Measuring Minimum Lot Size Regulation from the Housing Stock*

Jaehee Song†

September 1, 2026

Abstract

Zoning datasets rarely combine fine spatial resolution, national coverage, and temporal variation. I construct a national minimum lot size panel by detecting notches in lot-size distributions using tax assessor records for 92 million single-family and small multifamily parcels. I estimate neighborhood-level minimum lot sizes across construction cohorts from 1916 to 2025 and allow multiple minimum lot sizes within the same area and period. The estimates closely reproduce statutory minimums where regulations bind. The parcel-weighted median tract has an estimated minimum of 7,800 square feet. Minimum lot size is positively related to existing density restrictions and a lower multifamily share, but negatively related to broader regulatory stringency. I release the estimates for 76,770 Census tracts and corresponding metropolitan areas.

Keywords. Zoning, minimum lot size, land use regulation, housing supply, bunching estimation

JEL Codes. R14, R31, R52, R58

*The minimum lot size panel at the tract and metro levels developed in this paper will be made publicly available for research purposes.

†Leeds School of Business, University of Colorado Boulder. Email: jaehee.song@colorado.edu.

1 Introduction

Local zoning regulation is central to housing supply, yet it remains difficult to measure systematically. Zoning is administered by tens of thousands of local governments, with regulations embedded in codes and maps that are costly to collect and standardize. Existing measures rely primarily on surveys, most notably the Wharton Residential Land Use Regulation Index [Gyourko et al., 2008, 2021], hand-collected zoning codes and maps [Xu et al., 2023], and more recently automated text extraction [Mleczko and Desmond, 2023, Bartik et al., 2025]. These approaches have substantially expanded what researchers can observe, but coverage generally remains limited to jurisdictions that respond to surveys or digitize and publish their zoning codes. Measures that provide long-run variation in zoning at the submunicipal level remain particularly limited [Gallagher et al., 2026].

I use the distribution of constructed parcel sizes to build a national panel of minimum lot sizes. Minimum lot size is well-suited to this approach because it directly restricts residential density and, when binding, creates a detectable notch in the distribution of lot sizes. I estimate minimum lot sizes within zoning districts and Census block groups across construction-vintage windows from 1916 to 2025. The resulting panel covers 76,770 Census tracts in the contiguous United States and combines national coverage, submunicipal spatial resolution, and variation across construction cohorts. I publicly release the estimates as a Census tract panel with corresponding Core-Based Statistical Area (CBSA) aggregates, together with a zoning-authority crosswalk identifying the local government with zoning authority in each location under state enabling statutes.

I estimate minimum lot sizes by combining two complementary approaches. Both exploit the fact that effective minimum lot size requirements create a detectable notch in the empirical distribution of constructed lot sizes, but they differ in how they identify that notch. The first is a structural break detection method [Andrews, 1993], building on the implementation of Song [2025]. I refer to this as the structural break estimator. The second builds on Cui [2024] and identifies local missing mass below and excess mass above a candidate threshold. I refer to this as the local density jump estimator. Both baseline approaches assume a unique notch, meaning that a single minimum lot size governs each neighborhood within a given sample period. I report the gap between the two estimates, with a larger gap indicating that the regulation binds more weakly.

The assumption that a single minimum lot size governs a neighborhood within a given construction period is not always appropriate. A neighborhood can span zoning districts with different minimum lot sizes when zoning district boundaries are imperfectly observed. Even within a single zoning district, regulatory changes can place multiple thresholds. I therefore develop a menu estimator that allows multiple minimum lot sizes to coexist within the same neighborhood and construction period. The procedure first tests for multiple breaks and then fits a finite mixture of notched densities, extending the local density jump estimator to select the number, location, and weights of the thresholds. To limit the data requirements of estimating multiple thresholds, I restrict candidate values to a prespecified grid of conventional statutory minimum lot sizes consisting of common round values in square feet and acres.

I use a nationwide tax assessor compilation from ATTOM covering 92 million single-family and small multifamily parcels (duplexes, triplexes, and fourplexes) across 3,108 counties, combining current records with historical assessments. Each assessment reports lot size, construction year, property type, geocoordinates, and, for 46.4% of parcels, a zoning district name. I keep assessment records whenever I detect a change in lot size, construction year, property use, or district name, thereby preserving each parcel's characteristics prior to redevelopment or rezoning. Construction year provides the temporal dimension of the panel. Because existing development is generally grandfathered when zoning changes, lot sizes of homes constructed in different periods contain information about the constraints associated with those construction cohorts. I therefore estimate minimum lot sizes separately across overlapping construction-vintage windows.

Zoning authority is delegated by states to local governments, but the type of local government that exercises this authority varies across states. The relevant regulating jurisdiction is therefore not consistently a single government type, such as an incorporated place or county subdivision. I thus construct a parcel-to-zoning-authority crosswalk for the contiguous United States, extending the approach of Song [2025] by incorporating each state's zoning enabling statutes. The crosswalk assigns every parcel to one of 30.5 thousand municipal zoning authorities or, where no municipality holds zoning authority, to the county as the residual authority. Within each authority, the relevant spatial unit is the zoning district, observed directly from the zoning district name in the assessor record where available and proxied by the Census block group where it is not. In summary, I estimate minimum lot sizes within zoning authorities by zoning district and construction-vintage window, and separately by Census block group and construction-vintage window.

The minimum lot size estimates capture only constraints that bind strongly enough to be recovered from the constructed housing stock, making validation particularly important. I use four external benchmarks, including district-level statutory minimum lot sizes from the MAPC Zoning Atlas and the historical Cook County Longitudinal Zoning Database [Gallagher et al., 2026], as well as municipality-level measures from Bartik et al. [2025] and Mleczko and Desmond [2023]. The structural break estimator is the more accurate point estimator in the district-level comparisons. In recent MAPC districts, its median absolute error is 0.01 log points and 66% of estimates are within 10% of the statutory minimum. The corresponding median error is 0.08 log points in the dated Cook County panel. The menu estimator contains the statutory value within 10% of a menu estimate for up to 92% of MAPC districts. The validation also clarifies what the estimates measure. They identify lot-size thresholds in the built stock, which need not be generated by an ordinance. Notably, 43% of MAPC districts documented as having no minimum lot size nonetheless clear the significance cutoff. Such detections can arise from bunching at conventional lot sizes generated by subdivision platting and developer conventions rather than a statutory minimum lot size [Gallagher et al., 2026].

I document three main facts from the minimum lot size estimates. First, minimum lot sizes are substantial and highly dispersed. In the 2016–2025 construction window, the parcel-weighted median tract has an estimated minimum of about 7,800 square feet, and 8.5% of parcels are in tracts with an estimated minimum of at least one acre. Second, larger minimum lot sizes are more common in later construction cohorts. The one-acre share rises from 2.3% in 1916–1940 to 8.5% in 2016–2025, peaking at 17.0% in 1991–2000, with a similar increase among a fixed set of tracts. Third, minimum lot size captures a distinct dimension of zoning. The estimates are positively related to the minimum-lot-size component of the Wharton survey [Gyourko et al., 2021] but negatively related to its broader regulatory index. Additionally, within CBSAs larger minimum lot sizes are moderately associated with a lower multifamily share. Together, these patterns suggest that minimum lot size aligns with density restrictions, while other supply restrictions may be tighter in places with smaller minimum lot sizes.

The public panel contains 1.1 million tract and construction-vintage observations covering 76,770 Census tracts. The baseline tract-level measures combine estimates from the structural break, local density jump, and menu estimators where the structural break clears the significance cutoff. I also release corresponding measures using all estimates, including those that do not clear the cutoff, as well as coverage measures, estimator diagnostics, property-class-specific panels, CBSA aggregates, and the zoning-authority geography. These data allow researchers to assess the coverage underlying each estimate and to choose alternative coverage restrictions for particular applications.

Related Literature This paper contributes primarily to the literature measuring local zoning regulation. Land-use regulation is among the most important policy instruments shaping U.S. housing markets, but measuring it systematically and at scale has proved difficult. Gyourko and Molloy [2015] review the causes and consequences of land-use regulation and discuss the substantial measurement challenges facing this literature. One prominent approach relies on surveys of local governments. The Wharton Residential Land Use Regulation Index [Gyourko et al., 2008, 2021] provides the most comprehensive and widely used survey-based measures, covering thousands of jurisdictions. Other efforts focus on particular states, including Glickfeld and Levine [1992], Quigley et al. [2008], Mawhorter and Reid [2018], and Jackson [2018] in California and Ihlanfeldt [2007] in Florida. By directly surveying planners and local officials, these studies can capture multiple dimensions of the regulatory environment and administrative barriers.

Another approach focuses on zoning regulations as written. Glaeser and Ward [2009] combine locally collected regulatory information in Greater Boston and relate these measures to housing construction and prices. More recently, the National Zoning Atlas has grown from a single-state pilot into a standardized national mapping effort [Xu et al., 2023]. Most of these efforts emphasize broad spatial coverage at a point in time, while considerably fewer datasets recover how zoning changes over time. Historical studies typically require intensive collection and digitization of archival maps, ordinances, and surveys. Shertzer et al. [2018] digitize Chicago's first comprehensive zoning ordinance from 1923, and Gallagher et al. [2026] construct the Cook County Longitudinal Zoning Database, covering zoning regulations from 1921 to 1980.

Recent work has sought to scale the code-reading approach using automated text extraction. Mleczko and Desmond [2023] apply natural language processing to municipal codes to construct the public National Zoning and Land Use Database, covering minimum lot sizes and related regulations for more than 2,600 municipalities with released data and code. Bartik et al. [2025] use large language models to extract regulations from municipal code documents, reporting 96% accuracy on binary regulatory questions and a correlation of 0.87 for continuous measures. Together, these efforts have substantially expanded the geographic coverage and dimensional detail of zoning measurement. I complement this literature by using the constructed housing stock itself to recover effective minimum lot size restrictions at fine spatial resolution and over a long time horizon, providing a panel measure that can be combined with survey- and code-based measures of the broader regulatory environment.

The most closely related papers infer minimum lot size regulations from observed lot-size distributions. Zabel and Dalton [2011] apply a structural break method to estimate changes in minimum lot size regulations over time for 187 towns in eastern Massachusetts. More recently, Song [2025] applies structural break detection within zoning districts and validates the estimates against the MAPC Zoning Atlas and hand-collected ordinances, while Cui [2024] uses a discrete search algorithm to recover minimum lot size thresholds and their timing at the jurisdiction level. Macek [2026] infers minimum lot sizes nationally using a modal bunching approach, taking the smallest mode of the lot-size distribution within constructed zoning districts as the regulatory threshold. I build on these approaches by combining and comparing structural break and local density jump estimators within jurisdictions at the neighborhood level and across construction vintages. I also develop a menu estimator that restricts candidate values to a prespecified set of conventional statutory thresholds to allow multiple minimum lot sizes.

The rest of the paper proceeds as follows. Section 2 describes the data and the three dimensions that index a zoning regulation. Section 3 presents the estimators, validates them against external benchmarks, and discusses their limitations. Section 4 reports the estimates and characterizes their spatial and temporal patterns. Section 5 concludes.

2 Data

This section describes the construction of the input data used in Section 3 to infer minimum lot sizes. Section 2.1 describes the parcel-level base sample. Section 2.2 assigns each parcel to the zoning authority, district, and time period governing its regulation. Section 2.3 describes the four external benchmarks used for training and validation purposes.

