The Emergence of Exclusionary Zoning Across American Cities *

Tianfang Cui*

March 27, 2023

JOB MARKET PAPER

Latest version available here

Abstract

This paper identifies how Black migration into American cities caused their suburbs to adopt a widespread land use control — minimum lot sizes. I develop an algorithm detecting bunching on lot sizes, observable when governments' lot size controls bind developers from building denser housing. Applying the algorithm to national assessor records, I estimate for 7,000 local governments which lot size controls first came into effect and how they limited residential density over the last 80 years. Most suburbs adopted lot size controls from 1945–1970, the same period when four million Black Americans left the South for economic opportunity. I then use the "Second Great Migration" as a natural experiment that shifted central cities' racial composition toward Black Americans. From 1940–1970, the rise in central city Black composition in non-Southern central cities accelerated minimum lot size adoption while further explaining binding density controls applied to at least 830,000 housing units. Migration of lower-income whites into the same cities instead cause small and negative effects on suburban lot size outcomes. In states that passed early legislation to desegregate public schools, Black migration had the largest effects on lot size restrictiveness. Together, the results indicate that local governments designed land use controls to exclude Black migrants from neighborhoods and public goods.

*Wharton School, University of Pennsylvania. E-mail: `ttfcui@wharton.upenn.edu`. I thank Fernando Ferreira, Joseph Gyourko and Benjamin Keys for their extensive feedback and support to their advisee. I thank David Berger, Kirill Borusyak, Ellora Derenoncourt, Gilles Duranton, Nina Harari, Jessie Handbury, Samuel Hughes, Clemence Idoux, Gi Kim, Jacob Krimmel, Jeff Lin, Benjamin Lockwood, Philip Mulder, Elsie Peng, Diego Puga, Allison Shertzer, Todd Sinai, Eric Zwick and participants of the Wharton BEPP seminar, Wharton Urban seminar and the Urban Economics Association Meeting for their comments. I also thank Anne Leavitt-Gruberger for retrieving historical zoning records. Adam Sciara provided excellent research assistance.
I gratefully acknowledge financial support from the Wharton Zell-Lurie Real Estate Center.

People are still making zoning decisions without the slightest concern for the outside world...
They operate in the belief that their balkanized municipal universe is the happiest of worlds.
— Babcock (1966), *The Zoning Game*

1 Introduction

Local governments in the United States are empowered to shape urban form. Since the 1920s, U.S. local governments have had the right to regulate land use within their borders: the *zoning power*. In one application, localities dictate which neighborhoods do not allow dense housing development. Few legal avenues could overrule localities satisfying residents' demand for such *exclusionary zoning*. Past planning decisions could then have a legacy in today's residential segregation and unequal access to public goods.

Public finance offers a theory for why exclusionary zoning could enhance welfare. Any binding land use control sets an implicit price for access to local public goods, providing a fiscal motive for zoning. Local planning decisions summed together create menus of entry fees over different public goods, which could be an efficient mechanism for pricing out free riders (Hamilton (1975), Calabrese, Epple and Romano (2012), Fischel (2015)). Outside of economics, scholars argue instead governments first intended *racial exclusionary zoning*. Local actors assumed above all that postwar suburbanites placed less value on racially integrated neighborhoods. Limiting smaller homes was a means to preserve homogenous communities (Trounstine (2020), Winling and Michney (2021)).

Data availability complicates verifying either theory with empirics. Statistics are scarce on which governments adopted land use controls earlier or later. Unlike regulations in effect, historical records of local land use decisions are lost or too costly to retrieve.[1] This paper first fills the data gap by measuring the emergence and scope of a widely used regulation — the minimum lot size.

I construct the first national panel on the adoption and restrictiveness of minimum lot sizes, a key land use control for U.S. suburbs. Most Americans today live in suburbs,[2] where I quantify most governments first adopted minimum lot size controls from 1940 to 1970. The time frame of lot size adoption overlaps with one of the most important demographic shifts in U.S. history: the Second Great Migration of Black Americans out of the South (Collins (2021)). The arrival

[1] New surveys (Gyourko, Hartley and Krimmel (2021)) and text parsing of active ordinances (Shanks (2021), Bronin and Ilyankou (2021)) has substantially improved our understanding of today's local land use regulations.

[2] 54% of respondents in the 2020 wave of the American Housing Survey self-report living in suburban neighborhoods (Bucholtz, Molfino and Kolko (2020)).

of Black migrants to central cities beyond the South provide the largest natural experiment for racial exclusionary zoning: did suburban governments respond to postwar Black migration with restrictive lot size controls?

I estimate that from 1940–1970, local regulatory responses to Black migration caused at least 830,000 units outside the South to be built less densely than what market forces could have supplied. My empirical approach, which uses the lot size outcomes panel, addresses two further challenges. First, in addition to estimating when adoption happened, I also construct a measure of how restrictive lot size controls were over time. Second, lot size outcomes are less well explained by fiscal explanations to exclude only based on income.

Of all the tools to exclude dense housing units, the minimum lot size was the most popular among postwar suburbs. A neighborhood bound by the control requires each housing unit to occupy some minimum square footage of land. When development built after lot size adoption is seen as a whole, lot size controls have observable implications: development should bunch repeatedly on certain lot sizes. I introduce an algorithm that recovers when bunching on lot sizes first appears and persists over vintages of the housing stock. Applied to any American local government with lot size data, the algorithm outputs which lot size controls had been adopted and a measure of regulation adoption, based on the first homes showing bunching.

I apply the algorithm to administrative records across 86 million homes. Together, the data cover 407 U.S. metropolitan areas that contain 83% of America's 2020 population. I first improve the algorithm's accuracy by digitizing a national training sample of historical zoning records. After training the algorithm, I scale it to process lot size information for all U.S. cities and other governments with the zoning power.

Figure 1 shows one algorithm output: measures of lot size control adoption for 4,800 large and mid-sized U.S. zoning jurisdictions. For the first time in the literature. I calculate that from 1940–70, three-fifths of these cities adopted a minimum lot size exhibiting bunching behavior. The new zoning ordinances also included large lot size control exceeding 10,000 square feet, which features in three-fifths of jurisdictions today.

I also adapt methods from the bunching literature (Kleven (2016)) to introduce an excess mass measure of restrictiveness. The excess mass varies only based on how many developers on the margin upsize their lots to satisfy the minimum lot size. It excludes demand for the lot size in the absence of regulation. Excess mass is thus the appropriate measure of housing units that could have been built on smaller lots without binding controls. Summing up city-level calculations to the national level shows lot size controls constrained 7.5 million properties since the 1940s.

At the same time as suburbanization, Black migrants were moving into cities outside the South. More than four million Black migrants left the South during the Second Great Migration, resettling in Black neighborhoods near urban cores. I identify the Migration's effects on lot size control design in postwar suburbs. My empirical strategy uses the rising Black composition in central cities over time to explain the adoption of minimum lot sizes and those controls' restrictiveness.

Black migrants may have self-selected to cities with accelerated suburbanization patterns. This selection bias would confound migration's effects on lot size restrictiveness with non-racial factors. To address this concern, I construct shift-share instrumental variables for Black migration, following Boustan (2010) and Derenoncourt (2022). These instruments weigh the impact of migration rate shocks in Southern Black counties by cities' pre-1940 exposure to county migration flows. The IV strategy isolates variation in Black migration that does not correlate with how destination cities had grown.

For a panel of non-Southern metropolitan areas, I find a rise in the central city's share of Black residents caused early adoption and more restrictive design of lot size controls. Over 1940–70, a 10 percentage point rise in central city Black composition explains 18% of minimum lot size adoption across decades. Effects on lot size restrictiveness are greater, at about 37%. The estimate of 830,000 housing units constrained by racially motivated land use is the latter restrictiveness response converted into level terms. [3]

Postwar Black migrants differed with incumbents in both race and income. To verify that exclusion on race as opposed to income drive my main effects, I construct a separate migration instrument for shares of Southern white migrants in central cities. White migration out of the South was substantial, averaging 2.7 million migrants in each decade of my analysis period. As late as 1960, Southern white migrants earned less on average than non-Southern white incumbents. However, reduced form and instrumental variable regressions find null or negative effects of changes in Southern white composition on suburban lot size outcomes. The effect's direction is inconsistent with a predicted positive effect from localities practicing fiscal zoning.

Because poorer white migrants still earned more than Black migrants, a difference in lot size outcome responses by migrant race is not the perfect test for racial exclusionary zoning. However, I further explore a mechanism for the effects based on a theory of racial homophily affecting the valuation of local public goods. As the Great Migration accelerated, early civil rights activists proposed a legal agenda banning racial discrimination in American society. The

[3] It is beyond this paper's scope to estimate housing demand in the postwar decades over thousands of local governments. To interpret the estimate on constrained units, I later present how racial exclusionary zoning changed the *zoned capacity* of the built environment in the long run.

theory predicts that racial exclusionary zoning meets the demand of households who prefer to pay more for local public goods shared with racially homogeneous neighbors.

By 1950, certain states had adopted a model law that bans discrimination by race in public schools. I exploit these policy changes that desegregated access to a crucial local public good. Lot size controls for a neighborhood could circumvent *de jure* limits on school segregation, raising the neighborhood's amenity value for a type of household. Consistent with the theory, I find the highest causal effects of Black demographic change on lot size outcomes in states that adopted these laws early.

My research contributes to multiple literatures. First, I identify the significant contribution that racial exclusionary zoning played in shaping U.S. neighborhoods. Existing literature like Cutler, Glaeser and Vigdor (1999) document a rise and fall of Black-white residential segregation from 1940 to the present. Results in Cutler, Glaeser and Vigdor (1999) support a theory of *collective action* where the rise in segregation was due to white incumbents imposing barriers to Black homeownership. This explanation is usually contrasted with *white flight*, where white residents with distaste for living around Black families move out in unison after enough Black Americans move in Schelling (1971). [4]

My main results situate local government as a major contributor to collective action that excludes Black residents, as government actors responded to Black migration with exclusionary zoning. My empirical results suggest as local actors maintained exclusionary planning decisions, they also increased restrictiveness to capitalize on increased willingness to pay by households in white flight. This theory is consistent with results that lot size responses to Black migration were greater when state laws mandated desegregation. A first-order concern for racial integration policy, both in the past and today, could therefore be trading off letting some jurisdictions meet market demand for exclusions or restrict levers of local discretion in zoning. [5]

Existing research uses text data to define zoning outcomes or leverage property records to impute land use regulations applying in a residential zone (Nechamkin and MacDonald (2019), Song (2021)). These methods process present-day, not historical, zoning information, suitable to answer the consequences of present-day planning changes. To offer new facts on how land use controls were decided by U.S. governments and later perpetuated spatial inequality, I pro-

[4] Empirical evidence supports the explanation in recent times (Card, Mas and Rothstein (2008),Bayer et al. (2022)) as well as in the early 20th Century (Shertzer and Walsh (2019), Lee (2022)).

[5] Recent work by Sahn (2022) and Shi et al. (2022) also investigate the causal effects of Black migration on postwar urban planning, but on a sample of central cities rather than entire metropolitan areas. My work is complementary, showing the aggregate impact on urban form due to the Great Migration increases once I account for suburban zoning responses.

vide the first-ever panel of lot size controls for governments across the country.

The only paper I know of that estimates the adoption of lot sizes from deed records is Zabel and Dalton (2011), which covers one metropolitan area in the U.S. — Boston, MA — from a period starting in the late 1980s. My panel covers nearly every major U.S. metropolitan area in scope and has restrictiveness estimates spanning the past century. My restrictiveness measure is based on the degree of bunching on minimum lot sizes, which is observable in data and is well motivated by economic reasoning.

I show novel evidence related to the theoretical debate on how efficient zoning is in practice.[6] When data are only available on zoning decisions made decades after initial adoption, using regulatory outcomes that persist as an outcome cannot tease apart confounding explanations for why they had been adopted. One example of a confound is if after initial develpment, a suburb's zoning may have evolved to maintain the quality of local public goods, or capitalize households who prefer open space and natural amenities (Bogart (1993)). [7]

My empirical strategies use migration or policy shocks not dependent on the changing character of destination cities. Krimmel (2022) is the only other work I know taking this approach, where he estimates the adoption of residential land use controls before and after California equalized school funding in the 1970s. My lot size outcomes are also time-varying based on the housing vintage analyzed. Unlike existing work that use today's outcomes, my main empirical specification allows me to study how decade-long shocks caused decisions for homes built that same decade.

Finally, my results inform a growing literature on urban growth and spatial misallocation. Results in the literature indicate relaxing land use regulations in urban areas could cause considerable gains in land development (Turner, Haughwout and van der Klaauw (2014), Anagol, Ferreira and Rexer (2021)) and increase real per capita incomes (Hsieh and Moretti (2019), Duranton and Puga (2019)). What we observe, though, are local governments not pursuing these potential gains: zoning jurisdictions maintain or expand land use controls over time (Gyourko, Hartley and Krimmel (2021)).

I interpret my results as evidence for local governments using land use controls to not solely maximize land rents, but also to preserve endogenous amenities like racial composition. If urban migrants are more likely to be zoned out of housing due to race or demographic features, they redirect toward areas with lower price levels for housing. This could explain

[6]Mills, Epple and Vigdor (2006) and Zodrow (2007) assess the validity of theoretical assumptions while highlighting difficulties with testing assumptions directly.

[7] Dealing with how amenities today evolved jointly with present-day land use controls applies to both calibrated structural estimation (Calabrese, Epple and Romano (2007), Parkhomenko (2020)) as well as design-based approaches both past and present (Rolleston (1987), Hilber and Robert-Nicoud (2013)).

historic zoning decisions having persistent effects on regional economic divergence (Ganong and Shoag (2017)).

The paper proceeds as follows. Section 2 describes the estimation procedure, and the assumptions I make, to recover estimates of minimum lot size adoption. I define my principal regulation outcomes in Section 3 and present how they vary across time. Section 4 overviews which theoretical frameworks can or cannot rationalize a rise in regulation adoptions to specifically Black migration. Section 5 discusses my regression specification and identification strategy, with results and implications for land use restriction motives presented in Sections 6 and 7. Section 8 concludes.

2 Estimating Regulation Adoption

To jointly estimate both regulation levels and their respective adoption dates across zoning jurisdictions, I use the CoreLogic Tax Records, a national dataset collecting property-level information from county assessor offices. My algorithm has two steps, with the first step nested within a second model training step.

The first step conducts unsupervised learning over lot size distributions. Given parameters that characterize a kind of sizable bunching, the classifier defines a subset of *bunching bins* for each processed jurisdiction. To do this, I divide the distribution into conditional distributions based on *vintages* in year built, then partition the support into discrete *lot bins*. Calculating the magnitude of bunching at each lot bin, the algorithm isolates lot sizes on which many properties repeatedly bunch across time.

In the second step, I iterate over parameters that optimize over two model outputs. The first output is a mapping from time intervals in which lot size controls could have been first adopted into point estimates for lot size adoption. The second output pairs adoption dates with sets of bunching bins to interpret a detected bunching bin as a lot size control starting from a certain year.

My approach requires no assumptions on correctly identifying the boundaries of residential zones with minimum lot sizes. Instead, to identify lot size controls I assume they induce qualitatively distinct bunching behaviour that cannot be explained by buyer or development behaviour in unregulated housing markets. In Section 2.3, I microfound when bunching on lot sizes should appear and discuss which particular lot size controls are identified with my approach.

2.1 Data

From 2009 to 2019, CoreLogic has compiled records from tax assessor offices across the United States in the Tax Records file. Each observation in the file is a single property identified by its address and includes details on the parcel and the structure. In my analysis, I process 86 million properties classified as single-family homes or as duplexes in 426 Core-Based Statistical Areas as defined by the U.S. government. The data span 1151 counties and the aggregate count of structures is almost identical to the Census Bureau's 2019 estimates of 1-2 unit structures in those counties. I outline the data cleaning process below, while I provide full details in Appendix Section A.3.

I first match each property to a present-day zoning jurisdiction. A *zoning jurisdiction* is defined as an incorporated place as of 2010, plus the counties, townships or other local governments that have zoning powers within each state. For each incorporated place, I also separate out the area that was incorporated in 1980 and newly annexed areas from 1980 to 2010. In sum, I define 10364 zoning jurisdictions forming the observations in my analysis.[8]

With each property matched, I filter to the properties to the approximately five-sixths that have both the lot square footage and the year built of the property. The latter variable is my proxy for the year that the lot itself was platted, which is necessary information to estimate any regulation adoption.

Because of missing data and because the year built data may record only the most recent property on the lot as of the 21st Century, my sample has fewer properties recorded to be built during a past period than hypothetical records gathered at that period. To check attrition rates, I benchmark the CoreLogic records to historical Census estimates of properties built by the end of each decade. I find enough housing from the past endured to the present to conduct analysis. My sample contains 55% of all single-family homes that were built as of 1940 and 86% of all single-family homes that were ever built between 1940 and 1960. Appendix Section A.4 elaborates on the benchmarking procedure.

Since the incorporation of a city follows the initial development of the land, any jurisdiction incorporated after the constitutionality of zoning could have been developed under two jurisdictions: a prior zoning ordinance at the county or township level, followed by the post-incorporation ordinance. I split up development into pre-incorporation and post-incorporation samples based on year built for all places incorporated after 1930, so county or township-level regulations are estimated using these pre-incorporation subsamples.

With these data, I can reconstruct how development looked before and after a known min-

[8] Appendix Section A.1 describes the precise construction of the jurisdictions by state.

imum lot regulation adoption. Throughout this section, I use one jurisdiction as an illustrative example: the Philadelphia suburb of Lower Merion Township, PA. In 1939, the jurisdiction adopted 5 minimum lot size zones following a 1937 planning commission report. The effect of this ordinance is visually apparent in Figure 2. In the Figure, I use the CoreLogic records to construct the lot size distribution for properties built 20 years prior to the 1939 zoning ordinance, then for properties built 20 years after it.

Both distributions have similar sample sizes, but we can see properties after adoption have lot areas bunch at all but one of the minimum lot sizes. Few single-family homes occupied lots of less than 5,000 square feet following the 1939 ordinance, and we see the emergence of bunching at the 30,000 square feet minimum.

2.2 Inner Loop: Detecting Bunching Bins

My statistical method for detecting bunching is analogous to the estimation of bunching in the public economics literature (Kleven (2016)). That literature estimates the *excess mass* B of a distribution: the researcher has prior knowledge of what points on the support are bunching points, then assumes a *bunching region* where the distribution looks different from a counterfactual without incentives to bunch.

I have no knowledge beforehand of what minimum lot size controls are in each jurisdiction, but I can calculate excess mass statistics for lot bins and use that information for detection. Instead of limiting myself to the distribution of all observed properties, I can condition on intervals of years built and compare excess mass levels between older versus newer housing vintages.

However, another issue with using the excess mass statistic is that it is measured in levels of probability mass. Lot size distributions across different kinds of jurisdictions will vary in shape, and more generally have long right tails for expensive homes owned by rich households. Using excess mass alone leads to a test that has less statistical power to classify bunching in the right tail than around modal values. Detection therefore requires a statistic that is more robust to the underlying distributions.

My approach first fixes a jurisdiction j and a subset of *adoption times* τ over the support of years observed homes were built. Define

Definition 1. *A* ***housing vintage*** *in jurisdiction j after year τ, or a vintage indexed by (j,τ) denoted h^j_τ, is the collection of homes built in j between τ and $\tau+T$ for some integer T.*

The recorded lot sizes for h^j_τ form a lot size distribution, which is discretized into *lot bins*

$\{\ell\}$. In addition, let $\Delta(m,\mu)=\log(m(x)/m(x-\mu))$ be an operator that measures the gradient over the density m as the log change in the density between a point and a region of measure μ to its left.

If lot size controls were adopted after τ, observable bunching in h_τ^j reflects actual controls. As a first step, my algorithm maps lot sizes to whether they are *bunching bins* for a vintage indexed (j,τ). To do so, the algorithm follows a statistical decision rule making a classification based on rejecting the following null hypothesis:

$$H_0 : \Delta(h_\tau^j(\ell),\mu)-\Delta(h_{\tau,0}^j(\ell),\mu) \equiv G_\tau^j(\ell)=0.$$

where $h_{\tau,0}^j$ is a *counterfactual distribution* mapped to h_τ^j, representing lot development had there been no lot size controls.

$G_\tau^j(\ell)$ is what I call a *gradient statistic* for bunching around ℓ applied to a vintage indexed (j,τ). Intuitively, around ℓ it tests whether the distribution looks like what the theory of bunching predicts it should in presence of a notch point z: the density is high for values marginally larger than z and low for a bunching region of values lower than z. This statistic increases either from larger bunching levels, or from the low density in the bunching region reflecting density shifted to above the lot size control. The larger this statistic is, the likelier it reflects actual bunching around ℓ instead of noise.

Without prior knowledge of how the underlying lot size distribution looks like without controls, $\Delta(h_\tau^j,\mu)$ alone may be high due to increased demand for lots around size ℓ. Just like how excess mass in the bunching literature is calculated with respect to a counterfactual distribution, subtracting the statistic with an estimate of the counterfactual distribution centres the statistic around 0 when the null is true.

The three panels of Figure 3 visually demonstrate how to construct the gradient statistics $\hat{G}$ using Lower Merion data; details of the implemented algorithm are in Appendix Section B.1. In Panel (a), I show the histogram of properties after the adoption time $\tau = 1940$. I highlight two lot bins tested for significant bunching in orange, and highlight the measure of the bunching region to their left in blue. The log difference of the two densities form the first two terms of $\hat{G}_\tau^j(\ell)$. Panel (b) analogously shows statistic calculations over properties built a decade before τ, the 1930s. Any minimum lots adopted after 1940 would not bind over the pre-period sample, while the pre-period sample still indicates which lot sizes are in demand and which ones were not. The gradient statistic is evaluated over both samples and then takes their relative differences.

In Panel (c), I show how I classify the bunching bins. I plot the full gradient statistic for

Lower Merion, further standardized by a standard deviation of $\hat{G}$ estimated across jurisdictions. The standardized statistics oscillate around 0 except for statistics evaluated at some of the actual lot size controls, like at 5,000 and at 30,000 square feet. The algorithm concludes by classifying lot size controls with statistics above a critical value α as bunching bins starting in 1940.

Iterating over the set of vintages (j, τ), the algorithm outputs a collection of bunching bins for each vintage, $\underline{\mathbf{b}}(j, \tau)$. The algorithm also considers edge cases where the jurisdiction is entirely single-family residential, so strict enforcement of lot size controls means nothing is built below an arbitrary minimum lot size. The gradient statistic would be undefined in this case, but I use minimum and mode statistics to retrieve the minimum of minimum lot sizes. The decision rules are described below and explained in Appendix Section B.1.

To summarize, the algorithm maps information from lots in a jurisdiction into bunching bins for that jurisdiction using the following two definitions:

Definition 2. *Fix a set ℓ and a vintage of lot sizes $\{h^j_\tau\}$. A lot bin ℓ is a* ***bunching bin for a vintage*** *(j, τ), denoted $\underline{\mathbf{b}}(j, \tau)$, if one of the following holds:*

1. *ℓ is rejected by the gradient statistic decision rule explained by Figure 3;*
2. *ℓ is the minimum observed lot bin in the distribution and has a density mass of at least a threshold M^L;*
3. *ℓ is the modal value in the vintage distribution, and the empirical CDF evaluated at ℓ is no more than a threshold $\underline{F}$.*

To map this collection to estimated bunching bins *at the jurisdiction level*, I also consider that detecting bunching at the same lot size detected across multiple disjoint vintages provide information that the bunching is induced by persistent regulation.

Definition 3. *Fix a set ℓ and a collection of bunching bins for vintages $\{\underline{\mathbf{b}}(j, \tau)\}$. A lot bin ℓ is a* ***bunching bin for a jurisdiction****, denoted $\underline{\mathbf{b}}(j)$, if*

1. *ℓ was classified as a bunching bin over multiple disjoint vintages indexed by τ;*
2. *In at least one τ, ℓ was classified according to the gradient statistic decision rule.*

2.3 Identification Assumptions for Adoption Dates

While my algorithm flexibly classifies whether there is sizable bunching for lot bins, it cannot *identify* every lot size control coded in a zoning ordinance's text. I define and elaborate on the class of *regulatory notches* that the procedure can identify, which I claim are the distortionary lot size controls whose adoptions are worth understanding.

In the public finance literature, like Kleven (2016), a *notch* is a discontinuous rise in the tax schedule past some income threshold. Workers who would choose to earn more than the threshold absent taxes instead earn exactly the amount needed to avoid paying additional tax. Then, to the left of the notch there will be an anomalously high mass of workers.

To extend this reasoning to the housing market, let developers choose which lot sizes to build in a jurisdiction. Suppose the jurisdiction passes a minimum lot size control $\underline{\ell}'$, so no homes can be built on lots smaller than $\underline{\ell}'$ square feet per unit. When bidding for the land, developers whose cost efficiencies are in building small units (e.g. apartment developers) have no incentives to outbid other developers for lot size controled land.

If most of the jurisdiction has no minimum lot size but a portion of land is zoned for large lots $\underline{\ell}''$, the logic is similar. In a jurisdiction with undifferentiated developers, the land zoned for large lots where smaller lots would have been built in the counterfactual will lie empty, and a subinterval of housing segments built at certain densities will not be built.

In this sense, bunching around the minimum lot size is a form of *spatial arbitrage*: with the minimum lot size, developers who could outbid competing types absent regulatory constraints see falls in marginal revenue. Diminished competition for a parcel of land lowers the acquisition costs. Some developer may build on the land at the binding size, as long as for that developer the average reduction in land costs outweighs the losses in marginal revenue relative to the unregulated optimum.

Appendix Section C formalizes the above argument using a spatial equilibrium model with free entry into the housing sector but with heterogeneous types of developers. These model assumptions are sufficient to generate an endogenous land cost discontinuity and bunching behaviour in spatial equilibrium.

In the data, development is a dynamic process. Justifying my procedure therefore involves formalizing an econometric model with minimum lot regulations, bunching regions and sequential development.

Definition 4. *The data generating process (DGP) for each jurisdiction j is characterized by a collection $(\{n_t^j\}_t, L^j, \mathbf{Z}_t)$, where:*

- $\{n_t^j\}$ *is a distribution of heterogeneous developer and homebuyer types selecting into jurisdiction j during t.*
- $L^j : (\ell, t) \to \{0, 1\}$ *maps the time periods where a minimum lot size control is set at a lot range, if at all;*

- $\mathbf{Z}_t$ *is a vector of random variables capturing all remaining (time-varying) parameters of the development environment, like the value of a land plot.*

Developers form a competitive market in spatial equilibrium, profit maximizing subject to $(L^j, \mathbf{Z}_t)$.

There is a mapping from a lot size distribution without lot size controls, $\{\ell_t^j\}$ *to one that has them,* $\{\ell_t^{*j}\}$, *based on the following behavioural rule by developers. There exist minimum lot regulations* $\underline{L}$ *where at least one developer type building at* $\ell < \underline{L}$ *absent the control will bunch at* $\underline{L}$ *given the control. The set of these* $\underline{L}$ *form a subset of the image of* L^j.

The econometrician observes the distribution $\{\ell_t^{*j}\}$, the distribution of lot sizes developed in j based on $h_\tau^j(n_\tau^j)$.

The most important category of lot size controls without observable implications are "holding zone" regulations (Babcock and Bosselman (1973)). Vacant land is zoned requiring multiple acres of land per unit. No developer finds it economically feasible to build at those densities, but the jurisdiction lowers it once it finds a specific development project it likes. However, a consequence is that jurisdictions would if anything be estimated as less restrictive than they actually were: land zoned earlier on for unreasonably large lots will be interpreted as land infeasible for development at the time.

The next proposition formalizes what kind of "control jurisdiction" is needed for my identification strategy. Because I make minimal assumptions on where the bunching sizes have to be, the control jurisdiction's full lot size distribution — not just the moments of the distribution — must be similar to the tested jurisdiction's distribution.

Definition 5. *The* ***gradient statistic classifier*** $\hat{G}$ *of threshold* α *is defined over two distributions* $p(z), q(z)$ *and a function* μ, *using the operator* $\Delta(m,\mu) = \log(m(x)/m(x-\mu))$ *and the rule:*

$$\hat{G}(z^*) = 1 \quad \text{if} \;\; \Delta(p(z^*), \mu(z^*)) - \Delta(q(z^*), \mu(z^*)) \geq \alpha.$$

Proposition ID 1. *For a jurisdiction where the set of lot size controls* $\{\underline{L}\}$ *is nonempty, there is a threshold* δ *such that if the econometrician observes a collection of distributions* $\{q_t\}$ *for which the total variation distance between each* n_t^j, q_t *is less than* δ, *the gradient statistic classifier set identifies* $\{\underline{L}\}$.

Proof. See Appendix C. □

One can ask if the intuitive properties of bunching bins discussed in Section 2.2 can reflect actual lot size control adoption. Because the algorithm uses lot development patterns in years right before a potential adoption time as a control distribution, on which kinds of jurisdiction DGPs would this strategy work?

The set of bunching bins for a jurisdiction $\underline{\mathbf{b}}(j)$ are classified to approximate lot size controls satisfying

Assumption ID 1. *(Persistence) Conditional on causing bunching at ℓ over one vintage, a lot size control at ℓ causes bunching at a future vintage with probability 1. Any other reasons for bunching at some vintage — like developer preferences — has low conditional probability of reoccurring.*

The logic of how the gradient statistic was constructed in Figure 3 also assumes:

Assumption ID 2. *The econometrician observes at least one housing vintage built before any adoption of lot size controls.*

Assumption ID 3. *(Bounded demand adjustment) Excluding a fixed measure over the support of lot sizes, the distribution of lot sizes in j for τ and $\tau' = \tau + T$ satisfies a relation $|\ell^j_\tau - \ell^j_{\tau+T}| \leq T \cdot M^\ell$.*

One class of candidate DGPs appear to be jurisdictions where development is a mixture between *slow adjustment small development* and *stochastic planned developments*. On one hand, atomistic developers do not suddenly decide to all build on specific lot sizes which were not in large demand in the past. On another, I allow large-scale developers to build planned developments in a suburb where units all have a specific lot size. After building the development, they make an independent draw from a lot size distribution and decide to build another development with that standard lot size.

2.4 Outer Loop: Training Algorithm with Historical Regulations

A tractable classifier that can scale up to the nation deviate from arguments in Section 2.3. First, the assumptions for set identification is restrictive on the control distribution tagged to a distribution being classified, if such a control even exists. Second, arguments in Section 2.3 assume separate classifiers for *each jurisdiction*, whereas a tractable classifier keeps parameters constant *across jurisdictions*.

To implement a national classifier then involves a bias-variance tradeoff, common to machine learning problems. Variance appears as classification error: actual distortionary lot size controls can go undetected, while bunching due to statistical noise or nonstandard developer behavior are false positives in the data. These errors accumulate as the arguments in Section 2.3 make no guarantees on identifying the *adoption dates* of separate lot size controls. The gradient statistic classifier in Section 2.2 can infer a lot size control was adopted in some time period after a year τ, but even these interval estimates could be noisy.

The tradeoff to reduce noise is then to systematically not classify lot size controls that induce small levels of bunching, in terms of having low values of the gradient statistic. By raising classifier parameters like the critical value α, the density threshold M^L or the range of years built after adoption dates, the algorithm lowers conditional classification error at two costs: selecting on certain lot size controls and having the gradient statistic be less informative on precise adoption years.

To find an optimal classifer where classification bias can no longer fall without large costs in variance, I train an extended classifier to jointly optimize two sets of parameters. The first set are parameters that affect the bias of the bunching bin classification step. The second set are parameters that affect how to model the interval estimates for control adoption outputted by the classification step into point estimates for adoption.

Full details of the procedure are found in Appendix Section B.1, but I overview two main ideas: the additional model structure I impose to convert intervals for adoption into point estimates, as well as what data I use to train the algorithm.

With a set of bunching bins $\underline{\mathbf{b}}(j)$ identified, I assume they are in place for each possible year τ, then calculate excess mass and the gradient statistic for τ based off that assumption. The collection of these calculations forms an annual frequency time series of bunching magnitudes, measured two different ways. I fit a fully crossed collection of structural break models (Bai and Perron (1998)) across the two statistics and model specifications. The end result for each $\underline{\mathbf{b}}(j)$ is a set of estimates across models i for the "first appearance" of bunching around that bin.

