-\widehat{MLS}}{MLS}\right|$, is 11 percent. In 89 percent of zoning districts, the absolute error rate remains within 50 percent. Overall, the validation exercise demonstrates the accuracy of MLS estimation using the structural break detection method.[18]

When applying construction period restrictions different from the baseline, the precision of the estimated minimum lot sizes varies, as shown in Figure 2 and Table 2. While restricting to more recent constructions increases the likelihood of capturing current minimum lot size regulations, it also imposes stricter sample constraints, reducing the statistical power of structural break detection, particularly in districts with infrequent construction activity. When structural

15. Most other zoning datasets offer limited or aggregated information on minimum lot sizes. For example, the Wharton Residential Land Use Regulation Index reports only the largest minimum lot sizes in each municipality, categorized into ranges. Similarly, the Connecticut Zoning Atlas identifies whether minimum lot size regulations exist and classifies them into broad categories: less than 0.47 acre, 0.47 to 0.91 acre, 0.92 to 1.84 acres, or greater than 1.84 acres.

16. The additional sample municipalities are intentionally selected to ensure a broad geographic range and represent diverse municipality types across the U.S. The dataset includes five municipalities from each of the four U.S. regions, as well as the District of Columbia. Among these, seven are densely populated urban cities with populations exceeding 500,000: Phoenix, AZ; Philadelphia, PA; Jacksonville, FL; El Paso, TX; Seattle, WA; Washington, DC; and Milwaukee, WI. Nine are mid-sized cities with populations between 100,000 and 500,000: Tulsa, OK; Cincinnati, OH; Gilbert, AZ; Scottsdale, AZ; Montgomery, AL; Cary, NC; Bellevue, WA; Stamford, CT; and Fargo, ND. The remaining five are smaller towns with populations under 100,000: Canton, MI; Toms River, NJ; Nashua, NH; West Hartford, CT; and Edina, MN.

17. Note that $\log MLS$ is the main treatment variable in the border analysis in Section 4.

18. To provide benchmarks to evaluate the magnitude of the absolute error rates, Appendix Figure 2 presents the error distribution when minimum lot sizes are approximated using simple statistics, such as the 1st or 10th percentile of single-family constructed lot sizes.

breaks are detected among single-family homes constructed since 1970, for example, the median error rate is 3.33 percent, with a slightly lower but still strong correlation of 0.87 between MLS and $\widehat{MLS}$ of 0.87. When the sample is further restricted to homes constructed since 2000, it results in a lower correlation with actual MLS and higher absolute error rates, particularly in the higher deciles. Also, 50 districts are dropped due to insufficient data. These dropped districts have a median actual MLS of 10,000 square feet and a mean of 19,439 square feet. This pattern likely arises because districts with larger minimum lot sizes tend to show less frequent construction and redevelopment activity.

Overall, the structural break detection method accurately estimates current minimum lot sizes, though its precision can be limited by temporal variations in minimum lot sizes or inaccuracies in zoning district boundaries. In areas with frequent single-family home construction, the method can be applied using a granular spatial and temporal grid. However, in areas with fewer construction activities, such granular estimation may introduce excessive noise due to insufficient data. To balance precision and robustness, I use baseline minimum lot size estimates derived from single-family lots constructed since 1940 and conduct robustness checks using alternative minimum lot size estimates in Section 4.

3.3 Descriptive Statistics: The State of Local Zoning Laws in the United States

The constructed minimum lot size data offers a comprehensive overview of residential zoning regulations across the United States. First, I analyze the overall stringency of zoning, represented by median minimum lot sizes, and bunching rates around the regulated lot sizes across the nation. Next, I examine how zoning stringency varies within Core-Based Statistical Areas (CBSAs) and its correlation with neighborhood characteristics at the Census Block Group level.

I find that minimum lot size regulations appear to distort housing characteristics, with stringency varying widely across the U.S. Figure 3 illustrates the distribution of minimum lot sizes, with a median around 10,000 square feet and a notable concentration at 43,560 square feet (1 acre). Additionally, Figure 4 highlights state-level variation by presenting median estimated minimum lot sizes and bunching rates as proxies for the bindingness of the regulations. For example, New Hampshire and Vermont have median minimum lot sizes near 1 acre but relatively lower bunching rates of approximately 11%, indicating that these regulations are less restrictive given the local market conditions. In contrast, states like New Jersey, California, Florida, and Illinois enforce median minimum lot sizes between 7,000 and 9,000 square feet, with higher bunching rates indicating more restrictive zoning relative to local market conditions2: 23.7% in New Jersey, 23.6% in California, 22.8% in Florida, and 22.2% in Illinois. Nationally, approximately 18.5% of single-family homes built after 1940 cluster within 5% of their regulated lot sizes, highlighting the practical constraints imposed by these regulations.

Within CBSAs, zoning stringency exhibits significant variation based on proximity to the Central Business District (CBD). Figure 5 presents binscatter plots of allowed density, defined as the maximum number of homes per 10,000 square feet under minimum lot size restrictions $(10,000/\widehat{MLS})$. The left panel uses absolute distance to the CBD as the regressor, while the right panel uses relative distance normalized by the furthest Census Block distance within the CBSA. Both plots reveal that minimum lot size is generally larger farther from the CBD, with municipalities allowing greater residential density closer to city centers. This pattern seemingly contrasts with findings by Baum-Snow and Han (2024), which find that supply responses increase with CBD distance. However, this discrepancy can be explained by land availability: farther from the CBD, minimum lot sizes may be less binding due to more developable land available, enabling larger supply responses even under larger minimum lot size regulations.

Finally, I analyze the relationship between minimum lot sizes and neighborhood characteristics within CBSAs using 2020 American Community Survey data. In addition to incorporating CBSA fixed effects, one specification also includes distance to the CBD as a control variable, accounting for the strong correlation between zoning and proximity to the CBD, which shows consistent results. Figure 6 highlights several key findings. First, more stringent zoning (larger MLS) is associated with lower housing density, indicating that zoning regulations shift building characteristics. Second, higher zoning stringency correlates with increased home values and rents, demonstrating that zoning regulations impact housing prices. Finally, zoning stringency is closely related to neighborhood demographics. Specifically, stricter zoning is associated with a higher proportion of non-Hispanic white residents, lower proportions of Black, Asian, and

Hispanic residents, and higher median income and education levels. These findings suggest that higher-income, higher-educated, and white households are more likely to live in neighborhoods with stringent zoning, potentially driven by affordability barriers and/or preferences for such neighborhood environments.

4 Estimating the Effects of Zoning in Housing Markets

Zoning laws shape housing markets through several mechanisms. First, restrictive zoning shifts building characteristics toward larger, lower-density, and typically more expensive homes. Second, zoning regulations influence neighborhood amenities such as congestion, agglomeration, and local public good provisions. The neighborhood premiums associated with stricter zoning may be positive or negative, depending on household preferences for these endogenous amenities. Third, stricter zoning constrains housing supply, which increases prices, all else being equal. However, by limiting development options, stricter zoning may also reduce the unit price of land, holding other factors constant. Finally, zoning provides insurance value by protecting neighborhoods from undesirable future changes, potentially enhancing property values. Overall, zoning laws impact housing prices through various channels, making it crucial to estimate their overall effects and disentangle the underlying mechanisms as key empirical questions.

The descriptive analysis in the previous section underscores the correlation between zoning stringency, housing prices, and neighborhood demographics. However, the potential correlation between zoning stringency and unobserved amenities complicates causal interpretation. To address this issue, this section employs a spatial discontinuity design to mitigate endogeneity concerns and analyze the effects of residential zoning on housing prices. Furthermore, I disentangle the impact into two components: the direct effect, where restrictive zoning increases home sizes and consequently raises costs, and the neighborhood effect, where zoning alters neighborhood amenities. Lastly, I examine how residential zoning contributes to demographic sorting along dimensions of homeowner race and income.

4.1 Empirical Strategy

A major challenge in estimating the effects of residential zoning on housing market outcomes is the potential correlation between zoning stringency and unobserved location amenities. For example, suburban areas with high-quality residential environments may impose stricter zoning to discourage the construction of smaller, lower-value homes, while densely populated cities with well-developed infrastructure may allow denser development under more relaxed zoning regulations for reinforcing the agglomeration economies. Consequently, minimum lot sizes may be positively correlated with location amenities in the former case and negatively correlated in the latter. Since these locational amenities also affect housing market outcomes, such as prices and demographics, the OLS coefficients of minimum lot size on these outcomes are likely to be biased.

To mitigate the endogeneity concerns, I adopt a boundary discontinuity design developed by Black (1999). This approach compares housing market outcomes at border segments of neighboring municipalities (cities, towns, townships, and counties with functioning local governments) where zoning stringency differs. As described in Section 2, I identify single-family homes within 0.5 km of municipality boundaries to define border regions, with robustness checks provided in Section 4.4 for alternative border distance choices. To refine the analysis, I further divide each border region into up to four segments using a k-means clustering algorithm applied to the longitude and latitude of single-family parcels. This additional segmentation is particularly useful for long municipality borders, where parcels far apart may not be directly comparable even within the same border region. The number of clusters is determined for each border region to ensure at least 20 sales transactions per segment on both sides of the border. Figure 7 illustrates the construction of these border segments. On average, a border segment contains 1,044 sales transactions and 259 rental transactions. The minimum lot size on the more stringent side of a border segment is, on average, 71% larger and 38% larger at the median.