2.1 Parcel Records and Sample Construction

The inputs are parcel-level tax assessor records from ATTOM, combining the most recent snapshot as of April 2026 with historical assessment panels. Each parcel assessment record includes property type, lot size in square feet and acres, year built and effective year built, geocoordinates, and, for a subset of the sample, the zoning district name. I combine the current and historical records and collapse successive observations of each parcel when there is no detected change in lot size, year built, property use, or zoning district name.[1] Each parcel is spatially joined to a 2020 Census block group.[2]

The sample comprises three residential property classes identified through ATTOM's standardized property use codes, including detached single-family homes, attached single-family homes, and small multifamily structures. The base sample pools the three property classes, while I also estimate and report minimum lot sizes separately by property class. I drop observations with missing lot size or with lot sizes below 500 square feet or above 100

[1] I allow a 5% relative tolerance for changes in lot size and treat missing zoning district names as indicating no change.

[2] For the small share of parcels for which the spatial join fails, I use the Census block group reported by ATTOM.

acres.[3] I also drop observations with a year built before 1900 or after 2026. Parcels with missing year built are retained in the pooled cross-section but excluded from construction-year windows.[4]

Panel A of Table 1 shows summary statistics of the base sample. It covers 92.0 million parcels, or 109.6 million observations counting each parcel's deduplicated historical records, across 3,108 counties and 29,677 municipal authorities. Appendix Table A1 shows the summary statistics by property class. The median parcel sits on 10,600 square feet with quartiles at 6,800 and 28,200, and the median home was built in 1979.

2.2 Three Dimensions of Zoning

A minimum lot size regulation is indexed by three dimensions, the authority that enacts it, the district within that authority to which it applies, and the time at which it is in effect.

Authority. States delegate zoning authority to local governments under enabling acts modeled on the 1924 Standard State Zoning Enabling Act. Which class of government receives that power differs across states, so the regulating jurisdiction is not a fixed Census category. Song [2025] assigns each parcel to a jurisdiction by identifying general-purpose local governments based on the 2010 Census Guide to State and Local Census Geography. I follow the approach while resolving the cases where zoning power diverges from governmental function against each state's zoning enabling statutes, so that a civil township that governs but does not zone falls to its county, which I treat as the residual zoning authority.[5] The county serves as the residual authority for 29.8% of sampled parcels, across 3,043 counties.

The crosswalk unit is accordingly the authority empowered to zone residential land under state law. Specifically, incorporated municipalities enter with their full territory, towns and townships enter only in the states whose statutes grant them zoning power and only

[3] In many counties, assessors report lot area in acres rounded to two decimal places, with the square-footage field mechanically derived from the reported acreage. As a result, recorded lot sizes may lie on a lattice of approximately 436 square feet. For each estimation unit, I construct the share of parcels whose lot square footage is derived from reported acreage. The estimators adjust bin widths and candidate resolution based on this share (see Appendix Section B.2).

[4] Because construction year exhibits unusual bunching at exactly 1900, I treat 1900 as an assessor placeholder. These observations are retained in the pooled cross-section but excluded from construction-year windows.

[5] The residual assumption is imperfect where the county holds little or no zoning power. In Connecticut and Rhode Island, for example, county governments are abolished, but towns tile the entire state, so apart from a small number of boundary join artifacts no parcel falls to the residual and the assumption is essentially never invoked. In Texas, by contrast, counties hold essentially no residential zoning power over their large unincorporated territory, so residual parcels there are genuinely unzoned and the estimators should return no binding regulation.

over the unincorporated territory they actually zone, and Census designated places are excluded throughout. The resulting layer contains 30.5 thousand municipal zoning authorities. Municipalities spanning county lines, 126 nationally, are dissolved into single authorities. Appendix Section B.1 details the construction. Of the 5,417 municipalities that Bartik et al. [2025] report as having a positive minimum lot size, 5,401 match to an authority in the panel.

A small number of jurisdictions have no zoning, most prominently Houston and much of unincorporated Texas. Even without formal zoning, lot-size distributions need not be smooth because subdivision regulations, private deed restrictions, and developer conventions can generate notches. I discuss this further in Section 3.3.

District. A single authority typically maintains several residential districts, for example R-1, R-2, and R-3, each with its own minimum lot size regulation, so the correct spatial estimation unit is the district rather than the authority. I first observe it from the assessor's zoning district name, available for 46.4% of single-family parcels nationally. The data field is cleaned based on a simple string rule. If the zoning district name is missing, I use the block group as a spatial estimation unit. Among sampled block groups, the median spans two distinct named zoning districts, and roughly a third contain no coded parcel at all and rely entirely on Census block groups.

Time. Zoning data rarely report adoption dates, but construction year provides a useful proxy. Because existing structures are generally grandfathered, a home's construction year can reveal the minimum lot size governing its construction cohort. Cui [2024] uses this logic to identify adoption timing at the jurisdiction level. I apply it within jurisdictions at the district level using overlapping construction-vintage windows stepped five years apart from 1916 onward. The main specification uses 25-year windows through 1955 and 10-year windows thereafter, retaining roughly 16% more estimation units at the 30-parcel minimum than a uniform 10-year schedule while preserving 10-year resolution after 1960. Appendix Section B.2 reports additional statistics.

Construction year, however, does not precisely identify the date of regulatory adoption. Construction may occur several years after a regulation is adopted, and regulations already in place before the observed construction cohorts can be detected but not dated. Redevelopment also removes demolished parcels from the cross-section, although the historical assessor records partially mitigate this problem. Lot boundaries may predate construction. In suburban Cook County, for example, more than half of homes built in the 1950s occupy lots platted by the 1920s [Gallagher et al., 2026]. Ordinances can codify lot sizes already common in earlier subdivisions, causing bunching to appear before formal adoption. These factors can create a mismatch between observed bunching in lot sizes and the adoption dates of statutory minimum lot sizes. I therefore interpret the temporal patterns in Section 4.1 as variation across construction cohorts rather than precise adoption dates. Section 3.3 separately examines cases in the validation sample where a minimum lot size is detected without a documented statutory requirement.

Estimation unit. The three dimensions define an estimation unit as a spatial unit, which is either zoning district or Census block group, interacted with a construction year window within a zoning authority. I prefer the zoning district estimate when available, with the block-group estimate as the fallback. I keep estimation units with at least 30 parcels observed in the base sample. Overlapping construction year windows allow a parcel to enter multiple estimation units.

Panel B of Table 1 summarizes the resulting sample of estimation units. The median named district contains 90 parcels, while the median block group contains 70 parcels. The named channel covers 28,260 districts in 7,148 authorities, while the block-group channel covers 199,576 block groups in 26,606 authorities.

2.3 External Benchmarks

I use four external benchmarks reporting minimum lot sizes, two at the district level and two at the municipality level. A subset of the benchmark observations is used to tune the estimators, while the remainder is reserved for out-of-sample validation.

Metropolitan Area Planning Council (MAPC) Zoning Atlas. The MAPC Zoning Atlas provides detailed GIS zoning data for 101 municipalities in the Greater Boston area. The data map zoning district boundaries and record district-level regulations, including minimum lot sizes, permitted residential uses, and other dimensional requirements. I use the reported minimum lot sizes and district polygons.

Cook County Longitudinal Zoning Database (CCLZD). Gallagher et al. [2026] compile archival zoning maps and bylaws for more than 120 suburban Cook County municipalities, with uniquely detailed information on adoption dates. The digitized district polygons include bylaw parameters as attributes. In the version publicly available as of August 2026, the benchmark provides 975 residential district-decade observations with documented minimum lot sizes from 1920 to 2010 and 49 dated district-decade observations with no minimum lot size.

Bartik–Gupta–Milo LLM Measures. Bartik et al. [2025] use retrieval-augmented large language models to extract zoning regulations from municipal code documents, reporting 96% accuracy on binary regulatory questions and a correlation of 0.87 on continuous ones. Their released data cover 5,805 municipalities, of which 5,417 report a positive minimum lot size. The public data are available at the municipality level and report the lowest, mean, and highest minimum lot size across each municipality's residential districts.

National Zoning and Land Use Database (NZLUD). Mleczko and Desmond [2023] release NLP-extracted zoning measures for more than 2,600 municipalities nationwide, with both the data and construction code publicly available. They apply natural language processing to municipal zoning codes to extract multiple dimensions of land-use regulation, including minimum lot size and related use and density restrictions. The minimum lot size measure is categorical and records the largest minimum lot size in the municipality using four bins: less than one-half acre, one-half to one acre, one to two acres, and two acres or more.

Benchmark Linkage. I crosswalk each benchmark to the zoning-authority layer in Section 2.2 using Census place and county-subdivision identifiers. String matches are supplemented by hand verification. MAPC and CCLZD are also matched spatially at the district level by assigning parcels to the benchmark district polygons. MAPC parcels are assigned to current Atlas district boundaries, while CCLZD parcels are assigned to historical zoning polygons for each map year. The calibration and validation split is defined at the municipality level before benchmark linkage.

3 Methodology

3.1 Minimum Lot Size Estimators

A binding minimum lot size can generate a detectable notch in the distribution of constructed lot sizes. When the regulation binds, few lots are observed immediately below the threshold, while excess mass appears at or above it as developers subdivide land subject to the constraint. Panels (a) and (b) of Figure 1 illustrate this pattern for a district with a documented statutory minimum lot size. The measurement problem is therefore to recover the location of the threshold. This reverses the standard bunching problem, where the location of a kink or notch is known and the behavioral response around it is the object of interest [Chetty et al., 2011, Kleven, 2016, Bertanha et al., 2023]. Here, the threshold itself is the object of interest.

The primary estimand is the effective minimum lot size, defined as the smallest lot size governing construction in a district and construction-vintage window as revealed by the constructed housing stock. The statutory minimum is the corresponding value written in the zoning ordinance and serves as the ground-truth minimum lot size in validation. I begin with two complementary baseline estimators. The structural break estimator identifies a break in the empirical CDF, while the local density jump estimator identifies missing and excess mass in the density. Both assume that a single minimum lot size governs each estimation unit. The menu estimator developed below relaxes this single-threshold assumption.

The structural break estimator builds on Andrews [1993] and the implementation of Song [2025]. For each estimation unit j, defined by a district and construction-vintage window, I collapse constructed lot sizes to their distinct log values $\ell_1 < \cdots < \ell_{K_j}$. Let p_k denote the share of parcels with lot size exactly equal to ℓ_k. The estimator solves

$$\widehat{\underline{\ell}}_j^{\mathrm{SB}} = \arg\min_{c \in \mathcal{C}_j} Q_j(c), \qquad Q_j(c) = \min_{g_- \in \mathcal{P}_d} \sum_{\ell_k < c} p_k \big(F_k - g_-(\ell_k)\big)^2 + \min_{g_+ \in \mathcal{P}_d} \sum_{\ell_k \geq c} p_k \big(F_k - g_+(\ell_k)\big)^2, \tag{1}$$

where $F_k = \sum_{m \leq k} p_m$ is the empirical CDF and $\mathcal{P}_d$ is the set of polynomials of degree d in ℓ. For each candidate c, I fit the CDF separately below and above the candidate by weighted least squares using p_k as weights. The estimator selects the candidate that minimizes the total weighted squared error. The candidate set $\mathcal{C}_j$ combines up to 100 of the unit's most frequent exact lot-size values with the 5th through 95th percentiles in five-percentage-point increments. When nearby candidates produce similar fits, the selection rule favors prominent local mass points.