These estimates at the jurisdiction-bin level depend on parameters for the classification procedure in Section 2.2, each of which trades off detection of smaller bunching behavior with classification error: $\hat{\mathbb{B}} = \{\alpha, \mu^{miss}, M^L, \underline{F}, M^{Growth}\}$. Appendix Section B.1 goes into more details about the parameters $\mu^{miss} and M^{Growth}$.

For the training data, I collect both a panel of zoning ordinances across time for 17 jurisdictions across the United States, as well as a collection of zoning ordinance studies from the 1960s and 1970s. These data span over 380 zoning jurisdictions in 13 states covering every Census region. Appendix Section A.5 further elaborates on data sources.

Over this national dataset, I extract a set of three variables wherever data are available. I tally all minimum lot size controls recorded in the documents, as well as the earliest year in historical documents where they are observed. For jurisdictions with lot size information as well as jurisdictions where only the adoption year for the first zoning ordinance was recorded, I record the first year that lot size controls were adopted in the ordinances, $Year_j$. For a subset of these jurisdictions with available data, I record the first year a minimum lot size of at least 10,000 square feet were adopted, $Year_j^L$.

The large lot size is the key outcome for training the model. Fitting model parameters over training data means finding the best predictor of large lot adoption years $Year_j^L$ across jurisdictions through reweighing the structural break estimates over a vector $\hat{\mathbf{w}}$.

2.5 Validating the Algorithm

I randomly divide the jurisdiction with known minimum lot regulations into a training set and a test set, to prevent overfitting. On the training set, comprising 60% of jurisdictions with the target outcome, I find the $\{\hat{\mathbb{B}}, \hat{\mathbf{w}}\}$ minimizing a least squares objective defined in Appendix Section B.1.

Table 1 shows performance on the training and test sets. The goal of the algorithm is that by being trained on fitting large lot size adoption, it will also yield unbiased and variance minimizing estimates of adoption across all lot size controls. Unbiasedness would be seen as estimates of outcomes not systematically greater or smaller than real values, while variance minimization is observable by a lower mean average error (MAE) in outcome predictions.

The first row of Panel (a) shows performance on the only outcome on which the model is fit: large lot adoption year in the training set. Moments indicate the classifier does produce unbiased estimates, at the expense of higher noise.[9] The MAE statistics when weighed for population are about 10 years. To the extent that many observations in the training set are small cities or townships, it reflects that first appearance of bunching is still a noisy proxy for first adoption.

The following three rows show out-of-sample performance on adoption prediction. Across all outcomes, the estimates' central moments are slightly negatively, meaning estimated adoption is somewhat later than actual adoption. In the test set, which contains more major cities, the algorithm performs better and lead to a smaller population weighted MAE. Over all cities in the data, 63 out of 123, or about half, of all estimated adoption years are within 10 years of historically recorded adoption.

Panel (c) shows a further test out of sample, using lot size control level variation rather than jurisdiction variation. For each actual lot size control, I match to it all bunching bins estimated to have been adopted within 10 years of the date in records. 587 out of 792 lot

[9] The false negative rate – actual large lots going undetected and dropping out of the training set – was around 13%.

sizes, or 74%, successfully match under this criterion. I then look at the matched bunching bin which is closest in magnitude to the actual lot size control, and like Song (2021) compare the percentage difference between them.

Across the matched lot sizes, the procedure has a median error of 0. 310 out of 587, or 53%, of actual lot size controls have an estimated bunching bin within 25% of the value. Conditional on the lot sizes being at most half an acre (21,780 square feet), the statistic is 271 out of 444, or 61%. Along with Figure B.1, which plots a histogram of these variables, it shows that the negative mean and MAE levels are due to larger lot sizes in the ordinances that go undetected with bunching methods — the "holding zones" referenced in Section 2.3.

3 The Dynamics of Lot Size Regulation Adoption

3.1 Measuring Lot Size Adoption Over Time

With the algorithm described in Section 2, I scale up estimation of minimum lot size adoption to all major American metropolitan areas. My algorithm iterates over 407 CBSAs and process CoreLogic property records having lot square footage and the year the property was built.

There are 8422 zoning jurisdictions in total, and Table 2 breaks the number down into four types: counties, townships, incorporated cities at constant historic borders, as well as annexed territory for cities growing beyond historic borders. Compared to all zoning jurisdictions in metropolitan areas, my algorithm excludes or cannot find sizable bunching for the smallest suburbs or for many rural towns. However, bunching is detected for most of each metropolitan area. In aggregate, lot size controls are estimated over jurisdictions in which more than 80% of sample CBSA residents live.

Every jurisdiction in the sample has a set of bunching bins $\{\underline{\mathbf{b}}(j)\}$, each element of which is paired with an estimated year of adoption. Measures of lot size adoption outcomes at the jurisdiction level are simple aggregations of the jurisdiction — bunching bin data.

My primary measure of lot size adoption is to record the first estimated adoption year among bunching bins of at least 3,000 square feet.[10] As an outcome in the empirical specification, the point estimate is transformed into a dummy varying at the decade level. For decades t ending in zero, I define *adoption over decade t* if the point estimate is in the interval $[t-5, t+5)$.

Coarsening adoption status back into time bins may exaggerate prediction errors for some jurisdictions — as an example, if a jurisdiction was estimated to adopt in 1955 but actually

[10] Though lot size controls below this threshold exist, they appear seldomly in postwar zoning ordinances and their first adoption is confounded by lot subdivision standards for old homes.

did so in 1953 — while dampening error for others. Since Section 2.5 confirms the algorithm reports higher error for adoption year estimates, later regression results on lot size adoption will make corrections for nonclassical measurement error.

The disaggregated bunching bin data admit other adoption outcomes. First, I create a *large lot size adoption year*, which looks at the first estimated adoption year for a bunching bin greater than 10,000 square feet. Second, I create a *downzoning binary variable* which is based on the 10th percentile bunching bin over all properties at a vintage's bunching bins. The conditional mean of this outcome reflects a rise in the smallest lots constrained by controls, either because areas zoned for small lots have been built up or because an ordinance revision removed a residential zone with smallest minimum lot requirements.

3.2 Measuring Lot Size Restrictiveness Over Time

Both within and outside the economics literature, multiple approaches exist to convert a jurisdiction's full set of land use controls into a single index. To look specifically at non-apartment residential development, researchers can define the presence or absence of a particularly stringent control; calculate the average of interrelated controls, like averaging multiple minimum lot size controls; or use statistical decompositions like the first principal component in a large space of regulations, like the Wharton Residential Land Use Index (WRLURI, Gyourko, Saiz and Summers (2008)).

By combining the distribution of built homes' lots in a time period with algorithmic estimates from Section 2 of which lot size controls applied at the time, I can define measures of *lot size restrictiveness.* A desirable condition for the restrictiveness index would have a higher index for jurisdiction j compared to j' *only* because minimum lot sizes in j bind on relatively more land that could support denser development. This condition motivates what I call the *excess mass measure* of lot size restrictiveness. This measure is defined for each vintage (j, τ) in two steps.

The excess mass measure for vintage (j, τ) is constructed in two steps, following similar methodology in the public economics literature (Kleven (2016)). First, I generate a new lot size distribution which proxies as a counterfactual distribution for j if no lot size adoption happened. I invoke the same identifying assumptions for the gradient statistic and assume for some time interval around adoption, there is small bound on the rate at which underlying demand for lot sizes changed in j. I drop properties at each estimated $\underline{\mathbf{b}}$ adopted after τ, as well as those with lots in the bunching region to the left of each $\underline{\mathbf{b}}$ as defined in Appendix Section B.1. The new counterfactual lot size excludes lot sizes affected by bunching behaviour, defined over vintage τ as well as over a pre-period sample 10 years before τ.

With this lot size distribution, I use kernel density estimation methods to derive a smooth counterfactual mass estimate at each lot size, $\tilde{m}_t^j(\ell)$. The *excess mass measure* of j after τ is

$$\text{Excess}(j; \tilde{m}_\tau^j, \underline{\mathbf{b}}) = \sum_{\ell \in \{\underline{\mathbf{b}}\}} m_\tau^j(\ell) - \tilde{m}_\tau^j(\ell),$$

which I calculate separately for every year built observed in the data.

Under the algorithm's baseline specification, a housing vintage indexed by (j, τ) includes all houses in jurisdiction j built in years between τ and $\tau + 10$. For analysis purposes, I refer principally to the *excess mass measure* of *j over decade* τ, which equals the excess mass measure of j after $\tau - 10$.

By construction, the excess mass measure isolates variation across jurisdictions due only to constrained development on marginal land, zoned for a lot size control higher than what could be built there otherwise. It is an appropriate statistic for counting how many properties were constrained by lot size controls across jurisdictions and years, rather than counting properties that just happened to be built at the controlled lot sizes.

To be more specific, I claim the excess mass measure has two advantages over another intuitive measure, the *bunching share* measure. For a past housing vintage, I look at all properties around bunching bins $(\underline{b})$ in effect by then. The *bunching share* of j after τ is thus

$$\text{Bunch}(j; \underline{\mathbf{b}}) = \sum_{\ell \in \{\underline{\mathbf{b}}\}} m_\tau^j(\ell),$$

where $m_\tau^j \ell$ is the density mass in the lot size distribution of vintage (j, τ) at ℓ. Both the bunching share and excess mass measures range between 0 to 1. The first issue with the bunching share is that it is confounded by properties that would have been built at a lot size even if no lot size controls were in effect. Across time, a rise in the bunching share could be due both to more marginal units being constrained or from positive demand shocks for houses with certain lot sizes.

The second, more subtle issue is that when estimating causal effects of some explanatory variable on lot size outcomes, the effect could come through jurisdictions proposing more sophisticated planning and introducing more minimum lot sizes. If one channel for restrictiveness effects is a rise in the average number of imposed zones, that mechanically increases the bunching share measure even if those additional zones restrict density for few new developments.

In Appendix C, I show both issues above can be interpreted as forms of non-classical measurement error in the outcome. I show that the bunching share can be decomposed into two terms: a term capturing how much an explanatory variable explains variation in restrictiveness, as well as a term capturing average bias based on the conditional selection of minimum lot sizes by jurisdictions. I present results using the bunching share as an outcome, but the excess mass measure varies only based on marginal housing units constrained by lot size controls, which minimizes selection bias terms compared to alternative measures of lot size restrictiveness.

Of note is that counterfactual distributions are created excluding properties around bunching bins $\underline{\mathbf{b}}(j)$ regardless of year built. There is no deterministic relationship between excess mass levels and whether a bunching bin was detected for adoption in a decade.

3.3 Lot Size Adoption and Restrictiveness Over the Past Century

Using the Tax Data without missing variables across 1925 to 2010, a sample of 67.3 million properties, I calculate 12.1 million properties had lots around around bunching bins after regulation adoption. In other words, 18% of homes built over those 85 years were built at bunching bins. Figure F.1 plots the distribution of property-weighted bunching bins in full, which is a mixture of both standardized controls at fractions of an acre as well as further diffuse bunching (Anagol et al. (2022)) at lot sizes marginally larger than the most frequently occurring controls.

By matching bunching bin based estimates of lot size adoption in Section 3.1 to all large and mid-sized zoning jurisdictions with over 5,000 residents in 2010, I estimate when lot size controls spread across the U.S. Returning to Figure 1, I visualize 4,807 jurisdictions that have zoning powers and had at least 5,000 people by the 2010 Census. The orange series shows a fifth of jurisdictions were early adopters by 1940. However, 59% of jurisdictions adopted the controls from 1940–1970. These estimates corroborate historical anecdotes that the planning of postwar suburbs under lot size controls happened in tandem with the suburbanization process.

Figure 1 also conditions on bunching bins exceeding 10,000 square feet, my measure of large lot adoption: 38% of properties around bunching bins do so at bins greater than 10,000 square feet. Around 13% of jurisdictions had these large lot controls in 1940, but about 44% of all large and mid-sized jurisdictions added them too over 1940–1970, tapering off from 1975 onwards. [11] The general postwar strategy was for jurisdictions to zone for multiple "exclusive"

[11] When constructing the figure, I assume no jurisdictions repeal their minimum lot sizes in full, as well as repeal all large minimum lot sizes in the large lot adoption time series. It is unlikely this assumption significantly biases the time series, given such a reversion is not seen in recent surveys (Gyourko, Hartley and Krimmel (2021)).

non-overlapping zones, each limited to single-family home development and with their own minimum lot size. Of the zoning jurisdictions with detected lot sizes, over 79% of them have multiple bunching bins, with a mean of 3.7 bins and a standard deviation of 2.5.

I show the scale with which these lot sizes distort housing markets in Figure 4. For each vintage of housing built every decade, I aggregate up restrictiveness measures at the jurisdiction level to the nation. Dividing by total units in the national vintage results in a restrictiveness measure as a share of U.S. housing development. Once adoption of the lot size minimum regulation surged in the post-war years, I estimate a rise in the share of properties bunching at lots from 12% in the 1940-50 period to as high as 23% in the 1960-70 period. By the end of the 1950s, I also estimate the first nontrivial levels of large minimum lot controls above 10,000 square feet. Over the 1940–70 period, the average excess mass restrictiveness is 13.5%. Multiplied by the 20.1 million homes built during those years in my sample, I conclude 2.71 million postwar housing units were constrained by lot size controls.

Of additional interest is that after 1970, national lot size restrictiveness persists. From 1940 to 2010, national excess mass restrictiveness is 13.7% of all development. From 1970–2010, the same national rate stabilizes at 13.7%. Rising incomes and demand for larger homes influence more bunching around large lots, but continue to constrain homes to be larger than what market developers could subdivide and still satisfy demand.

4 Theoretical Predictions on Exclusionary Zoning

Explanations for why land use controls exist vary within economics, never mind between economics and other social sciences. An established economics literature considers the controls as a mechanism efficiently providing public goods by decentralized government under Tiebout sorting of households. I present a unified framework in which local governments may follow established theory, or decide based on racial endogenous amenities highlighted in political science or in history.

4.1 Potential motives for land use controls

Throughout this section, I consider a model where households are differentiated in incomes $y \in [\underline{y}, \overline{y}]$. Households of type y optimize over three kinds of goods, $U(c, h, g$: a numeraire c, a house size h with unit price p_h and a public good g.

In the model, public goods are produced by a set of jurisdictions $\{j\}$ in a regional economy. Each jurisdiction provides a specific level of public good per capita g. Residents decide the level of g to provide, usually following a median voter theorem assumption: in equilibrium, g per capita is based on the median willingness to pay among j's residents. [12]

Planners of j may choose to adopt a land use control that requires consumption of a minimum house size $\underline{h}(j)$ among all residents. I explain how results in the literature can be considered as further assumptions onto this general environment.

Fiscal zoning. The standard model assumes governments finance public good provision on a balanced budget. A marginal resident of j unwilling to pay the per capita tax financing the public good cause less per capita public good provision.

Hamilton (1975) shows that governments can use their right to minimum house sizes to implement efficient outcomes. A system of land use controls across jurisdictions equate to a menu of entry fees for any level of public goods. In spatial equilibrium, where households are indifferent between their residential choice and all alternatives, jurisdictions are perfectly stratified. All households demanding $\overline{g}$ level of public good live in the same jurisdiction j, have the same house size $\underline{h}(j)$, and earns utility

$$V(p_h(j), g(j)) = U(y - p_h(j)\underline{h}(j), \underline{h}(j), g(j)).$$

Calabrese, Epple and Romano (2012) generalizes this result by showing how Pareto optimal allocations of public goods can be implemented by decentralized governments each setting different entry fees.[13]

With spatial equilibrium and perfect stratification of jurisdictions, the hedonic price gradient for housing measured across jurisdictions reflect underlying marginal rates of substitution:

$$\frac{dp(j)}{dg(j)} = -\frac{dV/dg}{dV/dp} = \frac{1}{\underline{h}(j)} \times \frac{dU/dg}{dU/dy}.$$

Because fiscal zoning implements allocatively efficient public goods markets, loosening land use controls to integrate residents within government borders is justified only if equity concerns outweigh the loss in efficient implementation.

[12]In this discussion I abstract from local political economy factors, for which Hilber and Robert-Nicoud (2013) provide a theoretical framework. Land use adoptions are made by a planning board, which weighs the gain in land rents from new development along with residents' preferences for more or less development. The political economy mechanism could further affect the design of land use controls across jurisdictions.

[13]Fischel (2015) and Bogart (1993) consider household preferences where housing consumption is also a function of nearby natural amenities and negative externalities. For example, a residential neighborhood with open space and clean air is more desirable than one where residences are integrated with traffic congestion and industrial pollution. These are models of local public goods conceiving them as bundles of amenities rather than a single good. Calabrese, Epple and Romano (2012) shows decentralization implements efficient public goods provision even with differentiated willingness to pay for component amenities.

Endogenous amenities: racial homophily in location choice. An endogenous amenities theory considers racial homogeneity as a direct input into resident preferences. Evidence for endogenous amenities, even not explicitly racial homophily, belongs to an established literature beginning with Bayer, Ferreira and McMillan (2007) and Diamond (2016). Alternative methods in political science, as in Reny and Newman (2018) and Sahn (2022), find proximity to Black migration causes changing racial preferences among White incumbents or disapproval of anti-discrimination legislation.

A standard microfoundation for racial homophily is to add a separable disamenity term based on racial composition. Let there be a majority white group W and an excluded racial group E. Households now move between jurisdictions j to maximize

$$V(j;\underline{W}(j)) = \max_{c,h,g} \quad [\underline{W}(j)]^{\gamma} U(c,h,g)$$
$$s.t. \quad \underline{W}(j) = \frac{W(j)}{W(j)+E(j)}, \gamma > 1.$$

$\underline{W}(j)$ is the share of white residents in jurisdiction j, and a higher γ reflects greater unwillingness to live next to excluded group members, modeling a reduced form relationship of homophily preferences as in Schelling (1971).

Assume there exists one homogeneous jurisdiction j^*, where $\underline{W}(j^*) = 1$. Spatial equilibrium in the regional economy now implies relative to j^*,

$$U(p_h(j^*), g(j^*)) = [\underline{W}(j)]^{\gamma} U(c, p_h(j')h(j'), g(j')); \quad j' \neq j^*.$$

Willigness to pay for the endogenous amenity can be reflected in a steeper gradient for house prices p_h in homogeneous jurisdictions. As long as land use controls are still decided to price marginal households for entry, the gradient in housing expenditures $p_h h$ is also steeper across jurisdictions than in the fiscal zoning environment.

Under the condition $\frac{\partial p_h}{\partial (1-\underline{W})} < 0$, this economy may also exhibit "tipping point" behavior like in Card, Mas and Rothstein (2008). Stratification by income becomes unstable if members of the excluded group have income convergence with the majority group. Increasing migration of group E to the region, with nonnegative effects on $\underline{W}$ everywhere in equilibrium, leads to white flight from some jurisdictions to others.

Preferences for segregated public goods. A second way through which prefereces for racial composition alter the regional economy is if composition affects households' preferences for local public goods. Caetano and Maheshri (2017, 2022) and Angrist et al. (2022) document homophily preferences since the 2000s for a particular public good: public education.

For a more conceptual discussion, consider households choosing residence between jurisdictions under the preferences

$$V(j;\underline{W}(j)) = \max_{c,h,g} \quad U(c,h,[\underline{W}(j)]^{\gamma} g).$$

$\underline{W}$ remains the white share of a jurisdiction j in which the household resides. However, the white share affects residential choice through households less willing to finance public goods in more heterogeneous communities. In the baseline framework, there are also no way to add a surcharge just on residents of an excluded group, such as an impact fee proposed by Gyourko (1991).

Under the above assumptions, the median voter is less inclined to trade off higher housing prices for more public goods:

$$\frac{dp}{dg} = \frac{[\underline{W}(j)]^{\gamma}}{h(j)} \times \frac{dU/dg}{dU/dy}, \quad [\underline{W}(j)]^{\gamma} < 1.$$

The median voter is therefore incentivized to substitute away from public goods provision – in our environment, toward numeraire consumption and larger homes. Schwab and Oates (1991) and Alesina, Baqir and Easterly (1999) are also results in the literature along these lines. The end result is preferences for segregated public goods can lead to an aggregate loss in public good provision, causing allocative inefficiency beyond just equity concerns.

4.2 Theoretical predictions for empirics

Fiscal zoning, as mentioned in Section 4.1, can be called a *race-neutral motive for zoning* to contrast with the latter two *racial motives for zoning*. The two categories of theories differ on whether they can rationalize *racially differentiated causal effects*:

Hypothesis: Let the average treatment effect of demographic change in group X on land use controls be denoted ATE[X]. Consider demographic change in Black Americans $Black$ with average income level $\overline{y}^{Black}$, then with a subsample of white Americans with the same average income, $Wht \mid \overline{y}^{Black}$. Then

$$\text{ATE}\,[Black] - \text{ATE}\,[Wht \mid \overline{y}^{Black}] > 0.$$

Racial motive theories predict the above claim. Growth in the excluded group $Black$ means $\underline{W}$ decreases for at least some governments; growth in the white group with similar incomes does not. Racial motive theories allow for changes in $\underline{W}$ alone to shift the relative attractiveness of locations. Changes in equilibrium sorting then affect land use controls on the margin.

The claim is not as clear under race-neutral motives. A pure fiscal zoning theory cannot explain racially differentiated effects if income is controlled for entirely. A weaker race-neutral explanation must still assume Black Americans differ systematically in their willingness to pay for certain amenities then white Americans of similar incomes.[14]

I do not claim this paper shows results from the above ideal experiment. My claims are more modest: I aim to show a collection of causal effects where Black demographic change has much stronger causal effects than demographic change in lower-class Whites. The sum of the evidence should shift priors on whether postwar land use decisions were based only on race-neutral motives, or were best explained in part by racial motives.

5 Empirical Strategy

In an ideal experiment to estimate how land use controls respond to a metropolitan area's demographic change, different metropolitan areas see random shocks to demographic growth. Comparing how local governments design controls in high-shock areas versus low-shock areas identifies average treatment effects of the demographic change intervention. While these effects can arise out of several possible mechanisms, like fiscal zoning versus homophily, they inform on how much of the housing stock was constrained as a result of allowing local land use powers during a period of demographic transition.

Lacking such an experiment, using observed growth trends as the independent variable leaves open several confounding explanations for estimated effects. Cities with greater Black population growth may differ systematically from ones who don't, as they possess characteristics which pull greater migration toward them. Estimated effects then confound causal effects of demographic change with metropolitan area dynamics that continue to affect suburbanization and land use control design regardless of demographic transition.

The rest of this section describes how the endogenity challenge can be overcome by studying land use control responses to the Second Great Migration, from 1940–1970. As Black population growth in non-Southern cities during this time was mainly driven by Southern migration, I construct an instrument that isolates exogenous shocks in migration uncorrelated

[14]One thesis among many in Rothstein (2017) emphasizes the Federal Housing Administration (FHA) in designing residential segregation, where they refused to insure financing for racially integrated developments. A national policy directed by federal institutions, though, has difficulty explaining racial differentiation in effects identified from demographic change between cities across the country.

with the destination cities' urban growth dynamics. [15] An instrumental varibles regression, using variation in predicted migration across central cities and across time, approximates the ideal experiment of random demographic shocks.

5.1 Data

My empirical strategy requires panel variation in lot size control outcomes, observed demographic change in destination cities and Southern migration statistics needed to construct instrumental variables. For the analysis, I merge my measures of lot size adoption and restrictiveness detailed in Section 3, with Census demographic data defined over time-consistent city boundaries.

With every decennial Census since 1940, the U.S. Census Bureau tabulated demographic information across incorporated cities. I define a central city as the largest city of a Core-Based Statistical Area (CBSA), based off of 2013 definitions by the Office of Management and Budget (OMB). The remaining zoning jurisdictions j in the counties constituting the CBSAs are matched to a common shock in their central city $c(j)$.

For a central city incorporated since 1940, the Census provides levels and changes in population by race. Breakdowns by migration status are also available since 1950. However, these data are only digitized for all central cities in my sample following the 1970 Census. Additionally, the annexation of surrounding land by central cities (Alesina, Baqir and Hoxby (2004)) means boundaries also varied over the analysis period, necessitating constant boundary definitions.

To recover these city-level independent variables over 1940–1970 and construct constant boundaries for annexing cities, I harmonize five data sources: the City Data from the the 1944-1977 County and City Data Books; the 1950 and 1960 Elizabeth Bogue census tract level data along with 1970 Census place-level tabulations, as provided by Manson et al. (2021); the 1950 and 1960 Census public microdata, as provided by Ruggles et al. (2022); and the 1930 and 1940 full count censuses (Ruggles et al. (2021)), where individual records are aggregated to cities using the city crosswalk of Berkes, Karger and Nencka (2022). Appendix Section A.6 describes the harmonization procedure.

Throughout the paper, I use a measure of Black demographic change defined across central

[15] For the remainder of this paper, "Southern" refers to the 14 states in the Census Bureau's South region that are not Delaware, Maryland, or the District of Columbia.

cities c and decades t, central city Black composition change:

$$\Delta CC_{c,t}^{Black} = \Big(\log(\text{Black}_{c,t}) - \log(\text{Black}_{c,t-1})\Big) \times \frac{\text{s}_{c,t}^{b} + \text{s}_{c,t-1}^{b}}{2},$$

a product of Black population growth across decades and of Black population shares s^b across decades. A one percentage point increase in $\Delta CC_{c,t}^{Black}$, approximates a one percentage point increase in the central city's Black population as a share of total population, warranting the statistic as a measure of "composition change." [16]

If suburbanization trends were already accelerated in destination cities for Black migrants, outlying jurisdictions in metropolitan areas may systematically differ on characteristics. To control for these characteristics in my main specification, I also aggregate socioeconomic variables using the 1940 full count Census and the full Census Place Project crosswalk from Berkes, Karger and Nencka (2022). Appendix Section A.7 details the match process.

Table 3 shows counts and summary statistics for the main outcomes and right-hand side variables, pooled over three decades of observations. The leftmost four columns shows statistics for the analysis sample, limited to non-Southern states and where outcome, central city demographics and controls are all available. The rightmost columns show statistics for a full dataset, indicating if the analysis sample construction has caused significant selection on observables.

Panel (a) tabulates the distribution of outcomes across jurisdictions over postwar decades. About 5,000 of suburban cities outside metropolitan areas' central cities remain in the analysis sample. Over 1940–70, the pooled adoption rate in the sample is 77%; the average vintage size is around 700; and I estimate 6.3% of properties in those vintages were constrained by lot size controls. To account for certain jurisdictions not developing intensely until halfway through the sample, I drop vintages of 30 or fewer properties when looking at lot size restrictiveness outcomes.

Panel (b) shows 383 of the 407 CBSAs examined by my algorithm have central city demographic information from 1940–70. I observe a balanced panel of characteristics for 223 non-Southern metropolitan areas. Averaged across cities, central cities outside the South had 2% of their population being Black in 1940, but for the three decades afterwards this share

[16] ΔCC^{Black} also relates to the right-hand side variable in Derenoncourt (2022), where it is described as the "1940–70 increase in urban Black population, as a share of the 1940 urban population." Said statistic would be defined for central cities as: $\Delta \text{Black}_c^{1940\text{–}70} = \frac{\text{Black}_{c,1970} - \text{Black}_{c,1940}}{\text{Pop}_{c,1940}} = \frac{\text{Black}_{c,1970} - \text{Black}_{c,1940}}{\text{Black}_{c,1940}} \times s_{c,1940}^{b}$. Both our measures are therefore products of a Black population growth term and of a population composition term, but my measure express the growth term in natural logarithm points while her measure does so in percentage points.

increased by 2 points. Large standard deviations in the demographic change variables is due to a right skew in the distribution, driven by the most popular Black migrant destinations.

Panel (c) look at 1940 demographics for postwar suburban zoning jurisdictions matched to 1940 full count Census tabulations. Like their central cities before the Great Migration, the prewar suburbs outside the South had only 1.6% Black residents on average.[17] Apart from racial homogeneity, the suburbs had degrees of socioeconomic diversity. historical estimates of homeownership rates, average household earnings and professional status have sizable standard deviations relative to the mean. The sample also covers jurisdictions both near and far from the central city, and distance to the city will be used as a control in the main regression specification.

5.2 Second Great Migration: Historical Context

In the three decades up to the Second World War, Black Americans had already migrated out of the South and formed communities in America's largest cities. As important as the "First Great Migration" was in establishing migration networks between sending counties and destination cities, current historical evidence shows the "Second Great Migration" from 1940–70 involved much greater Black migration levels. Gregory (2005) estimates around 4 million Southern Black Americans migrated to the North from 1940–70, which was twice as many Black Americans who did during the preceding 1910–40 period. Collins (2021) also shows Southern Black Americans born from 1920–1950 left the South more quickly upon entering adulthood compared to prior cohorts.

Among additional work in the economics literature, Boustan (2016) shows the wage premia Black migrants could earn working in Northern industrial sectors, a strong incentive for southern Black adults to make the move. Stuart and Taylor (2021) further demonstrates initial Black migration networks accelerate future migration: social networks disseminate information about job opportunities in the North, such that multiple Southern Black adults follow the location choice of an initial migrant.

However, from the perspective of white incumbent residents in the destination cities, the growth in Black residents appeared to cause a permanent change in the urban fabric. Reynolds (2019) quotes an incumbent resident's account in Rochester, NY, suggesting a shift in beliefs that the divide between central city and suburbs turned racial:

> About 1950 Rochester became part of the greatest migration of human beings in

[17] 83% of residents in there are whites born in the U.S., so suburbanization had not been widespread for either Black or foreign-born households.

> the history of the world. *The city changed. It's now a Negro city.* [emphasis added] I think the gulf between the white man and the black man is wider [in 2000] than it was at any time in my lifetime ...

Boustan (2010) quantifies large effects on emigration from central cities caused by Black migration — the "white flight" channel of postwar suburbanization. The main results presented in the new section then provides novel evidence of how the suburbs were also reshaped in response to the outmigration, as measured by the lot size controls suburban governments exploited to stratify neighborhoods by income or by wealth.

To visualize any relationship between postwar Black migration and incumbents' response through land use regulations, Figure 5 maps the joint evolution of Black population growth in the central cities along with the restrictiveness of postwar lot size controls. Panel (a) plots the decade average of central city Black composition change from 1940-70, expressed in units approximating percentage point changes. A high value on the map corresponds to a central city receiving persistently high levels of Black migration, relative to the national average each decade, across the analysis time period.

Panel (a) confirms the growth in Black migration across 253 Northern and West Coast metropolitan areas during the postwar decades. America's most populous metropolitan areas at the time—New York, Chicago, Detroit and Philadelphia—all rank among the top receivers of Black migrants in their urban cores. Greater variation in Black composition change comes from regional cities: to list four examples, despite being the largest cities in their states, Portland, OR, Indianapolis, St. Louis and Salt Lake City had considerably different Black migration rates.

Panel (b), for a subset of 150 non-Southern metropolitan areas with sufficient data, plots an association with central cities growing more Black and the area's suburbs restricting lot sizes more. Lot size restrictiveness is calculated over the 1940–70 pooled vintage for a metro area's suburbs, plotted on the Y-axis. The X-axis variable is the Black composition change variable seen in Panel (a). When the identifying variation covers a national sample of cities, a positive association is clearly seen. While the rest of this paper teases out the causal channel of this association, my novel data advances understanding of lot size determinants beyond existing case studies of individual metropolitan areas. Subjects of those case studies, like Detroit, New York or Boston, are highlighted on the scatter plot in blue.

5.3 Main Specification and Instrument Construction

To estimate causal effects of demographic change on land use regulation, I begin with the following linear model on a national panel of jurisdictions j:

$$Reg_{jt} = \beta \Delta CC^{Black}_{c(j),t} + \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}. \tag{1}$$

For a regulatory outcome — lot size adoption and restrictiveness — measured in j over decade t, I estimate how the outcomes change with variation in their central city's Black composition change, $\Delta CC^{Black}_{c(j),t}$. In the baseline specification, I exclude the central city jurisdictions themselves, so the sample is limited to the surrounding suburban jurisdictions in the CBSA boundaries. I also include Census region–decade fixed effects $\delta_{r(j),t}$,[18] as well as a set of 1940 jurisdiction characteristics $\mathbf{X}_{j,pre}$, predetermined before the Second Great Migration started.