By comparing homes within close proximity, this boundary discontinuity design accounts for geographic conditions that simultaneously affect zoning and housing markets, such as land availability or distance to the Central Business District (CBD). For example, areas with limited land due to topographic constraints might impose smaller minimum lot sizes while maintaining high housing prices driven by land scarcity. Moreover, municipalities may allow denser construction closer to the CBD for agglomeration benefits while housing prices are high due to proximity to job opportunities. Controlling for these confounding factors is crucial, and the boundary discontinuity design achieves this by focusing on border segments within a small geographic area in the same metro area, with similar proximity to the CBD.

To further address the remaining endogeneity concerns, I incorporate pre-zoning municipality attributes, which may affect both zoning stringency and housing market conditions. Specifically, I use the 1940 Full-Count Census to construct an indicator variable identifying whether a municipality was developed before 1940 and include municipality-level demographic and housing market characteristics, such as population, racial composition, homeownership rates, home values, and rents, as described in Section 2. These controls account for pre-existing differences between municipalities prior to the adoption of zoning regulations.

Additionally, I subsequently replace pre-zoning municipality attributes with municipality-by-school district fixed effects to further address potential differences in local public services, such as trash collection, snow removal, and school district quality, which may correlate with zoning stringency. The inclusion of these granular fixed effects is feasible due to the variation in minimum lot size estimates, both across and within municipalities. However, it is important to note that local service quality may partially mediate the effects of zoning. For instance, stringent zoning may increase property tax revenue, thereby improving local services. While these fixed effects better control for locational amenities that change discontinuously at borders, they may also attenuate the coefficient estimates by capturing part of the zoning effect itself. Therefore, I present empirical results from two specifications, one including border segment fixed effects and 1940 municipality characteristics, and the other including border segment fixed effects and municipality-by-school district fixed effects throughout the paper.

In summary, I analyze housing market outcomes, such as sales price, rent, and homeowner demographics, in relation to zoning stringency at the transaction level within border segments, incorporating pre-zoning attributes or municipality-by-school district fixed effects as control variables. The key identification assumption is that unobserved amenities correlated with zoning stringency are effectively random, conditional on observable characteristics, within a pre-specified small region around the border.

Two important caveats should be noted when interpreting the estimated coefficients. First, the main treatment variable—zoning stringency, measured by $\log$ minimum lot size as estimated in Section 3—is subject to measurement error, which would attenuate the estimated effects of zoning laws. Second, the boundary discontinuity design may not fully account for unobserved discontinuities at borders. For instance, one side of a border segment might exhibit exclusive local politics that lead to stricter zoning regulations while also restricting new development through additional measures, thereby increasing housing prices. In such cases, the political leanings of the neighborhood could affect both zoning stringency and housing market outcomes, creating endogeneity concerns that are not entirely addressed. While specifications including municipality-by-school district fixed effects help mitigate these issues by controlling for municipality-level heterogeneity, they do not account for heterogeneity within municipalities across districts.

4.2 Effects on Housing Prices

My baseline transaction-level regressions take the form

$$\log p_{it} = \beta_{MLS} \log MLS_i + X_i \beta_X + \lambda_t + \lambda_{b(i)} + \lambda_{m\times s} + \varepsilon_{ibmt}, \tag{1}$$

where i denotes a single-family home transaction and t is the housing market, defined as county-by-transaction year. $\log p$ is log of sales price or rent, and $\log MLS_i$ is log of minimum lot size (MLS) applied to the single-family home, as constructed in Section 3. Because maximum housing density is constrained by $\frac{1}{MLS}$, $\log MLS$ scales with the restricted density and thus serves as the main treatment variable. In this specification, I assume the treatment effects of minimum lot size restrictions are linear in $\log MLS$. In Appendix Figure A1, I present alternative binscatter regressions, which yield consistent results. X_i includes a set of historic municipality characteristics and the distance to the border as controls. Appendix Table A3 provides the full list of historic characteristics sourced from the 1940 Census. λ_t denotes county-by-transaction year fixed effects, λ_b denotes border segment fixed effects, and $\lambda_{m\times s}$ denotes municipality (city, town, township, or county)-by-school district fixed effects. The inclusion of border segment fixed effects enables the spatial discontinuity design, comparing transactions within the same border segment. The main parameter of interest, β_{MLS}, measures the effects of zoning stringency, measured by the minimum lot sizes.

Table 4 reports the coefficient estimates using the baseline sample including observations within 0.5km of municipal borders. The OLS regressions with the full set of controls except for the border segment fixed effects in Columns (1) and (5) show a strong relationship between stringent zoning and higher home sales prices ($\hat{\beta}_{MLS} = 0.2637$, p-value < 0.001) and higher rents ($\hat{\beta}_{MLS} = 0.1473$, p-value < 0.001). Columns (2) and (6) add border segment fixed effects, and the results indicate that restrictive zoning (larger MLS) increases home sales prices ($\hat{\beta}_{MLS} = 0.1861$, p-value < 0.001) and rents ($\hat{\beta}_{MLS} = 0.1193$, p-value < 0.001). These estimates suggest that doubling MLS would increase home sales prices by approximately 14 percent and rents by 9 percent.

Columns (3) and (7) include municipality-by-school district fixed effects, which control for variations in municipal services and school district quality. However, as discussed in Section 4.1, β_{MLS} may be attenuated because municipal services and school district quality could partially reflect mechanisms of zoning laws. Despite this caveat, Columns (3) and (7) yield coefficient estimates statistically indistinguishable from those in Columns (2) and (6). Specifically, $\hat{\beta}_{MLS}$ is estimated at 0.1771 (p-value < 0.001) for sales prices and 0.1193 (p-value < 0.001) for rents. These estimates indicate that doubling MLS would increase sales prices by approximately 13 percent and rents by 8 percent.

Note that the outcome variables measure the sales price and rent of a home, rather than per square footage price or rent. As a result, the coefficient estimates in Columns (1), (2), (3), (5), (6), and (7) capture the combined effects of minimum lot size regulations—both by increasing home size, which in turn affect the quantity of housing consumption, and by influencing locational amenities, which in turn affect per square footage prices. To disentangle these effects, I next isolate the extent to which stringent zoning impacts building characteristics, which are regarded as the primary mechanism through which zoning laws operate.

Figure 8 presents the coefficient estimates of β_{MLS} with building characteristics as outcome variables, including lot size, building square footage, effective building age, number of bedrooms, and number of bathrooms. The results show that restrictive zoning (larger MLS) generally increases home size, as reflected in larger lot sizes, greater square footage, and more rooms. Additionally, restrictive zoning appears to increase the age of rental homes but does not significantly affect the age of sold homes, potentially indicating that older homes in strictly zoned neighborhoods are more likely to be listed as rentals.

As demonstrated above, stringent zoning shifts the housing supply toward larger homes. To examine how these shifts in building characteristics affect prices, I repeat the border analysis of housing prices, incorporating observed building characteristics as controls in Equation 1. Columns (4) and (8) in Table 4 report the results. Controlling for building characteristics reduces $\hat{\beta}_{MLS}$ by 78 percent for sales prices and 72 percent for rents. This suggests that 78 percent of the sales price effect and 72 percent of the rent effect can be attributed to zoning's direct impact on building characteristics, while the remaining effects are explained by other neighborhood factors.

The remaining effects, captured by $\hat{\beta}_{MLS}$ in Columns (4) and (8), reflect neighborhood environments shaped by stringent zoning, which are capitalized into housing prices. For example, low-density neighborhoods may offer quieter, more desirable living conditions. Additionally, stringent zoning may provide an "insurance value" by protecting neighborhoods from undesirable changes and preserving property values over time. However, I also acknowledge that unobserved building quality correlated with zoning stringency may confound these effects. For example, if homes in neighborhoods with large MLS requirements are more likely to feature high-end finishes or luxury interiors, this could bias the estimated neighborhood effects upward. Additionally, the mechanism by which MLS restricts supply may not be fully captured in the border analysis due to potential spillover effects on the opposite side of the border. Specifically, since border segments are in close proximity, the negative price effects of MLS stemming from restricted supply could extend to neighboring areas, thereby attenuating the supply channel captured in the border design.

The estimated neighborhood effects of stringent zoning are relatively modest. For example, doubling neighborhood zoning stringency while holding building characteristics constant is associated with a 2.8 percent increase in home prices and a 2.2 percent increase in rents. Although relatively small in magnitude, these positive neighborhood price effects suggest that existing homeowners have an economic incentive to advocate for more restrictive zoning. When zoning becomes more stringent, existing properties are grandfathered in, preserving their building characteristics. Nevertheless, homeowners still benefit from the neighborhood price premium associated with stringent zoning.

4.3 Effects on Residential Sorting

I examine whether zoning laws induce demographic sorting. To do so, I repeat the border analysis using Equation 1, with homeowner race and income from CoreLogic-HMDA data as the outcome variables. Panel A of Table 5 reports the baseline estimates of β_{MLS} for homeowner race and income. Consistent with the Census Block Group-level descriptive analysis in Section 3.3, Columns (1) and (5) indicate that more restrictive zoning (larger MLS) is associated with a higher share of non-Hispanic white homeowners ($\hat{\beta}_{MLS} = 0.0592$, p-value < 0.001) and higher-income homeowners ($\hat{\beta}_{MLS} = 0.2169$, p-value < 0.001), even when the analysis is restricted to the border sample and controls for 1940 neighborhood characteristics, distance to the border, and county-by-year fixed effects.