The identification comes from the sharp change in the CDF around a binding minimum lot size. A split near the threshold therefore improves the two-piece fit more than a split elsewhere. Panel (c) of Figure 1 shows the fit-improvement profile for the example district. The selected value, marked by the blue cross, coincides with the documented statutory minimum.

The local density jump estimator builds on Cui [2024]. For a candidate value L, let $\lambda = \log L$. I bin log lot sizes using width δ_L, placing a bin boundary at λ, and denote the resulting density by $\hat{f}_{j,L}$. Let $\lambda_L^- = \lambda - \delta_L/2$ and $\lambda_L^+ = \lambda + \delta_L/2$ denote the midpoints immediately below and above the candidate. I estimate a counterfactual density $\hat{f}^0_{j,L} \in \mathcal{P}_q$ using bins outside a neighborhood of the candidate. The estimator solves

$$\widehat{\underline{\ell}}_j^{\mathrm{N}} = \mathop{\arg\max}_{L \in \mathcal{G}_j :\, M_j(L) > 0,\, B_j(L) > 0} \big\{ M_j(L) + B_j(L) \big\}, \qquad \begin{aligned} M_j(L) &= \big(\hat{f}^0_{j,L}(\lambda_L^-) - \hat{f}_{j,L}(\lambda_L^-)\big)\, \delta_L, \\ B_j(L) &= \big(\hat{f}_{j,L}(\lambda_L^+) - \hat{f}^0_{j,L}(\lambda_L^+)\big)\, \delta_L, \end{aligned} \tag{2}$$

where $M_j(L)$ measures missing mass below the candidate and $B_j(L)$ measures excess mass above it. I require both terms to be positive so that a candidate must exhibit both features of the notch. The candidate grid $\mathcal{G}_j$ is constructed from the unit's observed lot-size distribution in the same manner as $\mathcal{C}_j$, except that it excludes percentile-based candidates. Because a statutory minimum should appear as a mass point in constructed lot sizes, the local density jump estimator searches only over observed atoms.

The local density jump estimator identifies the threshold from the local density around each candidate rather than from a break in the CDF. Panel (d) of Figure 1 shows the notch statistic for the same example district. The selected value again coincides with the documented statutory minimum.

I apply both estimators to every named zoning district and Census block group with at least 30 parcels in each construction-vintage window, with each district-by-window or block-group-by-window combination defining an estimation unit. The estimators depend on a small number of tuning parameters. I split the external benchmarks in Section 2.3 into calibration and validation samples at the municipality level. I use the calibration sample to select the structural break estimator's polynomial degree, the local density jump estimator's bin width and counterfactual polynomial degree, and the significance cutoff for detecting a minimum lot size. Estimation units are then classified as detected, not detected, or indeterminate when the data do not support a valid estimate.[6]

Several situations can generate multiple notches within a single estimation unit. A Census block group may span multiple zoning districts with different minimum lot sizes, creating multiple thresholds in space, as illustrated in Figure 2. A district may also be rezoned during the sample period, creating multiple thresholds over time. To accommodate these cases, I extend the local density jump estimator into a two-step menu estimator. The first step detects whether the estimation unit contains multiple significant breaks using a multiple-break extension of the structural break specification and sequential supF tests [Andrews, 1993, Bai and Perron, 1998]. The second step recovers the thresholds by fitting a finite mixture of notched densities, with each component assigned a candidate minimum lot size and mixing weight. I select the number of components K using the modified Bayesian information criterion of Liu et al. [1997], independently of the number of breaks selected by the structural break specification, with a maximum of nine components.

To preserve statistical power while allowing multiple thresholds, I prespecify $\mathcal{G}$ with 63 conventional statutory minimums. The grid includes square-foot values from 2,000 to 20,000, selected frontage-based values such as 4,800 and 9,600, common acre fractions beginning at one-quarter acre (10,890 square feet), an acre-based sequence up to forty acres, and round

[6] I release both tract-level and CBSA-level aggregates of all estimates and those classified as detected.

intermediate values such as 25,000 and 40,000 square feet.[7]

I estimate menus at two levels. At the zoning-authority-by-construction-vintage-window level, the estimator recovers the set of minimum lot sizes operating across districts and their associated weights. This allows areas spanning multiple districts to be represented by multiple thresholds rather than a single pooled estimate. I also estimate a menu for each Census block group using all construction vintages and all parcels, regardless of whether a zoning district name is observed.

3.2 Validation

I validate the estimates against four external benchmarks that differ in geographic coverage, spatial resolution, and availability of adoption dates. Table 2 reports the main validation results, while Appendix Table A2 reports the same statistics using all estimates regardless of the significance cutoff. These tables use district-level statutory minimum lot sizes from MAPC and CCLZD. Across all three district-level panels, the structural break estimator is the most accurate point estimator, with accuracy improving when construction vintages align with the regulatory regime. In the recent MAPC sample, where the constructed stock and current statute are most likely to reflect the same regime, its median absolute error is 0.01 log points, with 58% of estimates within 2% of the statutory value and 66% within 10%. In the pooled MAPC sample, the median error rises to 0.02 log points and 48% of estimates fall within 2%. In the dated CCLZD sample, the median error is 0.08 log points and 60% of estimates fall within 10%. The local density jump estimator is generally less accurate, although its median error is lower in the MAPC sample. The menu estimator contains the statutory value within 2% of a menu item in roughly 70% of MAPC districts and within 10% in up to 94%.

The two municipality-level benchmarks, the Bartik–Gupta–Milo and NZLUD measures, are evaluated separately in Appendix Table A3 because they summarize regulation at the municipality level and are themselves processed measures. The estimated menu contains the Bartik–Gupta–Milo lowest reported minimum within 10% of a menu item in 63% of validation municipalities, and the dominant estimate falls in or adjacent to the NZLUD-predicted bin in 59%. Agreement is weaker for the largest reported minimum. The menu's largest estimate misses the reported highest minimum by 0.9 log points at the median, consistent with the largest statutory minimums governing districts where residential construction is too sparse

[7] Frontage-based values arise from common combinations of minimum frontage and lot depth. For example, a 50-by-125-foot lot has an area of 6,250 square feet, a common lot size. Because ordinances often regulate frontage together with lot area or codify existing subdivision patterns, statutory minimums may be round in frontage-by-depth dimensions even when they are not round in square feet. The full set of conventional candidates is listed in Appendix Section B.2.

to reveal them.

Where a documented minimum exists, the structural break significance cutoff determines whether the estimation unit is classified as detected. The rule detects 59–61% of districts with positive documented minimum lot sizes, with detection concentrated in larger districts. Detected districts account for 84% of parcels in the pooled MAPC sample, 79% in the recent MAPC sample, and 78% in CCLZD. The rule trades coverage for precision. In the pooled MAPC sample, the share of estimates within 2% of the statutory value rises from 34% among all estimates to 48% among estimates that clear the significance cutoff, as shown in Appendix Table A2.

3.3 Detection Without a Statute

The previous section asked whether the estimators recover documented minimum lot sizes. A small set of benchmark districts has no documented statutory minimum, posing the reverse question of whether the estimators detect a structural break or local density jump where no minimum exists. I treat these cases as potential false-positive detections.

Districts documented as having no lot-size requirement can exhibit substantial structure in their lot-size distributions. Panel (a) of Figure 3 shows a representative case, a documented no-minimum district whose lot sizes bunch at a conventional value much like those of a regulated district. These districts are relatively small, with median sample sizes roughly half those of documented-positive districts, ranging from 65 to 333 parcels compared with 159 to 520 across the three panels. The significance cutoff is therefore harder to clear, but 43% of the MAPC and 72% of the Cook County no-minimum districts still clear it.

These false-positive detections highlight an important caution in interpreting the minimum lot size estimates. Two patterns are worth noting. First, the false-positive concern is concentrated in earlier periods. The CCLZD data in Panel (b) of Figure 3 show that no-minimum districts become increasingly rare over time. Among documented residential districts, the no-minimum share falls from 15.8% in the 1920s and 15.7% in the 1940s to 12.9% in the 1950s, 3.2% by 1960, and 0.9% by 2010. For modern construction-vintage windows, a detected threshold is therefore much more likely to correspond to a statutory minimum, while false-positive risk is concentrated in prewar windows and the relatively small set of unzoned areas.

Second, a nonstatutory lot-size convention can still impose a persistent constraint on development. Across fully disjoint construction-vintage windows, 38–39% of detected estimation units retain the same estimated value, indicating that these thresholds can persist for decades regardless of whether they originate in a zoning ordinance. This distinction also matters less where bunching is strongest, since conventional lot sizes can later be codified into formal regulation [Gallagher et al., 2026]. More generally, the platted parcel fabric itself can constrain subsequent development because building below an established lot size may require resubdivision and may also be limited by subdivision regulations or private deed restrictions. These mechanisms can generate persistent lot-size constraints even in places without conventional zoning, including Houston [Gray and Millsap, 2023].

3.4 Aggregation at the Tract and CBSA Level

I aggregate the estimates to Census tracts and CBSAs as follows. Within each tract and construction window, estimates are weighted by the number of contributing parcels. I report the weighted median, geometric mean, quartiles, and shares of parcels with estimated minimum lot sizes of at least one-half acre and one acre. The main measures used in Section 4 combine estimates that clear the significance cutoff, while the public panel also reports the same statistics using all estimates. The CBSA panel aggregates the tract measures using parcel weights. Geometric means and acreage shares aggregate directly, while the CBSA median is the weighted median of tract medians. The panel also reports the number of contributing tracts and flags CBSAs with fewer than five.

The value entering these aggregations is defined in Section 3.1. When the two single-threshold estimators agree, I use their common value. When they disagree, I use the structural break estimate, except for block groups whose pooled menu identifies more than one minimum lot size. In those cases, I take the weighted median of the structural break and local density jump estimates together with the menu candidates. Appendix Section C provides additional details on the released files.

4 Descriptive Analysis

This section describes the minimum lot size estimates. I first document the cross-sectional distribution using the most recent estimate for each district, block group, and tract, then patterns across construction vintages, and finally how the estimates relate to the 2018 Wharton survey and the use mix of the standing housing stock.

4.1 Distribution and Trends in Minimum Lot Sizes

Table 3 reports the distribution of estimated minimum lot sizes across zoning districts and block groups, using each unit's most recent estimate. The median zoning district, when observed in the assessor record, has an estimated minimum lot size of about 10,500 square feet, while the median block group is about 9,000 square feet. Both distributions are heavily right-skewed, with means roughly three times their medians and standard deviations near two acres. For zoning districts, the 75th percentile is exactly one-half acre and the 90th percentile exceeds one acre, while the block-group 90th percentile is exactly one acre.

Figure 4 turns to tract-level aggregates for the 2016–2025 window, weighting each tract's median estimate by its parcel count. The parcel-weighted median tract is about 7,800 square feet, the geometric mean is 9,400 square feet, and 8.5% of parcels are in tracts with an estimated minimum lot size of at least one acre. Regional differences are substantial. The tract median is about 10,000 square feet in the Midwest and 10,500 in the Northeast, compared with 7,800 in the South and 7,000 in the West. The one-acre share is 14.0% in the Northeast, compared with 6.4% in the Midwest, 8.8% in the South, and 7.5% in the West. The property-class panels show that detached single-family estimates are close to the pooled estimates, with a median tract of 7,900 square feet and a one-acre share of 9.1%. The attached single-family panel centers on 2,100 square feet and the small-multifamily panel on 5,700 square feet, with essentially no one-acre mass in either.