The exogeneity condition, $\mathbb{E}\left[\varepsilon_{j,c(j)t} \middle| \Delta CC^{Black}, \delta_{r(j),t}, \mathbf{X}\right] = 0$, implies no confounding explanations for why some metropolitan area have both more Black migration to their central cities and suburbs designing their lot size controls a certain way. Since standard theories for a location's land use regulation predicts greater restrictiveness in response to increasing demand for said location, there are numerous demand-side factors that could be a confounder and cause bias. [19]

Whatever the central city characteristics that can confound estimates, said characteristics introduce omitted variable bias if Southern Black migrants are more likely to move to those cities *on average.* In one direction, Black migrants may have moved systematically to cities with relative disamenities. Absent Black migration, over the decades there would have been accelerated suburbanization by households on the margin between staying in the central city versus moving out. Theory predicts accelerated suburbanization would increase suburban land use controls, all else held equal. Therefore, in this case estimating the effect β with ordinary least squares lead to positive bias.

Conversely, Black migrants may have moved systematically to cities with relative amenities, such as proximity to a variety of jobs and consumption amenities close to place of residence. Absent Black migration, households on the margin would continue to stay in the central city and postwar suburbanization would have decelerated over the decades. In this case, OLS estimates of β would be negatively biased.

A sufficient identification strategy is to compute counterfactual Black migration rates into

[18] Allowing for Census region fixed effects partials out geographical explanations for lot size restrictions. One example is to compare New England with the West Coast. Even if they received similar rates of Black migration, suburban planning on the West Coast may be more restrictive than New England because West Coast development requires denser water infrastructure (Burchfield et al. (2006)).

[19] Possible confounders include Black migrants selecting into cities with richer people, who could be early suburbanizers; cities with denser streetcar and transportation infrastructure allowing expansion of the urban area; or cities surrounded by natural amenities that are in demand by city-dwellers (Lee and Lin (2018))

destination cities that are exogenous to those cities' urban growth patterns. Following the existing literature in Boustan (2010) and Derenoncourt (2022), I construct these counterfactual Black migration rates from a regression of actual rates with a migration shift-share instrument. Because push factors due to the structural transformation of the rural Southern economy can explain rises in outmigration rates, the explainable growth in outmigration are tallied up in the shift-share construction. The rates, pooled across counties sending Black migrants, serve as the identifying variation for predicting counterfactual migration into central cities.

To construct the shift-share instrument, I first use the 1940 full-count census and filter to all Black residents in non-Southern central cities who report to have moved between 1935–1940. The 1940 Census is the first full-count census that asks for past migration activity up to the Census year, and 309 thousand Black residents report the state and county they moved from.

Across all non-Southern central cities c and potential sending Southern counties k, I define the shift-share instrument for Black migration Z_{ct}^{Black} as:

$$Z_{ct}^{Black} = \sum_{\text{Southern } k} \tilde{\omega}_{c,1940}(k) \times g_k(t),$$
$$g_k(t) = \frac{\overrightarrow{\text{Black}}_{k,t}}{\overrightarrow{\text{Black}}_{k,1940}} \times \frac{\widehat{\text{mig rate}}_{kt}}{\text{mig rate}_{kt}},$$

and $\tilde{\omega}_{c,1940}(k)$ is the share of all 1940 migrants into city c coming from a sending Southern county k. The term $g_k(t)$ denotes the "shocked" outmigration rate for a sending Southern county k, as a ratio of 1940 rates. In the migration rate term, $\overrightarrow{\text{Black}}_{k,t}$ is the level of Black migrants out of county k over decade t. mig $\text{rate}_{kt}, \widehat{\text{mig rate}}_{kt}$ are respectively the realized versus the projected migration rate for k.

Tabulating the 1940 full count census provides me with both $\tilde{\omega}_{c,1940}(k)$ as well as $\overline{\text{Black}}_{k,1940}$ for counties of origin reported by respondents. Data on migration levels and rates across Southern counties for ensuing decades come from Census records compiled by Boustan (2016) and used in Derenoncourt (2022). The outputted instrument has a natural interpretation as the predicted number of Black migrants into central city c, as a ratio of the 1940 pre-period Black migration level into c. [20]

[20] In Boustan (2010) and Derenoncourt (2022) the shift-share instrument is expressed as a linear combination of $\omega'_{k,1940}(c)$, migrants from county k to c as a share of all 1940 migrants out of k, multiplied by the level of predicted migrants out of k. Appendix D.2 shows the two expressions of the instrument are numerically equivalent up to a scaling factor, central city Black residents $\text{Black}_{c,1940}$.

5.4 Addressing Threats to Identification

The instrument Z_{ct}^{black} defined in Section 5.3 identifies the causal effect under relevance and the exclusion restrictions. Relevance is satisfied when $Cov(\Delta CC_{ct}^{black}, Z_{ct}^{black}) \neq 0$. In general, the exclusion restriction applied to data implies Z_{ct}^{black} must satisfy strict exogeneity: $\mathbb{E}[Z_{c(j)t}^{Black} \times \varepsilon_{jt'} | \mathbf{X}_{j,pre}] = 0$. The instrument for a decade t explains the outcome only through its effect on contemporary Black migration ΔCC_{ct}^{black}, and in particular not through any confounding effects that can appear in ensuing decades, $\varepsilon_{jt'}$. [21]

Recent literature on shift-share instruments separate out cases where the exclusion restriction is satisfied due to exogeneity of shares (Goldsmith-Pinkham, Sorkin and Swift (2020)) or due to exogeneity of shocks Borusyak, Hull and Jaravel (2022). In the context of the Great Migration instrument, the shares $\tilde{\omega}_{c,1940}(k)$ are a function of migration networks, developed due to historical railroad connections or social networks formed from the First Great Migration. As it is more difficult to argue that historical settlement patterns by Black migrants in central cities is conditionally exogenous with postwar urban planning, I argue the shift-share instrument identifies causal effects through exogenous shocks.

In Appendix Section D.2, I state the formal requirements on shock exogeneity based off of the results in Borusyak, Hull and Jaravel (2022). Intuitively, two conditions are sufficient for the shift-share IV to satisfy the exclusion restriction. First, each county-level migration rate shock $g_k(t)$ is the sum of a random component and of a fixed effect at some cluster of counties. The random component must also be mutually uncorrelated with shocks in other county-years (k, t). Second, instrument relevance for each destination city c must come from a variety of counties. As the number of cities c in the sample grows, the set of sending counties used in the shares construction also grows instead of staying fixed.

A clear threat to identification is if a common economic factor is changing a city's urban growth trajectory as well as the propensity to migrate. For example, an economic boom in Georgia can both increase suburbanization in Atlanta as well as within-state Black migration to Atlanta. This concern clarifies why I drop all Southern cities from my analysis sample, an example of a "split-sample" estimation strategy in the shift-share literature.

A second threat is that the observed migration rate, mig rate$_{kt}$, is driven by network effects that are functions of past migration *levels*. The growth in outmigration for a larger county over time, relative to a small rural county, is likelier to represent the acceleration of Black migrants

[21] For example, postwar cities in a metropolitan area could scale back public spending per capita given the rise in new Black residents, or pursue large scale urban renewal projects. While these policies cause displacement of Black residents, the exclusion restriction states predicted Black migration flows over time should be uncorrelated with the intensity of these policy changes.

informed of which cities have economic opportunities. As in Boustan (2010) and Derenoncourt (2022), I project migration rates for counties at each t, $\widehat{\text{mig rate}}_{kt}$, based off of cross-sectional relationships between county characteristics and migration.

Appendix Section D.1 describes the procedure in more detail; in sum, I confirm the literature's findings on agricultural sector factors determining Southern Black migration. Variation in the migration rate shock is predicted by the intensity of tenancy rates in Southern farms, as well as by the specific crops produced on those farms. The results are consistent with a story where agricultural mechanization after World War II rolled out at differential rates across the South. The technology shock disrupted the region's exploitative sharecropping system, then caused staggered exit of Black farmers to other occupations in American cities.

A final threat is the presence of unidentified fixed effects across clusters of counties. If predicted migration rates still estimate persistently high migration rates for specific counties, the instrument could have spurious correlation with the characteristics of cities to which those higher-rate counties send their migrants.

To address this, I assume the clusters are defined at the state level and construct state exposure shares as described in Borusyak, Hull and Jaravel (2022). For each c and Southern state s, I include the 1940 share of migrants to c from s, $\sum_{k' \in s} \tilde{\omega}_{c,1940}(k')$. Incluing this 14-element vector of controls for all Southern states, interacted with period fixed effects, removes contamination of migration shocks through state-period fixed effects.

In sum, addressing all threats to identification lead me to the final econometric model. With the Black migration shift-share $Z^{Black}_{c(j),t}$ defined as in Section 5.3, I estimate the following model using two-stage least squares:

$$\begin{aligned} Reg_{jt} &= \beta \Delta CC^{Black}_{c(j),t} + \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t} \\ \Delta CC^{black}_{c(j)t} &= \gamma Z^{Black}_{c(j)t} + \delta_{r(j),t} + \mathbf{X}_{c(j),pre}\Gamma + \nu_{c(j),t}, \end{aligned} \tag{2}$$

where the pre-period controls $\mathbf{X}_{c(j),pre}$ include both state exposure shares for Black migrants as well as 1940 central city and suburban jurisdiction characteristics. All pre-period controls are fully interacted with decade dummies to get decade-specific trends conditional on 1940 covariates.

To choose the 1940 characteristics to control for in $\mathbf{X}_{c(j),pre}$, I conduct covariate balance tests between the shift-share instrument and the characteristics. If the instrument significantly predicts jurisdiction characteristics, that demonstrates a channel for spurious correlation that must be closed by adding the characteristic as a control.

Table D.4 shows the result of the pre-period characteristic exercise across increasingly saturated specifications, from an OLS model with $\Delta CC^{Black}_{c(j),t}$ to using the shift-share instrument with controls. Overall, the addition of controls and the instrument attenuates biases and make some correlations insignificant, like manufacturing share and suburban native-born white share. However, in the main specification I include additional jurisdiction covariates which still have correlations with the residualized shift-share instrument: 1940 Black share and homeownership in outlying jurisdictions, as well as jurisdictions' distance from the central city. Appendix D.3 describes the balance tests in more detail.

Figure 6 presents an illustrative relationship between the instrument and the Black composition change explanatory variable. In the binned scatterplot and in the analysis, I make a further transformation on the variables by expressing them in percentiles relative to each decade's distribution. Some metropolitan areas like Washington, D.C. and Detroit received much higher Black migration rates than elsewhere, as shown in Derenoncourt (2022). The percentile transformation corrects for regression estimates overweighing for outlier metropolitan areas receiving the most Black migrants.

Figure 6 suggests the instrument satisfies relevance with endogenous migration, and in a persistent way. As the binned mean relationships are plotted separately by decade, the instrument maintains relevance across decades. Table 4 reports the actual first stage regression run over the jurisdiction-level panel. In this model and all ensuing models, I follow Abadie et al. (2022) and cluster all standard errors at the level of composition change variation: at the CBSA-decade level. Both visually and across several specifications, the shift-share instrument has clear relevance to the outcome, with cluster-adjusted first stage $F-$statistics exceeding 200.

6 Results

6.1 Effects of Black Composition Change on Lot Size Restrictions

Figure 7 shows how suburban lot size outcomes respond to within-decade variation in the migration shift-share instrument. As mentioned in Section 5.4, my main specifications make a further transformation on the composition change and migration variables, converting them into distribution percentiles calculated separately for each decade. Pooled over all three decades from 1940–70, there is a clear positive relationship with predicted Black migration to central cities with two measures of restrictiveness explained in Section 3.2. The slope of the relationships, reflecting the reduced form estimate of causal effects

$$\Delta\widehat{CC}^{black}_{c(j)t} = \gamma Z^{Black}_{c(j)t} + \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \nu_{j,c(j)t}$$

imply a move from the metro with the lowest to highest predicted Black migration explains a rise in 5.59 percentage points of excess mass restrictiveness, 88% of the baseline rate. The relationship of the instrument with the pooled adoption binary outcome is also significant and positive, as seen in Figure F.2.

Table 5 presents regression estimates of causal effects of central city Black composition change on lot size outcomes, across specifcations based on the OLS model (1) and the IV model (2). My preferred causal estimates are in Column (3) among IV models, which presents TSLS estimates of β conditional on predetermined controls selected in Section 5.4, interacted with decade dummies. A move from the metro with the lowest to highest actual Black composition change every decade explains a 28 percentage point rise in the propensity to adopt minimum lot size controls. On my preferred measure of restrictiveness, the same move across metropolitan areas explains a 7.95 percentage increase in excess mass.

Compared to Column (2) of the OLS model, where the controls are equally saturated, the OLS estimates are downward biased. Following the reasoning in Section 5.3, this means Southern Black migrants were selecting into central cities that would have had less suburbanization and restrictive land use absent the Great Migration. The economic reasoning is that new suburban residents in those cities are net fiscal gains for local governments' capacity to finance public goods, so local actors sought to attract the residents rather than exclude them.

Comparison with Column (2) of the IV estimates is also illustrative. Leaving out predetermined suburban characteristics sizably attenuates adoption effects and somewhat attenuates restrictiveness effects. This implies suburban characteristics control for suburban zoning practices prior to 1940. Even if suburbs did not adopt minimum lot sizes specifically, early zoning of single-family housing districts could still exclude higher density development. Not conditioning for the practices cause negative omitted variable bias in IV estimates.

Tables E.1 and E.2 show results using OLS and IV specification on the full set of outcomes defined in Section 3. In my preferred specification, Black composition change leads to *less* adoption of large lot size controls. However, this result is noisier across specifications: I show in Section 6.4 the effect changes sign after applying reweighing methods. I also find a null result on downzoning across decades, which occurred a third of the time. This implies planning restrictiveness was a matter of designing zones when lot size controls were first planned, rather than eased in over decades.

The causal estimates of Black composition change on bunching share restrictiveness in Table E.2 is higher than in the main specification, at 9.38 versus 7.95 percentage points. Following the discussion in Section 3.2, bunching share measures do not explicitly control for how strategic lot size adoption could cause sorting and future demand around properties at controlled lot sizes. The difference in estimates point to the mechanism causing positive bias, as expected.

6.2 Aggregation of Results

The regression results are expressed as comparisons between cities receiving minimal to maximal Black migration every decade, but a clearer estimand would be how much Black composition change can causally explain the lot size outcomes. I aggregate up decadal variation in three steps to derive the impact of a Second Great Migration shock, cumulated from 1940–1970, on postwar lot size outcomes.

First, for each decade I calculate the non-Southern standard deviation for the untransformed ΔCC^{Black} variable, weighing by CBSA single-family and duplex construction. The value, $\sigma_{\Delta CC^{black}}$, is steady at around 3.7 points per year: cumulated over three decades, it equates to around an 11 percentage point change in central city Black composition. Then, I scale the standard deviation by the causal effect and normalize with the non-Southern construction-weighted average in the lot size outcome:

$$Excl_t = \frac{\hat{\beta} \times s_{\Delta CC_t^{black}}}{E[\Delta Reg_t]}$$

With a set of $Excl_t$, I aggregate to the final Great Migration measure by weighing decadal estimates by non-Southern decadal single-family and duplex construction q_t:

$$Excl = \sum_{t=1950}^{1970} \frac{q_t^{natl} \times Excl_t}{\sum q^{natl}}.$$

Panel (a) of 8 visualize the aggregate measures for the two primary outcomes of lot size adoption and excess mass restrictiveness. I plot point estimates for outcomes across specifications, coupled with repeated simulations assigning influence weights across CBSA-decades to derive a Bayesian bootstrap confidence interval (Rubin (1981)). Figures F.3 and F.4 visualize the same for all five outcomes examined in Section 6.1.

For the adoption outcome of lot sizes, I derive a point estimate of share explained at 20.1% with a 95% bootstrapped confidence interval of [15.6%, 24.1%]. The share explained of restrictiveness — how local governments design where to place lot size controls after adoption — is significantly greater. 41.0% of excess mass restrictiveness is explained by the one standard deviation Great Migration shock, with a 95% bootstrapped CI of $[30.8\%, 51.0\%]$.

Yet another measurement of how lot size responses to race impact the housing market compares with a counterfactual where I shut down the relationship between the two variables. More precisely, I look at the change in lot size restrictiveness if all suburbs in my sample set lot sizes as if $\Delta CC^{Black} = 0$, and rescale the change in percentage terms by the realized number of housing units at the time. The $\Delta CC^{Black} = 0$ counterfactual should be thought of not just as a world where local governments have no right to control land use, but specifically one where land use planning was decided by race-neutral planners setting land use controls to follow the market. The average non-Southern change in restrictiveness would be the aggregate share of restrictiveness explained by the mean level of ΔCC^{black}, about 5.15 points per decade. I interpolate these effects in terms of total constrained housing units in my analysis sample of suburbs, over 1940–1970:

$$\begin{aligned}\text{Units Affected} &= \left(10.9M \text{ units} \times 13.4\ \%\right) \times \left(41.0\% \times \frac{\mu_{\Delta CC^{black}}}{\sigma_{\Delta CC^{black}}}\right) \\ &= 1.45M \times \left(41.0\% \times \frac{5.15}{3.7}\right) = \text{ 830,000 units.}\end{aligned}$$

I claim that at least 830,000 suburban units outside of the South were built in areas planned following racial exclusionary zoning and controlled by minimum lot sizes. If local governments could not exercise discretion to zone in response to Black migration, market forces would have provided denser homes on these marginal parcels of land. Just how many additional units would have been provided is difficult to predict, absent from a model of housing demand and location choice fit on historical prices and quantities. Instead, I consider simpler back-of-the-envelope calculations.

From 1940–1970, the average density per acre of properties built around bunching bins is 5.1 properties per acre. If developers were willing to subdivide smaller lots so the density increased to 7.5 properties per acre,[22] housing market forces would have provided 391,000 more housing units without local lot size control responses to Black composition change.

What if local governments acted altruistically and planned their land use correctly anticipating future affordable housing needs? I consider another calculation where instead of restricting

[22]Approximately 6,000 square feet, which is the modal minimum lot size nationwide relative to a property-weighted distribution plotted in Figure F.1.

land use on the margin, governments upzoned marginal parcels of land for denser development. If regulations allowed dense housing development at 15 dwelling units per acre,[23] that would increase the zoned capacity of marginal parcels by $(15/5.1-1)\times 830,000 = 1.61$ million units.

To interpret the size of this upper bound on permitted suburban density, the maximum amount of units built on the marginal parcels would reach over a quarter of the 1968 Housing Act's goal of 6 million affordable homes built in a decade. These calculations are uninformative on whether it would have enhanced welfare to upzone for such densities everywhere in the nation. Rather, it illustrates that by the time federal policy had ambitious goals for integrating suburban neighborhoods through increasing housing supply, local lot size controls were already binding constraints on any federal planning problem.

To think about which metropolitan areas contribute most to the aggregate effect, recall from Figure 4 that Black migration to central city is a right-skewed distribution. New York City, the Bay Area and Detroit are some of the metropolitan areas whose suburbs I predict were reshaped by local exclusionary zoning. Comparing the built form of Detroit—with a 31.2-point rise in central city Black composition over three decades—with Minneapolis—a 3.2 point cumulative rise —suggests the potential of the former to have had dense, accessible suburbs for its workers without local actors' zoning discretion.

6.3 Opposite Sign Effects with White Demographic Change

Even though they were less likely to leave the South from 1940–1970, Southern White Americans actually migrated to non-Southern cities at higher *levels* than Black migrants. Pooling Census records, I look at white residents living in non-Southern states who report their state of birth as one of 14 Southern states. I confirm estimates by Gregory (2005) from Census data that around 8.2 million Southern Whites left their region over my analysis period.

Southern white migrants, as another demographic shock to postwar U.S. cities, proxy for a group not differentiated from incumbents by race but is differentiated through income. Figure 8 contrast household income distributions for three groups in the 1960 Census data from Ruggles et al. (2022), conditional on living in identified central cities: incumbent non-Southern whites, self-reported white migrants from Southern states since 1955, and self-reported Black migrants from Southern states. While recent Southern white migrants earn more on average

[23] As is required by recent state-level mandates on affordable housing, like Massachusetts' upzoning laws (Kulka, Sood and Chiumenti (2022)).

than Black migrants, their mean annual household income is \$6,500 in 1960 dollars. When compared to an average of \$8,200 for incumbent whites, there is a 20% gap in mean incomes.

The potential for these lower earners to purchase homes beyond the central city, under a pure fiscal motive, should still cause a less pronounced increase in lot size restrictiveness in anticipation. To credibly estimate this causal demographic effect, I repeat the shift-share identification strategy for Black migrants, but now defining shares and shocks of a Southern white instrument $Z^{S-white}$ in terms of 1940 white migrant flows and Southern county white outmigration rates captured by Bowles et al. (2016).

I therefore run the instrumental variables regression

$$\begin{aligned} Reg_{jt} &= \beta \Delta CC^{S-white}_{c(j),t} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t} \\ \Delta CC^{S-white}_{c(j)t} &= \gamma Z^{S-white}_{c(j)t} + \delta_t + \mathbf{X}_{c(j),pre}\Gamma + \nu_{c(j),t}. \end{aligned} \tag{3}$$

Compared to the baseline Black composition change regression, I keep all controls except for the exposure shares by state variables calculated over historic Black migration flows.[24]Due to missing data across time, I also adjust the definition of $\Delta CC^{S-white}$ to be new Southern white migrants over the decade, in percent of the central city population:

$$\Delta CC^{S-white}_{ct} = \text{Migrants}^{S-white}_{c,t-10} / \text{Pop}_{c,t-10}.$$

Tables 6 and E.3 compares OLS and IV estimates across demographic changes by race, specifications (2) and (3) respectively. The main finding is that, unlike effects of Black migration on lot size outcomes, results over both restrictiveness and adoption using instrumented Southern white migration has null, if not negative effects. An additional OLS regression on the association with lot size outcomes of foreign-born share change in central cities—the variation coming mostly through differential rates of flight from the city — also finds null effects.

Differences in Southern migrant incomes by race mean the difference in effect sizes could be upward biased compared to the ideal test in Section 4.2. However, the magnitude of racially differentiated responses in land use controls, combined with nonpositive effects for two groups of lower-income white residents, offer additional evidence the results cannot be explained solely by fiscal zoning. Fiscal zoning would still predict a smaller, but positive response in land use regulation, rather than a reality where postwar suburbs did not behave in a way to exclude new white homeowners that were more working class than incumbents.

[24]For the Southern white specification, I also replace region-decade fixed effects $\delta_{r(j),t}$ with decade fixed effects δ_t to avoid a weak instrument problem with $Z^{S-white}_{c(j)t}$.

6.4 Effects Robust Across Specifications

Time-varying causal effects. Exclusionary zoning conducted by postwar local governments was part of wider practices by private actors discriminating on race. Some developers specialized in suburban housing with racial restrictive covenants, forbidding any individual on a street to deviate and sell to a Black homeowner (Sood, Speagle and Ehrman-Solberg (2019)); banks also practiced redlining, refusing to lend mortgages to Black renters willing to buy a home (Ross and Tootell (2004)). For private actors, these discriminatory practices had been justified as profit maximizing strategies. For example, Sood, Speagle and Ehrman-Solberg (2019) describes racial covenanted units advertised as an amenity, with developers marking up the units' prices for discriminatory buyers.

Economic justifications for a local government zoning for exclusion have analogies in private actors' arguments. In the standard framework of Section 4, local governments would benefit from marking up housing consumption through minimum lot sizes, giving the government more of a budget to expend. A racial exclusionary zoning explanation goes further: as long as the local government is providing the same bundle of public goods over time and there remain homeowners who pay premia for racial homophily, the local government will maintain exclusionary practices over the whole time period of analysis.

Racial exclusionary zoning therefore predicts a constant, or nondecreasing, effect of Black migration on lot size outcomes when effects are estimated separately across decades. Additional institutional context predicts that later in time, the effects of Black migration may even increase. After the 1940s, racial covenants were banned by the Supreme Court, taking away one form of private provision of exclusionary suburbs. A growth in local governments' market share for exclusionary suburbs would be reflected by a rise in causal effects estimated over post-1950 vintages compared to pre-1950 vintages.

Panel (b) of Figure 8 conducts the aggregation calculations in Section 6.2, but estimating a separate coefficient specific to outcomes for each decade. While these results have more noise than the pooled estimator assuming constant effects across decades, they conform with dynamic predictions on local government lot size predictions under racial exclusionary zoning.

Black composition change in central cities explains 22.7 % of lot size adoption from 1940-50, and a similar 16.2 % of adoption from 1960-70. Point estimates of Black composition change's effects on restrictiveness indicate increasing explanatory power of Black migration, from 27.7 % for 1940s homes to 50.5 % for 1960s homes. Wide confidence intervals imply I cannot reject the null that effects on restrictiveness are equal across decades. However, even constant effects over 30 years show racial exclusionary zoning was not a perfect complement of private sector discrimination that was outlawed early on in the analysis time frame.

Reweighing of jurisdictions. The main results in Section 6.1 are taken over all non-central city jurisdictions, as of 2010, in sample CBSAs. The sample of jurisdictions for CBSAs therefore include two kinds of "never taker" jurisdictions. Cities near the urban core may be entirely built up ahead of 1940 with no vacant lots, in which case a minimum lot size on paper has minimal effects on construction. Future exurbs may also not be developed until near the end or after the 1940–70 time period.

To find effects for a subpopulation of suburbs that are "compliers" to demographic change through altering lot size controls on developable land, I reweigh jurisdictions by the units of pre-1950 housing development in their borders, as a share of total extant pre-1950 development in CBSA borders. As an example, a suburb building 1000 units from 1930-1950 while all counties in the CBSA built 50,000 units over that period would have a share of 0.02 over all decades. These weights additionally correct for accidental overweighing of metropolitan areas where it was easier to incorporate and have fragmented governance than where it was not. The fragmentation could be endogenous to racial composition change (Alesina, Baqir and Hoxby (2004)).

Table E.4 present reweighed estimates for main results on adoption and restrictiveness in Section 6.1. Figures F.5 and F.6 repeat the aggregation steps in Section 6.2 with new reweighted estimates of causal effects.

These estimands, incorporating one way to correct for attenuation of results, imply a one standard deviation Great Migration shock across the nation led to adoption of lot size controls but also the extensive margin of having a restrictive large lot size control. They also imply a one-SD Great Migration shock caused 55% of postwar restrictiveness, or 1.1 million suburban postwar units outside the South constrained by lot size controls.

Tables E.5 and E.6 reweigh the specifications across racially differentiated demographic change in Section 6.3. While effect sizes are slightly more positive and less noisy across specifications, no non-Black demographic change have positive estimated effects on lot size outcomes.

Alternate specifications. Table 7 shows the main results on lot size restrictiveness are not due to the exact specification chosen. Relative to some specifications, the main results serve as a lower bound. For ease of reference, the main results are repeated in Column (1). Table E.7 runs similar alternatives for results on lot size adoption.

Column (2) shows including the Southern white migrant shift-share as a control variable, which also proxies for wider flight from central cities, does not affect estimates. Column (3) adds a CBSA-decade level control of suburban population growth, reflecting the relative attractiveness of a city's suburban amenities to central city ones. Adding the control only slightly decreases the magnitude of the main effects.

Columns (4)-(6) show the statistical significance of results are not sensitive to misspecification of the restrictiveness variable. In Column (4), I filter on a subsample which has at least one bunching bin detected at a higher critical value threshold of 1.96. Jurisdictions that remain had more constrained development: average restrictiveness over this subsample is 9.7% versus 6.1% in the full sample. Column (5) aggregates the regression to a MSA level by weighing jurisdiction-level outcomes by total units constructed from 1940–70. Column (6) corrects for left censoring of the excess mass variable by running an IV Tobit regression. All results show that, when conditioning on variation less likely to be noisy or censored, the effect sizes of Black migration on lot size restrictiveness can double.

Finally, Column (7) shows regressions run on the level of Black composition change, which is noisier than the within-decade percentile variable but fits more conventionally in the shift-share IV set up. Results remain significant at the 5% level in both OLS and IV specifications.

Nonlinearity in Black migration. To compare the IV estimates in Column (7) of Table 7 to estimates in Table 5, I derive a difference of 18 between the first and 99th percentiles of ΔCC^{Black}. Using the IV estimate, the product $0.35 \times 18.0 = 6.3$, which is comparable but smaller in magnitude to the main results.

Exploring if the Black migration effects are nonlinear both helps to explain the discrepancy as well as further validate the theory proposed in Section 4. If postwar suburbanites, as specified by the framework's utility function, have diminishing marginal disutilities from the Black share of a jurisdiction, restrictive lot sizes go up in greater degree in reaction to the first percent of Black migration than for subsequent migration growth. Least squares estimation of the best linear approximation for concave effects of Black migration, as is done in Table 7, will then have an attenuated estimate. To derive the linear approximation, least squares estimators overweigh the contribution from suburbs facing outlier Black migration shocks, like in New York or in Detroit, relative to a specification where the explanatory variables are scaled by a quantile transformation.

While the shift-share instrument lacks the power to detect nonlinear effects, a OLS specification that augments what is used in Column (7) with a quadratic term in Black composition change ΔCC^{Black} is suggestive of concave effects. The quadratic term is -0.0011 (s.e. 0.0005) when the outcome is lot size adoption, while it is -0.03 (s.e. 0.018) when the outcome is lot size restrictiveness.

In Appendix D.4, I report additional checks that show the main empirical results are not driven by outlier cities or by outcome mismeasurement issues. On the latter point, I use digitized zoning ordinances reported in Section 2.5 to argue any mismeasurement of outcomes downward biases on the true effect magnitudes.

7 Exclusionary Zoning Circumvents Desegregation Policies

The racially differentiated effects on Southern Black versus Southern white migration in Section 6 remain consistent with three potential theoretical mechanisms: fiscal zoning for excluding only the poorest residents, homophily preferences and demand for segregated public goods. In this section, I tease out the contribution of the latter mechanism through variation in the legal environment leading up to postwar suburbanization.

Prior to national Supreme Court rulings, states had differed in how intensely they outlawed discrimination by race. The policy changes affected the supply of segregated public goods unless a local government could ensure racial homogeneity through other means.

7.1 Desegregation Policy Heading into Postwar Suburbanization

As wartime policies ended after 1945, Americans focused again on the domestic issues around them. Legal restrictions on racial integration in services were legitimized by the Supreme Court decision in *Plessy v. Ferguson* (1896), but a postwar political coalition around desegregation launched a new attack on legalized segregation. A coalition of political and civil society groups — driven by the National Association for the Advancement of Colored People (NAACP), labor unions and the Truman Administration — campaigned on the passage of anti-discrimination law and an end to legal segregation (Sullivan (2009)).

Apart from filing due process suits against segregation in the courts, a well-staffed NAACP and fellow liberals lobbied for desegregation in schools, fair employment practices and restricting segregation in housing. The campaign would be remembered for two major legal victories. In *Shelley v. Kraemer* (1948), the Supreme Court declared the unconstitutionality of racially restrictive covenants, which bound a housing development to be sold only to other white buyers. Six years later, the Court ruled segregated public education was unconstitutional in *Brown v. Board of Education* (1954).

The period between these cases also saw a progression of cases ruling in favor of civil rights.[25] The same political coalition scored incremental victories at the state level, passing their proposed anti-discrimination legislation in certain non-Southern states.

Referring to the framework in Section 4.1, legal changes set constraints on permissible levels of racial homogeneity $\underline{W}$ in public goods when the overall Black population increases. These additional constraints will not have a first-order effect on location choices for residents who support integration, always unwilling to pay premia for racial homogeneity. Those who do have racial homophily preferences, however, may have paid premia to access exclusive neighborhoods in central cities prior to the legislation. Seeing policy change and anticipating a fall in $\underline{W}$, residents with exclusionary preferences exploit a new way they can access their endogenous amenities. They sort into new suburban neighborhoods that are zoned to ensure provision of racial endogenous amenities, paying a premium to do so.