When border segment fixed effects are included in Columns (2) and (6), $\hat{\beta}_{MLS}$ is estimated at 0.0329 for the likelihood of a homeowner being non-Hispanic white and 0.0160 for log homeowner income. These coefficients change slightly to 0.0306 and 0.1546, respectively, when municipality-by-school district fixed effects are added in Columns (3) and (7), with all p-values remaining below 0.001. Quantitatively, the results indicate that doubling MLS increases the probability of a homeowner being non-Hispanic white by approximately 2 percentage points (where the mean of this outcome variable in the analysis sample is 0.67) and raises homeowner income by about 11 percent.

Since race and income are highly correlated, I additionally report the estimated effects on race by income quartiles in Panel B and the estimated effects on income by race and ethnicity in Panel C of Table 5. Panel B demonstrates that racial sorting with respect to zoning stringency is strong across all income groups. Specifically, non-Hispanic white homeowners exhibit a disproportionately strong preference for living in neighborhoods with stringent zoning. This may be attributed to their preferences for larger homes, low-density neighborhoods, or other endogenous amenities associated with stringent zoning. Similarly, the results in Panel C indicate that income sorting with respect to zoning stringency is strong across all racial and ethnic groups. Together, these findings suggest that residential sorting is significant along both dimensions—race/ethnicity and income levels—rather than one solely driving the other due to their correlation.

Demographic sorting with respect to zoning stringency is critical to understanding exclusionary zoning. Exclusionary zoning refers to a discriminatory practice in which communities impose restrictive zoning laws to exclude racial minorities and low-income households. My analysis of demographic sorting finds that stringent zoning disproportionately attracts white and high-income homeowners. This suggests that zoning laws can potentially serve as a mechanism for the implicit exclusion of racial minorities and low-income households, as has been historically documented.

4.4 Robustness Checks

In Figure 9, I present the border analysis results under alternative specification choices. Overall, the results are consistent, with a few statistically significant changes in coefficient estimates.

First, I test the sensitivity of the analysis to using an alternative minimum lot size (MLS) estimate. The baseline MLS is derived from a structural break in constructed single-family lots built since 1940, as described in Section 3. Alternatively, I use an MLS estimate based on all single-family construction, which provides broader geographic coverage, and post-1970 construction, which offers a more precise estimate of current MLS but with reduced coverage, as robustness checks. As shown in Panel (a) of Figure 9, using MLS from all construction yields a statistically indistinguishable result across all outcome variables—sales price, rent, homeowner race/ethnicity, and homeowner income. In contrast, the MLS estimate from post-

1970 construction produces a slightly smaller coefficient estimate for sales price and homeowner income. This difference may reflect the fact that post-1970 MLS estimates are limited to areas with more frequent development, which may be less mature and thus exhibit smaller premiums from stringent zoning laws.

Second, I repeat the analysis with a subset of border segments. I first select borders where municipalities on both sides do not appear in the 1940 Census, indicating newly established borders. When restricted to these borders, the coefficient estimate remains stable, except for an increased estimated price premium, as illustrated in Panel (a) of Figure 9 ("New borders" color code). This may indicate that when zoning laws are adopted in an existing neighborhood, existing construction is grandfathered in, attenuating its price effect. In contrast, the new border specification may capture the longer-term effect of zoning laws, where an entire area is newly developed or reconstructed in compliance with zoning regulations, maximizing the premium. Next, I analyze a border within a single school district and present the result in Panel (a) of Figure 9 ("Within SD" color code). The coefficient estimate shows a slight reduction when municipality-by-school district fixed effects are not included, but it remains statistically indistinguishable when these fixed effects are included. This pattern reflects that school district quality tends to be higher in areas with more restrictive zoning, although the coefficient estimate is generally robust to school district sample restrictions.

Finally, I present the result using observations at varying distances from the border in Panel (b) of Figure 9. Specifically, I analyze 10 subsamples based on proximity to the border: 0 to 0.1 km, 0.1 to 0.2 km, ..., and 0.9 to 1 km, with the baseline sample consisting of the first five subsamples. Across these border distances, the coefficient estimate remains largely stable, except for a gradual increase in the sales price premium as the distance from the border increases. This finding suggests that the price premium associated with stringent zoning is larger in a city center than in a peripheral area. One possible explanation is that as the distance from the border increases, the neighborhood is more likely to be surrounded by a low-density environment resulting from stringent zoning, amplifying the price premium.

5 Conclusion

Residential zoning has long been criticized for limiting housing supply and exacerbating the housing affordability crisis. Despite the growing body of empirical literature on zoning laws, the lack of comprehensive zoning data remains a significant challenge. In this paper, I construct a nationwide dataset on minimum lot size regulations and examine their impact on housing markets. My findings show that stringent minimum lot size restrictions significantly increase housing prices, primarily by shifting building characteristics, and intensify segregation by disproportionately attracting high-income white homeowners.

One major contribution of this paper is the creation of a geographically detailed, nationwide dataset of minimum lot sizes—one of the most prevalent types of zoning laws that restrict housing density. I propose a novel approach to estimate dimensional requirements by identifying structural breaks in observed lot sizes of construction. Using this structural break detection algorithm, I generate zoning district-level minimum lot size estimates that offer extensive geographic coverage and fine granularity. Leveraging this dataset, I conduct a nationwide analysis to examine the effects of minimum lot size regulations on housing markets.

To estimate the impact of minimum lot size restrictions, I adopt a spatial discontinuity design that exploits variations in zoning stringency at municipal borders. This design helps address the endogeneity concerns that unobserved locational amenities may simultaneously influence zoning stringency and housing market outcomes. Additionally, I control for pre-zoning municipality characteristics from 1940, before zoning laws were widely adopted. My analysis finds that larger minimum lot sizes increase both home sales prices and rents. For instance, doubling minimum lot sizes increases home sales prices by 14 percent and rents by 9 percent. I also find that zoning laws intensify residential segregation, as restrictive zoning disproportionately attracts high-income white homeowners. These results are robust to the inclusion of municipality-by-school district fixed effects and various subsample analyses.

Zoning laws are increasingly scrutinized as many regions grapple with housing affordability crises. By defining minimum requirements for housing construction, restrictive zoning eliminates affordable housing options and raises housing prices. This paper contributes to the literature by providing a comprehensive measure of zoning stringency across the U.S., enabling future empirical research on residential zoning. Furthermore, this research quantifies the impact of zoning laws on housing markets, offering valuable insights for policymakers and guiding ongoing zoning reform efforts in the U.S. and globally.

Table 1 — Descriptive Statistics of Border Analysis Sample

A. Deed sample statistics (within 0.5km of borders)

# obs.	9,177,086
# parcels	5,011,216
# borders	8,792

Variable	Q1	Q2 (Median)	Q3	Mean	SD
Sales price (in $1000s)	114	185	295	244	227
Min lot size (in SF)	6,843	8,930	12,650	12,438	14,752
Lot size (in SF)	6,400	8,751	13,620	24,326	1,763,504
Building footage (in SF)	1,304	1,700	2,293	1,910	1,748
(Effective) Year built	1967	1990	2002	1983	25
# Bedrooms	3	3	4	3.26	0.81
# Bathrooms	2	2	3	2.32	0.93

B. Deed+HMDA sample statistics (within 0.5km of borders)

# obs.	4,009,732
# parcels	2,937,190
# borders	6,281

Variable	Q1	Q2 (Median)	Q3	Mean	SD
Sales price (in $1000s)	140	207	321	271	223
Min lot size (in SF)	6,750	8,712	12,075	11,682	13,058
1 (Non-Hispanic white)				0.672	
1 (Asian Pacific)				0.063	
1 (Black)				0.065	
1 (Hispanic)				0.098	
Income (in $1000s)	49	74	114	100	777

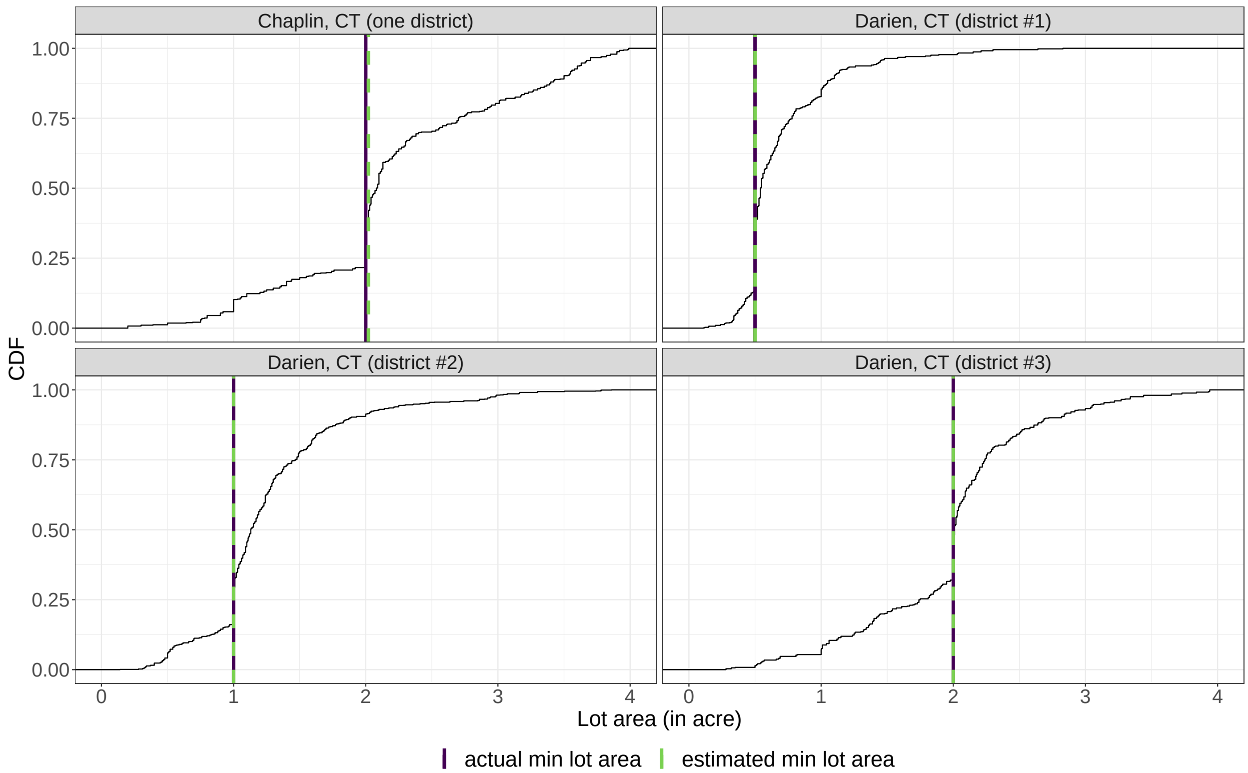
C. MLS rental listing sample statistics (within 0.5km of borders)