Figure 5 plots the estimates by construction-vintage window, with each window shown at its midpoint year. Panel (a) shows the pooled geometric mean rising from about 6,000 to 9,400 square feet. The detached single-family series closely tracks the pooled series, while the attached single-family and small-multifamily series remain well below it. Panel (b) shows that the share of parcels with an estimated minimum of at least one acre rises from 2.3% in the 1916–1940 window to 8.5% in the 2016–2025 window, a factor of 3.6, with a peak of 17.0% in the 1991–2000 window. The dashed line repeats the calculation for the fixed set of 7,838 tracts with estimates in both the 1941–1950 and 2016–2025 windows. The two series move closely together, indicating that the increase is not simply driven by construction shifting toward large-lot places. Panels (c) and (d) report the regional series, with the Northeast showing the sharpest increase in both mean minimum lot size and the share above one acre.

As discussed in Section 2.2 and Section 3.3, the temporal patterns reflect construction cohorts rather than precise adoption dates. Lot boundaries may predate construction, and regulations may later codify lot sizes already common in existing subdivisions. These features can make regulatory changes appear later or produce bunching before formal adoption. The series should therefore be read as changes in lot-size patterns across construction cohorts, not as a dated history of statutory minimums.

4.2 Comparison with Other Measures of Land-Use Regulation

Table 4 compares the estimates with other measures of land-use regulation and with the use composition of the standing stock. Panel A compares the CBSA-level aggregates with the 2018 Wharton survey [Gyourko et al., 2021]. The survey's density-restriction subindex, DRI, is based on its question about the largest minimum lot size in the community, reported in four categories from less than one-half acre to more than two acres, and is therefore the closest comparison. The overall index, WRLURI, additionally incorporates other dimensions of regulation, including political and court involvement, the number of required approvals, restrictions on permits and units, open-space and exaction requirements, and approval delay.

The panel's 2016–2025 mean has a Spearman rank correlation of 0.10 with DRI across all 382 matched CBSAs and 0.29 across the 104 where the survey covers at least five municipalities. The all-vintages one-acre share shows a similar pattern, with correlations of 0.09 across 470 CBSAs and 0.25 across 111. The correlation strengthens when the survey measure is based on more responding municipalities, consistent with greater sampling error in thinly covered metros. By contrast, the mean is negatively correlated with the overall WRLURI, at -0.22 and -0.33 in the corresponding samples. This sign reversal within the same survey indicates that estimated minimum lot size captures a specific regulatory margin rather than general regulatory restrictiveness. The 2016–2025 one-acre share is essentially uncorrelated with both indices.

Panels B and C compare the minimum lot size estimates with the tract-level use mix of the standing stock, measured from the full assessor snapshot. Parcels are classified as single-family, multifamily, nonresidential, or other, with the last category dominated by vacant and agricultural land. These are observed land-use outcomes that proxy for use permissions rather than direct measures of use restrictions. Panel B reports rank correlations between the tract-level geometric mean for 2016–2025 and the one-acre share across all construction vintages and five use measures. Correlations are reported nationally, within Census regions, and within CBSAs. For the within-group measures, both variables are ranked within each region or CBSA and rescaled to $[0, 1]$ before pooling.

Estimated minimum lot size is essentially uncorrelated with the single-family share of all parcels, with within-CBSA correlations of -0.04 for the one-acre share and 0.01 for the mean. Restricting the denominator to residential parcels, the single-family share correlates at 0.33 with the one-acre share within CBSAs, closely matching the -0.34 correlation with the multifamily share. The multifamily correlation is also -0.38 nationally and -0.37 within regions. By contrast, the indicator for tracts with no multifamily parcels correlates at only 0.10.

Panel C reports the same minimum lot size and use measures in levels by metropolitan status. Across metropolitan, micropolitan, and non-CBSA tracts, the mean minimum lot size rises from about 10,000 square feet to 21,800 to 39,100, and the one-acre share rises from 7.4% to 23% to 39%. Over the same categories, the multifamily share falls from 10% to 3% to 1.6%, while vacant and agricultural land rises from 16% to 33% to 46%. The levels show that larger minimum lot sizes are concentrated in areas with less multifamily housing and more undeveloped land. These differences occur largely across metropolitan status categories, while the correlations within CBSAs in Panel B characterize variation within housing markets.

Together, these comparisons suggest that minimum lot size aligns with density restrictions but does not necessarily track other dimensions of regulation. It is higher where multifamily housing is less prevalent and more land remains undeveloped, negatively related to composite regulatory stringency, and only moderately related to the use margin within metropolitan areas.

5 Conclusion

This paper constructs a national panel of minimum lot sizes from the distribution of lot sizes in the built housing stock. Using tax assessor records covering 92 million single-family and small multifamily parcels, I estimate minimum lot sizes within zoning districts and Census block groups across construction-vintage windows from 1916 to 2025. The approach combines structural break and local density jump estimators with a menu estimator that recovers multiple minimum lot sizes coexisting within the same lot size distribution. I aggregate the estimates into a public panel covering 76,770 Census tracts, with corresponding CBSA aggregates and a national zoning-authority geography.

Validation against documented zoning regulations shows that the constructed housing stock contains substantial information about minimum lot size requirements. In recent MAPC districts, the structural break estimator has a median absolute error of 0.01 log points and 66% of estimates fall within 10% of the statutory minimum. The corresponding median error is 0.08 log points in the historical Cook County data, while the menu estimator contains the statutory value within 10% of a menu estimate for up to 94% of MAPC districts. The validation also establishes an important limitation. Districts without a documented statutory minimum can exhibit similar bunching because subdivision platting, developer conventions, and other constraints can generate persistent lot-size thresholds. The estimates therefore measure minimum lot size patterns visible in the built stock rather than provide definitive evidence that a particular ordinance produced them.

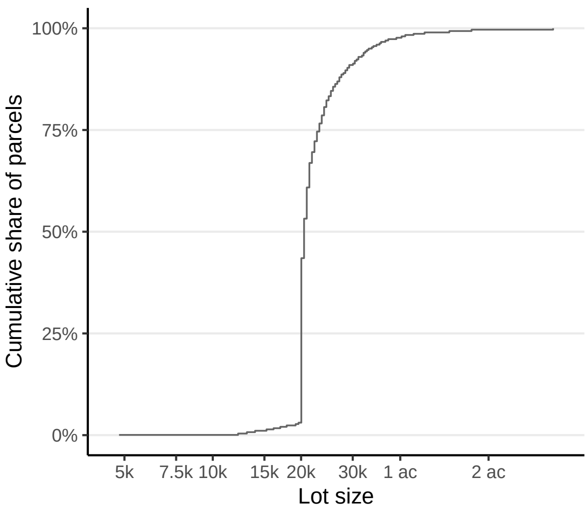
The estimates reveal substantial variation across both space and construction cohorts. In the 2016–2025 window, the parcel-weighted median tract has an estimated minimum lot size of about 7,800 square feet, and 8.5% of parcels are in tracts with an estimated minimum of at least one acre. The one-acre share rises from 2.3% in the 1916–1940 window to 8.5% in 2016–2025 and peaks at 17.0% in 1991–2000. Minimum lot size also captures a distinct dimension of zoning. It is positively related to existing measures of density restriction and to lower multifamily prevalence, but negatively related to broader regulatory stringency measured in survey data. Other supply restrictions may therefore be tighter in places with smaller minimum lot sizes.

The public panel provides a new measure of a specific zoning margin at a spatial and temporal resolution that has been difficult to obtain from surveys or zoning codes alone. It can be used alongside existing zoning measures to study housing supply and other outcomes for which variation in allowable residential density is economically important.

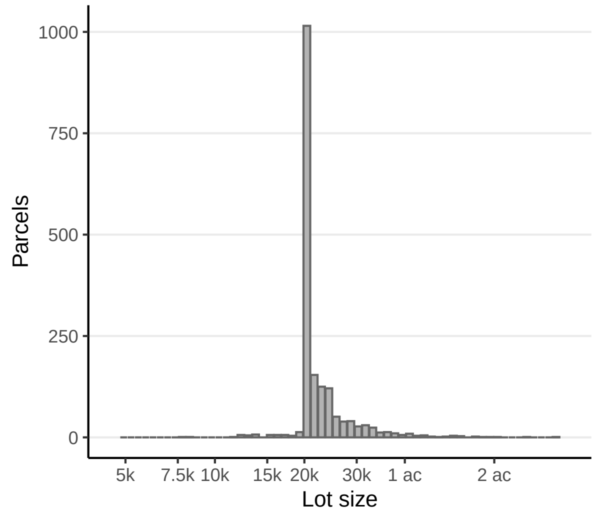
Table 1. Sample Statistics

	Mean	SD	Q1	Median	Q3
Panel A. Base sample					
Lot size (1,000 sq ft)	79.4	313.5	6.8	10.6	28.2
Year built	1977	29	1956	1979	2000
Building size (1,000 sq ft)	1.84	0.80	1.26	1.66	2.24
# observations			109,554,140		
# parcels			91,998,687		
# block groups			230,539		
# municipalities			29,677		
# counties			3,108		
Panel B. Estimation units					
Zoning districts					
Parcels per cell	317.7	1,174.5	49.0	90.0	223.0
# estimation units			229,791		
Share of estimation units retained			19.4%		
Share of parcels in retained cells			94.0%		
# districts			28,260		
# authorities represented			7,148		
Block groups					
Parcels per cell	107.5	117.4	45.0	70.0	126.0
Share of parcels with named code	0.39	0.42	0.00	0.07	0.88
# named codes spanned per cell	1.6	3.1	0.0	1.0	3.0
# named codes per block group (pooled)	3.0	11.6	0.0	2.0	4.0
# estimation units			1,554,292		
Share of units with ≥ 30 parcels			41.1%		
Share of parcels in such units			88.6%		
# block groups			199,576		
# authorities represented			26,606		

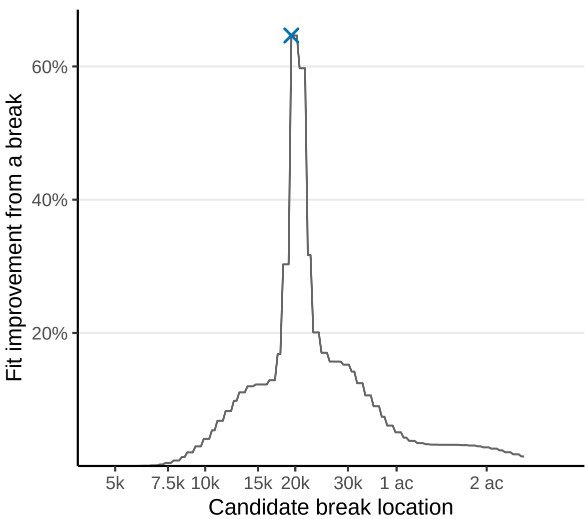
Notes: Panel A reports the base parcel sample after validity filters. Year built statistics exclude missing values and the 1900 placeholder, while building size statistics exclude missing, zero, and outlier values. Municipality counts refer to municipal zoning authorities, with remaining parcels assigned to county residual authorities. Panel B reports statistics for estimation units, defined as zoning districts observed in assessor records or, where unavailable, Census block groups, interacted with construction-year windows.