In particular, the theory predicts the marginal effects of Black migration would differ systematically by the legal environment. Land use control responses to Black composition change are stronger in a more anti-discriminatory legal environment. A large migration elasticity to policy changes corresponds to inelasticity for racial composition γ in the theoretical framework, adding an incentive to set lot size controls and raising jurisdiction-level bunching mass on land controlled by lot sizes.

7.2 Heterogeneous Effects by Legal Environment

I use a standard reference for anti-discrimination laws featured in recent work (Cook et al. (2022)); *States' Laws on Race and Color* by Pauli Murray (1950), Along with cataloguing the scope of legal segregation, it also breaks down a variety of anti-discrimination laws passed between states as of 1950. Variables constructed from this data then represents early adoption of anti-discrimination laws at the start of postwar suburbanization.

The body of Murray's book includes nine categories of anti-discrimination laws, but for analysis I consider three: anti-discrimination in public education (hereafter "ADE") laws, fair employment laws and anti-discrimination in public accomodation (hereafter "ADP") laws. The first two categories tend to follow model legislation at the time, especially the policy proposals offered by President Truman's Committee on Civil Rights (1947). An example of an ADE law for Pennsylvania states:

> Hereafter it shall be unlawful for any school director, superintendent, or teacher to make any distinction whatever. . . by reason of the race or color of any pupil. . . seeking

[25] Supreme Court cases like *McLaurin v. Oklahoma State Regents* (1950) previewed the *Brown v. Board* decision, as the Court unanimously rendered unconstitutional segregation policies in higher education.

admission to any public school maintained [under Pennsylvania law].

A comprehensive fair employment law, like one passed in New Jersey in 1949, empowers a commission to investigate any refusal to hire, any denial of labor union entry, and any discriminatory job posting based on race.

Based off of Chart 2 in Murray (1950), I code 16 states that have or have not passed an ADE law in public education by 1950. I then count the number of occupations in each state with legal rights against racial discrimination as a proxy for the comprehensiveness of the fair employment law. I finally use the count of total anti-discrimination laws by state coded by Cook et al. (2022), growing in both comprehensiveness of protections as well as in the scope of ADP laws.

For each measure of a type of state anti-discrimination law, State Law$_s$, I estimate the main specification's effects interacted with the legal environment variable:

$$
\begin{aligned}
Reg_{jt} = \beta \Delta CC^{Black}_{c(j),t} + \alpha^{SL} \text{ State Law}_{s(j)} + \beta^{SL} \Delta CC^{Black}_{c(j),t} \times \text{State Law}_{s(j)} \\
+ \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}.
\end{aligned} \tag{4}
$$

Presence of effect heterogeneity for a state law measure, defined as $\beta^{SL} \neq 0$, indicates a policy environment affecting certain incumbents' preferences drive the average marginal effects. In particular, the coefficient on the interaction term with presence of ADE laws, β^{ADE}, would be positive if preferences for segregated local public goods drive the urban area's land use adoption equilibrium. Any residual causal effect, captured by β, is attributable to either fiscal zoning or homophily motives.

I reuse the shift-share identification strategy in Section 5.3 to estimate both β and β^{SL}. Identification of any β^{SL} requires an exogeneity condition on the policy variable of interest with the identifying variation: $\mathbb{E}[Z^{Black}_{c(j)t} \text{State Law}_{s(j)} \,|\mathbf{X}_{j,pre}] = 0$. One interpretation of this exogeneity condition is that the timing of when new laws were adopted was uncorrelated with the predicted intensity of migration based off of Southern structural change.

Falsification tests for the assumption may not be easy to construct without a clear model on what political variables enabled the legislation's passage. However, given historical evidence that the NAACP played a leading role with legislation, the timing of legal environment change would be associated more with organizational strength across states than future migration trends.[26] Figure F.7 shows that presence of legal rights against segregated schooling was found

[26]Historical records from NAACP-allied organizations list successful initiatives to desegregate local school districts before and after legislation passage (Congress (1952)).

in states in all non-Southern regions, and is not tightly correlated with the number of laws passed as collected by Cook et al. (2022).

7.3 Results: The Role of Demand for Segregated Education

Table 8 presents OLS and IV estimates of the legal interactions model in 7.2, where the legal measure is presence of an anti-discrimination in public education law. For excess mass restrictiveness, the coefficient β^{ADE}, capturing the causal effect of Black composition change explained by presence of ADE laws, is highly significant along with the uninteracted effect β. [27] Following theoretical predictions, the effect magnitude is stronger in ADE states than non-ADE states by a factor of $1+\beta^{ADE}/\beta = 1+6.17/4.18 = 2.5$. For adoption, ADE states have stronger Black composition change responses by a factor of 1.8. Just as how the main causal effects of migration are stronger on lot size restrictiveness, heterogeneity by early ADE adoption is strongest when restrictiveness is the outcome. Results on restrictiveness point specifically at the role preferences for segregated education could play in designing postwar suburbs.

One concern is that policy variation at the state level is correlated with other policies adopted at the same time or with state characteristics. In Table E.8, I look at two other sources of legal variation between states. In Panel (a), I interact Black composition change across states with total anti-discrimination laws presented in Cook et al. (2022). Effects are both statistically insignificant and nearly zero in magnitude. A state like New York, which passed 54 of these laws, did not shift local incentives to adopt lot size controls like one law on desegregation of schools did.

In Panel (b) , I interact Black composition change with the number of fair labor laws across different work categories, which was the classification in Murray (1950). The average non-Southern state had laws across two categories, with a standard deviation of 2.1. Results show more fair labor laws decelerated lot size adoption rates in metropolitan areas, while having insignificant positive effects on restrictiveness.

Even if neither source of legal variation explains additional variation like ADE law adoption , it does not mean there was no role for fiscal zoning or homophily. Rather, results reflect the salience of Black homeownership affecting school homogeneity above other consequences of residential integration.

The results by ADE laws are also robust to alternative specifications. Table E.9 shows that a reweighted sample in the manner of Section 6.4 output similar results for lot size restrictive-

[27] Equivalently, conditional on being in an ADE state, the same change in Black composition change led to stronger effects. Figure F.8 plots nonparametric evidence for this claim over both outcomes.

ness. Add-on effects on lot size adoption, which were already small in Table 8, are negligible in the reweighed specification.

Separating out certain states or periods in the analysis sample retains the results on restrictiveness. Columns (2) and (3) of Table E.10 additionally compares the main results in Table 8 without two sets of states that passed the strongest "Fair Education" laws by 1950. Column (4) drops 5 states who last revised their ADE laws during 1920-1940 up to 1950; a concern might be their policy adoption was endogenous to Black migration conditions during the First Great Migration, 1910–1940. The result of β^{ADE} being large and statistically significant does not disappear across any alternate model.

When limiting the analysis sample to only lot size outcomes and Black migration in the 1950s – the post-period of state-level policies, but prior to the most expansive Civil Rights laws – I find a larger and statistically significant value of β^{ADE} : 7.522 (s.e. $= 3.736$). The point estimate of β^{ADE} is comparable in a sample covering the 1950s and 1960s, as well as for the 1960s alone[28].

Using state-level variation in anti-discrimination laws, I find that only early adoption of laws banning discrimination in public education independently explains the strength of migration's effects on lot size outcomes. While policy adoption is not randomly assigned, results from the adoption design indicate the role preferences for segregated public education could have played on determining suburban residential design. A remaining positive effect in non-ADE states suggests a remaining role played by fiscal zoning and homophily preferences.

8 Conclusion

In this paper, I develop a novel algorithm to estimate when and how U.S. suburban land use controls emerged over time. While U.S. suburbs used zoning ordinances for a century to regulate a variety of land uses, my algorithm focuses on the intensity with which suburbs controlled residential density with minimum lot size controls.

The algorithm's results pinpoint how suburbs already restrictive land use in the postwar years. Using a new excess mass measure of lot size restrictiveness, I estimate 2.7 million housing units from 1940–1970 were built on land that could have supported more density without decentralized land use planning. It is beyond this paper's scope to estimate exactly how many units could have been built, but a rise in density of 40% from actual development would have produced over a million more single-family units.

[28] As the 1960s saw civil rights legislation adopted nationally, the point estimate on β^{ADE} is no longer significant.

I use time-varying outcomes in lot size adoption and restrictiveness, the first in the literature, to estimate how much the adoption of restrictive lot size controls can be explained by postwar Black migration to non-Southern cities. I find sizable effects: from 1940–1970 around 800 thousand to over a million units during these postwar decades would have been developed at higher densities absent local government land use controls.

A similar research design on white demographic change finds null effects, while the effects of Black composition change are strongest in states passing a model law banning segregation in public education. In combination, the effects of Black demographic change are not well explained by traditional fiscal zoning motives for regulation. Results are more consistent with a model where land use controls were planned in a way mutually desirable to incumbent and new residents: to preserve communities' racial composition when new migrants are separated through race.

Only near the end of my analysis period did public debate stir over the reality of what I call racial exclusionary zoning. A year after the 1965 Watts Riots in Los Angeles's Black neighborhoods, Babcock (1966) provided the first use of the term "exclusionary zoning" in popular press. A year after the Long Hot Summer of 1967 and at President Johnson's behest, the Kerner Commission provided the following explanation on the causes of 1967's civil disturbances:

> White racism is essentially responsible for the explosive mixture which has been accumulating in our cities since the end of World War II. At the base of this mixture are three ... bitter fruits ... The first [of which] is surely the continuing exclusion of great numbers of Negroes from the benefits of economic progress[,] through . . . their enforced confinement in segregated housing and schools. (1968)

The Civil Rights Act was passed in 1968, part of a series of laws aiming to integrate cities and reverse declining Black social mobility. My results imply that regardless of policy objectives, postwar local governments had already planned land use to shield themselves from future demographic change. Legal challenges to governments' land use intentions on constitutional grounds also floundered since *Warth v. Seldin* (1975), where the Supreme Court ruled arguments on how land use controls cause residential exclusion lacked standing for federal judicial remedies.

Analyzing the patchwork of laws, federal programs and policies to integrate racial groups after 1970 is beyond this paper's scope. My novel data on lot size outcomes would be a key part of approaching those questions in future work. First, knowing where land use controls bind today involves identifying jurisdictions where lot size restrictiveness persisted beyond their

1970 levels. Both postwar levels and post-1970 levels of lot size restrictiveness can be used to explain Black-Nonblack segregation both within and between government borders. Such estimates could tease apart homophily explanations for racial sorting with the role of durable housing and path-dependent urban planning decisions.

Next, measures of lot size adoption and restrictiveness across jurisdictions can be matched to other jurisdictional data on house prices, taxation, public goods provision and demographics. The panel data would cover all the main mechanisms through which U.S. local governments adjust the quality and consumer base of their public goods, allowing estimation of local governments' behavioral responses to equalization policies across their full choice sets.

My outcomes of interest are most relevant for suburban jurisdictions and not for the urban core of metropolitan areas.[29] Lot size regulations are also only one of several zoning regulations used by zoning authorities: In practice, developers must also satisfy parking requirements for multifamily buildings (Shoup (2017)), subject architectural designs to environmental impact statements (Mangin (2014)) and more.

The way my results speak to a variety of urban environments is quantifying which motives matter for local authorities who had maintained land use controls. During the postwar decades when zoning became a standard U.S. government practice, I argue authorities prioritize land use to preserve endogenous racial amenities.

Though the homes constrained by local discretion persist, social attitudes on race relations have improved compared to the past. Recent policy proposals, like California's Senate Bills 9 and 10 in 2021 or new upzoning regulations in Massachusetts in 2022 (MilNeil (2022)) advocate for "housing abundance" through legislation overriding local density controls in targeted neighborhoods. My results indicate these policies are in part reversing planning decisions attributable to prejudice, and suggest reforming local discretion today have both equity gains without disincentivizing public good provision.

Policy reforms grow in urgency as America anticipates higher internal mobility. In recent years, households from the largest cities were on the move in the wake of remote work (Ramani and Bloom (2021)). Historical data inform that climate change in later decades could increase migration out of areas with disaster risk (Boustan, Kahn and Rhode (2012), Hornbeck (2022)). I show that local discretion on land use in the past worked to preserve demographics over a growing urban area's housing needs. Insofar as the same motives are in play today, state and federal decision makers should consider how to align local governments' incentives away from restrictive land use to ensure growing cities remain integrated and equitable.

[29] In the urban core, zoning parameters are based much less on minimum lot sizes and more on floor-to-area ratios (FAR), as investigated by Brueckner and Singh (2020) and Anagol, Ferreira and Rexer (2021).

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2022. "When Should You Adjust Standard Errors for Clustering?*." *The Quarterly Journal of Economics*, qjac038.

Alesina, Alberto, Reza Baqir, and Caroline Hoxby. 2004. "Political Jurisdictions in Heterogeneous Communities." *Journal of Political Economy*, 112(2): 348–396.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *The Quarterly Journal of Economics*, 114(4): 1243–1284.

Anagol, Santosh, Benjamin B Lockwood, Allan Davids, and Tarun Ramadorai. 2022. "Diffuse Bunching with Frictions: Theory and Estimation."

Anagol, Santosh, Fernando V. Ferreira, and Jonah M. Rexer. 2021. "Estimating the Economic Value of Zoning Reform."

Angrist, Joshua, Peter Hull, Parag A Pathak, and Christopher R Walters. 2022. "Race and the Mismeasure of School Quality."

Babcock, Richard F. 1966. *The Zoning Game: Municipal Practices and Policies.* University of Wisconsin Press.

Babcock, Richard F., and Fred P. Bosselman. 1973. *Exclusionary Zoning. Praeger Special Studies in U.S. Economic, Social and Political Issues*, Praeger.

Bai, Jushan, and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66(1): 47.

Bai, Jushan, and Pierre Perron. 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics*, 18(1): 1–22.

Baum-Snow, Nathaniel. 2020. "Urban Transport Expansions and Changes in the Spatial Structure of U.S. Cities: Implications for Productivity and Welfare." *The Review of Economics and Statistics*, 102(5): 929–945.

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

Bayer, Patrick, Marcus D. Casey, W. Ben McCartney, John Orellana-Li, and Calvin S. Zhang. 2022. "Distinguishing Causes of Neighborhood Racial Change: A Nearest Neighbor Design."

Berkes, Enrico, Ezra Karger, and Peter Nencka. 2022. "The Census Place Project: A Method for Geolocating Unstructured Place Names."

Bogart, William T. 1993. "'What Big Teeth You Have!': Identifying the Motivations for Exclusionary Zoning." *Urban Studies*, 30(10): 1669–1681.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *The Review of Economic Studies*, 89(1): 181–213.

Boustan, Leah. 2016. "Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets." National Bureau of Economic Research, Inc NBER Books.

Boustan, Leah Platt. 2010. "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration*." *The Quarterly Journal of Economics*, 125(1): 417–443.

Boustan, Leah Platt, Matthew E. Kahn, and Paul W. Rhode. 2012. "Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century." *American Economic Review*, 102(3): 238–244.

Bowles, Gladys K., James D. Tarver, Calvin L. Beale, and Everette S. Lee. 2016. "Net Migration of the Population by Age, Sex, and Race, 1950-1970."

Broderick, Tamara, Ryan Giordano, and Rachael Meager. 2021. "An Automatic Finite-Sample Robustness Metric: When Can Dropping a Little Data Make a Big Difference?"

Bronin, Sara C., and Ilya Ilyankou. 2021. "How to Make a Zoning Atlas: A Methodology for Translating and Standardizing District-Specific Regulations."

Brueckner, Jan K. 1987. "The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model." In *Handbook of Regional and Urban Economics*. Vol. 2, 821–845. Elsevier.

Brueckner, Jan K., and Ruchi Singh. 2020. "Stringency of Land-Use Regulation: Building Heights in US Cities." *Journal of Urban Economics*, 116: 103239.

Bucholtz, Shawn, Emily Molfino, and Jed Kolko. 2020. "The Urbanization Perceptions Small Area Index: An Application of Machine Learning and Small Area Estimation to Household Survey Data."

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner. 2006. "Causes of Sprawl: A Portrait from Space*." *The Quarterly Journal of Economics*, 121(2): 587–633.

Caetano, Gregorio, and Vikram Maheshri. 2017. "School Segregation and the Identification of Tipping Behavior." *Journal of Public Economics*, 148(C): 115–135.

Caetano, Gregorio, and Vikram Maheshri. 2022. "Explaining Recent Trends in US School Segregation." *Journal of Labor Economics*.

Calabrese, Stephen, Dennis Epple, and Richard Romano. 2007. "On the Political Economy of Zoning." *Journal of Public Economics*, 91(1): 25–49.

Calabrese, Stephen M., Dennis N. Epple, and Richard E. Romano. 2012. "Inefficiencies from Metropolitan Political and Fiscal Decentralization: Failures of Tiebout Competition." *Review of Economic Studies*, 79(3): 1081–1111.

Card, David, Alexandre Mas, and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics*, 123(1): 177–218.

Collins, William J. 2021. "The Great Migration of Black Americans from the US South: A Guide and Interpretation." *Explorations in Economic History*, 80: 101382.

Committee on Civil Rights. 1947. *To Secure These Rights: The Report of the President's Committee on Civil Rights.* Washington, D.C.:U.S. Government Printing Office.

Congress, American Jewish. 1952. *Civil Rights in the United States in 1951: A Balance Sheet of Group Relations.* National Association for the Advancement of Colored People.

Cook, Lisa D, Maggie E C Jones, Trevon D Logan, and David Rosé. 2022. "The Evolution of Access to Public Accommodations in the United States*." *The Quarterly Journal of Economics*, qjac035.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, 107(3): 455–506.

Derenoncourt, Ellora. 2022. "Can You Move to Opportunity? Evidence from the Great Migration." *American Economic Review*, 112(2): 369–408.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *American Economic Review*, 106(3): 479–524.

Ditzen, Jan, Yiannis Karavias, and Joakin Westerlund. 2021. "Xtbreak: Estimating and Testing for Structural Breaks in Stata". Working Paper,."

Duranton, Gilles, and Diego Puga. 2019. "Urban Growth and Its Aggregate Implications." National Bureau of Economic Research w26591, Cambridge, MA.

Fischel, William A. 2015. *Zoning Rules! : The Economics of Land Use Regulation.* Lincoln Institute of Land Policy.

Ganong, Peter, and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the U.S. Declined?" *Journal of Urban Economics*, 102: 76–90.

Glaeser, Edward L. 2008. *Cities, Agglomeration, and Spatial Equilibrium.* Oxford University Press.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review*, 110(8): 2586–2624.

Gregory, James N. 2005. *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America.* Chapel Hill ::University of North Carolina Press,.

Gyourko, Joseph. 1991. "Impact Fees, Exclusionary Zoning, and the Density of New Development." *Journal of Urban Economics*, 30(2): 242–256.

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

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

Hamilton, Bruce W. 1975. "Zoning and Property Taxation in a System of Local Governments." *Urban Studies*, 12(2): 205–211.

Hilber, Christian A. L., and Frédéric Robert-Nicoud. 2013. "On the Origins of Land Use Regulations: Theory and Evidence from US Metro Areas." *Journal of Urban Economics*, 75: 29–43.

Hornbeck, Richard. 2022. "Dust Bowl Migrants: Environmental Refugees and Economic Adaptation."

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

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

Klimek, Amanda, Christopher Mazur, William Chapin, and Ellen Wilson. 2018. "Housing Administrative Records Simulation."

Krimmel, Jacob. 2022. "Reclaiming Local Control: School Finance Reforms and Housing Supply Restrictions."

Kulka, Amrita, Aradhya Sood, and Nicholas Chiumenti. 2022. "How to Increase Housing Affordability? Understanding Local Deterrents to Building Multifamily Housing."

Lee, Sanghoon, and Jeffrey Lin. 2018. "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income." *The Review of Economic Studies*, 85(1): 663–694.

Lee, Sun Kyoung. 2022. "When Cities Grow: Urban Planning and Segregation in the Prewar US."

Leys, Christophe, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. 2013. "Detecting Outliers: Do Not Use Standard Deviation around the Mean, Use Absolute Deviation around the Median." *Journal of Experimental Social Psychology*, 49(4): 764–766.

Mangin, John. 2014. "The New Exclusionary Zoning." *Stanford Law & Policy Review*, 25(91).

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. *IPUMS National Historical Geographic Information System: Version 16.0 [Dataset].* Minneapolis, MN:IPUMS.

Meyer, Bruce D., and Nikolas Mittag. 2017. "Misclassification in Binary Choice Models." *Journal of Econometrics*, 200(2): 295–311.

Mills, Edwin S., Dennis Epple, and Jacob L. Vigdor. 2006. "Sprawl and Jurisdictional Fragmentation [with Comments]." *Brookings-Wharton Papers on Urban Affairs*, 231–256.

MilNeil, Christian. 2022. "New State Rule Would Force Suburbs to Legalize Thousands of New Apartments Near T Stops."

Murray, Pauli. 1950. *States' Laws on Race and Color.* [Cincinnati,:Woman's Division of Christian Service, Board of Missions and Church Extension, Methodist Church].

Nechamkin, Emma, and Graham MacDonald. 2019. "Predicting Zoned Density Using Property Records."

Parkhomenko, Andrii. 2020. "Local Causes and Aggregate Implications of Land Use Regulation."

Ramani, Arjun, and Nicholas Bloom. 2021. "The Donut Effect of Covid-19 on Cities."

Reny, Tyler T., and Benjamin J. Newman. 2018. "Protecting the Right to Discriminate: The Second Great Migration and Racial Threat in the American West." *American Political Science Review*, 112(4): 1104–1110.

Report of the National Advisory Commission on Civil Disorders. 1968. *Report of the National Advisory Commission on Civil Disorders.* United States, Kerner Commission : U.S. G.P.O.,.

Reynolds, Conor Dwyer. 2019. "The Motives for Exclusionary Zoning."

Rolleston, Barbara Sherman. 1987. "Determinants of Restrictive Suburban Zoning: An Empirical Analysis." *Journal of Urban Economics*, 21(1): 1–21.

Ross, Stephen L., and Geoffrey M. B. Tootell. 2004. "Redlining, the Community Reinvestment Act, and Private Mortgage Insurance." *Journal of Urban Economics*, 55(2): 278–297.

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America.* . First edition. ed., W.W. Norton.

Rubin, Donald B. 1981. "The Bayesian Bootstrap." *The Annals of Statistics*, 9(1): 130–134.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. 2021. *IPUMS Ancestry Full Count Data: Version 3.0 [Dataset].* Minneapolis, MN:IPUMS.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. 2022. *IPUMS USA: Version 12.0 [Dataset].* Minneapolis, MN:IPUMS.

Saavedra, Martin, and Tate Twinam. 2020. "A Machine Learning Approach to Improving Occupational Income Scores." *Explorations in Economic History*, 75: 101304.

Sahn, Alexander. 2022. "Racial Diversity and Exclusionary Zoning: Evidence from the Great Migration."

Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *The Journal of Mathematical Sociology*, 1(2): 143–186.

Schwab, Robert M., and Wallace Oates. 1991. “Community Composition and the Provision of Local Public Goods: A Normative Analysis.” *Journal of Public Economics*, 44(2): 217–237.

Shanks, Brendan. 2021. “Land Use Regulations and Housing Development: Evidence from Tax Parcels and Zoning Bylaws in Massachusetts.”

Shertzer, Allison, and Randall P. Walsh. 2019. “Racial Sorting and the Emergence of Segregation in American Cities.” *The Review of Economics and Statistics*, 101(3): 415–427.

Shi, Ying, Daniel Hartley, Bhash Mazumder, and Aastha Rajan. 2022. “The Effects of the Great Migration on Urban Renewal.” *Journal of Public Economics*, 209: 104647.

Shoup, Donald C. 2017. *The High Cost of Free Parking: Updated Edition.* Taylor & Francis.

Song, Jaehee. 2021. “The Effects of Residential Zoning in U.S. Housing Markets.”

Sood, Aradhya, William Speagle, and Kevin Ehrman-Solberg. 2019. “Long Shadow of Racial Discrimination: Evidence from Housing Covenants of Minneapolis.” *SSRN Electronic Journal*.

Stuart, Bryan A., and Evan J. Taylor. 2021. “Migration Networks and Location Decisions: Evidence from US Mass Migration.” *American Economic Journal: Applied Economics*, 13(3): 134–175.

Sullivan, Patricia. 2009. *Lift Every Voice: The NAACP and the Making of the Civil Rights Movement.* New York:New Press : Distributed by Perseus Distribution.

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

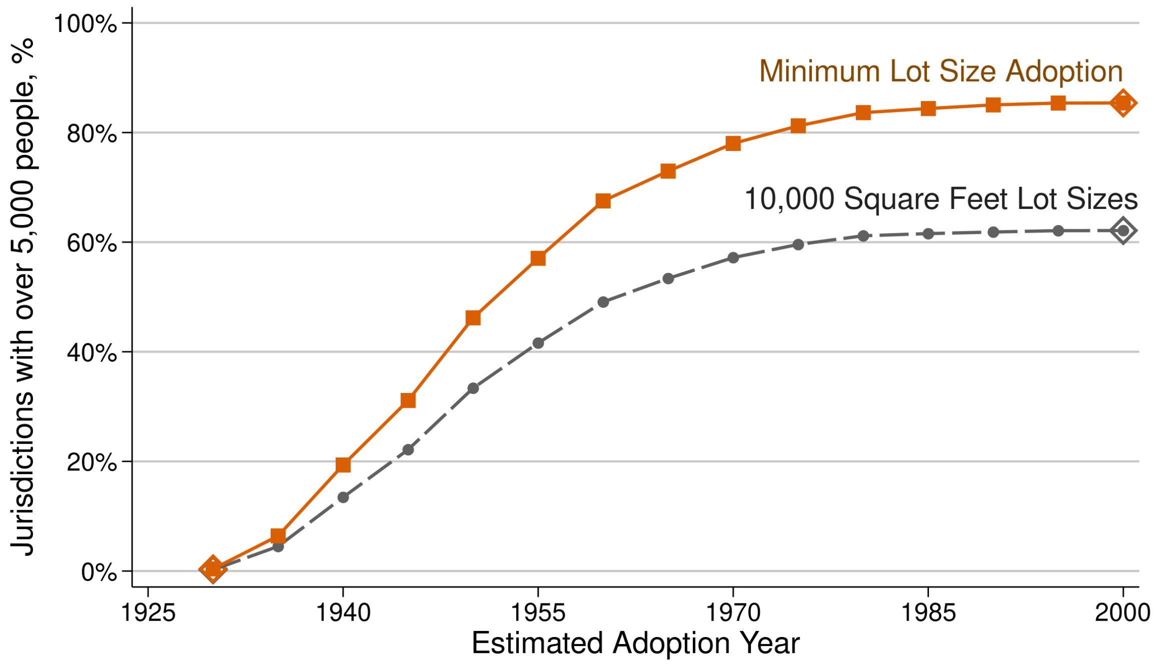
Turner, Matthew A., Andrew Haughwout, and Wilbert van der Klaauw. 2014. “Land Use Regulation and Welfare.” *Econometrica*, 82(4): 1341–1403.

Winling, LaDale C, and Todd M Michney. 2021. “The Roots of Redlining: Academic, Governmental, and Professional Networks in the Making of the New Deal Lending Regime.” *Journal of American History*, 108(1): 42–69.

Zabel, Jeffrey, and Maurice Dalton. 2011. “The Impact of Minimum Lot Size Regulations on House Prices in Eastern Massachusetts.” *Regional Science and Urban Economics*, 41(6): 571–583.

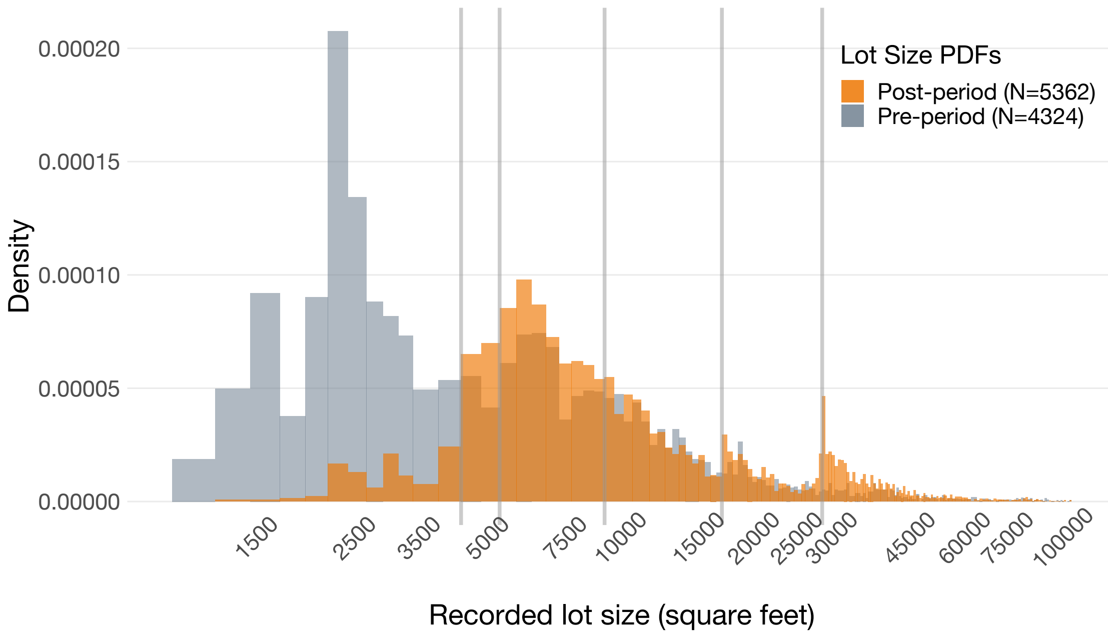
Zodrow, George R. 2007. “The Property Tax Incidence Debate and the Mix of State and Local Finance of Local Public Expenditures.” *CESifo Economic Studies*, 53(4): 495–521.

Figure 1: Estimated Trends in Minimum Lot Size Adoption



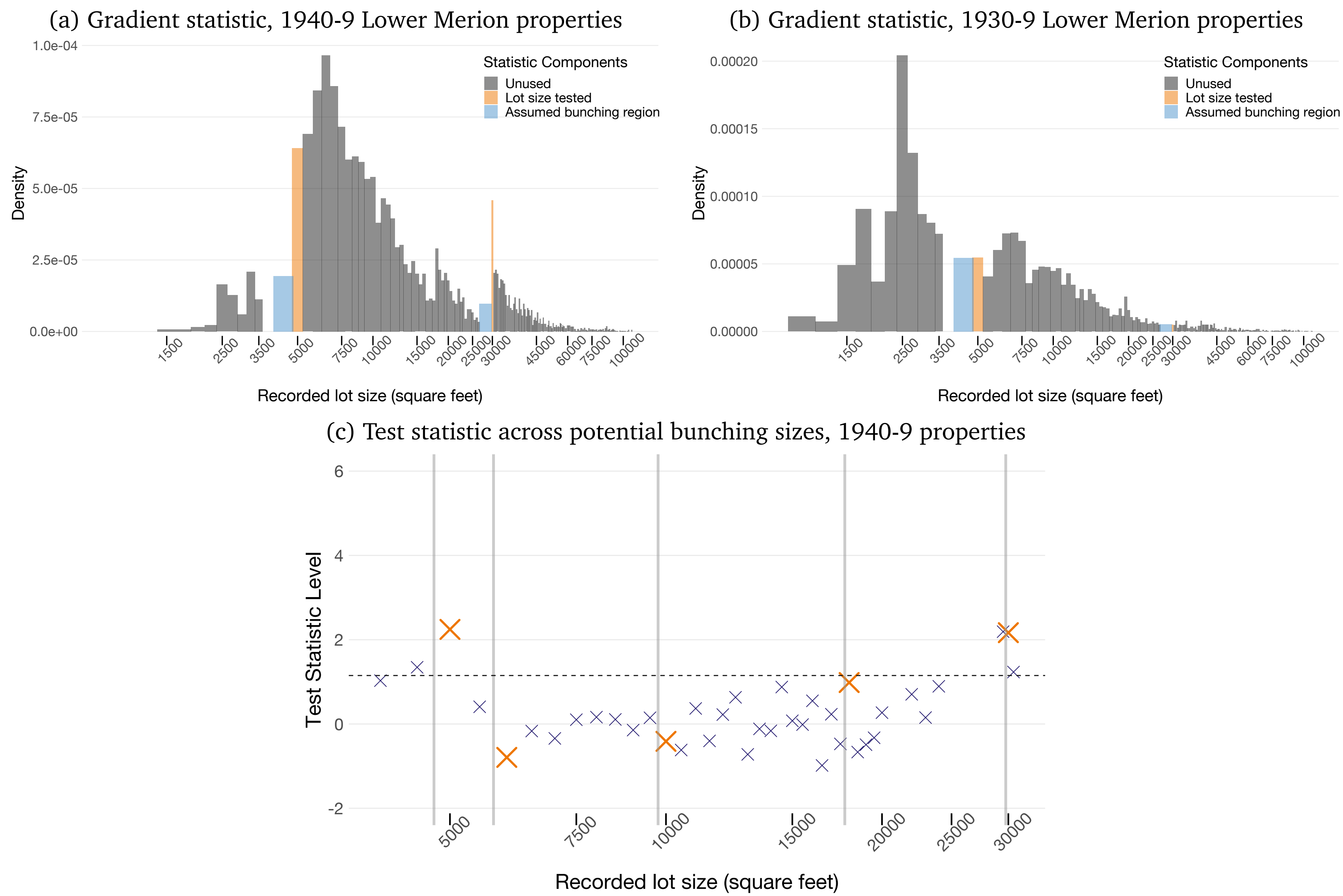
Notes: This figure plots the first adoption estimates defined in Section 3.1 based on the lot size detection algorithm in Section 2. Adoption rates are calculated over all cities and counties with zoning powers with at least 5,000 residents as of the 2010 census. The main series considers adoption of any minimum lot size, while the dashed series considers adoption of large minimum lots greater than 10,000 square feet ($\sim 1/4$ acre). Displayed rates are based on values at the end of five-year time bins. Diamond markers note the start and end of the time series in 1930 and 2000, respectively.
Sources: Calculations from CoreLogic Tax Records, 2010 Census.