# obs.	185,458
# parcels	101,928
# borders	717

Variable	Q1	Q2 (Median)	Q3	Mean	SD
Monthly Rent (in $)	1,175	1,473	1,992	6,179	26,565
Min lot size (in SF)	6,125	7,797	10,200	9,420	7,336
Lot size (in SF)	5,793	7,350	9,677	14,201	499,132
Building footage (in SF)	1,438	1,829	2,365	1,979	784
(Effective) Year built	1981	1998	2005	1991	18
# Bedrooms	3	3	4	3.33	0.76
# Bathrooms	2	2	3	2.35	0.79

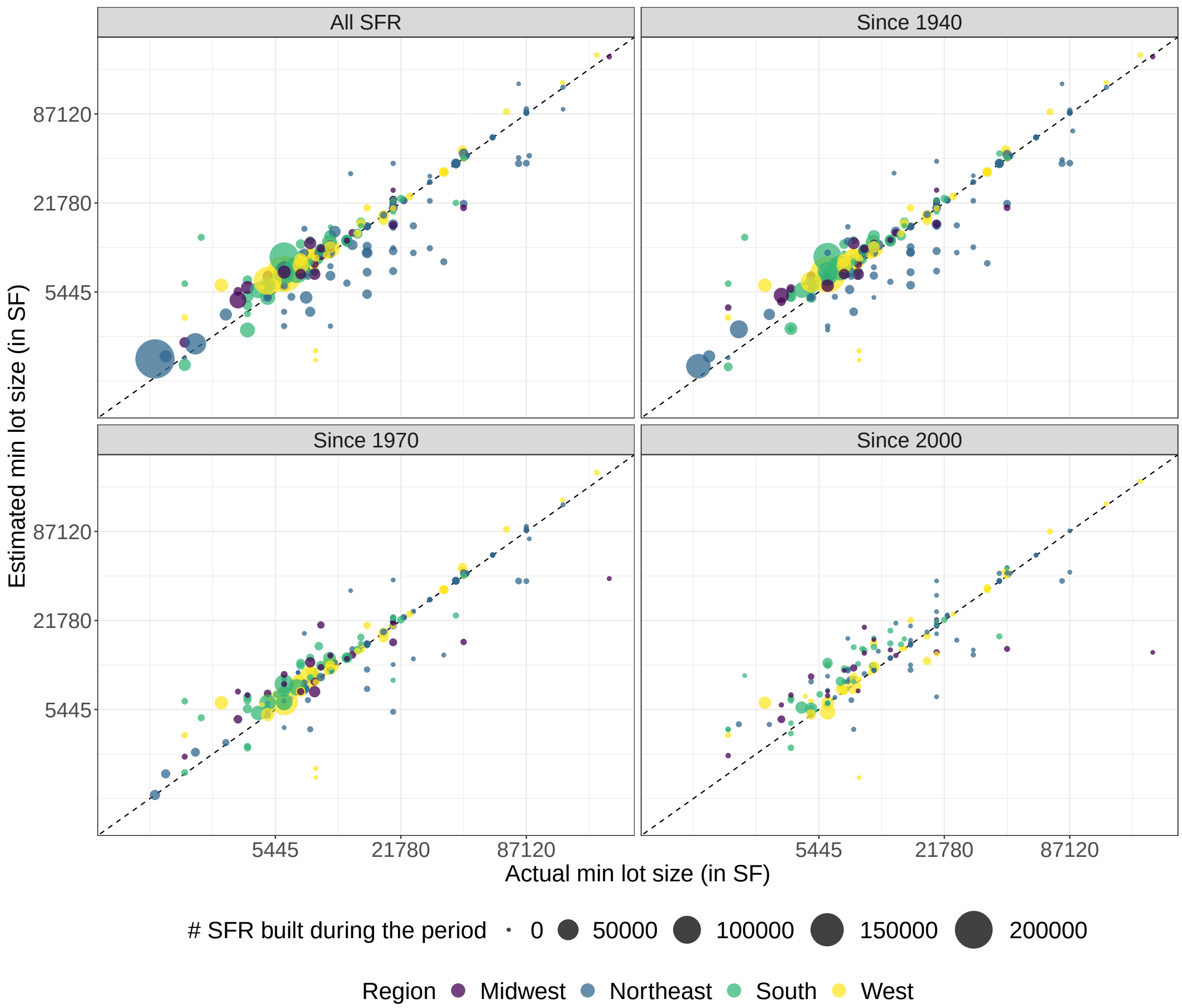
Note. This table reports sample statistics. Panel A describes the sales price data from CoreLogic deed data. Panel B describes the homeowner demographics data from CoreLogic deed data merged with Home Mortgage Disclosure Act data. Panel C describes the rental listing data from CoreLogic Multiple Listings Services. All samples are restricted to single-family homes within 0.5km of municipal borders with minimum lot size estimates and structure characteristics (lot size, building footage, year built, and the number of bedrooms/bathrooms).

Figure 1 — Illustration of Structural Break Detection for Estimating Min Lot Sizes



Note. The figure depicts the distribution of lot sizes of single-family homes built after 1940 in four example zoning districts. For each district, actual minimum lot sizes are denoted by purple solid lines, and estimated minimum lot sizes are denoted by dashed green lines. In Chaplin, the actual minimum lot size 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 sizes. The estimated minimum lot sizes are constructed by detecting structural breaks in the lot size distribution.

Figure 2 — Actual vs. Estimated Min Lot Sizes By Construction Period Restriction



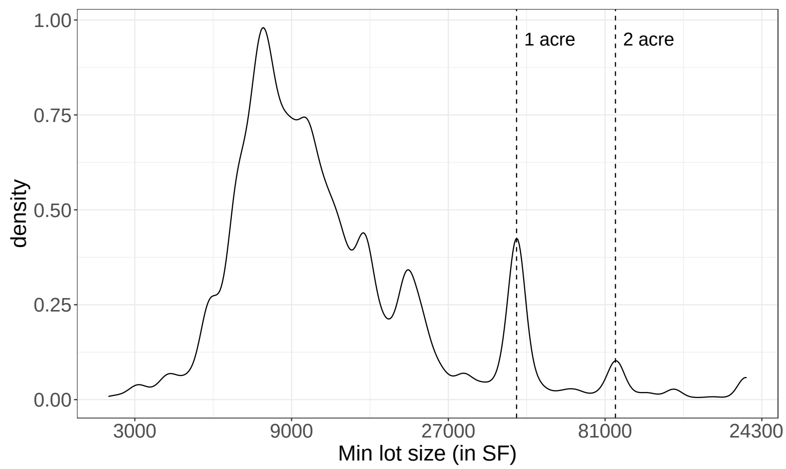
Note. The figure illustrates the relationship between actual minimum lot sizes (MLS) on the x-axis and estimated minimum lot sizes on the y-axis at the zoning district level. Each panel applies a different construction year restriction in the structural break detection: the top-left panel includes all single-family homes, the top-right panel includes homes constructed since 1940, the bottom-left panel includes homes constructed since 1970, and the bottom-right panel includes homes constructed since 2000. The actual MLS data covers 228 single-family residential zoning districts from 34 municipalities in the MAPC Zoning Atlas and 21 additional municipalities. All panels are color-coded by region (Midwest, Northeast, South, and West), and point size corresponds to the number of single-family residential parcels used for minimum lot size detection in each zoning district. Both axes are log-scaled.