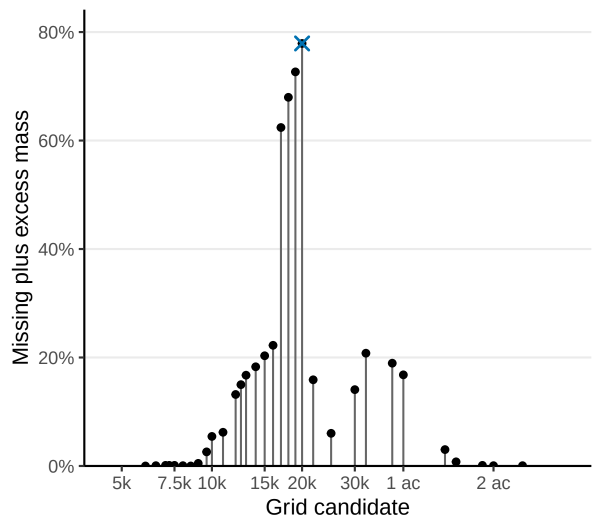
(a) Empirical CDF



(b) Histogram



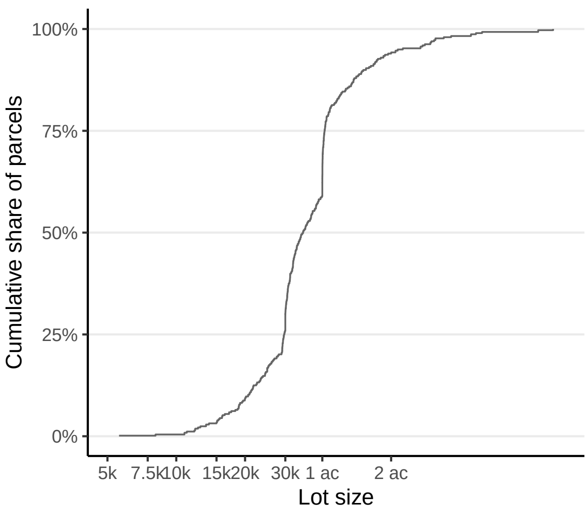
(c) Structural break



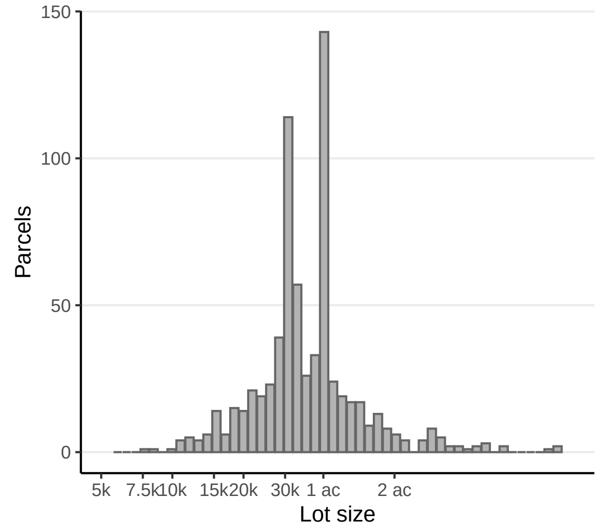
(d) Grid search

Figure 1. Illustration of the Minimum Lot Size Estimators

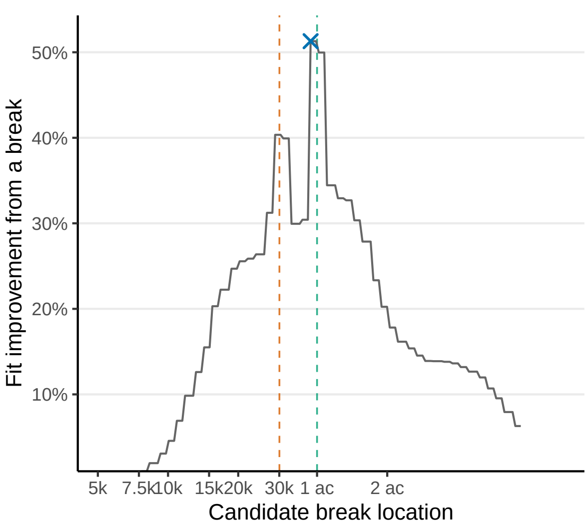
Notes. The figure illustrates a district with a documented statutory minimum lot size of 20,000 square feet. Panels (a) and (b) show the lot size histogram and empirical CDF, scaled in log. Panels (c) and (d) show the structural break estimator's bounded fit-improvement profile and the local density jump estimator's notch statistic. Blue crosses mark the estimated thresholds in Panels (c) and (d).



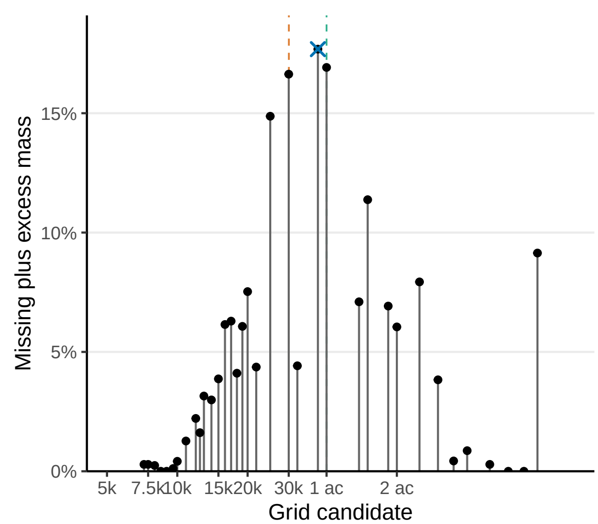
(a) Empirical CDF



(b) Histogram



(c) Structural break



(d) Grid search

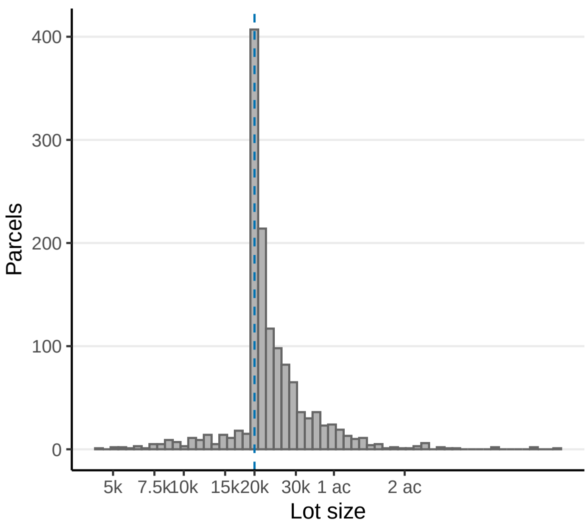
Figure 2. Illustration of Multiple Minimum Lot Size Case

Notes. This figure shows a Census block group that spans two zoning districts with different statutory minimum lot sizes. Panels (a) and (b) show the empirical CDF and histogram of lot sizes on a logarithmic scale. Panel (c) shows the structural break estimator's fit-improvement profile, and Panel (d) shows the local density jump estimator's notch statistic. Blue crosses mark the estimated thresholds, and dashed lines mark the two statutory minimum lot sizes.

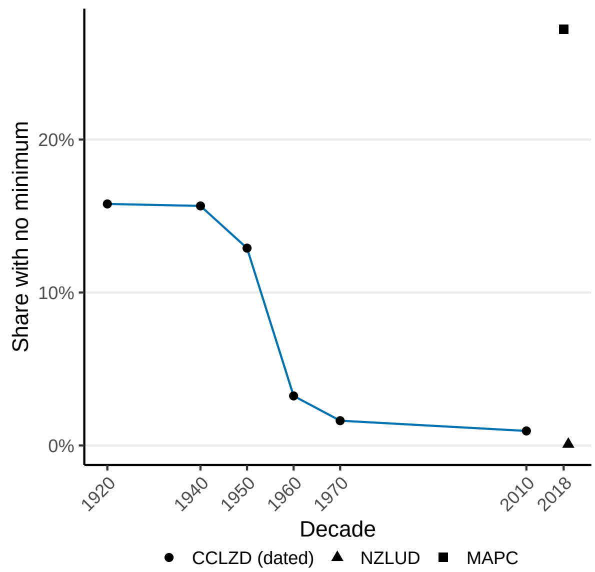
Table 2. Validation Against Statutory Minimum Lot Sizes

	Within 2%	Within 5%	Within 10%	Median |log error|
Panel A. MAPC districts, pooled vintages ($N_+ = 142$)				
Structural break	0.476	0.537	0.585	0.024
Local density jump	0.293	0.317	0.329	0.228
Menu	0.683	0.780	0.854	0.000
Panel B. MAPC districts, built $\geq$ 1980 ($N_+ = 92$)				
Structural break	0.580	0.640	0.660	0.010
Local density jump	0.460	0.480	0.480	0.002
Menu	0.720	0.780	0.920	0.000
Panel C. Cook County district-decades, dated ($N_+ = 156$)				
Structural break	0.333	0.419	0.602	0.076
Local density jump	0.108	0.161	0.237	0.228
Menu	0.484	0.667	0.742	0.024

Notes. This table reports validation statistics on benchmark estimation units in municipalities never used in calibration. N_+ counts benchmark units with a positive documented minimum lot size. The first three columns report the share of estimates within the stated tolerance of the statutory value, and the final column the median $|\log(\text{estimate}/\text{statute})|$. For the menu estimator, each documented statutory value is compared with the closest estimated menu item. All columns condition on retained estimates. Appendix Table A2 reports all units without the retention restriction.



(a) Example of False-Positive Detection



(b) Share of No-Minimum Districts

Figure 3. Detection Without a Statutory Minimum Lot Size

Notes. This figure illustrates detected lot-size structure in the absence of a documented statutory minimum. Panel (a) shows the lot-size distribution for a MAPC district with no documented minimum lot size but a detected threshold at a conventional value. Panel (b) reports the share of documented residential districts with no minimum lot size over time in the CCLZD, together with the corresponding cross-sectional shares in MAPC and NZLUD.

Table 3. Cross-Sectional Distribution of Estimated Minimum Lot Sizes

	p10	p25	p50	p75	p90	Mean	SD	N
Zoning districts	5,020	7,181	10,454	21,780	54,886	32,076	101,000	17,895
Block groups	4,606	6,372	8,999	15,682	43,560	24,707	86,647	151,318

Notes. This table reports the distribution of minimum lot size estimates in square feet using the most recent one in each geographic unit. N is the number of units with minimum lot size estimates meeting the significance cutoff, representing 65% of the 27,575 estimated named zoning districts and 77% of the 196,329 estimated block groups. Block group distribution includes every block group, not only those routed as fallbacks in the tract panel without named zoning districts.