Figure 2: Lower Merion: Case Study for Minimum Lot Regulation



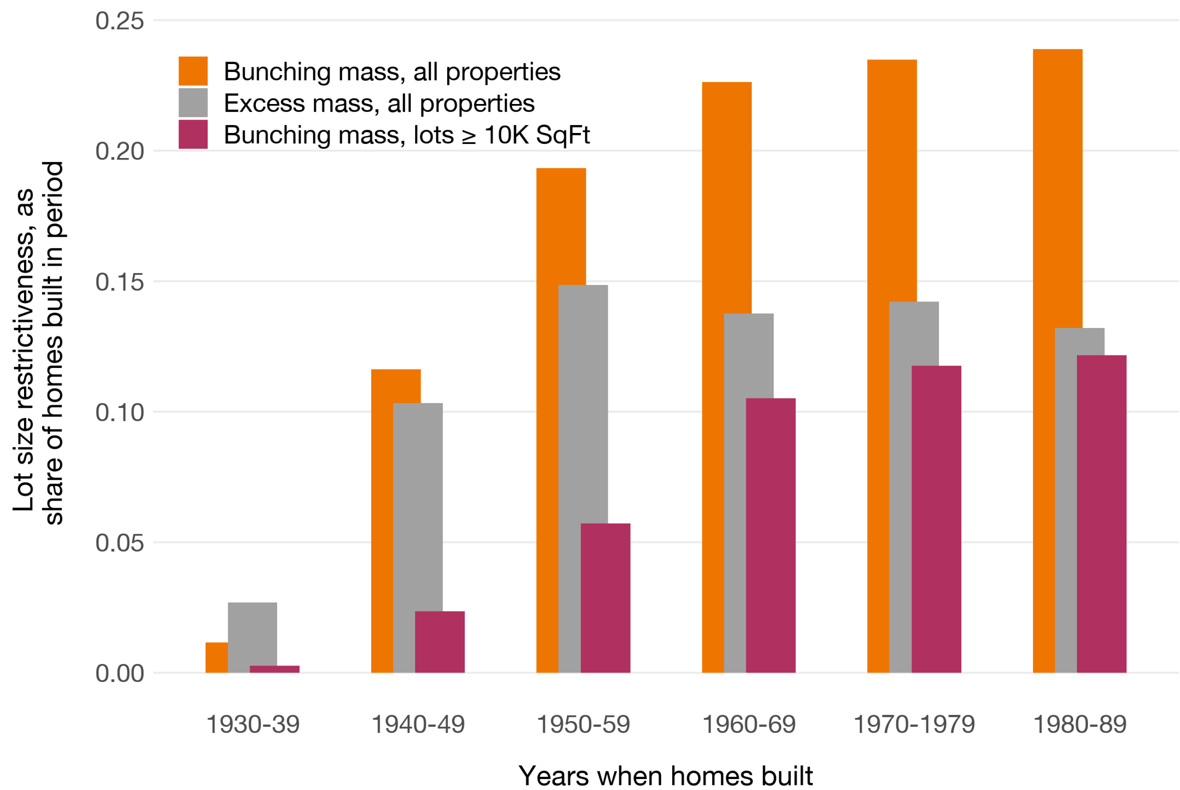
Notes: Two histograms of properties in Lower Merion Township are shown, with the five minimum lot size zones enacted in the 1939 zoning ordinance overlaid as gray lines. The gray "pre-period" sample contains all properties built in the years between 1920 to 1939. The orange "post-period" sample contains all properties built in the years between 1940 to 1959.
Sources: CoreLogic Tax Records, Lower Merion Township, Pennsylvania (1939)

Figure 3: Visualization of Locally Estimated Bunching Model



Notes: This figure visualizes the construction of the gradient statistic $\hat{G}$, defined in Section 2.2, in steps. For lot ranges around 2 actual lot size regulations, Panel (a) highlights two histogram density estimates that go into calculating the two terms of the statistic defined on the post-period sample. Panel (b) does the same but for the remaining two terms defined on the pre-period sample.
Panel (c) takes the calculated $\hat{G}$ and displays it for every lot range for which the statistic is defined. Lot ranges at which there are actual lot size regulations are highlighted in orange. Values are compared to the critical value, defined as the 90th percentile of the Chi-squared distribution with one degree of freedom.
Sources: Calculations from CoreLogic Tax Records.

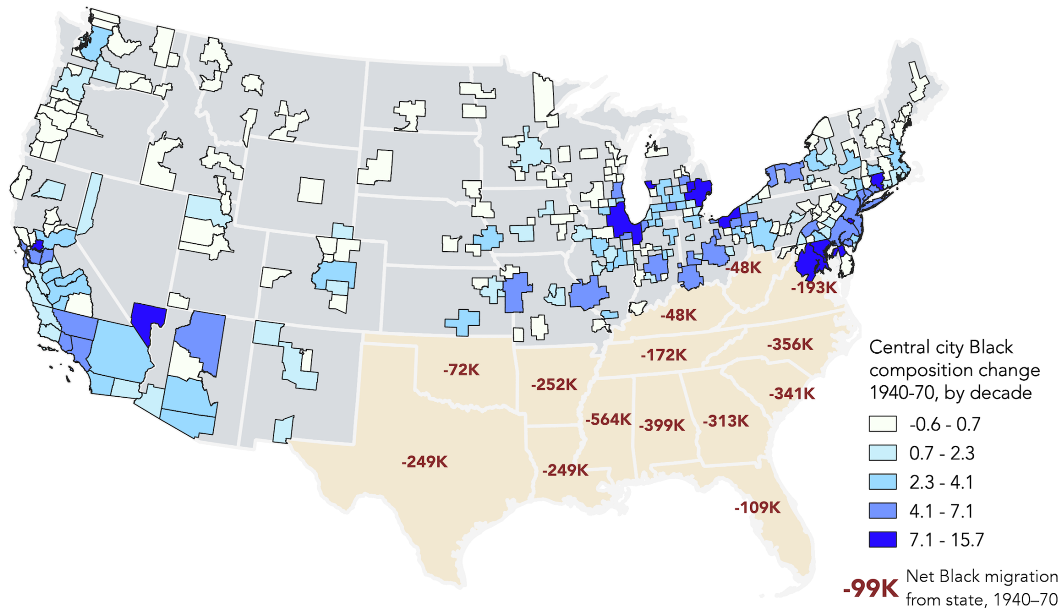
Figure 4: Estimated Trends in Lot Size Restrictiveness



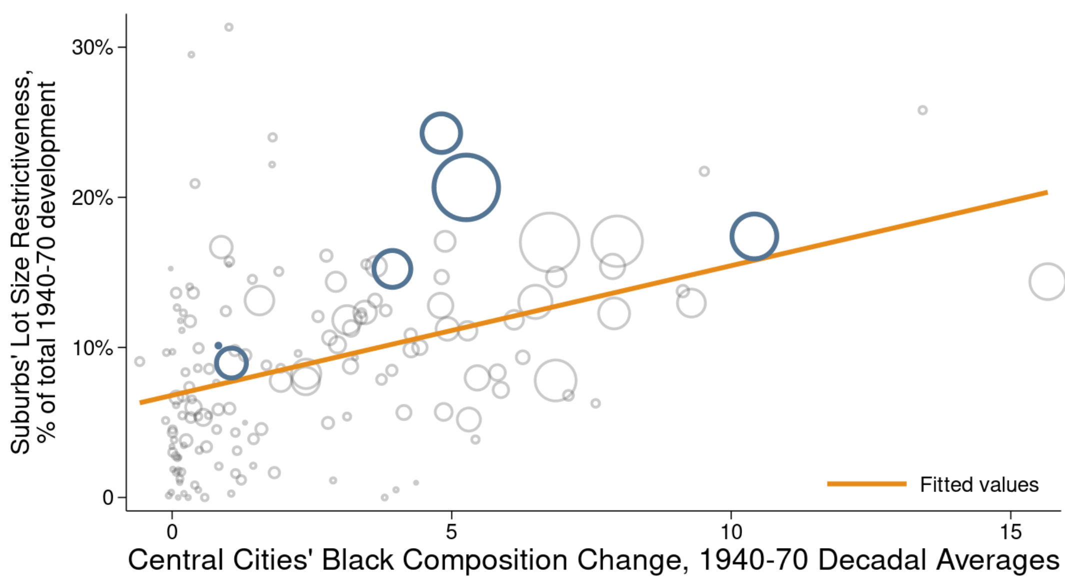
Notes: This figure plots how three measures of lot size restrictiveness, described in Section 3.2, change across homes built in different decades. Over a decadal vintage of single-family and duplex homes, the bunching mass and excess mass measures over all properties sums up the respecive measures over all detected bunching bins. The bunching mass measure for lots over 10,000 square feet sums up mass around bunching bins over 10,000 square feet, then express it as a ratio of all properties that decade.
Sources: Calculations from CoreLogic Tax Records.

Figure 5: The Geography of Postwar Minimum Lot Sizes and Black Migration

(a) Black demographic change during the Second Great Migration

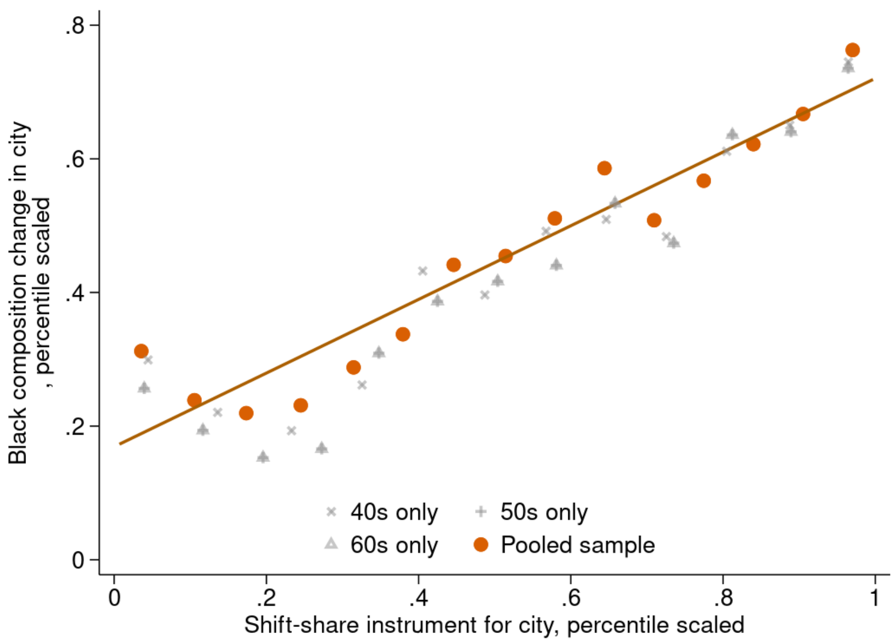


(b) Black migration associates with lot size restrictiveness



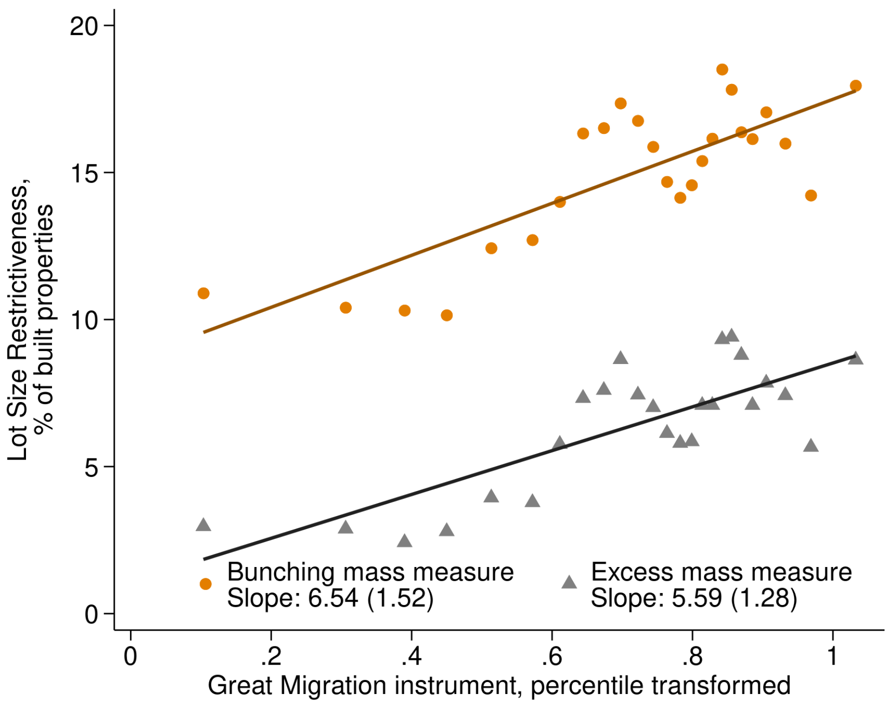
Notes: Panel (a) of this figure plots the cumulated migration out of 14 Southern states, as well as the variation across CBSA central cities in Black migration growth. Each non-Southern CBSA is marked by its central city's Black composition change variable ΔCC_{ct}^{Black} as defined in Section 5.1, transformed to the percentile in the decade distribution and averaged over 1940–70. State-level outmigration estimates are based off of methods in Gregory (2005) and are further explained in Appendix Section A.6.
Panel (b) of this figure aggregates the excess mass measure of lot size restrictiveness, as defined in Section 3.2, over 1940–70 and all detected jurisdictions in CBSA borders. Aggregation is done only for CBSAs where at least 10 jurisdictions, or the metropolitan area's major city and counties, have detected minimum lot sizes.
Sources: Calculations from NHGIS Tables (Manson et al. (2021)), Ruggles et al. (2022), Gregory (2005) and CoreLogic Tax Records.

Figure 6: First Stage Relationship on Central City Black Composition



Notes: This figure plots nonparametric relationships between the shift-share instrument described in Section 5.3 and the endogenous variable of central city Black composition change. The source data is a panel of 253 central cities outside of 14 Southern states, with demographic changes over three decades of Census reports, 1940–1970. The outcomes are first transformed from levels to percentiles of each decade's distribution, as in Derenoncourt (2022), then residualized on share exposure variables as described in Section 5.3.
Sources: Calculations from NHGIS Tables (Manson et al. (2021)), Ruggles et al. (2022), CCDB, IPUMS 1940 full count Census (Ruggles et al. (2021)), Boustan (2016), Derenoncourt (2022) and CoreLogic Tax Records.

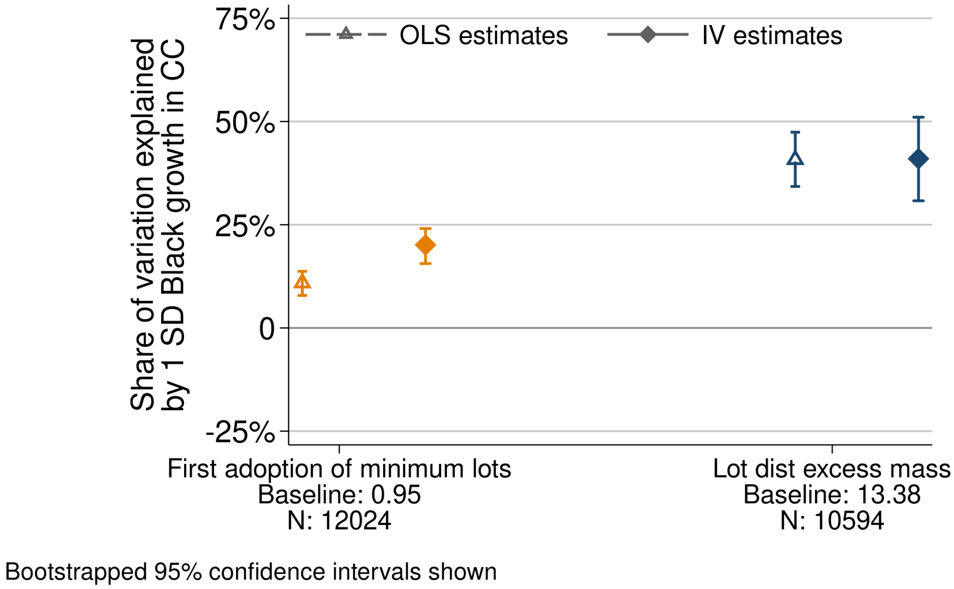
Figure 7: Lot Size Restrictiveness Linked with Black Composition Change



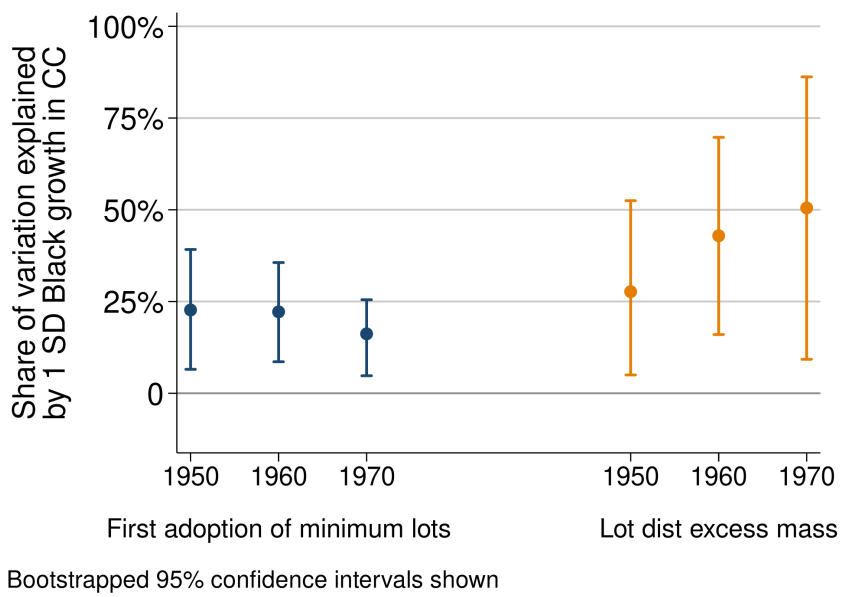
Notes: This figure plots reduced form, nonparametric relationships between the shift-share instrument described in Section 5.3 and two outcomes of interest: measures of lot size restrictiveness from 1940–1970. The source data is a panel of non-central city jurisdictions in CBSAs outside of 14 Southern states. The instrument is first transformed from levels to percentiles of each decade's distribution, as in Derenoncourt (2022), then residualized on share exposure variables as described in Section 5.3 and additional controls. Control variables include the CBSA central city's manufacturing share, and analysis sample cities' 1940 black share, homeownership rates and distance to CBD, interacted by period. Reported standard errors are clustered at the CBSA-decade level.
Sources: Calculations from NHGIS Tables (Manson et al. (2021)), Ruggles et al. (2022), CCDB, IPUMS 1940 full count Census (Ruggles et al. (2021)), Boustan (2016), Derenoncourt (2022) and CoreLogic Tax Records.

Figure 8: Calculating The Impact of Racial Exclusionary Zoning

(a) Lot Size Outcomes Explained by Black Migration, Constant Effects

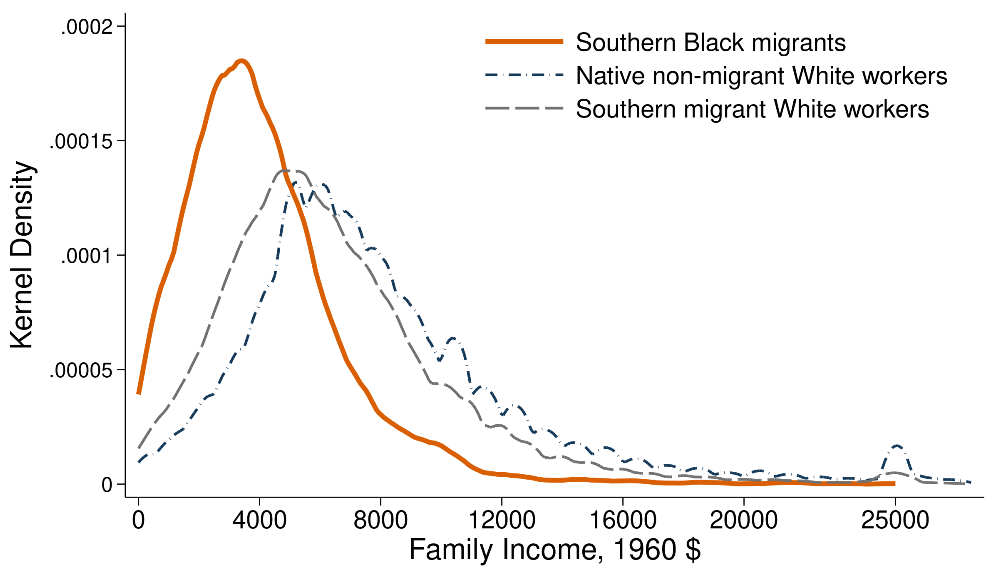


(a) Outcome Explained by Black Migration, Decade Varying Effects



Notes: This figure presents an aggregation exercise, converting the regression coefficients estimated in Table 5 into how much the Second Great Migration explained lot size outcomes in non-Southern metropolitan areas. The aggregation is a three-step procedure detailed in Section 6.2. Both panels plot two outcomes: a binary variable for lot size adoption and a ratio for lot size restrictiveness. Panel (a) assumes constant effects across the 1940–70 analysis period, pooling Black migration over three decades to estimate the share of lot size outcomes explained by the Great Migration. Panel (b) estimates a separate causal effect of Black migration for each decade, then conducts the conversion to share of outcome explained for each decade by itself. 95% confidence intervals are bootstrapped using random weights on CBSA-decade clusters, following the Bayesian bootstrap of Rubin (1981). *Sources:* Calculations from NHGIS Tables (Manson et al. (2021)), Ruggles et al. (2022), CCDB, IPUMS 1940 full count Census (Ruggles et al. (2021)), Boustan (2016), Derenoncourt (2022) and CoreLogic Tax Records.

Figure 9: Household Incomes Within and Between Racial Groups



Notes: This figure plots household income distributions in 1960 across three racial and demographic groups, for Northern central cities where the data are available. Household income is defined as self-reported family income in the 1960 Census and is not adjusted for inflation. Southern migrant workers are defined as any worker who reports having moved from one of 14 Southern states in the last 5 years.
Sources: Calculations from IPUMS, 1960 Census 5% sample (Ruggles et al. (2022).

Table 1: Evaluation of Algorithm on Training and Test Data

	N	Mean	Median	MAE	Pop-weighted MAE
Panel A: Training Set					
Difference in first large lot adoption year	55	0.17	0	10.9	9.89
Difference in first adoption year	76	-0.30	0	9.74	12.0
Panel B: Test Set					
Difference in first large lot adoption year	15	-2.10	-3.0	8.14	5.43
Difference in first adoption year	32	-2.48	-3.5	10.6	5.68
Panel C: Actual Minimum Lot Sizes Matched to Estimates					
Log difference between matched lot sizes	587	-0.195	0	0.42	0.47

Notes: This table reports metrics for model evaluation over the three outcomes matched through optimizing tuning parameters. The mean and median statistics measure unbiasedness. Both are taken over differences between estimated values and values coded in historical zoning records. The mean absolute error (MAE) terms measure fit. The population-weighted MAE uses 2010 population weights on each jurisdiction when calculated the weighted mean of absolute errors. The sample size varies between outcomes, as not all historical zoning records record all three outcomes.

Table 2: Coverage of Jurisdictions with Estimated Adoption

	N	% All Jurisdictions	Population-weighted %
Incorporated Cities	4270	49%	92%
Counties With Zoning Power	658	72%	87%
Townships With Zoning Power	2225	41%	74%
Newly Annexed Portions of Cities	1269	—	—
Distinct Jurisdictions	7153		

Notes: This table reports type-specific counts of all zoning jurisdictions over which a lot size adoption is estimated. The second column lists the estimation rate over the universe of all such jurisdictions (e.g. all counties making up CBSAs in the states counties have zoning power). The third column recalculates the estimation rate but weighs different jurisdictions by their 2010 Census population.

Table 3: Summary Statistics

Variable Name	Analysis Sample				Full Data		
	N	Juri.	Mean	SD	Juri.	Mean	SD
Panel A: Outcome variables							
Minimum lots adopted around end of decade	15105	5069	0.773	0.419	8059	0.715	0.451
Large lot minimums around end of decade	15105	5069	0.483	0.500	8059	0.464	0.499
Units of housing in decade-long vintage	15105	5069	718.4	2250.9	8059	833.8	3064.0
Bunching share measure, lot size restrictiveness	12902	5069	14.59	14.77	8059	13.55	14.47
Excess mass measure, lot size restrictiveness	12902	5069	6.118	10.24	8059	5.524	9.604
Panel B: Measures of demographic change							
End of decade population, 000s	654	223	146.2	360.7	384	127.1	299.1
CC Black composition change	654	223	1.984	3.032	384	2.823	4.823
GM shift-share, black migrants	654	223	0.138	0.260	384	0.808	1.581
Quantile-transformed Black composition change	654	223	0.450	0.263	384	0.503	0.290
1940 worker share in manufacturing	652	223	0.218	0.0647	384	0.212	0.0653
1940 Black household head share	654	223	0.0236	0.0374	384	0.114	0.149
Panel C: Pre-period (1940) demographic variables							
Census household count		3542	2076.2	5016.4	5185	2062.4	4553.7
Black household head share		4237	0.016	0.047	6598	0.074	0.143
Distance from central city CBD, km		4546	29.84	19.31	6745	31.11	19.89
Household homeownership rate, %		4237	58.23	10.81	6598	54.32	13.05
Median household income, est. 1950 000$		4237	2.635	0.454	6598	2.364	0.635
Workers in white-collar industries		4237	0.375	0.133	6598	0.353	0.140
Workers in agricultural industries		4237	0.207	0.200	6598	0.247	0.225

Notes: This table reports summary statistics of all variables used in the regression specification of Section 6.1, plus additional demographic variables for context. I report the number of observations over the panel dataset, as well as the number of time-invarying jurisdictions that are included in the data. For variables defined at a central city level, observations are counted at a central city–time period level.

Table 4: First Stage Estimation Results

	(1)	(2)	(3)
Percentile of $Z^{Black}_{c(j)t}$	0.754	0.680	0.694
	(0.0345)	(0.0519)	(0.0467)
Panel N	15759	15759	12668
First-stage F	478.0	171.3	220.6
R^2	0.591	0.638	0.657
Region–Decade FE	X	X	X
Migration exposure shares		X	X
Pre-period controls			X

Notes: This table reports results from the first stage regression of the shift-share Black migration instrument in Section 5.3 on central city Black composition change,

$$\Delta CC^{black}_{c(j)t} = \gamma Z^{Black}_{c(j)t} + \delta_{r(j)t} + \mathbf{X}_{c(j),pre}\Gamma + \nu_{c(j),t},$$

at the jurisdiction-decade level. For each decade, both the outcome and instrument are rescaled based on their quantile in the between-city distribution. Across specifications, I add Census region–decade fixed effects and a central city's 1940 exposure to Black migrants over different Southern states, constructing the shares as described in Section 5.4. I also control for decade-specific trends based on cities' 1940 rates of Black households, distance from the CBD, homeownership rates and central city manufacturing worker share. Standard errors and first stage F-statistics are calculated clustering at the CBSA–decade level.

Table 5: Black Demographic Change Causes Lot Size Controls

	OLS		IV		
	(1)	(2)	(1)	(2)	(3)
Panel A: Effects on Lot Size Regulation Adoption					
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.162***	0.152***	0.272***	0.203***	0.280***
	(0.0276)	(0.0312)	(0.0379)	(0.0435)	(0.0513)
Panel *N*	15105	12024	15105	15105	12024
Baseline mean	0.770	0.782	0.770	0.770	0.782
R^2	0.0841	0.133	0.0801	0.0938	0.130
Panel B: Effects on Excess Mass Restrictiveness					
Percentile of $\Delta CC^{Black}_{c(j)t}$	8.008***	7.905***	8.568***	7.757***	7.948***
	(0.986)	(0.984)	(1.370)	(1.634)	(1.660)
Panel *N*	12902	10594	12902	12902	10594
Baseline mean	6.102	6.347	6.102	6.102	6.347
R^2	0.0658	0.101	0.0657	0.0790	0.101
Region–Decade FE	X	X	X	X	X
Migration exposure shares		X		X	X
Pre-period controls		X			X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents regressions of central city Black composition change, as defined in Section 5.1, to lot size outcomes defined in Section 3,

$$Reg_{jt} = \beta \Delta CC^{Black}_{c(j),t} + \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t},$$

using both OLS specifications and instrumenting for Black composition change with shift-share Black migration instruments. In addition, Black composition changes and the migration instruments are normalized to their quantile within each decade's distributions. Across specifications, I add Census region–decade fixed effects and a central city's 1940 exposure to Black migrants over different Southern states, constructing the shares as described in Section 5.4. I also control for decade-specific trends based on cities' 1940 rates of Black households, distance from the CBD, homeownership rates and central city manufacturing worker share. Standard errors are calculated clustering at the CBSA-decade level.

Table 6: Effects of Demographic Changes on Lot Size Restrictiveness

	Δ Black		Δ South. White		Δ Foreigner
	Reduce.	IV	Reduce.	IV	OLS
Percentile of $\Delta CC^{Dem}_{c(j)t}$	5.586***	7.948***	-2.779**	-5.045	-4.904***
	(1.279)	(1.660)	(0.934)	(2.383)	(1.427)
Panel N	10632	10594	10632	8733	10632
Baseline mean	0.698	0.647	0.421	0.364	0.397
R^2	0.0896	0.101	0.0451	0.0289	0.0544
Region–Decade FE	X	X			X
Decade FE			X	X	
Migration exposure shares		X			X
Pre-period controls		X		X	X

Significance levels: $^{*} = 5\%$; $^{**} = 1\%$; $^{***} = 0.1\%$.