Table 2 — Error Distribution of Estimated Min Lot Sizes By Construction Period Restriction

Period	Districts	Cor		Deciles of Abs Error Rate (in %)								
		level	log	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
All SFR	228	0.95	0.92	0.08	0.53	1.00	6.99	10.75	15.34	24.42	33.33	51.18
$\geq$1940	226	0.96	0.92	0.26	0.66	1.12	5.02	10.99	15.20	23.30	32.80	51.55
$\geq$1970	217	0.87	0.93	0.20	0.42	0.75	1.23	3.33	9.75	19.86	31.34	54.54
$\geq$2000	178	0.78	0.89	0.19	0.49	1.08	4.94	10.24	17.19	32.85	49.52	67.14

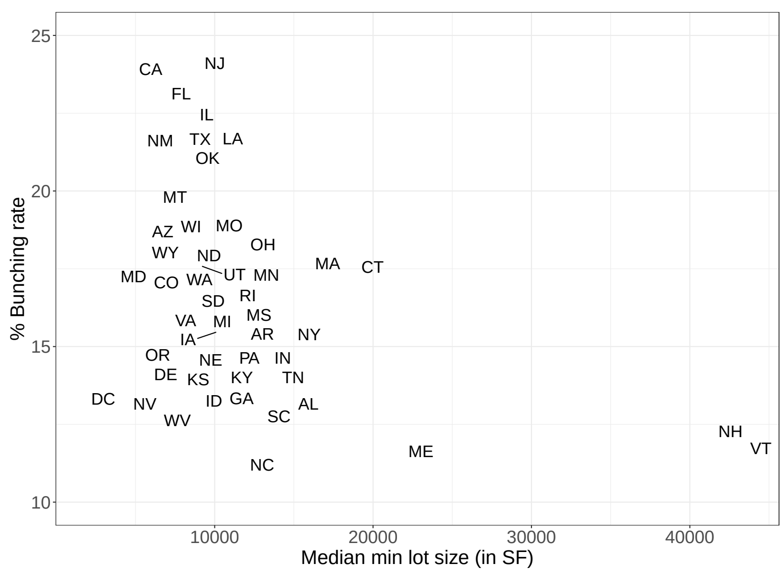
Note. This table reports the correlation and error distribution of estimated minimum lot sizes (estimated MLS) compared to actual minimum lot size regulations (actual MLS) at the zoning district level. The actual MLS data includes 228 single-family residential zoning districts from 34 municipalities in the MAPC Zoning Atlas and 21 additional municipalities. Each row corresponds to estimated MLS based on single-family homes constructed within specific time restrictions: all single-family residences, homes built since 1940, since 1970, and since 2000, respectively. The table reports the number of zoning districts where minimum lot sizes are estimated, the correlation between estimated MLS and actual MLS (labeled as "level"), and the correlation between the log of estimated MLS and the log of actual MLS (labeled as "log"). Additionally, it presents the deciles of the absolute error rate (in %), defined as $\left|\frac{\text{actual MLS - estimated MLS}}{\text{actual MLS}}\right|$.

Figure 3 — National Variation in Minimum Lot Size



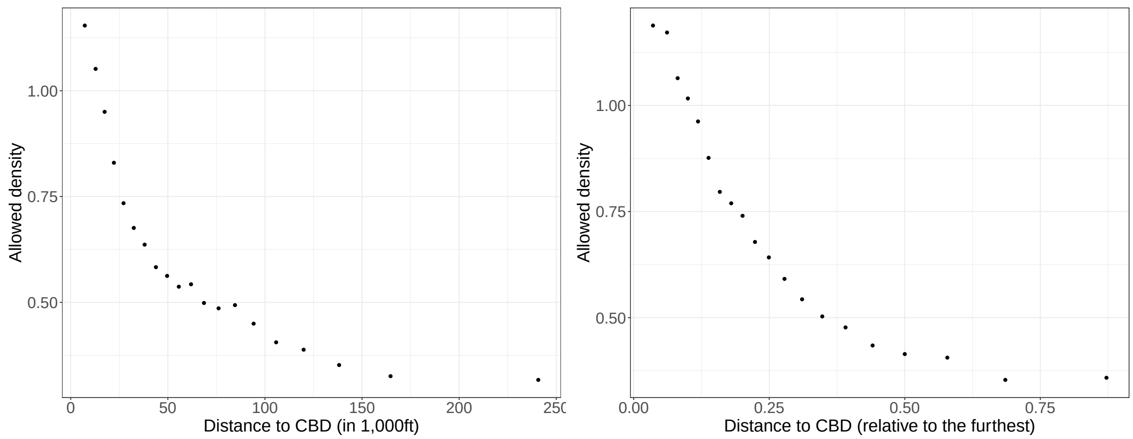
Note. The figure illustrates the density function of Census Block Group-level median minimum lot sizes across the United States, restricted to Core Based Statistical Areas. The two dashed lines represent one acre (43,560 square feet) and two acres (87,120 square feet), respectively. The x-axis is presented on a log scale.

Figure 4 — State-Level Description of Zoning Stringency



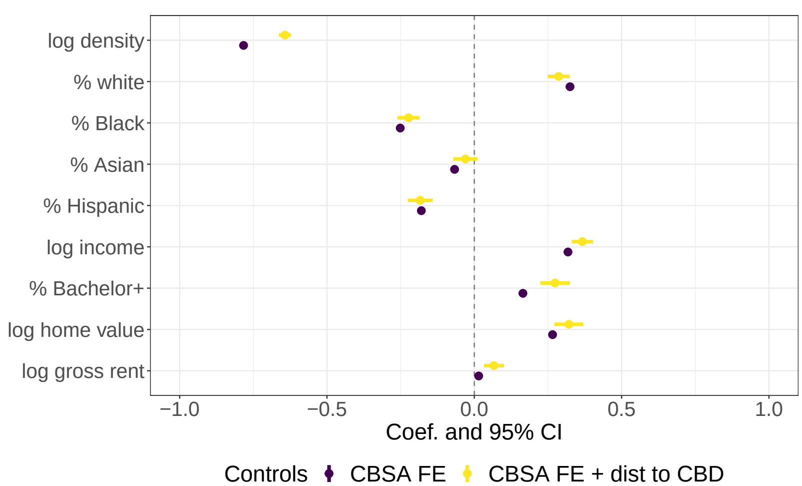
Note. The figure depicts the relationship between the median estimated minimum lot sizes (x-axis) and the bunching rate of single-family constructions around regulated density (y-axis) for each state. The medians are calculated by weighting each district-level estimated minimum lot size by the number of single-family parcels within the district. The bunching rate is defined as the probability of constructed lot sizes falling within 5 percent of the minimum lot sizes among post-1940 single-family homes. The statistics are restricted to Core-Based Statistical Areas within each state.

Figure 5 — Allowed Density by Distance to the Central Business District



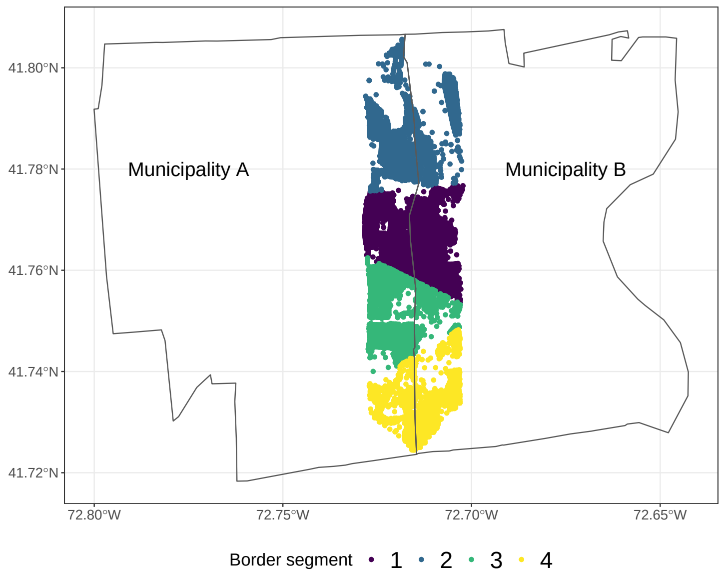
Note. The figure depicts the relationship between the allowed density and distance to the Central Business District (CBD) using binscatter regressions with 20 bins at the Census Block Group level. The x-axis represents the distance to the nearest CBD. In the left panel, distance is measured in 1,000 feet, while in the right panel, it is expressed as a relative proportion of the furthest block group within the CBSA. The y-axis represents allowed density, measured by the maximum number of homes allowed in 10,000 square feet $(10,000/\text{median } \widehat{MLS})$. Both regressions include CBSA fixed effects.

Figure 6 — Correlation Between Zoning Stringency and Census Block Group Characteristics



Note. The figure depicts the coefficient estimates (dots) and their 95% confidence intervals (lines) from Census Block Group (CBG)-level univariate OLS regressions. The regressor is the log of the median minimum lot size in each CBG. The outcome variables include log population density (measured as population divided by CBG land area), % non-Hispanic white, % Black, % Asian, % Hispanic, log median income, % of the population with a Bachelor's degree or higher, log median home value, and log gross rent, respectively. All outcome variables are normalized by their standard deviation. All regressions include CBSA fixed effects.

Figure 7 — Illustration of Border Segmentation



Note. This figure illustrates the construction of border segments. The boundaries of two bordering municipalities (A and B) are depicted as lines, and single-family homes in the border region are represented as points. The border region is divided into four segments, with homes in each segment shown in a different color.

Table 3 — Balance Test of 1940 Municipality Characteristics

Outcome variable	(1)	(2)	# obs.
log population	0.2333	0.0569	2,178,354
	(0.2232)	(0.2033)	
avg. household size	−0.0164	−0.0211	2,178,354
	(0.0309)	(0.0289)	
log mean income	−0.0162	−0.2874	2,175,192
	(0.0246)	(0.4287)	
% white	−0.0159	−0.0063	2,178,354
	(0.0112)	(0.0055)	
% owner	0.0217^{*}	0.1283	2,177,932
	(0.0121)	(0.1155)	
log mean home value	−0.1710	0.9641	2,169,652
	(0.0119)	(1.272)	
log mean rent	−0.0313	-3.4061^{***}	2,168,175
	(0.0617)	(1.1375)	
Controls			
border seg FE	✓	✓	
SD FE		✓	

Robust standard-errors in parentheses clustered at the municipality level
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note. This table presents the regression results using various 1940 municipality-level demographic characteristics as outcome variables, based on the deed border sample restricted to municipalities observed on both sides in the 1940 Census. The treatment variable is an indicator for the side of the border segment with a larger median MLS. Column (1) includes county-by-year fixed effects and border segment fixed effects, while Column (2) additionally incorporates school district fixed effects.