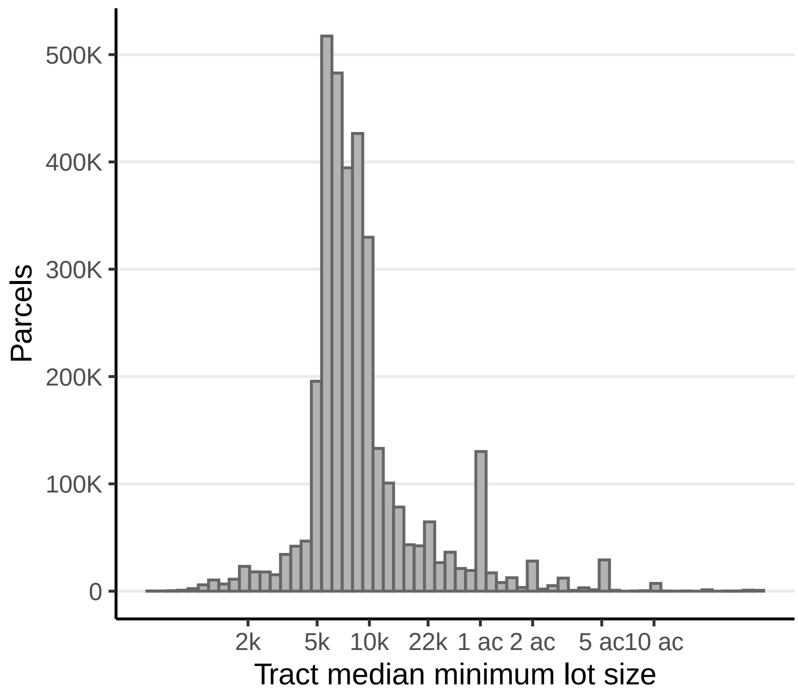
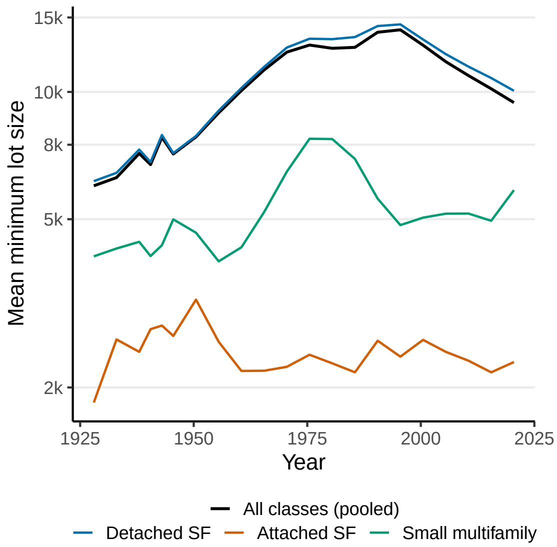
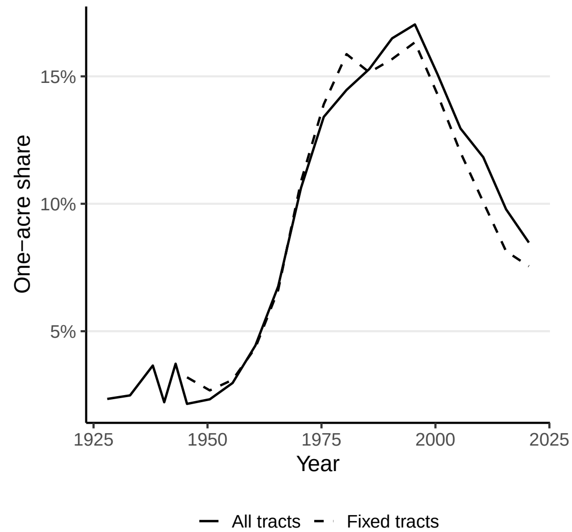


Figure 4. Distribution of Recent Minimum Lot Size Estimates

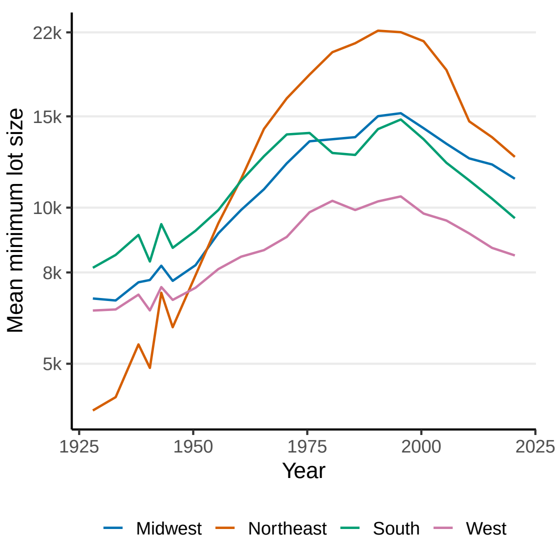
Notes. This figure shows the parcel-weighted distribution of tract-level minimum lot size estimates for the 2016–2025 construction-vintage window. Each tract is assigned the parcel-weighted median estimate across its constituent estimation units, and observations are weighted by the number of parcels in those units. The vertical axis therefore reflects the parcel distribution rather than the tract distribution. The horizontal axis is logarithmic, with tick labels in thousands of square feet through one-half acre (21,780 square feet) and in acres thereafter. The sample includes 22,160 Census tracts in the contiguous United States and the District of Columbia.



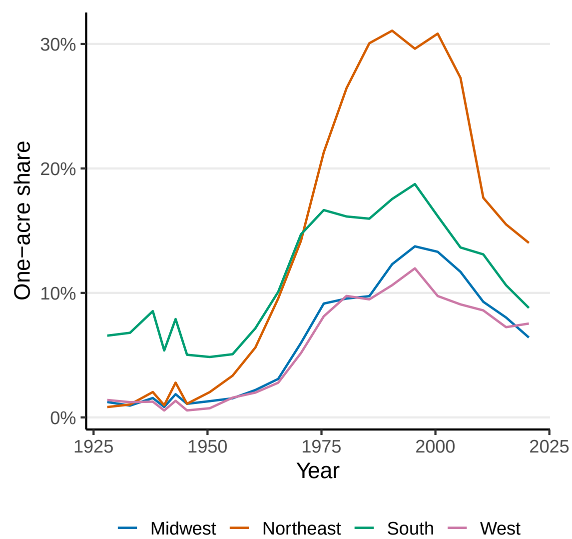
(a) Mean minimum lot size by property class



(b) One-acre share, national



(c) Mean minimum lot size by region



(d) One-acre share by region

Figure 5. Minimum Lot Size Estimates by Construction Vintage

Notes. This figure plots parcel-weighted statistics of estimated minimum lot sizes by construction-vintage window, each window at its midpoint year. Panels (a) and (c) plot the geometric mean on a logarithmic scale. Panel (a) pools all property classes in black, with class-specific series estimated on class-restricted parcels. Panels (b) and (d) report the one-acre share, the share of parcels in estimation units with an estimated minimum of at least 43,560 square feet. In panel (b), the dashed line restricts to the 7,838 tracts with estimates in both the 1941–1950 and 2016–2025 windows. Regions follow Census definitions.

Table 4. Comparison with Other Measures of Regulation and the Use Mix

Panel A. Rank correlation with the 2018 Wharton survey, across CBSAs

	All matched CBSAs		≥ 5 surveyed munis	
	DRI	WRLURI	DRI	WRLURI
Mean, 2016–2025	0.10	−0.22	0.29	−0.33
One-acre share, 2016–2025	0.11	0.01	0.07	0.09
One-acre share, all vintages	0.09	−0.10	0.25	−0.08
CBSAs	382	373	104	104

Panel B. Rank correlation with the use mix, across tracts

	National	Within region	Within CBSA
Mean, 2016–2025			
Single-family share, all parcels	−0.09	−0.09	0.01
Single-family share, residential parcels	0.31	0.30	0.29
No multifamily parcel	0.10	0.11	0.11
Multifamily share	−0.32	−0.32	−0.30
Non-residential share	−0.14	−0.16	−0.20
One-acre share, all vintages			
Single-family share, all parcels	−0.09	−0.11	−0.04
Single-family share, residential parcels	0.37	0.35	0.33
No multifamily parcel	0.10	0.09	0.10
Multifamily share	−0.38	−0.37	−0.34
Non-residential share	−0.20	−0.19	−0.20

Panel C. Levels of minimum lot size and the use mix, by metropolitan status

	Metropolitan	Micropolitan	Non-CBSA
Tracts	63,097	6,001	3,935
Mean, 2016–2025 (sq ft)	10,033	21,796	39,119
One-acre share, all vintages	0.074	0.234	0.387
Single-family share, all parcels	0.706	0.590	0.488
Single-family share, residential parcels	0.883	0.956	0.975
Share of tracts with no multifamily parcel	0.078	0.071	0.115
Multifamily share	0.099	0.031	0.016
Non-residential share	0.038	0.044	0.037

Notes. Panel A reports Spearman rank correlations across CBSAs between the panel's parcel-weighted measures and municipality-level survey means, weighted by the survey's CBSA weights. DRI is the density-restriction subindex constructed from the survey's minimum-lot-size questions; WRLURI is the overall regulatory index. The all-vintages row pools all construction-vintage windows and includes 470 CBSAs for DRI and 457 for WRLURI, of which 111 have at least five surveyed municipalities. Panels B and C use the use mix of the standing stock, computed from parcel counts in the full assessor snapshot of 158.8 million parcels; parcels with use codes outside the bucket map account for 0.4% of the sample and are assigned to other. The mean is geometric. Panel B correlations use 22,157 tracts for the 2016–2025 measure and 73,033 for the all-vintages measure; within-region and within-CBSA correlations rank both variables within group, normalized to $[0, 1]$, before pooling, with the within-CBSA sample restricted to 176 and 209 CBSAs with at least 50 panel tracts, respectively. Panel C reports parcel-weighted levels by metropolitan status using the Office of Management and Budget's March 2020 CBSA delineations; counties outside any CBSA are classified as non-CBSA.

References

Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

Alexander Bartik, Arpit Gupta, and Daniel Milo. The costs of housing regulation: Evidence from generative regulatory measurement. *Available at SSRN 4627587*, 2025.

Marinho Bertanha, Andrew H McCallum, and Nathan Seegert. Better bunching, nicer notching. *Journal of Econometrics*, 237(2):105512, 2023.

Raj Chetty, John N Friedman, Tore Olsen, and Luigi Pistaferri. Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The quarterly journal of economics*, 126(2):749–804, 2011.

Tianfang Cui. Did race fence off the american city? the great migration and the evolution of exclusionary zoning. *Tom Cui*, 2024.

Ryan Gallagher, Allison Shertzer, and Tate Twinam. Zoning and the american suburb. 2026.

Edward L Glaeser and Bryce A Ward. The causes and consequences of land use regulation: Evidence from greater boston. *Journal of urban Economics*, 65(3):265–278, 2009.

Madelyn Glickfeld and Ned Levine. *Regional growth–local reaction: The enactment and effects of local growth control and management measures in California*. Lincoln Institute of Land Policy, 1992.

M Nolan Gray and Adam A Millsap. Subdividing the unzoned city: an analysis of the causes and effects of houston's 1998 subdivision reform. *Journal of Planning Education and Research*, 43(4):990–1006, 2023.

Joseph Gyourko and Raven Molloy. Regulation and housing supply. In *Handbook of regional and urban economics*, volume 5, pages 1289–1337. Elsevier, 2015.

Joseph Gyourko, Albert Saiz, and Anita Summers. A new measure of the local regulatory environment for housing markets: The wharton residential land use regulatory index. *Urban studies*, 45(3):693–729, 2008.

Joseph Gyourko, Jonathan S Hartley, and Jacob Krimmel. The local residential land use regulatory environment across us housing markets: Evidence from a new wharton index. *Journal of Urban Economics*, 124:103337, 2021.

Keith R Ihlanfeldt. The effect of land use regulation on housing and land prices. *Journal of urban economics*, 61(3):420–435, 2007.

Kristoffer Kip Jackson. Regulation, land constraints, and california's boom and bust. *Regional science and urban economics*, 68:130–147, 2018.

Henrik Jacobsen Kleven. Bunching. *Annual Review of Economics*, 8(1):435–464, 2016.

Jian Liu, Shiying Wu, and James V Zidek. On segmented multivariate regression. *Statistica Sinica*, pages 497–525, 1997.

James Macek. Housing regulation and neighborhood sorting across the united states. *Available at SSRN 7035859*, 2026.

Sarah Mawhorter and Carolina Reid. Terner california residential land use survey. *University of California, Berkeley*, 2018.