Notes: This table presents regressions of central city composition change on lot size restrictiveness, defined in Section 3.2, across measures of composition change for multiple demographic groups,

$$Excess_{jt} = \beta \Delta CC^{Dem}_{c(j),t} + \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}.$$

Columns (1) and (2) repeat reduced form and instrument variable estimates of Black composition change on lot size restrictiveness. Columns (3) and (4) show specifications using Southern white demographic change in central cities, defined by and instrumented with shift-share instruments as defined in Section 6.3. Column (5) define central city composition change on foreign-born White residents in central cities analogously to what I do for Black residents. In addition, I add Census region–decade fixed effects and normalize demographic composition change variables and the migration instruments to their percentile within each decade's distributions. For Black composition change, I add a central city's 1940 exposure to Black migrants over different Southern states, constructing the shares as described in Section 5.4. I also control for decade-specific trends based on cities' 1940 rates of Black households, distance from the CBD, homeownership rates and central city manufacturing worker share. Standard errors are calculated clustering at the CBSA-decade level.

Table 7: Robustness of Migration's Effects on Lot Size Restrictiveness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pctile. $\Delta CC_{c(j)t}^{Black}$ (OLS)	7.905***	7.503***	6.996***	11.79***	7.122***	16.31***	
	(0.984)	(0.956)	(1.085)	(1.291)	(1.129)	(1.848)	
$\Delta CC_{c(j)t}^{Black}$ (OLS)							0.647***
							(0.0788)
Panel N	10594	10594	10513	5591	691	11227	10476
R^2	0.101	0.105	0.106	0.123	0.201		0.104
Pctile. $\Delta CC_{c(j)t}^{Black}$ (IV)	7.948***	8.636***	7.255***	13.06***	13.86***	14.21***	
	(1.660)	(1.497)	(1.919)	(2.265)	(2.494)	(3.099)	
$\Delta CC_{c(j)t}^{Black}$ (IV)							0.351*
							(0.147)
Panel N	10594	10594	10513	5591	691	11227	10476
R^2	0.101	0.104	0.106	0.123	0.159		0.0994
Main specification controls	X	X	X	X		X	X
Southern white IV control		X					
Suburban growth control			X				
Most anomalous bunching				X			
CBSA level regression					X		
Tobit censored regression						X	
No pctile transformation							X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents alternate specifications showing results in Table 5 are robust across regression specifications as well as outcome misspecification. Column (1) displays again the OLS and IV results of Black composition change on lot size restrictiveness,

$$Excess_{jt} = \beta \Delta CC_{c(j),t}^{Black} + \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}.$$

In addition, Black composition changes and the migration instruments are normalized to their percentile within each decade's distributions. Columns (2) and (3) report robustness to changes in right-hand side controls in the specifications. Columns (4) to (6) report robustness to definitions and scale of observation for the lot size restrictiveness outcome. Column (7) reports a specification using levels of Black composition change not normalized by the within-decade percentile function. Details of the specifications are described in Section 6.4. Main specification controls include a central city's 1940 exposure to Black migrants over different Southern states. They also include decade-specific trends based on cities' 1940 rates of Black households, distance from the CBD, homeownership rates and central city manufacturing worker share. Standard errors are calculated clustering at the CBSA-decade level.

Table 8: Migration's Effects Where Anti-Discrimination Education Laws Adopted Early

	Lot Size Adoption			Excess Mass Restrictiveness		
	OLS	IV		OLS	IV	
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.0596	0.135**	0.224***	5.782***	0.818	4.177*
	(0.0369)	(0.0523)	(0.0547)	(1.059)	(1.789)	(1.629)
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.109*	0.136	0.186*	3.541*	7.773**	6.172**
$\times \mathbf{1}$[ADE state]	(0.0526)	(0.0698)	(0.0776)	(1.581)	(2.408)	(2.239)
$\mathbf{1}$[ADE state]	-0.116**	-0.117*	-0.129*	-2.208*	-5.457**	-4.196**
	(0.0449)	(0.0494)	(0.0557)	(1.024)	(1.796)	(1.575)
Panel N	11958	15030	11958	10537	12842	10537
Baseline mean	0.783	0.771	0.783	6.352	6.105	6.352
R^2	0.129	0.0755	0.102	0.0988	0.0496	0.0788
Region–Decade FE	X	X	X	X	X	X
Migration exposure shares	X	X	X	X	X	X
Pre-period controls	X		X	X		X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents heterogeneous effects by 1950 state-level public education policies on the lot size restrictiveness effects in Table 5. I define 16 states that were early adopters of an anti-discrimination law in public education ("ADE state"), as recorded by Murray (1950). I then estimate the interacted the regression model:

$$\begin{aligned} Reg_{jt} = {} & \beta \Delta CC^{Black}_{c(j),t} + \alpha^{ADE} \text{ ADE State}_{s(j)} + \beta^{ADE} \Delta CC^{Black}_{c(j),t} \times \text{ ADE State}_{s(j)} \\ & + \delta_{r(j),t} + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}. \end{aligned}$$

With rescaled Black composition change, rescaled shift-share Black migration instruments and the early adopter dummy. Across specifications, I add Census region–decade fixed effects and a central city's 1940 exposure to Black migrants over different Southern states, constructing the shares as described in Section 5.4. I also control for decade-specific trends based on cities' 1940 rates of Black households, distance from the CBD, homeownership rates and central city manufacturing worker share. Standard errors are calculated clustering at the CBSA-decade level.

For Online Publication

A Details of Data Build

A.1 Primer on Zoning Jurisdictions

In the United States, local government powers to regulate land use vary by state. Every state in the United States gives incorporated cities the right to zone, but states vary in both the extent of unincorporated land as well as treatment of zoning on said land.

34 states, most of which are in the American West and South, give zoning powers to both incorporated cities and counties. Outside of the urban core, urban areas are likely to have been zoned under two jurisdictions as they developed: as part of a county and as part of an incorporated city. Two additional states, Texas and Oklahoma, do not assign zoning powers to counties; zoning applies on unincorporated land only due to state laws.

The six New England states have only nominal powers for county governments, so the zoning power rests with the cities and towns that incorporate all land. 8 other states in the Mid-Atlantic and Midwest issue zoning powers to three tiers of government: incorporated cities, townships and counties. [1] The legal relationship between competing jurisdictions is complex, but in general zoning varies across incorporated cities in more populated areas and county planners are likelier to handle all zoning in rural areas.

My definition of zoning jurisdictions varies by state and incorporate all the above cases. In every state, I consider incorporated places, as defined by the Census Bureau, as zoning jurisdictions. In the 14 states that give zoning powers to cities and townships, I additionally consider townships as having zoning power over land not covered by incorporated places. In all remaining states, land outside of incorporated cities is unincorporated and default to being zoned at the county level.

A.2 Time-Consistent Jurisdiction Boundaries

A zoning jurisdiction, the unit of analysis in my empirical results, includes every incorporated city as of 2010, plus townships and counties that have zoning powers over unincorporated land. For rural townships where few lot records are available, I consolidate all townships into one county and proceed as if lot size controls are implemented by the county.

The County and City Data Books confirm certain central cities of CBSAs annexed land during the postwar years, which means development within modern central city borders are a mix of land use laws applied to the urban core versus in outlying suburbs. To separate out the cases, I fix central cities based on their 1960 boundaries. I use tract data from Baum-Snow (2020) to identify which 1960 tracts are labelled to be within central cities. Census blocks further out than 1960 borders but are within 1980 central city borders are separated out as a separate jurisdiction.

Improved data quality in Census TIGER files means I separate out annexed land between 1980 and 2010 for both central city and non-central cities. In this paper, I consolidate an-

[1] Michigan, Minnesota, New Jersey, New York, North Dakota, Pennsylvania, Ohio and Wisconsin.

nexed territory over this period into the county jurisdiction. Figure A.1 illustrate how these jurisdictions are constructed from 2010 Census Block boundaries.

A.3 CoreLogic Tax Records

My estimation procedure using the CoreLogic data relies on three variables: the zoning jurisdiction it is in, the square footage of a property's lot and year built. This subsection elaborate on how those variables were cleaned.

Matching properties to zoning jurisdictions. Each property in the CoreLogic Tax Records has the county it is in (based off of the assessor office where the data were retrieved) as well as various types of unstructured address data. Both the property's mail address and address in the assessor records give a definition of the local government it falls under. The data are also geocoded by CoreLogic and matched to a Census block.

The city in the mailing address is not always the name of the zoning jurisdiction the property falls under, like when a property is in a township outside of the city or when a property is on unincorporated land. On the other hand, the deed municipality name may be missing or mark that the property is on unincorporated land in idiosyncratic ways.

For each property, I check for local governments that both encompass the territory the property is geocoded to be in and whose name matches with one of the mail or deed municipality names. If the local government is also one of the zoning jurisdictions defined in Section A.1, that is the zoning jurisdiction matched to the property.

From a raw sample of over 100 million properties, 92% of properties are matched to either a valid census tract or block. The low attrition rate suggests only a minority of misspecified records were dropped in this step. I then use the Census TIGER files processed by [NHGIS] to match the tract or block to a zoning jurisdiction encompassing it.

Around 96% of geographically matched properties are validated in one of two ways: either the municipality the property is matched to has a name that is the same as what's given in the CoreLogic address variables, or the municipality is matched to unincorporated territory. In the latter case, I assign them to the county as the property's relevant zoning jurisdiction.

The remaining 4% of properties are in tracts and blocks that are part of incorporated territory but do not have a name match. I assume the geocode is more accurate and assign the property to the city to which it was geographically matched.

Cleaning lot square footage. The vast majority of counties in the Tax Records data has full coverage of lot square footage for their properties. To control for any coding errors that would be conflated as bunching, I trim the top and bottom percentiles of the CBSA-wide lot size distribution. I also verify lot size with another variable in the Tax Records that measure size in units of acres. I drop properties whose lot sizes are measured inconsistently, defined as having significant differences in the square footage and acre variables.

Throughout the 2010s, developers could have torn down older properties and redeveloped or subdivided the lot they were on. I observe these events in the Tax Records by seeing if a parcel with the same identifier saw significant changes in the year property built variable or in the lot size. Wherever possible, I use the oldest assessor record in CoreLogic for a parcel to retrieve year built and square footage data. However, as long as the parcel is consistently identified, I impute other variables from later data waves like the property geocode or building characteristics.

Year built data and data quality. The availability of year built data is less consistent across counties. 85.4% of total properties matched to zoning jurisdictions have a year built measure. The attrition rate is evenly spread across counties. The mean share of properties missing year built across counties is is 75.5%, 85% of counties have year built information for at least half of all properties and 63% of them have the data for more than 80% of all properties. The counties where more than half of all properties lack year built data make up 8.2% of all housing units in my data.

Where data are available, year built may be a poor proxy of when the lot was first developed in two ways: due to recoding of the year built variable following replacement of the housing unit on the lot, or due to mismeasurement of the correct year. I discuss the former issue in more detail in Section A.4, when benchmarking the CoreLogic records to Census data.

On the issue of mismeasurement, the primary issue with the CoreLogic data is that the records feature bunching of the year built dates on round numbers. Figure A.2 visualizes this for two CBSAs. Panel (a) shows a CBSAs that has very limited bunching on years: the Sacramento–Roseville–Arden-Arcade Metropolitan Statistical Area (MSA). The magenta line plots the density of properties in year built over the whole region, with its constituent counties' densities in gray.

Panel (b) visualizes the same objects but for the Philadelphia-Camden-Wilmington MSA. Across most counties, the year built variable has visible bunching on years ending in 0 or 5.

To be conservative, in counties where there is bunching on a year ending in 0, I interpret that as saying the structure was built at some point in that decade (e.g., bunching at 1950 means the properties could have been built from 1950-1959). I detect bunching on years ending in 0 by linearly interpolating an estimate based on neighbouring year built's densities, and then calculating the excess mass of the observed density less the interpolated estimate.

Properties that are estimated to have year built bunched on a 0 are recoded to have been built by the decade's end, rather than by the decade's start. This avoids potential bias in estimating zoning adoption earlier than expected (i.e., ahead of true adoption by 5 years). In Appendix Section B.1, I also discuss how I adapt the structural break procedure to account for this kind of bunching.

Reconciling past and present zoning jurisdictions. In Section 2.1, I mention that two kinds of zoning jurisdictions — cities incorporated after zoning was ruled constitutional and territory annexed by cities between 1980 and 2010 — are places that could have been zoned by two jurisdictions over its development. To account for this, I split up all properties in those jurisdictions into two. Properties built before incorporation or before 1980 (for newly annexed territory) are part of a county or township level sample when estimating minimum lot size adoption. Properties built afterwards are part of a separate incorporated place sample.

For these jurisdictions, I also assume by default that the new jurisdiction carries over any minimum lot sizes adopted at the county or township level. Any new bunching bins detected after incorporation then reflect additions to a pre-incorporation set of lot size controls. [2]

[2]There are examples of such carrying over of codes for cities I investigated, like [. . .]. Generically, however, this assumption places a lower bound on how many lot size controls a newly jurisdictions will have, even if they

This means that a jurisdiction can have an estimated first zoning adoption date before its incorporation date, and that outcomes that measure the degree of bunching for a zoning jurisdiction can be nonzero prior to incorporation.

This reconciliation process assumes I have incorporation dates for all U.S. cities. To construct such a dataset at a national level, I combine incorporation years reported in the 1987 Census of Governments with an imputation procedure based on modern Censuses. For cities not incorporated until after 1987, I use Census Bureau place-level population estimates from 1990 onwards. The first appearance of a city in these data serves as a proxy for decade of incorporation. [3]

With information on incorporation year, I define the pre-incorporation period for an applicable jurisdiction as all properties before the first decade of incorporation. This definition would assign homes built after incorporation but before a decade's end as part of the pre-incorporation period.

A.4 Census Housing Unit Records

In addition to data on properties bunching at minimum lot sizes, I collect Census Bureau records to get more accurate snapshots of past housing production rates. Historical estimates of single-family starts are used to benchmark the CoreLogic data, which I explain here.

Since the introduction of the Census of Housing in 1940, the Census Bureau has produced housing unit estimates for selected incorporated cities and townships. In particular, the Bureau have always broken down housing units by the type of structure (single-family, duplex, up to 20+ unit apartments) as well as by decade or half-decade of year built. Given a vintage of housing, I can access the earliest Census tables in which that vintage would have been built and recorded. These contemporary records make up my estimates of housing units ever built over some time interval.

I access the structure type-by-year built tables in the 1970 and 1980 Censuses digitized by Manson et al. (2021). I identify the number of housing units ever built for five-year intervals starting with 1960-64 and up to 2005-2009. I also observed a more lagged estimate of housing units built over 1940-49 and over 1950-59 using the 1970 data.

For housing units built as of 1940, I use the IPUMS 100% Full Count data from the 1940 Census (Ruggles et al. (2021)) in combination with the Census Place Project crosswalk from Berkes, Karger and Nencka (2022). These files allow me to tabulate the number of occupied housing units and the ownership rate for all counties and almost all incorporated cities at the time. I then impute the number of housing units by structure type in each jurisdiction by matching jurisdictions to state-urban status-tenure cross-tabulations of structure type shares in the Census tables. The jurisdiction's imputed 1940 structure type distribution comes from the linear combination of the Census table shares with the homeowner and renter rates in the jurisdiction.

Table A.1 displays the main benchmarking results referenced in the main text. I take the log difference in units build for a vintage over an entire county or city, as reported in the Census

move to simplify the zoning code.

[3]Because the Northeastern states are known to have had borough and town governments before zoning was popularized, I assume every such government I observe today were incorporated before the start of my sample.

tables, with what is reported in the CoreLogic data. Panel (a) reports log difference averages equally weighing counties and places, while Panel (b) weighs observations by population.

Due to imprecision with matching postwar suburbs in 1940 data, in the paper I report the benchmarking results matching CoreLogic counties with 1940 Census counties. The paper reports the "Up to 1940" and "1940–1969" rows in the County column of Table A.1. Results are similar in a test matching 2010 cities to 1940 data, but as the data is more biased toward inner cities there is greater attrition of older buildings.

Table A.2 conducts a robustness test by looking at CoreLogic units relative to only units still existing in the 2000 Long-form Census. Unlike Table A.1, which benchmarks how well existing properties in CoreLogic records approximates longitudinal housing data, this table compares how comprehensive CoreLogic records are to a recent snapshot from the Census. I find aggregates in the CoreLogic records are close to Census counts, and that CoreLogic records classify an older year built for many properties compared to the Census. This discrepancy reflects, if anything, recency bias in the Census's measure as it comes from self-reported dates rather than historical documents (Klimek et al. (2018)).

A.5 Historical Zoning Information

I collect historical zoning information, through ordinances and planner reports, for over 400 jurisdictions in 15 states covering all regions in the United States. For 17 of these jurisdictions, I acquire multiple zoning ordinances from the first zoning ordinance to the latest one (c. 2010s). To the best of my knowledge, this panel of zoning ordinances includes an observation after each major revision of the ordinance; jurisdictions, if they revise their ordinances at all, do so after multiple decades from first passage and use the occasion to define new zones. Panel (a) of Figure A.3 show a table of lot size controls in the 1939 Lower Merion Zoning Ordinance.

The remaining data are collated from planners' reports in the 1960s and 1970s, where regional authorities surveyed individual local governments under their purview about zoning practices. Unlike the panel of zoning ordinances, not every jurisdiction has the same common set of variables used to construct the moment conditions. For each of the three variables, I keep the jurisdictions for which data are not missing. Panel (b) of Figure A.3 show a table of lot size controls surveyed by New Jersey planners in 1960.

Table A.3 cites the ordinances and planners' reports in full.

A.6 Central City Demographic Change

In the paper, I calculate demograpic change variables for three groups in non-Southern central cities: Black population, Southern white population and non-Black foreign born population. Decadal counts for each group are tabulated from either place-level counts in the CCDB or Census, or from tract-level counts in the Census for cities I keep fixed at 1960 boundaries. I list below the specific sources for each group and for each decade:

- **Black population and foreign-born population:** 1940 full count Census (Ruggles et al. (2021)); 1950-1960 entries in the City and County Data Books; 1970 place and tract level data (Manson et al. (2021)).

- **Southern white migrants:** 1950-1960 public-use microdata, for identified cities (Ruggles et al. (2022)); 1970 place and tract level data (Manson et al. (2021)).

To construct the shift-share migration instruments, I need relative changes in migration rates from 1950–1970 for Southern Black and Southern White Americans. I retrieve these data from the following sources:

- **Black outmigration:** Replication dataset from Boustan (2016).
- **Southern white outmigration:** Bowles et al. (2016).

Figure 5 featured state-level estimates of Black outmigration from the South. I create these estimates by replicating calculations by Gregory (2005), using IPUMS data to derive an estimate for four million Black Americans leaving the South from 1940–1970. By scaling using relative changes in the compositions of Black migrants by state, I convert national estimates to state-level ones.

A.7 1940 Outlying City Characteristics

The Census Place Project crosswalk (Berkes, Karger and Nencka (2022)) assigns to every 1940 census respondent a match to a city or a minor civil division (MCD), like towns in the Northeast and civil townships in the Midwest. The crosswalk matches respondents to a county even if they lived in unincorporated areas.

I exploit this crosswalk and clean it to create prewar sociodemographic variables not just for central cities, but for suburbs existing by 1940 and for postwar suburbs undeveloped until future Censuses. I know the location of suburbs that had not incorporated by 1940, so I impute 1940 characteristics onto it by matching the suburb's centroid to the nearest 1940 MCD or county geography.

When tabulating the demographics, my demographic variables are shares of heads of households living in the jurisdiction. I also include estimated income levels using the machine learning algorithm of Saavedra and Twinam (2020).

Table A.1: Benchmarking CoreLogic to Historical Census Bureau Data

Panel A: Log differences in vintages

	County-level		City-level	
	N	Mean	*N*	Mean
Up to 1940	1056	-0.885	7694	-0.42
1940-1949	1049	-0.375	7242	-0.05
1950-1959	1056	-0.238	7700	-0.02
1960-1969	1056	-0.177	7802	0.02
1970-1979	1062	-0.293	7874	-0.06
1980-1989	1064	-0.134	7603	0.14
1990-1999	1065	-0.135	10883	0.13
2000-2009	1064	-0.152	6325	0.02

Panel B: Log difference of aggregates

	County-level	City-level
Up to 1940	-0.593	-0.665
1940—1969	-0.148	-0.259
1970—	-0.121	-0.050

Notes: This table matches historical Census estimates of single-family homes in each housing vintage, aggregated at two geographic levels: counties and cities with zoning powers. For each geography and vintage, I take log differences of Census counts with reported units in CoreLogic. The table reports the unweighted mean of the log differences, over each decade and over multi-decade analysis periods.

Table A.2: Benchmarking CoreLogic to 2000 Census Data

Panel A: Log differences in vintages

	County-level		City-level	
	N	Mean	*N*	Mean
Up to 1940	1056	-0.11	7681	0.09
1940-1949	1049	-0.27	7262	-0.10
1950-1959	1056	-0.20	7698	-0.05
1960-1969	1056	-0.17	7799	-0.05
1970-1979	1062	-0.18	7868	-0.07
1980-1989	1064	-0.13	7619	-0.03
1990-1999	1065	-0.13	10883	0.13

Panel B: Log difference of aggregates

	County-level	City-level
Up to 1940	0.161	0.015
1940—1969	-0.109	-0.295
1970—1999	-0.111	-0.150

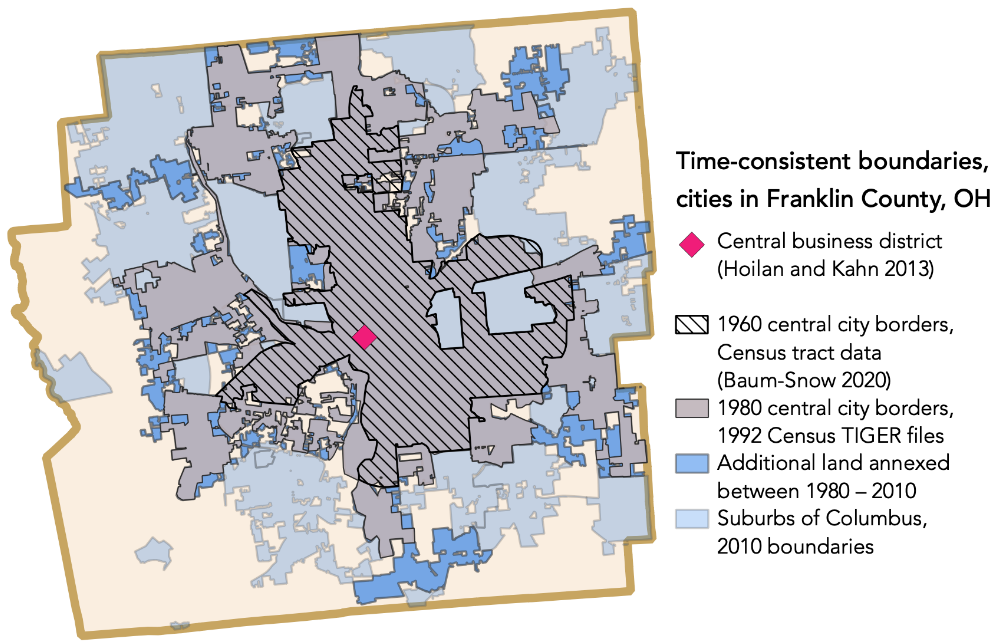
Notes: This table conducts the same matching and averaging procedure as Table A.1, but using one fixed wave of Census estimates. I compare the relative sizes of vintages in the 2000 Census estimates with reported units in CoreLogic built before 2000.

Table A.3: Inventory of Historical Zoning Ordinance Records

Panel A: Planners' reports on lot size controls Name of jurisdiction	State	First & Last Ordinance
Atherton	CA	1955 & 2007
Broward County	FL	1952 & 2022
Durham	NH	1934 & 2006
East Lansing	MI	1926 & 2006
Fulton County	GA	1946 & 2011
Grand Rapids	MI	1952 & 2012
Indianapolis	IN	1922 & 2016
Lake Forest	IL	1923 & 2020
Lakewood	CO	1969 & 2019
Lower Merion Township	PA	1930 & 2014
Marin County	CA	1954 & 2016
Memphis	TN	1922 & 1981
Mercer Island	WA	1937 & 2010
Plainfield City	NJ	1923 & 2019
Portland	OR	1957 & 2003
Seattle	WA	1923 & 2005

Panel B: Planners' reports on lot size controls Name of historical document	State	*N* Jurisdictions
New Jersey State Planning Bureau (1960). "Zoning in New Jersey, 1960."	NJ	290
Montgomery County Planning Commission (1971). "Zoning Ordinance Study."	PA	55
State Planning Division (1972). "Michigan Planning and Zoning Survey."	MI	20

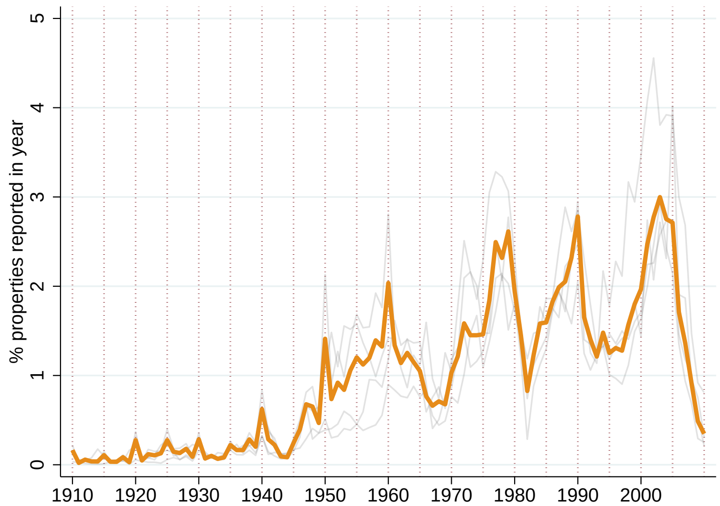
Figure A.1: Sample Jurisdiction Borders in A Metropolitan County



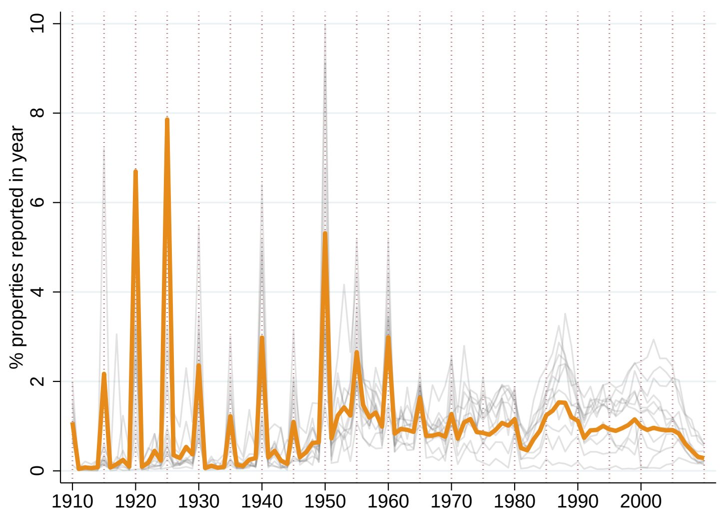
Notes: For Columbus, Ohio and the county it is in — Franklin County, Ohio — I visualize the time-consistent boundaries for the central city that control for postwar annexation. Central city composition change is calculated based on demographics within the tracts making up Columbus's 1960 boundaries. The extra land annexed by Columbus from 1960–80 as well as from 1980–2010 are counted as separate zoning jurisdictions. Non-central city jurisdictions around Columbus are displayed based on their 2010 borders.

Figure A.2: Degree of Bunching on Years Built Variable in Tax Records

(a) Minimal bunching for Sacramento CBSA properties



(b) Large degree of bunching for Philadelphia CBSA properties



Notes: Over two panels, this figure plots the distribution across year built for all properties in two CBSAs in orange. Years ending in 5 or 0, on which year built data may bunch, are highlighted with dotted lines. Gray lines represent subsample distributions over component counties for the CBSA.

Figure A.3: Sample Lot Size Data from Historical Zoning Records

(a) Example of lot size control data from ordinances

Summary of Zoning Requirements

	R 1	R 1-A	R 2	R 3	R 4
Minimum Lot Area per Family, in square feet.......	30,000	18,000	10,000	6,000	5,000
Maximum Building Area, Percentage of Lot Area..	15%	18%	20%	30%	35%
Front Yard or Setback, in feet	50	40	40	30	25
Side Yards, Single-Family Detached Dwellings					
Minimum width each yard, in feet.............	15	12	10	8	8

(b) Example of lot size control data from planners' reports

FEATURES OF NEW JERSEY

MUNICIPAL ZONING ORDINANCES

NEW JERSEY STATE DEPARTMENT OF CONSERVATION & ECONOMIC DEVELOPMENT STATE PLANNING BUREAU	ALLENHURST BORO	ALLENTOWN BORO	ASBURY PARK CITY	ATLANTIC TWP.
1. OFFICIAL PLANNING BOARD		55	45	54
2. ZONING ORDINANCE (YEAR ADOPTED)	29	41	45	54
3. YEAR OF LATEST REVISION	57		58	55
4. ZONING ORDINANCE ON FILE	29	41	58	55
5. DEFINITIONS	Y	Y	Y	Y
6. NUMBER OF RESIDENTIAL ZONES	9	1	4	2
EXCLUSIVE	9	1	4	2
NON-EXCLUSIVE				
12. MINIMUM RESIDENTIAL LOT SIZES	Y		Y	Y
SMALLEST	60 W			20,000 100 W
LARGEST (in square feet or in acres)	100 W		5,000	40,000 200 W

B Details of Adoption Estimation

B.1 Algorithm Mechanics

Defining Lot Bins. To detect bunching regions in the lot size distribution, I fix a partition of the distribution's support that stays constant across jurisdictions. This partition of the support into lot bins discretizes the learning problem of finding the bunching bins. Throughout the paper, properties that "bunch at lot sizes" are defined as properties within the lot bins that are classified as bunching bins.

Each lot bin is centred at a round number on the lot size support, and the bandwidth of the lot bin varies by what value at which it is centred. Table B.1 defines the different bandwidths and the size thresholds at which one takes effect over another. For example, from 5000 square feet onwards the bandwidth is 250 feet, so the next lot bin is centred at 5500 square feet, then 6000, and so on.

The choice behind these lot bins is to maintain proportionality between the reference size level and the bandwidth, while also keeping the lot bins as centred on round numbers as possible. By changing the bandwidth at discrete points, the lot bin's bandwidth hovers around 4-6% of the value it's centred on no matter what the size. The proportionality can be seen in Figure B.1, which plots lot square footage on a logarithmic scale; the proportionality is visualized as linear growth in the lot bin value curve.

The variable-bandwidth lot bins are used in all steps except for when calculating the gradient statistic, where I use fixed bandwidths of 500 square feet from the smallest to largest lot bin. Smaller bandwidths are needed for this one case to capture the anomalous densities due to bunching on specific round numbers. I then match any detected bunching bins to the larger variable-width lot bins it belongs to, and proceed with the variable-width detected bunching bins.

Bunching Bin Detection. The algorithm takes all jurisdictions in a CBSA, then cycles over detecting bunching bins at the jurisdiction-vintage level for a fixed set of adoption times $\{\tau\}$.

First, the algorithm decides for each j the size of a post-τ sample $T^j(\tau)$, or the value of T as defined in the paper. I use an adaptive procedure, where $T = 10$ if the vintage over $[\tau, \tau + T]$ has at least 500 observations. If not, $T = 20$, no matter how many observations are over $[\tau, \tau + T]$.

The algorithm then creates a pre-period sample $C^j(\tau)$ defined for years before τ. If adoption occurred sometime between $[\tau, \tau + T]$, the gradient statistic at certain lot bins will be anomalously high to this pre-period sample. If not, anomalously high statistics are due to statistical noise. The pre-period sample spans 10 or 20 years depending on sample size, like with the post-period.

Once $T^j(\tau), C^j(\tau)$ are selected, the gradient statistic calculated using the pre-period sample as a control distribution is

$$\hat{G}^j(\tau,\ell) = \left(\log m_\ell^{T^j(\tau)} - \log m_{[\ell-\mu(\ell),\ell)}^{T^j(\tau)}\right) - \left(\log m_\ell^{C^j(\tau)} - m_{[\ell-\mu(\ell),\ell)}^{C^j(\tau)}\right)$$

where $m_\ell^{T^j(\tau)}, m_\ell^{C^j(\tau)}$ are respectively the histogram density estimate at ℓ over the post-period and pre-period samples, and $m_{[\ell-\mu(\ell),\ell)}^{T^j(\tau)}, m_{[\ell-\mu(\ell),\ell)}^{C^j(\tau)}$ are respectively the density estimates over an interval of measure μ to the left of ℓ.