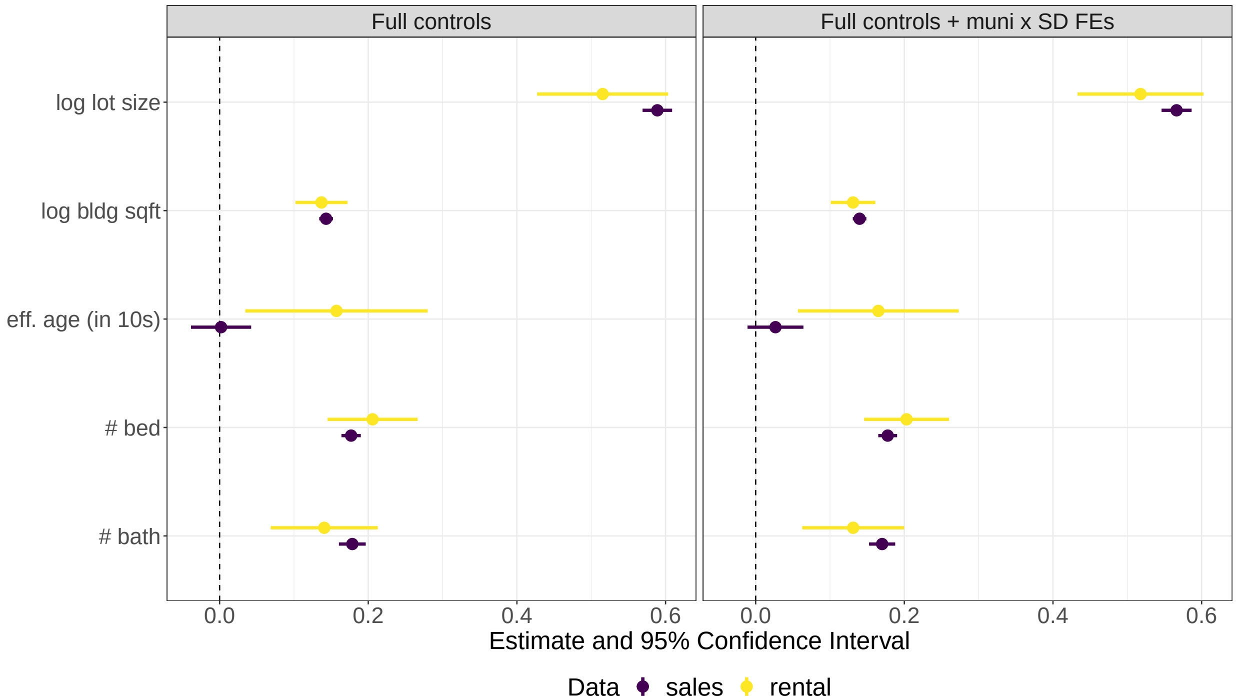
Table 4 — Estimated Effects of Zoning on Housing Prices

	log sales price				log rent			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log MLS	0.2637***	0.1861***	0.1771***	0.0393***	0.1473***	0.1193***	0.1110***	0.0315***
	(0.0099)	(0.0070)	(0.0067)	(0.0031)	(0.0200)	(0.0178)	(0.0139)	(0.0070)
log lot size				0.0743***				0.0324***
				(0.0031)				(0.0049)
log bldg sqft				0.6317***				0.4796***
				(0.0083)				(0.0130)
(effective) age				-0.0051***				-0.0025***
				(0.0001)				(0.0003)
# bed				-0.0157***				-0.0040
				(0.0018)				(0.0031)
# bath				0.0681***				0.0374***
				(0.0019)				(0.0037)
Controls								
border seg FE		✓	✓	✓		✓	✓	✓
loc gov't × SD FE			✓	✓			✓	✓
Observations	9,177,086	9,177,086	9,177,086	9,177,086	185,458	185,458	185,458	185,458
R^2	0.38914	0.50126	0.51306	0.61409	0.73783	0.77045	0.77592	0.81541

Robust standard-errors in parentheses clustered at the municipality level
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note. This table reports baseline coefficient estimates of regression 1. The outcome variable in Columns (1) to (4) is log sales price, and the outcome variable in Columns (5) to (8) is log rent. All columns include county-by-transaction year fixed effects, 1940 neighborhood characteristics, and distance to the border as control variables. Columns (2) to (4) and (6) to (8) implement the border discontinuity design by including border segment fixed effects. Columns (3), (4), (7), and (8) include municipality-by-school district fixed effects. Finally, Columns (4) and (8) additionally include building-level characteristics (lot size, building square footage, age, number of bedrooms, and number of bathrooms).

Figure 8 — Estimated Effects of Zoning on Building Characteristics



Note. The figure depicts the coefficient estimates (dots) and 95% confidence intervals (lines) from the border regressions where the outcome variables are building characteristics (y-axis), and the main explanatory variable is log min lot size. Both panels include county-by-transaction year fixed effects, 1940 neighborhood characteristics, and distance to the border as control variables. The right panel additionally includes municipality-by-school-district fixed effects. For each outcome variable in each panel, regression results using two sample data are shown: sales data from CoreLogic deed database (purple), and rent data from CoreLogic MLS database (yellow). All sample is restricted to transactions within 0.5 km of municipality borders.

Table 5 — Estimated Effects of Zoning on Demographic Sorting

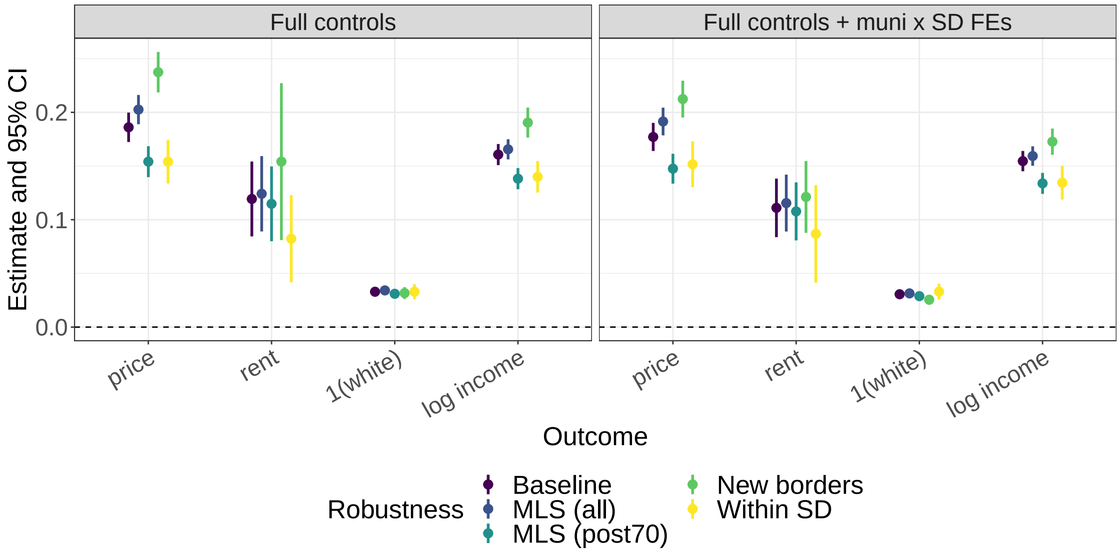
	1(race = non-Hispanic white)				log income			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Baseline								
log MLS	0.0592***	0.0329***	0.0306***		0.2169***	0.1607***	0.1546***	
	(0.0032)	(0.0021)	(0.0020)		(0.0069)	(0.0050)	(0.0048)	
Panel B. By Income Quartile								
Q1 ($N = 810,780$)				0.0285***				
				(0.0029)				
Q2 ($N = 761,476$)				0.0238***				
				(0.0024)				
Q3 ($N = 752,527$)				0.0228***				
				(0.0021)				
Q4 ($N = 777,095$)				0.0226***				
				(0.0019)				
Q5 ($N = 767,616$)				0.0209***				
				(0.0021)				
Panel C. By Race/Ethnicity								
Non-Hispanic white								0.1512***
($N = 2,620,238$)								(0.0045)
Hispanic								0.1511***
($N = 379,128$)								(0.0080)
Black								0.1066***
($N = 252,163$)								(0.0116)
Asian/Pacific								0.1430***
($N = 247,155$)								(0.0089)
Native								0.1564***
($N = 15,243$)								(0.0445)
Other								0.1561***
($N = 355,567$)								(0.0054)
Controls								
border seg FE		✓	✓	✓		✓	✓	✓
loc gov't × SD FE			✓	✓			✓	✓
Observations	4,009,732	4,009,732	4,009,732	.	3,869,494	3,869,494	3,869,494	.
R^2	0.12866	0.18468	0.18982	.	0.22280	0.32364	0.33426	.