Matthew Mleczko and Matthew Desmond. Using natural language processing to construct a national zoning and land use database. *Urban Studies*, 60(13):2564–2584, 2023.

John M Quigley, Steven Raphael, and Larry A Rosenthal. Measuring land-use regulations and their effects in the housing market. 2008.

Allison Shertzer, Tate Twinam, and Randall P Walsh. Zoning and the economic geography of cities. *Journal of urban economics*, 105:20–39, 2018.

Jaehee Song. The effects of residential zoning in us housing markets. *Journal of Urban Economics*, 149:103784, 2025.

Wenfei Xu, Scott Markley, Sara C Bronin, and Diana Drogaris. A national zoning atlas to inform housing research, policy, and public participation. *Cityscape*, 25(3):55–72, 2023.

Jeffrey Zabel and Maurice Dalton. The impact of minimum lot size regulations on house prices in eastern massachusetts. *Regional science and urban economics*, 41(6):571–583, 2011.

Appendix for

"Measuring Minimum Lot Size Regulation from the Housing Stock"

A Supplementary Tables and Figures

Table A1. Sample Statistics by Property Class

	Mean	SD	Q1	Median	Q3
Panel A. Detached single-family					
Lot size (1,000 sq ft)	83.7	322.2	7.1	11.2	31.4
Year built	1977	28	1957	1979	2000
Building size (1,000 sq ft)	1.84	0.80	1.26	1.66	2.24
# observations			102,365,686		
# parcels			86,438,607		
# block groups			228,599		
# municipalities			29,676		
# counties			3,107		
# states			49		
Panel B. Attached single-family					
Lot size (1,000 sq ft)	7.1	66.6	1.4	2.0	3.2
Year built	1986	30	1975	1992	2007
Building size (1,000 sq ft)	1.51	0.43	1.20	1.44	1.75
# observations			3,716,479		
# parcels			2,965,135		
# block groups			39,366		
# municipalities			4,963		
# counties			1,473		
# states			44		
Panel C. Small multifamily					
Lot size (1,000 sq ft)	16.0	98.2	3.9	6.1	9.5
Year built	1953	33	1924	1951	1979
Building size (1,000 sq ft)	2.26	0.84	1.66	2.11	2.70
# observations			3,471,975		
# parcels			2,594,945		
# block groups			116,449		
# municipalities			13,507		
# counties			2,225		
# states			48		

Notes. This table reports sample statistics by property class. It uses the same definitions and filters as Panel A of Table 1, which pools the three property classes.

Table A2. Validation Against Documented District Minimums, All Estimates

	Within 2%	Within 5%	Within 10%	Median |log error|
Panel A. MAPC districts, pooled vintages ($N_+ = 142$)				
Structural break	0.338	0.387	0.430	0.182
Local density jump	0.218	0.232	0.282	0.308
Menu	0.669	0.739	0.824	0.000
Panel B. MAPC districts, built ≥ 1980 ($N_+ = 92$)				
Structural break	0.435	0.533	0.565	0.039
Local density jump	0.359	0.380	0.380	0.020
Menu	0.630	0.717	0.815	0.000
Panel C. Cook County district-decades, dated ($N_+ = 156$)				
Structural break	0.288	0.372	0.506	0.094
Local density jump	0.096	0.147	0.205	0.237
Menu	0.462	0.647	0.744	0.028

Notes. This table reports validation statistics using the same external benchmarks as Table 2, without applying significance cutoffs. N_+ counts benchmark units with a positive documented minimum lot size. The first three columns report the share of estimates within the stated tolerance of the statutory value, and the final column the median $|\log(\text{estimate}/\text{statute})|$. For the menu estimator, each documented statutory value is compared with the closest estimated menu item.

Table A3. Agreement with Municipality-Level Benchmarks

Menu statistic	Benchmark statistic	N	≤2%	≤5%	≤10%	Median |log err|
Panel A. Bartik–Gupta–Milo municipal minimum lot size statistics						
Closest menu estimate	Lowest reported	1,272	0.389	0.502	0.634	0.045
Largest menu estimate	Highest reported	1,263	0.064	0.071	0.108	0.916
Dominant estimate	Mean reported	641	0.034	0.080	0.131	0.429
Panel B. NZLUD predicted largest-minimum bin						
Menu statistic		N	Exact	Adjacent		
Largest menu estimate		686	0.351	0.656		
Dominant estimate		325	0.412	0.585		

Notes. Both benchmarks are municipality-level summaries rather than district statutes, and each is itself a measurement: the Bartik–Gupta–Milo values are language-model readings of ordinance text reporting the lowest, mean, and highest residential minimum in the municipality, and NZLUD is a model prediction of the bin of the municipality's largest minimum (<0.5, 0.5–1, 1–2, >2 acres). Panel A pairs each menu statistic with its analogous benchmark statistic; the tolerance columns report the share of municipalities within 2, 5, and 10% of the benchmark value, and reported values outside the estimator's support are excluded from the corresponding comparison (3, 44, and 50 of 5,401 matched municipalities for the lowest, mean, and highest). In Panel B, "Exact" and "Adjacent" are the shares in the predicted bin and within one bin of it. The extreme-to-extreme comparisons diverge by construction: the highest statutory minimum typically governs districts with little residential construction, so the construction-revealed menu rarely contains it, and both systems' extreme statistics carry their largest measurement error. The dominant estimate is the parcel-weighted modal structural break estimate across the authority's districts.

B Minimum Lot Size Estimation Details

B.1 The Municipality Crosswalk

The authority layer is constructed from the 2020 TIGER/Line place and county-subdivision files using a state-specific rule table that identifies which units hold zoning authority under each state's enabling statutes. Incorporated municipalities enter with their full territory in every state, using function-status codes to exclude inactive and nonfunctioning governments. County subdivisions, including towns and townships, enter only in states whose statutes grant them zoning authority, primarily in New England, the Mid-Atlantic, and the upper Midwest. In these states, incorporated places are removed from the county-subdivision polygons before the layer is assembled, so each location is assigned to only one authority. Census designated places are excluded because they are statistical rather than governmental units. For each state, the rule table records which geographic layer enters, the relevant class and function-status codes, and the statutory basis for the assignment. I release the rule table together with the authority layer.

Three construction steps require additional care. First, 126 municipalities span county lines and therefore appear as multiple TIGER fragments. I dissolve these fragments into single authorities keyed on the statewide county-subdivision identifier and release the fragment-to-authority crosswalk so that county-level joins remain reversible. Second, because the authority layer is constructed rather than directly observed, I validate state-level authority counts against Census counts of incorporated places and minor civil divisions. The benchmark crosswalks in Section 3.2 provide an additional check where independently compiled jurisdiction lists are available. Third, the Census Bureau replaced Connecticut's eight legacy counties with nine planning regions in 2022, changing county and tract identifiers statewide. The pipeline uses the pre-change geography throughout. The authority layer is based on 2020 TIGER files, parcel Census identifiers use the same vintage, and metropolitan areas follow the March 2020 OMB delineations. Connecticut therefore enters under its legacy county codes, and planning-region identifiers do not appear in the released data.

The resulting municipality layer contains 30 thousand municipal zoning authorities. Parcels outside all municipal authority polygons are assigned to the county as the residual zoning authority. This assignment matters primarily in states without township zoning, where unincorporated territory is regulated, if at all, at the county level. Parcels are assigned by point-in-polygon spatial joins using their coordinates, and each parcel carries a flag indicating whether it is assigned to a municipal authority or the county residual.

B.2 Specification Choices

Candidate values. The menu estimator begins with a grid $\mathcal{G}$ of 63 conventional minimum lot sizes. The first tier covers 2,000 to 10,000 square feet in 500-square-foot increments and adds 2,400, 3,200, 4,800, 7,200, and 9,600 square feet, which correspond to common combinations of lot width and depth. The second tier covers 12,000 to 20,000 square feet in 1,000-square-foot increments and adds one-quarter acre (10,890), 12,500, one-half acre (21,780), 25,000, 30,000, three-quarter acre (32,670), 40,000, one acre (43,560), 60,000, and 80,000 square feet. The acre-based tier is $43{,}560 \times \{1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 10, 12, 15, 20, 25, 30, 35, 40\}$. I add four values identified from the calibration sample, yielding a final menu of 62 candidate values. The complete candidate set and the source of each value are reported in the replication files.

Acreage rounding. In many counties, assessors record lot area in acres rounded to two decimal places and mechanically convert that value to square feet. Lot sizes in these counties therefore lie on a lattice of 0.01 acre, or 435.6 square feet. For example, an 8,000-square-foot minimum can appear in the assessor data as $\text{round}(8{,}000/435.6) \times 435.6 = 7{,}841$ square feet. I flag parcels whose reported square footage is derived from acreage or lies exactly on this lattice and calculate the corresponding share for each estimation unit.

When more than half of an estimation unit's parcels are acreage-derived, I make two adjustments. First, I impose a lower bound on the local density jump estimator's bin width. With the baseline width $\delta = 0.025$ log points, the adjusted width at candidate L is $\delta_{\text{eff}}(L) = \max\left\{\delta, \log\left(1 + \frac{435.6}{L}\right)\right\}$. This prevents mechanical alternation between empty and populated bins from driving the estimated counterfactual density. Second, I treat conventional values and their acreage-rounded counterparts as the same candidate. Estimated values may therefore appear at lattice points such as 7,841 square feet even when the corresponding statutory value is 8,000 square feet.

Construction window. The main specification uses 25-year construction-vintage windows for earlier cohorts, with start years from 1916 through 1931, and 10-year windows for later cohorts, with the two window families overlapping from 1936 through 1955. Narrower windows provide finer resolution but leave more estimation units below the 30-parcel minimum. Wider windows retain more estimation units but combine construction cohorts across longer periods.

Table A4 quantifies this tradeoff. For each schedule, the table reports the number of estimation units with at least 30 parcels and the share of parcel-window observations they contain. The hybrid schedule retains 16.0% more estimation units than the uniform 10-year schedule, with most of the gain occurring before 1960. In that period, the uniform 10-year schedule retains only 67.4% as many units as the hybrid schedule. The uniform 25-year schedule retains 49.5% more units than the hybrid schedule, but most of the additional units occur after 1960, when the hybrid already covers 90.8% of parcel-window observations. I therefore use the hybrid schedule to preserve 10-year resolution after 1960 while retaining substantially more observations in earlier construction cohorts.

Table A4. Estimation Sample under Alternative Window Schedules

	Hybrid		Uniform 25-year		Uniform 10-year	
	Units	Cov. (%)	Units	Cov. (%)	Units	Cov. (%)
Named zoning districts						
Prewar (<1941)	49,830	93.3	59,216	94.6	24,977	87.0
Midcentury (1941–1970)	60,759	94.2	105,803	97.1	60,759	94.2
Late (1971–2005)	88,489	95.0	138,416	97.2	88,489	95.0
Modern (2006+)	30,713	91.4	37,900	93.3	30,713	91.4
Census block groups						
Prewar (<1941)	374,338	88.3	462,492	91.9	153,235	67.9
Midcentury (1941–1970)	449,103	86.4	826,881	97.1	449,103	86.4
Late (1971–2005)	581,662	91.2	851,789	96.9	581,662	91.2
Modern (2006+)	149,189	85.1	184,591	89.7	149,189	85.1
All periods	1,784,083	90.2	2,667,088	95.9	1,538,127	88.6

Notes. This table reports estimation-unit counts and parcel-window coverage ("Cov.") under alternative construction-vintage window schedules. All schedules use a five-year step, begin in 1916, and are truncated at the extract year. An estimation unit is included when it contains at least 30 parcels. Because windows overlap, a parcel may contribute to multiple estimation units. Parcel-window coverage is the share of these observations contained in estimation units meeting the 30-parcel minimum. The 1900–1915 pre-zoning baseline is excluded. The sample covers the contiguous United States and the District of Columbia.