I also calculate a version of the gradient statistic where the counterfactual density is zero everywhere, i.e. I make no correction for counterfactual gradients. I use this version of the statistic where there was a surge of development in the post-sample relative to the chosen pre-period sample: $N^{T^j(\tau)}/N^{C^j(\tau)} > M^{Growth}$ for some parameter M^{Growth}.

I allow the measure function $\mu(\ell)$, which determines the width of the bunching region with missing mass displayed in Figure 3, to be a function of ℓ. The function takes the form $\mu(\ell) = \mu^{miss} \times \nu(\ell)$, where $\nu(\ell)$ is the width of the lot bin at ℓ described in Table B.1. Testing bunching at, for example, 30000 versus 30500 square feet will both have a bunching region of width $\mu^{miss} \times 2000$ to each bin's left.

The estimated $\hat{G}$ across vintages (j, τ) are standardized using a standard deviation estimated across lot bins over jurisdictions in the CBSA. The set of all lot bins for one city is appended to all lot bins for another city in the CBSA, and so forth: the standard deviation is estimated over the collected sample.

Before being detected, lot sizes where controls apply will have outlier gradient statistics affecting calculation of moments. Following Leys et al. (2013), I estimate the standard deviation in the presence of outliers by first calculating the median absolute deviation, then multiplying it by a factor of 1.4826 to account for normal error in the histogram estimates.

All lot bins ℓ for which the standardized $\hat{G}^j(\tau, \ell) > \alpha$ for the critical value parameter α are bunching bins for (j, τ). I then collapse diffuse bins which are all part of the same larger lot bin into the larger bin (i.e. bunching bins at 29500, 30000 and 30500 are all collapsed to 30,000).

The bunching bins detected with the gradient statistic rule are combined with bins detected using the other two methods. It is worth discussing the final method, which relates to the detection method in Zabel and Dalton (2011). Those authors look at a fixed quantile of the lot size distribution, then take the minimum of two elements for a housing vintage to get the minimum lot size at (j, τ): the known minimum lot size and the observed lot size at that fixed quantile. My method differs because I make a guess at a minimum lot size by identifying the modal lot bin, then verify that it is a *minimum* lot bin for housing development by seeing if it falls below a fixed quantile.

Once all the bunching bins over vintages are retrieved — with each (j, τ) using one of two definitions of the gradient statistic depending on the condition $N^{T^j(\tau)}/N^{C^j(\tau)} > M^{Growth}$ — The bunching bins at the jurisdictions are defined as specified in Section 2.2.

Adoption Year Estimation. I first detail how the annual time series of excess mass and the gradient statistic are computed. For every τ in the support of year built, I construct the gradient statistic $\hat{G}$ a post-τ and a pre-τ sample using adaptive sample size methods. I also construct excess mass over a post-τ and a pre-τ sample covering 10 years each, following procedures in Section 3.2. The major difference in this step is that since τ is annual, the samples used to calculate different statistics will overlap with each other. An example of the time series for the excess mass statistic is shown in Panel (a) of Figure B.2.

Using these statistics, I allow for two specifications of a structural break model. Specification 1 processes the time series for each bunching bin one by one, where a structural break for the time series of excess mass at just one bin is detected using the algorithm developed in Bai and Perron (2003) and implemented in Ditzen, Karavias and Westerlund (2021). Structural breaks are stored only if the algorithm in Bai and Perron (2003) considers the break to be statistically significant. Panel (b) of Figure B.2 shows how a structural break is identified on the same time series in Panel (a).

Specification 2 allows for common adoption between different bunching bins detected to have bunched over the same vintage. The pooled statistic, which sums up excess mass across bunching bins and averages the gradient statistic across bunching bins, has more statistical power to identify a common adoption date under prior information that the jurisdiction could have adopted multiple minimum lot sizes at the same time. Because demand around pooled bunching bins may not be stationary over the whole sample period, break detection is limited to a window around the first τ for which a bunching bin was detected.

The fully crossed set of specifications uses two outcomes for the break detection over individual bunching bins: the gradient statistic and excess mass. It uses three outcomes for break detection over common adoption bins: Gradient statistic, only within 10 years of first adoption; excess mass, within 10 years of first adoption; and excess mass within 20 years of first adoption (a longer time frame).

Therefore, five estimated structural breaks under different assumptions lead to five adoption year estimates for each bunching bin at a jurisdiction. Denote, for structural break model i, the estimated structural break $T_i(\underline{\mathbf{b}})$.

Choosing Tuning Parameters. Ahead of training the classifier on real historical records, I loop the algorithm across a set of possible parameter values $\hat{\mathbb{B}} = \{\alpha, \mu^{miss}, M^L, \underline{F}, M^{Growth}\}$. For each parameter vector, I get five estimated structural breaks for each jurisdiction, from which I derive first large lot adoption years $\min\{T_i(\underline{\mathbf{b}})\}$.

Taking the minimum of adoption years over bunching bins outputs first adoption year estimates for jurisdictions. Fitting model parameters over training data means finding the best predictions of large lot adoption years $Year_j^L$, or minimizing the objective using least squares:

$$\min_{\hat{\mathbb{B}},\hat{\mathbf{w}}} \sum_j \left(Year_j^L - T_{\underline{\mathbf{b}}}' \hat{\mathbf{w}} \right)^2,$$

$$T_{\underline{\mathbf{b}}} = [\min\{T_i(\underline{\mathbf{b}})\}]_i, \quad \underline{\mathbf{b}} \geq 10{,}000.$$

Appendix Tables B.2 and B.3 lists the final vector of parameters and weights. The rightmost column of Table B.3 also shows how each individual structural break specification correlates with the final ensemble estimate reweighing multiple specifications.

Table B.1: Width of Lot Bins by Lot Size Intervals

Range (000s sq. ft.)	1K-3.75K	4K-14.5K	15K-27K	28K-58K	60K-112K	$\geq$ 116K
Width of each bin (sq. ft.)	250	500	1000	2000	4000	$\geq$ 8000

Table B.2: Parameters For Bunching Detection Algorithm

Notation	Name	Value
α	Critical value for bunching bin classification	1.10
μ^{miss}	Multiplicative factor for bunching region	1.6
M^{L}	Minimum density on bin needed for classification	0.025
$\underline{F}$	CDF threshold for modal value classifier	0.25
M^{Growth}	Growth factor threshold for pre-period violations	2.5

Table B.3: Weights Placed By Ensemble Model

Model	Weight	Corr. With Final Estimate
Common adoption, $\hat{G}$	0.428	0.78
Common adoption, Excess mass	0.677	0.93
Common adoption, Excess mass, long window	0.021	0.67
Individual, $\hat{G}$	-0.036	0.67
Individual, Excess mass	-0.091	0.67

Figure B.1: Visualizing Grid of Lot Bins

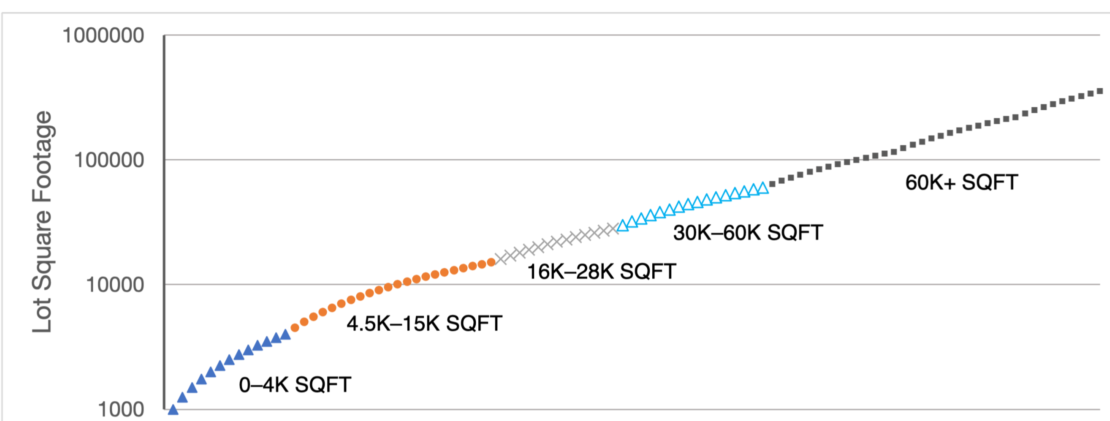
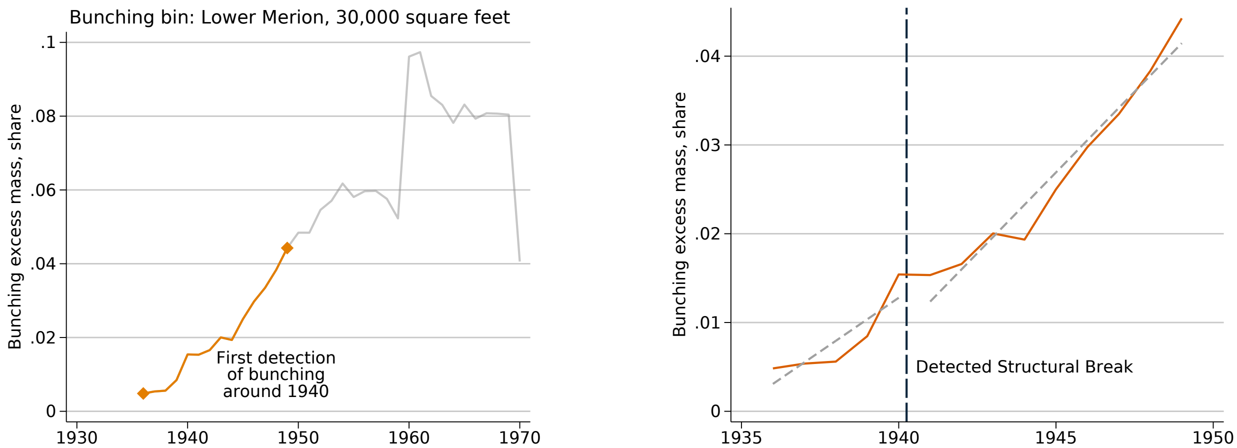


Figure B.2: Visualization of Structural Break Detection

(a) Examples of excess mass time series

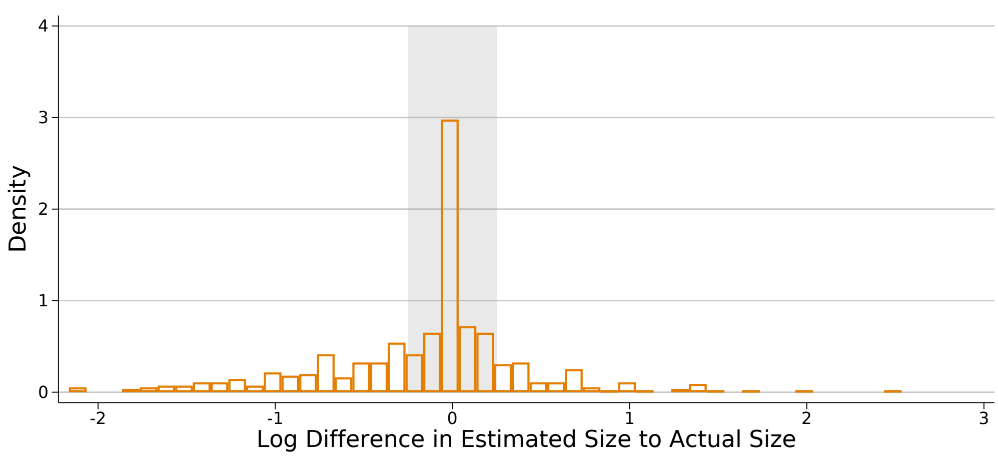
(b) Application of structural break test around 1940



Notes: This figure visualizes the estimation of a zoning adoption date in Lower Merion Township across three models. The gray line marks the year 1939, the adoption of the first minimum lot regulations in Lower Merion and the target of all the algorithms. The orange line marks the structural break year using the empirical CDFs in each panel and the Bai-Perron error-minimizing estimate.

Panel (a) visualizes the sample of properties with local bunching detected around 1940. Panel (b) visualizes how the algorithm filters to just the time interval around 1940 when detecting breaks.

Figure B.3: Histogram of Lot Size Estimate Errors



Notes: This figure plots a histogram of 587 actual minimum lot sizes recorded in the data that were matched to at least one estimated bunching bin around the year of adoption. For each lot size, I take the logarithm of the actual minimum lot value less the logarithm of the closest estimated bunching bin. A 0 reflects perfect matching of values, while a negative value reflects the closest bunching bin was smaller than the actual lot.

C Technical Proofs

C.1 Extension of bunching to monocentric city models

To formalize the spatial arbitrage argument presented in Section 2.3, I solve and interpret a monocentric city model with a competitive housing sector, differentiated developers and minimum lot size constraints.

As in Brueckner (1987) and Glaeser (2008), let there be homogeneous residents in a monocentric city maximize consumption C and the housing size, H. Consumption is location-specific as it is wealth net of the housing value $p(d)H$ as well as by commuting costs td.

The city adopts a minimum lot size constraint $\underline{\ell}$; for an acre of land, at most $\underline{N}$ units can be built on said land. Wherever the constraint is applied, it translates one-to-one to a minimum consumption of housing, $\underline{H} = L/\underline{N}$. The residents then solve the maximization problem:

$$\begin{aligned} &\max_{C,H} u(C,H) \\ &s.t.\ C = W - td - p(d)H \\ &\quad H \geq \underline{H}. \end{aligned}$$

The city is in spatial equilibrium, so the Alonso-Muth-Mills spatial indifference equation applies here in two forms. For locations d that are unconstrained and where $\underline{H}$ does not bind, $t = -p'(d)H$. For locations $\underline{d}$ that do, observe that the Lagrange multiplier enters into the problem through H alone, so differentiating with respect to d obtains $t = -p'(\underline{d})\underline{H}$.

It follows that relative to the unconstrained pair gradient $(H^*(\underline{d}), p^*(\underline{d})), p'(\underline{d}) <= p^{*\prime}(\underline{d})$ as $H^*(\underline{d}) <= \underline{H}$.

Suppose the central area $[0, \Delta]$ of the city's land was zoned with a minimum lot size. There is a point d' where housing demand is $\underline{H}$, so $p^*(d') = p(d')$. It may either be in the interval, in which case $p(d) < p^*(d)$ over $[0, d']$, or it is outside of it, in which case $p(d) < p^*(d)$ over $[0, \Delta]$.

Suppose the interval zoned was on the periphery, $[\Delta, \infty)]$. The argument above is analogous if d' is in the interval, and for $d \in [\Delta, d']$ we have $p(d) < p^*(d)$. The price gradient is right-continuous up to Δ, at which point the gradient jumps up to the unconstrained upper envelope p^*. This case also applies to where the interval zoned is fully within the city limits itself, $[\Delta^1, \Delta^2]$, as the price gradient will be continuous with p^* at Δ^2 but discontinuous at Δ^1.

I now introduce differentiated developers with discrete types $\{n\}$, each of which is solving the per-acre profit maximization problem of units to built N while obeying the zero-profit condition,

$$\begin{aligned} &\max_{N} p(d)N - c_n(N) - r(d) \\ &s.t.\ p(d)N = c_n(N) + r(d), \\ &\quad \overline{N} \geq N. \end{aligned}$$

The cost function $\{c_n(N)\}$ is differentiated as follows: the functions satisfy single crossing differences them. This implies that for each $\{n\}$, the interval over which a particular $c_n(N)$ is the minimal value, C_n^*, is a connected subset of the reals.

With this configuration of developer types, the unconstrained city has a sorting equilibrium. A developer of type n builds only over the subset of land supporting densities $N^*(d)$ in the subset where the type has minimal cost. Suppose this were not true at some distance d: in that case, the developer whose cost function is cost-minimizing can produce the same amount of N at a lower $p(d)$, causing other types to have negative profits and exit production at that distance.

With the consumer and developers' problems in the unconstrained equilibrium explained, I provide a formal definition of spatial arbitrage with lot size constraints. Suppose the left endpoint of the lot size constrained area $\Delta^1 > 0$, and there is a subinterval $[\Delta^1, d']$ where the constraint is binding on residents. For the type n' that developed at $d \in [\Delta^1, d']$ at a density $N^* > \underline{N}$, whether it continues developing at d depends on if in equilibrium,

$$[p^*(d) - p(d)]N^* - \int_{\underline{N}}^{N^*} [c'(h) - p(d)]dh > r^*(d) - r(d),$$

where the left hand side represents the net per-acre loss in revenue (foregone producer surplus net of the costs of producing units costing more than market price) and the right hand side are the per-acre decreases in land costs. This condition is not necessarily true, since there are other developer types n'' who are not constrained by the lot size minimum and up-bids $r(d)$ accordingly.

The subinterval of land with a lot size minimum developed by the subset of developer types still entering the housing sector forms the bunching mass in an overall lot size distribution.

C.2 Proof of Proposition O1.

The proof is in two parts. First, define

Definition 6. *The* ***excess mass classifier*** $\hat{L}$ *of threshold* δ *is defined over two distributions* $p(z), q(z)$ *and uses the rule:*

$$\hat{L}(z^*) = 1 \quad \textit{if} \;\; \tilde{B}(z^*) \equiv p(z^*) - q(z^*) \geq \delta.$$

Lemma. For a nonempty set of $\{\underline{L}\}$, there is a real number M such that if the econometrician observes a collection of distributions $\{q_t\}$ for which the total variation distance between each n_t^j, q_t is less than $\delta = M/3$, the excess mass classifier set identifies $\{\underline{L}\}$.

Proof. As each $\underline{\ell}$ is a regulatory notch for some t', to each $\underline{\ell}$ on the distribution $\{\ell_{t'}^{*j}\}$ there is a marginal bunching developer type $\Delta\underline{\ell}^*$ and a bunching mass $B(\underline{\ell}) = \int_{\underline{\ell}-\Delta\underline{\ell}^*}^{\underline{\ell}} \ell_0(s)ds$, where ℓ_0 is the observed distribution if developer types maximize subject to only $\mathbf{Z}_t$. When developers maximize subject to $(L^j, \mathbf{Z}_t)$, we have: $\ell_t^{*j}(\underline{\ell}) = B(\underline{\ell}) + \ell_0(\underline{\ell})$, both terms on the right being nonnegative.

Let $M = \min\{B(\underline{\ell}) \,|\, \underline{\ell}$ is minimum lot size $\}$. Choose $\delta = M/2$. If ℓ^* is not in $\{\underline{L}$, $\tilde{B} < \delta$ by assumption on q_t.

For $\underline{\ell}$ that is in $\{\underline{L}\}$,

$$\tilde{B} = \left(\int_{\underline{\ell}-\Delta\underline{\ell}^*}^{\ell^*} \ell_0(s) ds \right) - (q_t(\ell^*) - \ell_0(\ell^*)) \geq (M - \delta) = \delta.$$

This concludes the proof. □

Remainder of proof. Note that the gradient statistic, as a function of $(\tilde{B}(\ell), \log(p(\ell-\mu)) - \log(q(\ell-\mu))) \equiv (\tilde{B}, z)$, is submodular in the final argument z wherever $G > 0$;

$$G(\tilde{B}', z') - G(\tilde{B}, z') \leq G(\tilde{B}', z) - G(\tilde{B}, z)$$

For $\tilde{B}' \geq \tilde{B}, z' \geq z$.

Take the minimizer of $G(p(\ell), z)\big|_{z=\underline{z}}$, where $\underline{z}$ is the lowest valued z, to be $\tilde{M}$. Topkis's theorem on $-G$ and $-z$ implies the minimizer of the sets $G(p(\ell), z)\big|_{z=z'}$ across the arguments z' is nonincreasing in z'. Therefore, $\tilde{M}$ is no greater than any element of the observed values $G(\tilde{B}, f(\ell-\mu)$, conditional on $G > 0$. Setting δ to be at most $\tilde{M}$ implies the argument in the lemma can be repeated to show the result.

C.3 Details of Lot Size Restrictiveness Decomposition

Begin, as in Section 3.2, with the *bunching share* of residential development within jurisdictions:

Definition 7. *For a discrete distribution of lot sizes in jurisdiction j, $\{m^j(\ell)\}$ and identified bunching bins $\{\underline{\mathbf{b}}\}$, the bunching share of development in j with respect to $\underline{b}$ is*

$$Bunch(j; \underline{\mathbf{b}}) = \sum_{\ell \in \{\underline{\mathbf{b}}\}} m^j(\ell).$$

We can view the bunching share as both a function of known levels, or as a function of two random variables: random realizations of jurisdictions as well as random sets of bunching bins. The latter view helps formalize the "desirable condition" on what drives variation in the measure.

Consider a collection of jurisdictions $j \in J$ each with bunching bins $\underline{\mathbf{b}}(j)$, along with an explanatory variable X. Define the following random variable: $\underline{\mathbf{B}}$ maps to each distinct element in the collection of all $\underline{\mathbf{b}}(j)$ with equal probability. For ease of notation, also denote the conditional random variable $\text{Bunch} \mid \underline{\mathbf{b}}(j') \equiv \text{Bunch}(j; \underline{\mathbf{b}}(j'))$. Then, the law of total covariance implies

$$\begin{aligned} \text{Cov}_J\left[\text{Bunch}(j; \underline{\mathbf{b}}(j)), X\right] = & \underbrace{\text{Cov}_{\underline{\mathbf{B}}}[\mathbb{E}(\text{Bunch} \mid \underline{\mathbf{b}}(j')), \mathbb{E}(X \mid \underline{\mathbf{b}}(j'))]}_{(1)} \\ & + \underbrace{\mathbb{E}_{\underline{\mathbf{B}}}[\text{Cov}(\text{Bunch} \mid \underline{\mathbf{b}}(j'), X \mid \underline{\mathbf{b}}(j'))]}_{(2)}, \end{aligned}$$

On the left hand side, the statistical relationship between jurisdictions $\mathrm{Cov}_J\left[\mathrm{Bunch}(j;\underline{\mathbf{b}}(j)),\, X\right]$ is what conventional least squares estimators use to identify causal effects of X on outcomes. This relationship can then be decomposed into two components. Component (1) should be thought of as the relationship driven by variation in only regulation adoption between jurisdictions. If we knew the expected amount across many counterfactual jurisdictions by which a lot size control we sample binds on development, $\mathbb{E}(\mathrm{Bunch}\,|\,\underline{\mathbf{b}}(j'))$, we can see how that correlates with conditional variation in explanatory variables.

Component (2) should be thought of as a form of *selection bias*. Realistically, the way jurisdictions choose which minimum lot sizes to adopt is endogenous to a jurisdiction's underlying characteristics. Where the choice of which bins is influenced both by characteristics and by which lots had the highest demand, the conditional covariances will be nonzero and component (2) will have nonzero bias. A positive effect of X on bunching mass may be due more to better manipulation of where jurisdictions place their lot sizes, not necessarily that the lot size controls were restrictive on denser development.

This contamination problem could apply to other outcomes used in the literature, like the share of vacant land zoned for a large lot size. To sketch this argument, we can redo the decomposition replacing the random variable $\mathrm{Bunch}\,|\,\underline{\mathbf{b}}(j')$ with the random variable of land development probability $\mathbb{P}(\mathrm{Develop})\,|\,\underline{\mathbf{b}}(j')$. Land zoned for the largest lots could also be systematically unattractive to be developed during a time frame of interest. This systematic relationship could be a source of mismeasurement contaminating the outcome.

Unlike the bunching share for a jurisdiction j, the excess mass $\tilde{B}_j$ as a random variable has lower variance. This is a consequence of the bunching share being the sum of $\tilde{B}_j$ with prior demand for lots at the lot size control. Because

$$\left|\mathrm{Cov}(Y \mid \underline{\mathbf{b}}(j')\,,\, X \mid \underline{\mathbf{b}}(j'))\right| \leq \sqrt{\mathrm{Var}(Y \mid \underline{\mathbf{b}}(j'))\mathrm{Var}(X \mid \underline{\mathbf{b}}(j'))},$$

and $\mathrm{Var}(\tilde{B}\,|\,\underline{\mathbf{b}}(j'))$ is lower than $\mathrm{Var}(\mathrm{Bunch}\,|\,\underline{\mathbf{b}}(j'))$, using the excess mass measure of restrictiveness provides tighter bounds on the bias component (2) provided the excess mass estimates $\tilde{B}$ well in finite sample.

D Details of Identification

D.1 LASSO model selection procedure

To construct county level migration shocks for the instrument Z_{ct}^{Black}, I follow the model selection procedure in Derenoncourt (2022) with some modifications. Recall that the goal of the model selection step is to generated projected migration rates from counties $\widehat{\text{mig rate}}_{kt}$ using only Southern county characteristics, so the instrument varies between counties only due to predetermined economic characteristics that saw structural transformation in the postwar decades.

For each decade t, I fit observed outmigration rates from a Southern county k using the linear model

$$\text{mig rate}_{kt} = \alpha + \mathbb{X}'_{k,t-10}\Lambda + \upsilon_{kt},$$

where the explanatory variables $\mathbb{X}$ are chosen from a set of candidate Southern variables using a LASSO regression. When mig rate$_{kt}$ is overidentified through many instruments, and a subset of those instruments of at most rank s approximates the true CEF of migration rates, Belloni et. al. (2012) shows instruments using predictions from a LASSO regression consistently estimate effects for the endogenous variable — in this case, the county migration rates which may be contaminated by pull factors of other U.S. cities.

Derenoncourt (2022) makes two decisions to make the prediction problem more tractable. First, she chooses instruments from a smaller set of 8 county level variables proposed in Boustan (2010). The model selection step ends up verifying that those variables all have robust positive weights in the prediction problem. Second, she conducts a post-LASSO procedure where the predicted outcomes come not from the LASSO specification, but from a second OLS regression of the variables chosen by LASSO.

I do not restrict myself to the two researcher decisions. I use a larger set of county level variables recorded in Boustan (2016) to obtain the optimal instruments projecting onto mig rate$_{kt}$. I also output the LASSO estimates without post-estimation conditioning, which consistently estimates effects conditional on some further restrictions on the dimension of controls in the main specification first stage. To minimize overfitting in the LASSO step, I also estimate the LASSO weights on only 80% of all Southern counties in my data, extrapolating predictions for the remaining 20% in the test set.

Table D.1 records both the total dimension of $\mathbb{X}_{k,t-10}$ chosen over each decade of migration rates, as well as the variables that have the largest model weights. I allow for three variables for which historians believe have predictive power to always remain in the model: share of workers in agriculture, share of agricultural production in cotton and the percent of farms who had tenant workers. Together, the three variables measure a common factor: the size of sharecropper farms as a share of the county's labor market.

These three variables alone have high negative weights in the LASSO model, and so are predictive of outmigration. The weights reassure that the "approximate sparsity" assumption in Belloni et. al. (2012) applies. Table D.1 also includes additional variables in Southern counties which retains Southern Black workers on average (like mining intensity in Texas and Oklahoma) as well as further push them out of the South (like average precipitation, and intensity of tobacco production in 1970).

Figure D.1 visualizes both how my predicted outmigration rates compare to actual rates, as well as how the original Derenoncourt (2022) model selection performs on outmigration prediction. Intuitively, a LASSO specification which well specifies the CEF should see significant positive relationships between the CEF of migration rates with predicted values.

Both my model and the model in Derenoncourt (2022) exhibit the positive relationship across every decade separately fit onto the models. Of the 2×2 specifications I visualize, I remark both my model and Derenoncourt's model project attenuated migration rates than actual ones by 1970. I also note that over 1940–1950, my more flexible LASSO procedure chooses Southern county predictors that correct for a slope shift in Derenoncourt's model, which predicts net migration *towards* Southern counties when outmigration was takig place.

To conclude, the LASSO procedure chooses sensible characteristics of Southern counties to instrument for outmigration rates of Southern counties, but model misspecification cannot be ruled out due to relative attenuation of estimates between decades. Therefore, the main specification in Section 6 normalize all demographic change variables by percentiles across cities in each decade, in an attempt to correct for instrumental variable misspecification.

D.2 Additional propositions

Proposition on SSIV identification. An application of Proposition 4 in Borusyak, Hull and Jaravel (2022), which gives a general treatment of consistent estimation using shift-share instruments, yields:

Given a panel of counties k and decades t, a fixed vector of clusters of counties across decades $\mathbf{C}(k,t)$, county-level unobservables $\overline{\nu}$ and the following assumptions:

1. The instrument is relevant in finite sample;
2. (Conditional quasi-random shocks) $\mathbb{E}[g_k(t)\,|\,\overline{\nu}, \tilde{\omega}(k), \mathbf{C}(k,t)] = \mathbf{C}(k,t)'\mu$, for all (k,t);
3. (Instrument relevance from many shocks) $\mathbb{E}\left[\sum_n s_n^2\right] \to 0$ as the treated units c grow large;
4. (Uncorrelated shocks) $\text{Cov}[g_k(t), g_{k'}(t')\,|\,\overline{\nu}, \tilde{\omega}(k), \mathbf{C}(k,t)] = 0$.

Then, provided the regression model contains the saturated cluster share exposures $\{\sum_{k^c \in C(k,\tau)} \tilde{\omega}(k^c)\mathbf{1}[t = \tau]\}$, the conventional IV estimator satisfying the moment condition $\mathbb{E}[\sum_{j,t} Z^{black}_{c(j),t} \hat{\varepsilon_{j,t}}] = 0$ consistently estimates the causal effect β.

SSIV definitions are numerically equivalent. I first define some notation for how the migration shift-share is presented in Boustan (2010). Let $\overrightarrow{\text{Black}}_{k,1940}$ be the number of net migrants out of county k, and $\overrightarrow{\text{Black}}_{k\to c,t}$ be the observed flow of Black migrants from county k to city c in time t. Without arrows, Black_{kt} represents the number of Black residents in k (and analogously for destination cities c) in t.

The migration shift-share in Boustan (2010) has the shares equal to "the share of Blacks who left county k after 1935 and reside in city c in 1940." in my terminology, this defines shares

$$\omega_{kc,1940} = \frac{\overrightarrow{\text{Black}}_{k\to c,1940}}{\overrightarrow{\text{Black}}_{k,1940}}.$$

The shifts are the predicted *levels* of migrants out of county k, denoted $\widehat{\text{mig rate}}_{kt} \times \text{Black}_{kt}$.

Numerical equivalence follows from the following algebra:

$$\begin{aligned}
&\sum_k \omega_{kc,1940} \times \widehat{\text{mig rate}}_{kt} \times \text{Black}_{kt} \\
=&\sum_k \frac{\overrightarrow{\text{Black}}_{k\to c,1940}}{\overrightarrow{\text{Black}}_{k,1940}} \times \widehat{\text{mig rate}}_{kt} \times \text{Black}_{kt} \\
=&\sum_k \frac{\overrightarrow{\text{Black}}_{k\to c,1940}}{\text{Black}_{c,1940}} \times \frac{\text{Black}_{c,1940}}{\overrightarrow{\text{Black}}_{k,1940}} \times \widehat{\text{mig rate}}_{kt} \times \text{Black}_{kt} \\
=&\text{Black}_{c,1940} \sum_k \frac{\overrightarrow{\text{Black}}_{k\to c,1940}}{\text{Black}_{c,1940}} \times \frac{\text{mig rate}_{kt} \times \text{Black}_{kt}}{\overrightarrow{\text{Black}}_{k,1940}} \times \frac{\widehat{\text{mig rate}}_{kt}}{\text{mig rate}_{kt}} \\
=&\text{Black}_{c,1940} \sum_k \tilde{\omega}_{c,1940} \times \frac{\overrightarrow{\text{Black}}_{k,t}}{\overrightarrow{\text{Black}}_{k,1940}} \times \frac{\widehat{\text{mig rate}}_{kt}}{\text{mig rate}_{kt}}.
\end{aligned}$$

D.3 Further validation of instrument

Added value of controlling for migration exposure shares. The utility of using migrant exposure shares as controls — clustered at the state level in this paper, $\sum_{k'\in s} \tilde{\omega}_{c,1940}(k')$ — is to control for any fixed effect differences in the migration shock terms even after the LASSO procedure.