Robust standard-errors in parentheses clustered at the municipality level
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

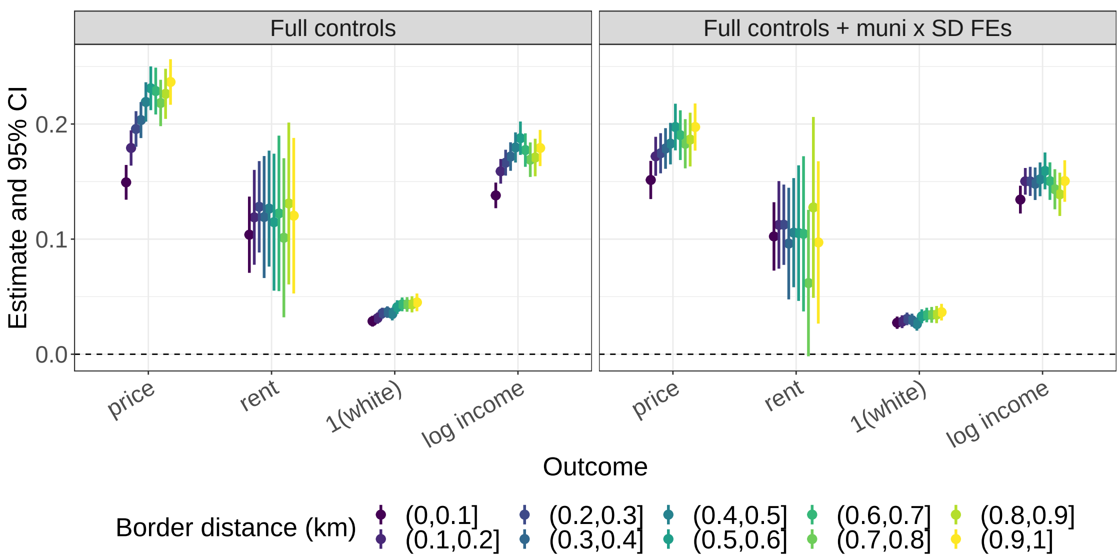
Note. This table reports coefficient estimates of β_{MLS} from the border regressions using homeowner demographics from CoreLogic-HMDA data. The outcome variable in Columns (1) to (4) is an indicator variable of homeowner being non-Hispanic white, and the outcome variable in Columns (5) to (8) is homeowner income. All columns include county-by-transaction year fixed effects, 1940 neighborhood characteristics, and distance to the border as control variables. Columns (2) to (4) and (6) to (8) implement the border discontinuity design by including border segment fixed effects. Columns (3), (4), (7), and (8) include municipality-by-school district fixed effects. Finally, Columns (4) and (8) run the regressions by income quartiles and by race/ethnicity, respectively, and report the estimates and standard errors of β_{MLS}.

Figure 9 — Robustness Checks for Border Analysis

(a) Robustness to Alternative MLS and Selection of Borders



(b) By Distance to the Border



Note. The figure displays the coefficient estimates (dots) and 95% confidence intervals (lines) from the border regression with alternative specifications. The x-axis represents the outcome variables: sales price, rent, indicator for homeowner being non-Hispanic white, and log homeowner income. The left subplots include county-by-transaction year fixed effects, 1940 neighborhood characteristics, and distance to the border as control variables, while the right subplots additionally include municipality-by-school district fixed effects. Panel A presents the results from five different specifications for each outcome variable: baseline (purple), alternative MLS estimates using all constructions (navy), alternative MLS estimates using post-1970 constructions (blue), a sample restricted to municipalities not listed in the 1940 Census (green), and a sample restricted to within-school district border segments (yellow). Panel B shows the results using 10 subsamples based on proximity to the border: 0 to 0.1 km, 0.1 to 0.2 km, ..., and 0.9 to 1 km from the border.

References

Albouy, David, and Gabriel Ehrlich. 2018. "Housing productivity and the social cost of land-use restrictions." *Journal of Urban Economics* 107 101–120.

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.*

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

Baum-Snow, Nathaniel, and Lu Han. 2024. "The microgeography of housing supply." *Journal of Political Economy* 132 (6): 1897–1946.

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.

Brueckner, Jan K, and Kala Seetharam Sridhar. 2012. "Measuring welfare gains from relaxation of land-use restrictions: The case of India's building-height limits." *Regional Science and Urban Economics* 42 (6): 1061–1067.

Cui, Tom. 2023. "The Emergence of Exclusionary Zoning Across American Cities."

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.

Fee, Kyle, Daniel Hartley et al. 2013. "The relationship between city center density and urban growth or decline." *Revitalizing American Cities* 3 45–64.

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

Freemark, Yonah. 2023. "Zoning Change: Upzonings, Downzonings, and Their Impacts on Residential Construction, Housing Costs, and Neighborhood Demographics." *Journal of Planning Literature* 08854122231166961.

Fulton, Bill, David Garcia, Ben Metcalf, Carolina Reid, and Truman Braslaw. 2023. "New pathways to encourage housing production: A review of California's Recent Housing Legislation." *Terner Center for Housing Innovation.*

Ganong, Peter, and Daniel Shoag. 2017. "Why has regional income convergence in the US declined?" *Journal of Urban Economics* 102 76–90.

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.

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.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing constraints and spatial misallocation." *American Economic Journal: Macroeconomics* 11 (2): 1–39.

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.

Kulka, Amrita, Aradhya Sood, and Nicholas Chiumenti. 2023. "Under the (neighbor) Hood: Understanding Interactions Among Zoning Regulations." *Available at SSRN 4082457.*

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.*

Mleczko, Matthew, and Matthew Desmond. 2023. "Using natural language processing to construct a National Zoning and Land Use Database." *Urban Studies* 00420980231156352.

Nechamkin, Emma, and Graham MacDonald. 2019. "Predicting Zoned Density Using Property Records." *Washington, DC: Urban Institute.*

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."

Resseger, Matthew. 2022. "The impact of land use regulation on racial segregation: Evidence from Massachusetts zoning borders." *Mercatus Research Paper.*

Rothwell, Jonathan, and Douglas S Massey. 2009. "The effect of density zoning on racial segregation in US urban areas." *Urban Affairs Review* 44 (6): 779–806.

Ruggles, Steven, Catherine A Fitch, Ronald Goeken, J David Hacker, Evan Roberts, and Matthew Sobek. 2017. "IPUMS Restricted Complete Count Data. Version 2.0." *University of Minnesota.*

Sahn, Alexander. 2021. "Racial diversity and exclusionary zoning: Evidence from the great migration."Technical report, Working Paper. Online: https://drive. google. com/file/d/10 . . . .

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.

Zabel, Jeffrey, and Maurice Dalton. 2011. "The impact of minimum lot size regulations on house prices in Eastern Massachusetts." *Regional Science and Urban Economics* 41 (6): 571–583.

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 Figures and Tables

Table A1 — Municipality definition by state

	Description	2010 Census count	# identified
AL	All incorporated places (167 cities, 293 towns)	460	461
AR	All incorporated places (311 cities, 191 towns)	502	501
AZ	All incorporated places (45 cities, 45 towns)	90	91
CA	All incorporated places (459 cities, 21 towns)	480	482
CO	All incorporated places (75 cities, 196 towns)	271	271
CT	All county subdivisions (169 towns)	169	169
DE	All incorporated places (10 cities, 44 towns, 3 villages)	57	57
DC	District of Columbia (1 city)	1	1
FL	All incorporated places (268 cities, 124 towns, 19 villages)	411	412
GA	All incorporated places (425 cities, 105 towns, 2 balances)	535	536
ID	All incorporated places (201 cities). Lost River city is inactive and not included in the data	201	200
IL	1 incorporated place independent of any township (Chicago). 12 incorporated places in counties where county subdivisions are nonfunctioning election precincts. 1432 functioning Census county subdivisions (townships) excluding the 13 incorporated places	1445	1444
IN	2 incorporated places independent of any county subdivisions (Indianapolis, Terra Haute). 1009 county subdivisions that are not undefined or unorganized territories excluding Indianapolis and Terra Haute cities (1005 townships, 3 towns, 1 cities)	1011	1008
IA	County subdivisions excluding unorganized territories (1598 townships, 59 cities that are either wholly or partially independent of MCDs that create 62 CCDs)	1660	1661
KS	All county subdivisions (1403 townships while 129 of them are inactive, 120 cities creating 127 MCDs)	1530	1530
KY	All incorporated places (420 cities, 1 urban county, 1 balance)	422	417
LA	All incorporated places (69 cities, 128 towns, 107 villages)	304	304
ME	County subdivisions that are not plantations, gore, American Indian reservations, or unorganized territories (433 towns, 22 incorporated places)	455	453
MD	All incorporated places (29 cities, 123 towns, 5 villages)	157	157
MA	53 incorporated places independent of MCDs. 298 county subdivisions that are towns	351	349
MI	County subdivisions that are not undefined MCDs (1123 townships, 117 charter townships, 275 incorporated places that are cities creating 293 MCDs)	1533	1540