B.3 Sensitivity

Polynomial degree. The structural break estimator fits separate polynomials to the empirical CDF on either side of each candidate break. The main specification uses degree 7, selected using the calibration sample. I assess sensitivity to this choice by re-estimating the validation sample using degrees 3, 5, 7, and 9, holding all other settings fixed. The comparison includes MAPC districts using pooled construction vintages and post-1980 construction, and dated CCLZD district-decade observations.

Table A5 reports, for each degree, the share of documented-positive estimation units that clear the significance cutoff and the accuracy of the resulting estimates. In MAPC, lower polynomial degrees produce more detections but less accurate estimates. In the pooled sample, moving from degree 3 to degree 7 lowers the detection share from 72% to 61%, while the median absolute log error falls from 0.18 to 0.01 and the share within 10% of the statutory value rises from 42% to 60%. In the post-1980 sample, the detection share falls from 68% to 59%, while the share within 10% rises from 63% to 74%. Degree 9 produces little further change. The CCLZD results are less sensitive to polynomial degree. The detection share falls from 63% at degree 3 to 57% at degree 9, while the median absolute log error remains between 0.075 and 0.087 and the share within 25% remains between 73% and 74%. I therefore retain degree 7 as selected by the calibration procedure.

The final column reports the corresponding detection share among estimation units with no documented minimum lot size. Between degrees 3 and 7, this share falls from 65% to 50% in the pooled MAPC sample and from 47% to 29% in the post-1980 sample. It remains between 72% and 78% in CCLZD. These comparisons are based on only 34, 17, and 18 no-minimum estimation units, respectively. Moreover, as discussed in Section 3.3, documented no-minimum districts can exhibit substantial bunching even without a statutory minimum. I therefore report these detection rates as a sensitivity check rather than interpret them as conventional false-positive rates.

Prespecified candidates. The baseline local density jump estimator takes its candidates from the estimation unit's own modal lot sizes, while the menu estimator uses a prespecified set of candidates. As a sensitivity check, I re-estimate the local density jump model using the prespecified round-value grid. Table A6 compares the two versions across 1,794,177 estimation units. The prespecified version produces an estimate for 94% of units, compared with 76% for the baseline, because the grid supplies candidates even when the unit's modal set is sparse. Among the 1,300,279 units estimated by both versions, the two agree within the stated tolerance in 44% of cases and in 38–40% of units whose estimates clear the significance cutoff. The disagreements are generally small and symmetric. The median absolute log difference is 0.04–0.06 across all estimated units and 0.08 among units clearing the significance cutoff, while the grid estimate is lower in 41–43% of cases. The two versions of the local density jump estimator therefore recover broadly similar minimum lot sizes.

Table A5. Sensitivity to Structural Break Polynomial Degree

Degree	N_+	Detected (%)	Within tolerance ≤2%	≤10%	≤25%	Median |log err|	No-min. (%)
MAPC, pooled							
3	301	72.4	32.6	41.7	56.9	0.182	64.7
5	301	66.8	47.3	55.7	62.7	0.044	61.8
7	301	61.5	51.9	60.0	64.9	0.011	50.0
9	301	57.8	51.7	59.8	65.5	0.015	50.0
MAPC, built ≥ 1980							
3	215	67.9	48.6	63.0	74.7	0.022	47.1
5	215	57.2	62.6	71.5	75.6	0.010	29.4
7	215	58.6	63.5	73.8	77.8	0.009	29.4
9	215	58.1	68.0	76.0	80.0	0.005	35.3
CCLZD, dated							
3	394	63.5	30.4	54.4	72.8	0.080	72.2
5	394	61.7	30.9	52.7	73.3	0.087	72.2
7	394	61.2	29.9	55.2	74.3	0.077	72.2
9	394	57.4	31.4	55.8	73.9	0.075	77.8

Notes. This table reports sensitivity of the structural break estimator to polynomial degree using the validation sample. All estimation units satisfy the 30-parcel minimum. N_+ is the number of estimation units with a documented positive minimum lot size. "Detected" is the share whose estimates clear the significance cutoff. The three tolerance columns and the median absolute log error are calculated among detected estimation units. "No-min." is the corresponding detection share among estimation units with no documented minimum lot size, based on 34 MAPC pooled, 17 MAPC post-1980, and 18 CCLZD observations. Degree 7 is the main specification selected using the calibration sample.

Table A6. Sensitivity to Modal and Prespecified Candidates

Spatial unit	N	Share estimated: Modal	Share estimated: Prespec.	Both	Agree	Disagreement: Prespec. lower	Disagreement: Median $\lvert\Delta \log\rvert$
All estimated units							
District	231,435	0.78	0.94	0.75	0.43	0.41	0.063
Block group	1,562,742	0.76	0.94	0.72	0.44	0.41	0.044
Units clearing the significance cutoff							
District	85,783	0.86	0.93	0.82	0.40	0.43	0.083
Block group	549,711	0.84	0.92	0.78	0.38	0.42	0.083

Notes. This table compares two candidate sets for the local density jump estimator. "Modal" uses modal lot sizes from each estimation unit, as in the baseline specification; "Prespecified" uses the prespecified round-value candidates. "Share estimated" is the share of estimation units for which each version returns a valid estimate, and "Both" is the share estimated by both. "Agree" is the share of jointly estimated units whose estimates fall within the agreement tolerance used elsewhere in the paper. "Prespec. lower" is the share of disagreements for which the prespecified estimate is smaller, and the final column reports the median absolute log difference among units estimated by both versions. The lower panel restricts to estimation units whose estimates clear the significance cutoff. The value 0.083 in the lower panel corresponds to one step of the candidate grid.

C The Public Minimum Lot Size Panel

The estimates are released at the tract and metropolitan-area levels, with national summaries. This section describes what the release contains, how its columns relate to the exhibits in the paper, and what is withheld.

Data files. The primary data file is a tract panel with one row per 2020 Census tract and construction-vintage window, comprising 1,114,603 rows across 76,770 tracts in the contiguous United States and the District of Columbia. Windows follow the hybrid schedule, with a five-year step, 25-year widths for the four windows beginning from 1916 through 1931, and 10-year widths from 1936 onward. The two window families therefore overlap over 1936–1955, and a parcel can contribute to multiple rows. The 1900–1915 pre-zoning windows are included as a reference and are not interpreted as estimates of regulation. A CBSA panel aggregates the tract rows to 377 metropolitan and 501 micropolitan areas, yielding 18,913 rows, and a national file reports by window the number of estimation units and the shares with an estimate and clearing the significance cutoff. The pooled panel, which combines detached single-family, attached single-family, and two-to-four-unit parcels, underlies most exhibits in the paper unless mentioned otherwise. The same four files are released separately for each property class. A class appears in a tract and window only when it supports an estimation unit on its own, so these panels are sparser (Table A7).

The release also includes the zoning-authority geography described in Section B.1, with one GeoPackage per state containing incorporated-place and county-subdivision polygons. Counties serve as the residual authority for parcels outside both, and a crosswalk identifies the 261 jurisdictions assembled from multiple source polygons. These files contain no estimates. Each root panel is released in both Parquet and CSV formats with identical values, while the property-class and estimator-specific panels are released in Parquet format.

Two versions of every estimate. Each row reports each aggregate in two versions. The plain-named columns (`mls_sf_wmedian`, `mls_sf_geomean`, the quartiles, and the half-acre and one-acre shares) are computed over estimation units in the tract and construction-vintage window that clear the significance cutoff of the multiple-break test. This is the criterion used in Section 3, and Section 3.2 shows that estimates clearing the cutoff more closely reproduce documented ordinances. These columns underlie every result in the paper. The corresponding columns with an `_all` suffix use every estimation unit with a valid estimate, regardless of whether it clears the cutoff. Both versions are released so users can assess the effect of this restriction. The variable `share_we_flagged` reports the share of estimated weight in units clearing the cutoff. Nationally, 65% of estimated zoning-district units (17,895

Table A7. Files in the Public Release

File	Rows	Columns
Pooled classes		
`min_lot_size_tract_panel`	1,114,603	39
`min_lot_size_cbsa_panel`	18,913	42
`national_rates_by_window`	46	20
By class: tract panel; CBSA panel		
`by_class/detached_sf/`	1,084,035; 18,894	39; 42
`by_class/attached_sf/`	24,626; 1,525	39; 42
`by_class/small_mfam/`	106,604; 2,876	39; 42
By estimator: tract panel; CBSA panel		
`by_estimator/`	1,114,603; 18,913	44; 45
Zoning authorities		
49 state GeoPackages		
`dissolve_map`	261	7

Notes. This table reports the files included in the public release and their row and column counts. Parquet and CSV versions for the pooled classes contain identical rows and columns.

of 27,575) and 77% of block-group units (151,318 of 196,329) clear it when each unit is counted at its most recent nontruncated window. Column names retain the internal tokens `flagged` for units clearing the cutoff and `cell` for estimation units so that they match the replication code.

Coverage and diagnostics. Weights count parcel-window contributions assigned to each tract. Three weights and three unit counts report, respectively, all construction in the tract and window, the portion in estimation units with an estimate, and the portion in units clearing the significance cutoff. The identities $w_{\text{flagged}} \leq w_{\text{est}} \leq w_{\text{parcels}}$ and the corresponding count inequalities are verified before release. The ratio of the last weight to the first, `share_w_flagged`, measures the share of a tract's construction represented by the reported estimate, and the paper's exhibits condition on this measure. `share_w_named` reports the share of weight contributed by named zoning districts rather than block-group proxies.

Additional diagnostics report tract-weighted values of the estimation-unit measures used in Section 3.2, Section 3.3, and Section 3.4. These include the share of weight for which the menu enters the combined estimate, the share with a notch at the estimate, whether the estimate belongs to the recorded-value menu, the missing-to-excess mass ratio, the smoothness test, the share of district weight in planned-unit and flexible-standard districts, the code composition of block-group units, and the share of parcels with lot size recorded in acres. The by-estimator panels additionally report the agreement share and median absolute log difference between the two component estimators. The CBSA panel adds the number of contributing tracts and a `few_tracts` indicator for areas with fewer than five tracts carrying an estimate. Its unit counts are summed across tracts, so an estimation unit spanning multiple tracts is counted once in each tract, as indicated by the column names.

What is not included. No parcel records or fields from the assessor extract are released, under its license. No estimates are released for individual zoning districts or block groups. Districts are the level at which an estimate would read as a claim about a particular ordinance, which the attribution discussion does not support; block groups are small relative to the 30-parcel minimum, and a block-group value with its location would come close to identifying the parcels behind it. The tract is the smallest geography at which the estimates are both useful and safely aggregated. Alaska and Hawaii are excluded throughout.

Using the panel. Three applications cover most uses. For a cross section, choose one window, restrict to rows with `share_w_flagged` above a chosen coverage threshold, and join `tract_geoid` to Census, permit, or price data. For a series within a tract, follow `mls_sf_wmedian` across windows. Because adjacent 10-year windows share half their parcels under the five-year step, difference designs should use nonoverlapping windows. For a metropolitan series, use `mls_sf_geomean` from the CBSA panel by window and exclude rows marked `few_tracts`. Figure 5 presents the national version of this series. The release README provides these instructions and defines every column.