	Description	2010 Census count	# identified
MN	County subdivisions that are not unorganized territories or undefined MCDs excluding 23 nonfunctioning townships in Lake of the Woods County (1785 active townships, 845 incorporated places creating 893 MCDs)	2678	2672
MS	All incorporated places (110 cities, 169 towns, 19 villages)	298	298
MO	331 active townships, 959 incorporated places (637 cities, 110 towns, 212 villages while 8 of the incorporated places are inactive) that are not dependent on 331 active townships	1087	1087
MT	All incorporated places (52 cities, 75 towns, 1 city with no description, 1 balance)	129	129
NE	County subdivisions that are not election precincts or election districts (435 active townships, 77 incorporated places that are independent of MCDs creating 79 CCDs)	512	501
NV	All incorporated places (18 cities, 1 place with no legal descriptor)	19	19
NH	County subdivisions that are not townships, locations, purchases, grants, or undefined, excluding Livermore town (inactive). 13 incorporated places that are independent	234	234
NJ	County subdivisions that are not undefined MCDs (242 townships, 324 incorporated places that are independent of MCDs comprised of 254 boroughs, 52 cities, 3 towns, and 3 villages)	566	563
NM	All incorporated places (35 cities, 19 towns, 48 villages)	102	105
NY	County subdivisions that are not boroughs, American Indian reservations, or undefined MCDs (932 towns, 61 cities creating 62 MCDs)	994	994
NC	All incorporated places (76 cities, 456 towns, 21 villages)	553	552
ND	County subdivisions that are not unorganized territories (1317 townships, 357 incorporated places that are independent of MCDs creating 364 CCDs)	1681	1673
OH	County subdivisions that are not undefined MCDs (1324 townships, 258 incorporated places that are wholly or partially independent creating 274 CCDs)	1598	1601
OK	All incorporated places (164 cities, 433 towns while four are inactive)	597	590
OR	All incorporated places (233 cities, 9 towns while 1 city is inactive)	242	240
PA	County subdivisions that are not undefined MCD (1547 townships while 1 of them is inactive, 1015 incorporated places creating 1027 CCDs)	2574	2572
RI	County subdivisions that are not undefined MCDs (31 towns, 8 incorporated cities independent of MCDs)	39	39
SC	All incorporated places (69 cities, 200 towns)	269	270
SD	County subdivisions that are not unorganized territories (915 townships, 311 independent incorporated places creating 320 CCDs)	1234	1222

	Description	2010 Census count	# identified
TN	All incorporated places (182 cities, 162 towns, 1 metropolitan government, 1 with no descriptor, 1 balance)	347	345
TX	All incorporated places (956 cities, 234 towns, 24 villages while 2 places are inactive)	1214	1218
UT	All incorporated places (144 cities, 101 towns)	245	250
VT	County subdivisions that are not gores or grants (237 towns actively functioning, 9 cities independent of MCDs)	246	251
VA	All incorporated places (39 cities, 190 towns)	229	228
WA	All incorporated places (208 cities, 73 towns)	281	281
WV	All incorporated places (77 cities, 148 towns, 6 villages, 1 corporation)	232	232
WI	County subdivisions that are not undefined MCDs (1257 towns, 594 incorporated places that are independent of MCDs creating 651 CCDs)	1908	1911
WY	All incorporated places (19 cities, 80 towns)	99	99
Total		32252	32232

Note. This table presents types of municipalities by state with local government entities according to the 2010 Census Guide to State and Local Census Geography. Note that the counts of municipalities include both municipalities in Core-Based Statistical Areas and those not part to any Core-Based Statistical Areas.

Table A2 — Error Distribution of Simple Proxies Compared to MAPC Zoning Atlas

A. Error rate in MLS, when proxied by 1st percentile constructed lot sizes

Decile	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
	50.1 %	55.9 %	61.6 %	68.0 %	73.8 %	75.5 %	77.6 %	82.1 %	86.4 %

B. Error rate for in MLS, when proxied by 10th percentile constructed lot sizes

Decile	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
	14.3 %	32.0 %	33.3 %	42.1 %	45.6 %	53.5 %	58.9 %	66.2 %	74.0 %

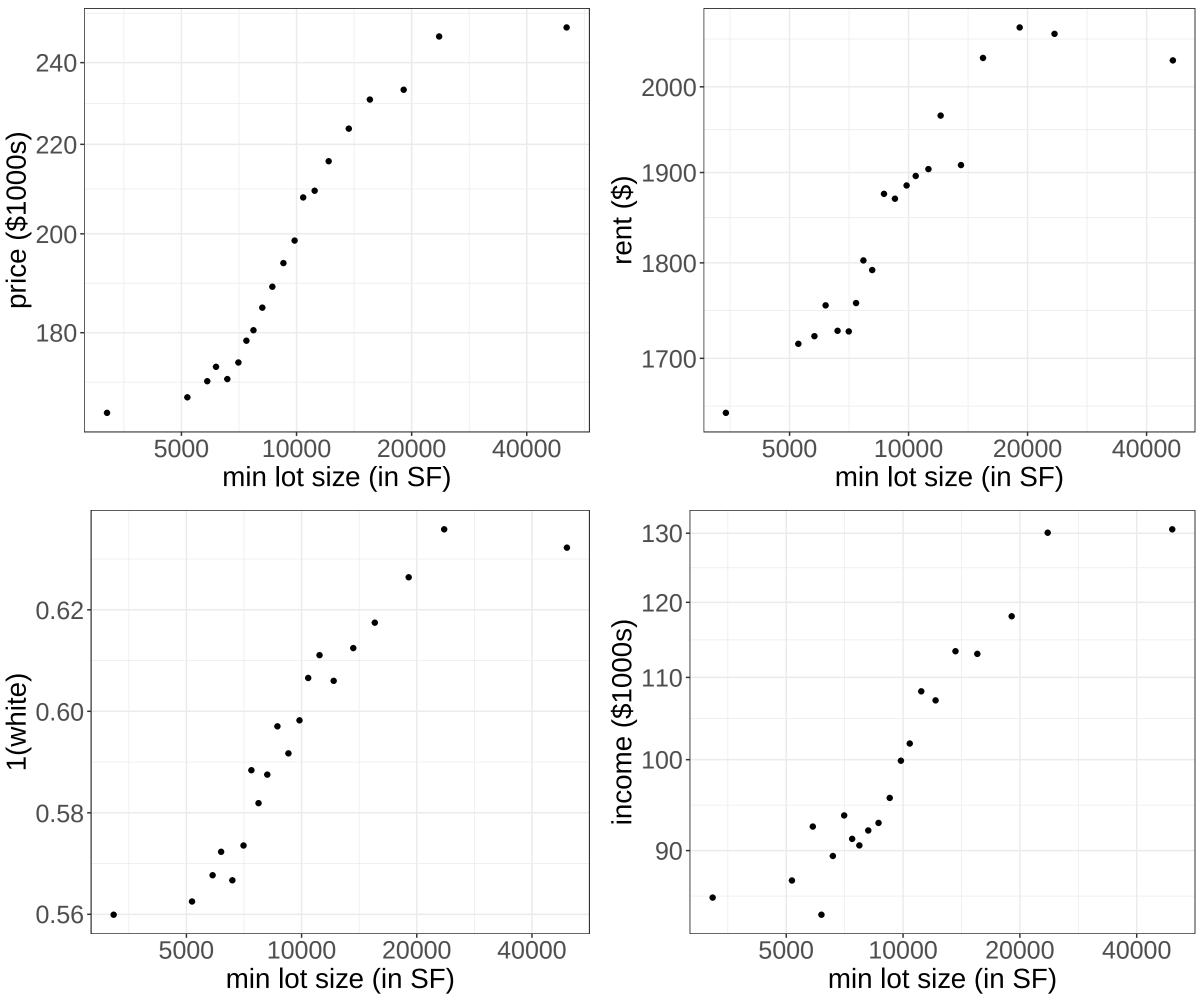
Note. This table reports the deciles of absolute error rate (in %) in other possible minimum lot size proxies, defined as = $\left|\frac{\text{estimated MLS - actual MLS}}{\text{actual MLS}}\right|$, when it is compared to actual minimum lot size regulations (actual MLS) found in MAPC Zoning Atlas, excluding outliers with minimum lot sizes larger than or equal to 5 acres. The data is restricted to single-family homes subject to minimum lot size regulations (actual MLS > 0). Panel A uses the 1st percentile of single-family lot sizes in each proxy zoning district, constructed since 1940, as proxies for minimum lot sizes. Panel B uses the 10th percentile of single-family lot sizes in each proxy zoning district, constructed since 1940, as proxies for minimum lot sizes.

Table A3 — Full list of control variables and their data sources

Variable Description	Source
Parcel characteristics	
Lot area	Corelogic Tax Assessor
Building square footage	Corelogic Tax Assessor
Effective building age	Corelogic Tax Assessor
# Bedrooms	Corelogic Tax Assessor
# Bathrooms	Corelogic Tax Assessor
1940 municipality characteristics	
Total population	Full-Count 1940 Census
Average household size	Full-Count 1940 Census
Mean household wage	Full-Count 1940 Census
% white	Full-Count 1940 Census
% Homeownership	Full-Count 1940 Census
Mean home value	Full-Count 1940 Census
Mean rent	Full-Count 1940 Census

Note. This table lists the control variables used in the border analysis in Section 4. For variables with missing values, I include a dummy variable indicating missing data and replace the missing values with -99 in the original variable. Lot area, building square footage, population, and mean household wage are included in the specifications in log-linear form, while all other variables are entered linearly. Effective building age is calculated using the effective construction year (year of substantial improvement) from the tax assessor data and, when this is unavailable, the original construction year.

Figure A1 — Binscatter Plots of Border Analysis



Note. The figure depicts the binscatter plots with 20 bins with various outcome variables. The x-axis is minimum lot size (in square feet). The y-axis is sales price (in 1000 dollars), rental price (in dollars), indicator for homeowner being white, and homeowner income (in 1000 dollars), respectively. All four binscatter regressions include 1940 IPUMS control variables, distance to borders, border region fixed effects, county-by-transaction year fixed effects, and municipality-by-school district fixed effects. Minimum lot size, prices, and homeowner income are transformed on a logarithmic scale in binscatter regressions.