If the shock terms with predicted migration rates, $g_k(t)$, varies between states as well as within states, controlling for state clustered exposure shares should improve estimates. Figure D.2 visualizes this variation by stacking within-state migration shock histograms on top of each other. In addition, the mean level of migration shocks in each state and decade are plotted as separate blue lines. Because overlap of the shocks is not perfect across the states, especially going out into 1970 (Panel (b)), adding state clustered exposure shares would bring the identifying variation closer to an ideal where the distributions fully overlap.

To confirm that different cities also drew in migrants from different states, I produce a histogram of 1940 Black migrant shares by state for major non-Southern cities along the lines of Figure C.1 in Derenoncourt (2022). The histograms are plotted in Figure D.4.

Falsification test of instrument on 1940 suburban demographics. I describe the regression coefficients and hypothesis tests conducted in Table D.4. Column (1) presents correlations with Black population growth from 1940–70 and pre-period characteristics for central cities and suburbs. Each row presents a different outcome from the pooled regression.

The results confirm findings from Derenoncourt (2022) that Black migrants selected more towards high manufacturing cities, based on the positive coefficient with manufacturing employment share as the outcome. It also confirms destination cities for Black migrants have suburbs that are less native-born white and 0.2% more homeowners on average in suburbs than in the central city. The positive association between Black migrants and the distance of zoning jurisdictions from central cities additionally indicates Black migrants selected to metropolitan areas with preexisting suburbanization trends and incorporation further out of the central city.

Across the columns, Column (2) runs the same outcomes on the shift-share instrument. Column (3) adds the state exposure controls, while Column (4) excludes central cities from the test dataset. Column (4) is the closest specification and sample to the one used in the results.

Southern white migration shift-share instrument. In Section 6.3, I also construct a shift-share migration instrument for Southern white migrants analogously to the Black migration instrument, but without a projection of migration rates to predicted levels from Southern county characteristics:

$$Z_{ct}^{S-white} = \sum_{\text{Southern } k} \tilde{\omega}_{c,1940}^{S-white}(k) \times g_k^{S-white}(t), \quad g_k^{S-white}(t) = \frac{\overrightarrow{\text{S-white}}_{k,t}}{\overrightarrow{\text{S-white}}_{k,1940}}.$$

The shares $\tilde{\omega}_{c,1940}^{S-white}$ can be defined using 1940 full-count Census data, and the shifts from migration rates in Bowles et al. (2016). Missing data limitations come from the observed composition change in Southern whites in central cities. Composition change for Southern whites is observed only in the largest non-Southern central cities in 1950 and 1960, though it is available for all 1970 cities.

I note that in Figure D.3, migration shocks across Southern counties for white net migration, without any corrections using predicted rates, exhibit high overlap and common means. This implies on the margin, adding exposure shares for Southern whites would not correct for any sources of bias.

Table D.3 runs two specifications of the first stage for Southern white composition change. Column (1) reports first stage relevance when adding decade fixed effects while Column (2) reports relevance adding in predetermined controls interacted by decade dummies, in the same way as with the saturated IV models for Black composition change. The instrument remains relevant but with lower first stage F-statistics. One reason for instrument weakness for another demographic could be that, unlike Black migrants, Southern whites did not base their migrations on the strength of past migrant networks. Evidence for this hypothesis is found again in Stuart and Taylor (2021).

D.4 Other sources of regression model misspecification

Measurement error. The discussion in Section 3.1 considered misclassification error when the outcome is a binary measure of lot size adoption. If actual adoption for a jurisdiction was in 1954 but estimated adoption was in 1956, the definitions in Section 3.1 would lead to a false negative outcome for 1950. I assess the magnitude and direction of bias caused by this error using suggested corrections applicable to linear probability models in Meyer and Mittag (2017).

The idea of Meyer and Mittag (2017)'s proposed correction is to derive expectations of covariates X conditional on there being a false positive or a false negative. If these conditional expectations differ, taking that difference and projecting it on the matrix $(X'X)^{-1}$ identifies bias due to misclassification in expectation. I set the covariate to be the exogenous shift-share migration instrument Z_{ct}^{Black}, and conduct the correction by constructing the same adoption outcomes based on recorded lot size adoption years recorded in my sample of historical zoning data (Section 2.5). The results suggest the reduced form estimate of Z_{ct}^{Black} on adoption is downward biased by -0.049, or that the IV effect size is 37% lower than it may have been absent misclassification.

I also match the Black migration shift-share instrument on the pooled sample of lot size records in historical data. I then regress the percentile-transformed instrument on the log difference between actual recorded lot sizes with their closest bunching bins. The coefficient on this regression is -0.313 with a 95% confidence interval of [-0.77, 0.15]. This finding has ambiguous implications on how residual measurement error in lot size restrictiveness affects the direction of estimated effects. However, negative effects would be driven by larger holding zones in ordinances that cannot be detected from observed bunching. A negative result implies jurisdictions are more willing to exercise discretion in land use controls when racial composition changes in central cities.

Influential metropolitan areas. If average lot size outcomes are driven by a few very large CBSAs where the instrument is confounded with central city characteristics, $t-$tests based on standard errors may overstate the statistical significance of effects. I follow a suggested procedure in Broderick, Giordano and Meager (2021), where dropping random shares in the data conducts a sensitivity analysis of results to outliers.

I run simulations where I reestimate results on adoption and restrictiveness, but every time drop 5% of CBSA-year clusters from the sample. Dropping data at this rate reflects the upper bound of sensitivity analyses conducted in Broderick, Giordano and Meager (2021). Out of 200 simulations, 0 regressions with adoption as the outcome see estimates lose significance at a 95% confidence level. Loss of such significance is also 0 when the outcome is excess mass restrictiveness.

Table D.1: Southern County Determinants of Black Outmigration, 1940–70

	1940–1950	1950–1960	1960–1970
% LF in agriculture	-30.72	-113.01	117.83
Agriculture: % tenant farms 1959	-19.33	-22.63	-95.87
% of planted acres in cotton	-10.85	-12.62	-43.55
Mineral states (TX, OK) × % LF in Mining	190.58	229.48	
Dustbowl county dummy	42.86	12.80	
1930s average precipitation	-6.557		-0.871
# rivers, in total passing through 11-20 k		-1.656	
Tobacco states × % LF in Agriculture			-39.51
Total variables selected	8	8	7

Notes: This table plots LASSO weights for every variable over three separate decade specifications, as long as it received a weight of magnitude greater than 1 in one specification. The total number of variables included in the LASSO model selection step is reported at the bottom. Standard errors and confidence intervals are omitted due to issues with their interpretation for LASSO weights.

Table D.2: Covariate Balance Tests for Right-hand Side Specifications

	(1)		(2)		(3)		(4)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Native white share	-0.066	(0.017)	-0.030	(0.017)	-0.014	(0.017)	-0.015	(0.018)
P-value, Coef. $\neq 0$	0.000		0.075		0.416		0.395	
Black share	0.029	(0.008)	0.035	(0.007)	0.047	(0.007)	0.047	(0.007)
P-value, Coef. $\neq 0$	0.000		0.000		0.000		0.000	
CC manufacturing share	0.042	(0.011)	0.043	(0.010)	-0.012	(0.012)	-0.014	(0.013)
P-value, Coef. $\neq 0$	0.000		0.000		0.337		0.285	
Distance from CBD, km	18.58	(2.18)	15.52	(2.063)	15.14	(2.229)	13.04	(2.287)
P-value, Coef. $\neq 0$	0.000		0.000		0.000		0.000	
Homeownership rate	-0.830	(1.168)	-1.245	(1.043)	-2.519	(1.254)	-4.232	(1.343)
P-value, Coef. $\neq 0$	0.477		0.233		0.045		0.002	
Instrument on RHS			X		X		X	
Migration exposure shares					X		X	
Excluding central cities							X	

Notes: This table compares covariance balance tests, where the percentile transformed right-hand side variables are regressed on 1940 jurisdiction characteristics, across a variety of specifications. Column (2) uses the percentile of shift-share migration instruments instead of endogenous composition change. Columns (3) add state clustered migration exposure shares as controls. Column (4) excludes central cities, or only on the analysis sample of suburbs. Standard errors and P-values for hypothesis testing use clustering at the CBSA-decade level.

Figure D.1: Comparing Realized vs. Projected Southern Migration

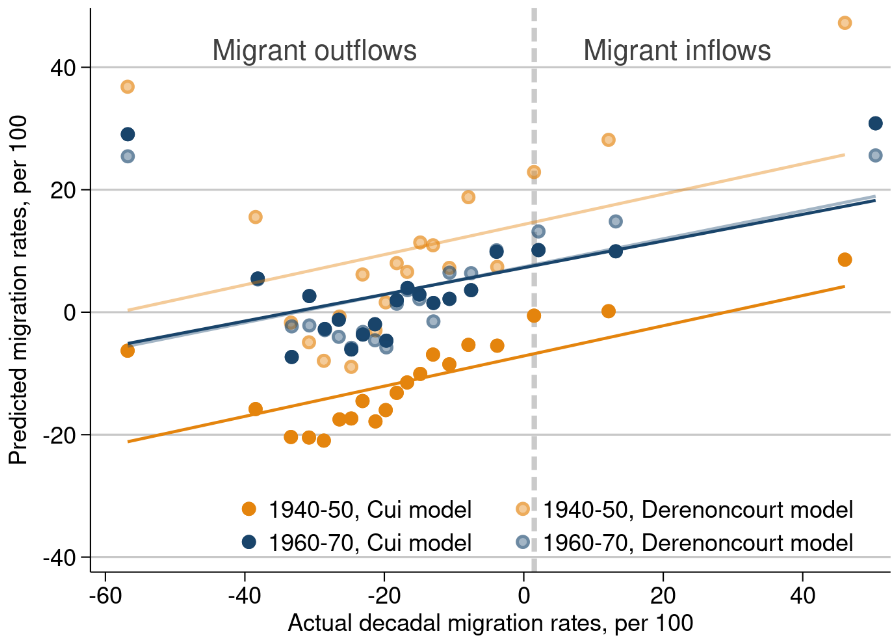


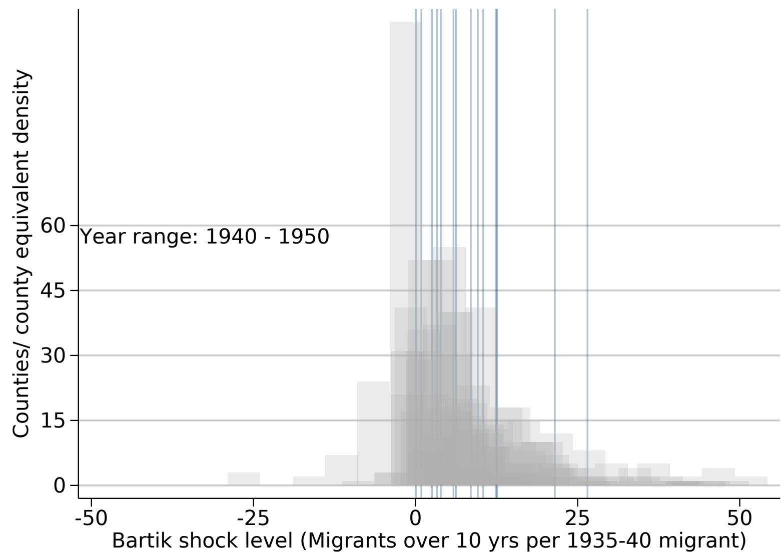
Table D.3: First Stage for Southern White Migration Instrument

	(1)	(2)
Percentile of $Z^{S-white}_{c(j)t}$	0.470	0.546
	(0.0715)	(0.0734)
Panel N	11802	9478
First-stage F	43.28	55.24
R^2	0.216	0.322
Decade FE	X	X
Pre-period controls		X

Notes: This table reports results from the first stage regression of the shift-share Southern white migration instrument in Section 6.3 on central city Southern white composition change. In column (2), I add 1940 rates of Black households, distance from the CBD, homeownership rates and central city manufacturing worker share as pre-period controls, all interacted by decade fixed effects. Standard errors and first stage F-statistics are calculated clustering at the CBSA-decade level.

Figure D.2: Black Migration Shock Levels Within and Between States

(a) Stacked state-level distributions, 1940-50 migration in numerator



(b) Stacked state-level distributions, 1960-70 migration in numerator

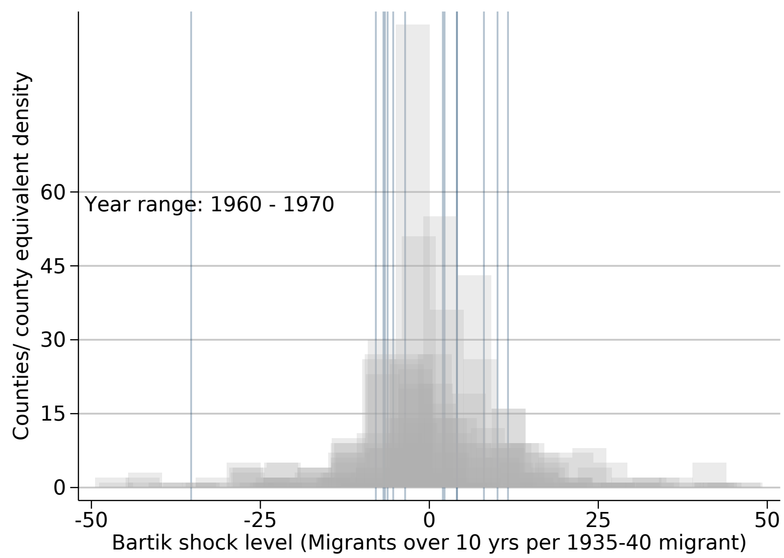
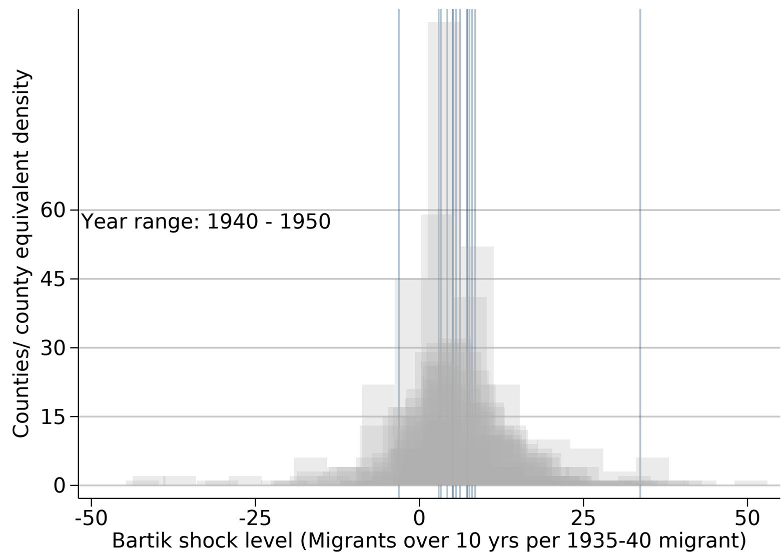


Figure D.3: Southern White Migration Shock Levels Within and Between States

(a) Stacked state-level distributions, 1940-50 migration in numerator



(b) Stacked state-level distributions, 1960-70 migration in numerator

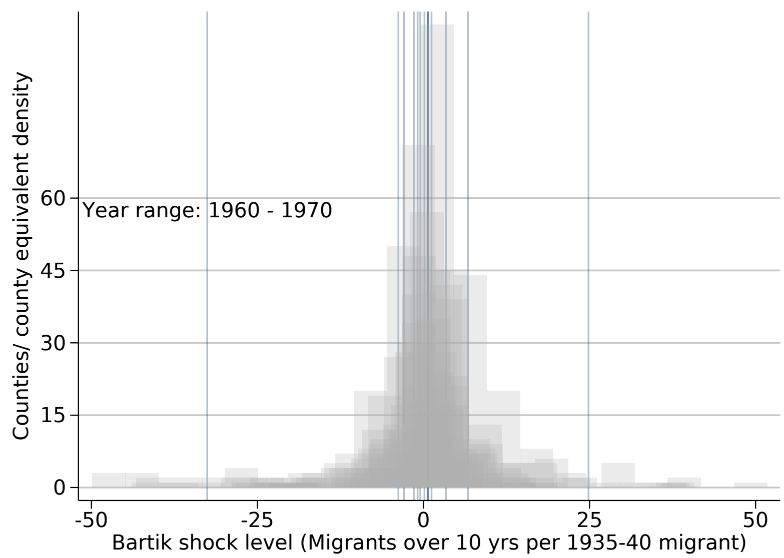
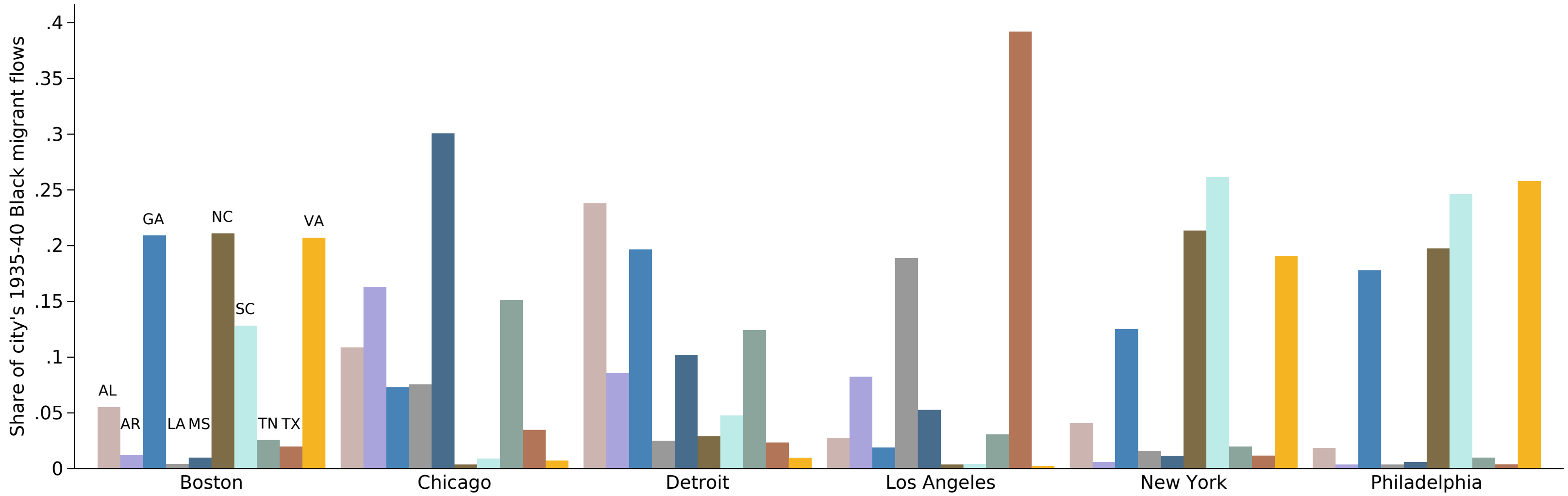


Figure D.4: Central Cities Had Varying 1940 Black Migrant Exposure by State



E Appendix Tables

Table E.1: Migration's Additional Effects on Adoption Outcomes

	OLS		IV		
	(1)	(2)	(1)	(2)	(3)
Panel A: Effects on Large Lot Lot Size Regulation Adoption					
Percentile of $\Delta CC^{Black}_{c(j)t}$	-0.0514	-0.0838*	-0.0996*	-0.224***	-0.162*
	(0.0368)	(0.0402)	(0.0485)	(0.0668)	(0.0650)
Panel N	15105	12024	15105	15105	12024
Baseline mean	0.482	0.486	0.482	0.482	0.486
R^2	0.0942	0.134	0.0937	0.102	0.133
Panel B: Effects on Downzoning Between Periods					
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.0580*	0.0517	0.0176	-0.0466	-0.0327
	(0.0246)	(0.0279)	(0.0333)	(0.0442)	(0.0437)
Panel N	14192	11476	14192	14192	11476
Baseline mean	0.333	0.337	0.333	0.333	0.337
R^2	0.00623	0.0182	0.00581	0.0113	0.0170
Census Region–Year FE	X	X	X	X	X
Migration exposure shares		X		X	X
Pre-period controls		X			X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents regressions of central city Black composition change, as defined in Section 5.1, to additional binary lot size outcomes defined in Section 3.1,

$$Reg_{jt} = \beta \Delta CC^{Black}_{c(j),t} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t},$$

using both OLS specifications and instrumenting for Black composition change with shift-share Black migration instruments. The notes for Table 5 contain further details on the regression specifications.

Table E.2: Migration's Additional Effects on Bunching Mass Restrictiveness

	OLS		IV		
	(1)	(2)	(1)	(2)	(3)
Panel A: Effects on Bunching Share Restrictiveness					
Percentile of $\Delta CC_{c(j)t}^{Black}$	8.518***	7.964***	10.26***	9.277***	9.378***
	(1.242)	(1.218)	(1.737)	(2.025)	(2.020)
Panel N	12902	10594	12902	12902	10594
Baseline mean	14.56	14.91	14.56	14.56	14.91
R^2	0.102	0.138	0.101	0.112	0.138
Census Region–Year FE	X	X	X	X	X
Migration exposure shares		X		X	X
Pre-period controls		X			X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents regressions of central city Black composition change, as defined in Section 5.1, to additional continuous measures of lot size restrictiveness defined in Section 3.2,

$$Reg_{jt} = \beta \Delta CC_{c(j),t}^{Black} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t},$$

using both OLS specifications and instrumenting for Black composition change with shift-share Black migration instruments. The notes for Table 5 contain further details on the regression specifications.

Table E.3: Causal Effects of Demographic Changes On Lot Size Adoption

	Δ Black		Δ South. White		Δ Foreigner
	Reduce.	IV	Reduce.	IV	OLS
Percentile of $\Delta CC^{Dem}_{c(j)t}$	0.195***	0.280***	-0.0619	-0.0715	-0.178***
	(0.0343)	(0.0513)	(0.0326)	(0.0657)	(0.0387)
Panel N	12083	12024	12083	9772	12083
Baseline mean	0.698	0.647	0.421	0.364	0.397
R^2	0.140	0.130	0.0893	0.0991	0.105
Census Region–Year FE	X	X	X	X	X
Central city pop. exposure	X	X	X	X	X
Pre-period controls	X	X	X	X	X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents regressions of central city composition change on lot size adoption, a binary variable defined in Section 3.1, across measures of composition change for multiple demographic groups,

$$Excess_{jt} = \beta \Delta CC^{Dem}_{c(j),t} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}.$$

The notes for Table 6 contain further details on the regression specifications.

Table E.4: Results in Table 5, With Early Postwar Weights

	OLS		IV		
	(1)	(2)	(1)	(2)	(3)
Panel A: Effects on Lot Size Regulation Adoption					
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.162***	0.0426	0.231***	0.202**	0.266**
	(0.0276)	(0.0324)	(0.0523)	(0.0754)	(0.0865)
Panel N	15105	11425	14093	14093	11425
Baseline mean	0.770	0.782	0.770	0.770	0.782
R^2	0.0841	0.135	0.0780	0.108	0.112
Panel B: Effects on Excess Mass Restrictiveness					
Percentile of $\Delta CC^{Black}_{c(j)t}$	8.008***	5.274***	9.547***	10.20***	10.65***
	(0.986)	(1.109)	(1.672)	(2.598)	(2.874)
Panel N	12902	10577	12879	12879	10577
Baseline mean	6.102	6.347	6.102	6.102	6.347
R^2	0.0658	0.0999	0.0582	0.0701	0.0864
Census Region–Year FE	X	X	X	X	X
Migration exposure shares		X		X	X
Pre-period controls		X			X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents regressions of exposure to central city Black migration on lot size outcomes, following the specifications in Table 5:

$$Reg_{jt} = \beta \Delta CC^{Black}_{c(j),t} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t},$$

additionally reweighing the sample by surviving units in the jurisdiction built from 1930-1950, as a share of properties from that period across the CBSA. The notes for Table 5 contain further details on the regression specifications.

Table E.5: Effects in Table 6, With Early Postwar Weights

	Δ Black		Δ South. White		Δ Foreigner
	Reduce.	IV	Reduce.	IV	OLS
Percentile of $\Delta CC^{Dem}_{c(j)t}$	5.087***	10.65***	-0.795	-1.367	-2.330
	(1.331)	(2.874)	(0.883)	(3.644)	(1.294)
Panel N	10615	10577	10615	8716	10615
Baseline mean	0.698	0.647	0.421	0.364	0.397
R^2	0.0988	0.0864	0.0426	0.0205	0.0532
Census Region–Year FE	X	X	X	X	X
Central city pop. exposure	X	X	X	X	X
Pre-period controls	X	X	X	X	X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents regressions of central city composition change on lot size restrictiveness, a continuous measure defined in Section 3.2, across measures of composition change for multiple demographic groups,

$$Excess_{jt} = \beta \Delta CC^{Dem}_{c(j),t} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}$$

additionally reweighing the sample by surviving units in the jurisdiction built from 1930-1950, as a share of properties from that period over a CBSA. The notes for Table 6 contain further details on the regression specifications.

Table E.6: Effects in Table E.3, With Postwar Weights

	Δ Black		Δ South. White		Δ Foreigner
	Reduce.	IV	Reduce.	IV	OLS
Percentile of $\Delta CC^{Dem}_{c(j)t}$	0.124***	0.266***	-0.0277	-0.0605	-0.100**
	(0.0386)	(0.0865)	(0.0277)	(0.0431)	(0.0362)
Panel *N*	11473	11425	11473	9292	11473
Baseline mean	0.698	0.647	0.421	0.364	0.397
R^2	0.152	0.112	0.108	0.0694	0.113
Census Region–Year FE	X	X	X	X	X
Central city pop. exposure	X	X	X	X	X
Pre-period controls	X	X	X	X	X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents regressions of central city composition change on lot size adoption, a binary variable defined in Section 3.1, across measures of composition change for multiple demographic groups,

$$Excess_{jt} = \beta \Delta CC^{Dem}_{c(j),t} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}$$

additionally reweighing the sample by surviving units in the jurisdiction built from 1930-1950, as a share of properties from that period across the CBSA The notes for Table 6 contain further details on the regression specifications.

Table E.7: Robustness of Migration's Effects On Lot Size Adoption

	(1)	(2)	(3)	(4)
Pctile. $\Delta CC^{Black}_{c(j)t}$ (OLS)	0.152***	0.157***	0.153***	
	(0.0312)	(0.0324)	(0.0347)	
$\Delta CC^{Black}_{c(j)t}$ (OLS)				0.0118***
				(0.00237)
Panel N	12024	12024	11921	11884
R^2	0.133	0.134	0.135	0.134
Pctile. $\Delta CC^{Black}_{c(j)t}$ (IV)	0.280***	0.264***	0.316***	
	(0.0513)	(0.0469)	(0.0600)	
$\Delta CC^{Black}_{c(j)t}$ (IV)				0.0264**
				(0.00645)
Panel N	12024	12024	11921	11884
R^2	0.130	0.131	0.130	0.127
Main specification controls	X	X	X	X
Southern white IV control		X		
Suburban growth control			X	
No pctile transformation				X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents alternate specifications showing results in Table 5 are robust where lot size adoption is the outcome variable. I show three alternative regression specifications also in Table 7 in which main results remain statistically significant. I exclude misspecification models that do not apply to the binary adoption outcome. The notes for Table 7 contain further details on the regression specifications.

Table E.8: Other Anti-Discrimination Laws Do Not Explain Effects

	Lot Size Adoption			Excess Mass Restrictiveness		
	OLS	IV		OLS	IV	
Panel A: Total Anti-Discrimination Laws						
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.120***	0.203***	0.329***	7.647***	3.926***	6.782***
	(0.0327)	(0.0402)	(0.0451)	(0.955)	(1.140)	(1.241)
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.000	0.000	-0.002	-0.0382	0.074	-0.005
×Anti-Discrim Laws	(0.00135)	(0.00147)	(0.00159)	(0.0418)	(0.0465)	(0.0448)
Anti-Discrim Laws	0.001	0.001	0.002	0.0488	-0.0347	0.0223
	(0.00117)	(0.00115)	(0.00131)	(0.0324)	(0.0367)	(0.0357)
Panel N	11958	15030	11958	10537	12842	10537
Baseline mean	0.783	0.771	0.783	6.352	6.105	6.352
R^2	0.127	0.0769	0.104	0.0988	0.0496	0.0788
Panel B: Total Fair Labor Laws						
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.132***	0.249***	0.367***	6.972***	2.860**	6.170***
	(0.0369)	(0.0406)	(0.0469)	(0.980)	(1.015)	(1.181)
Percentile of $\Delta CC^{Black}_{c(j)t}$	-0.0103	-0.0345**	-0.0359*	0.407	1.440**	0.814
×# Labor Laws	(0.0104)	(0.0108)	(0.0145)	(0.367)	(0.493)	(0.537)
# Labor Laws	0.0219**	0.0303***	0.0362***	-0.160	-0.792**	-0.476
	(0.00787)	(0.00752)	(0.00961)	(0.236)	(0.293)	(0.322)
Panel N	11958	15030	11958	11963	10537	12842
10537						
Baseline mean	0.783	0.771	0.783	0.783	6.352	6.105
6.352						
R^2	0.130	0.0791	0.107	0.108	0.0989	0.0545
0.0798						
Census Region–Year FE	X	X	X	X	X	X
Central city pop. exposure	X	X	X	X	X	X
Pre-period controls	X		X	X		X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents heterogeneous effects by 1950 anti-discrimination legislation across non-Southern states. I estimate interacted effects in the regression model:

$$Reg_{jt} = \beta \Delta CC^{Black}_{c(j),t} + \alpha^{ADE} \text{ ADE State}_{s(j)} + \beta^{ADE} \Delta CC^{Black}_{c(j),t} \times \text{ ADE State}_{s(j)} + \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t},$$

on the lot size adoption and restrictiveness effects in Table 5. I use two measures of 1950 legislation: the number of fair labor laws recorded by Murray (1950) and the total number of anti-discrimination laws in the same book, as recorded by Cook et al. (2022). The notes for Table 8 contain further details on the regression specifications.

Table E.9: Effects in Table 8, With Postwar Weights

	Lot Size Adoption			Excess Mass Restrictiveness		
	OLS	IV		OLS	IV	
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.00984	0.180*	0.243**	3.696**	3.699	4.739
	(0.0402)	(0.0707)	(0.0779)	(1.304)	(2.516)	(2.708)
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.093	0.0939	0.0834	4.596*	9.622**	10.04**
$\times\mathbf{1}$[ADE state]	(0.0612)	(0.097)	(0.107)	(1.818)	(3.364)	(3.364)
$\mathbf{1}$[ADE state]	-0.0801	-0.0569	-0.0793	-2.208	-4.616*	-4.906**
	(0.0542)	(0.0652)	(0.0687)	(1.244)	(1.745)	(1.854)
Panel N	11365	14030	11365	10520	12819	10520
Baseline mean	0.783	0.771	0.783	6.352	6.105	6.352
R^2	0.134	0.0783	0.0856	0.0972	0.0508	0.0601
Census Region–Year FE	X	X	X	X	X	X
Central city pop. exposure	X	X	X	X	X	X
Pre-period controls	X		X	X		X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents heterogeneous effects by 1950 state-level public education policies on the lot size restrictiveness effects in Table 5. I define 16 states that were early adopters of an anti-discrimination law in public education ("ADE state"), as recorded by Murray (1950). I then estimate the interacted the regression model:

$$Reg_{jt} = \beta\Delta CC^{Black}_{c(j),t} + \alpha^{ADE}\text{ ADE State}_{s(j)} + \beta^{ADE}\Delta CC^{Black}_{c(j),t} \times \text{ ADE State}_{s(j)}$$
$$+ \delta_t + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}.$$

additionally reweighing the sample by surviving units in the jurisdiction built from 1930-1950, as a share of properties from that period across the CBSA. The notes for Table 8 contain further details on the regression specifications.

Table E.10: Table 8 Effects Not Driven by Particular States

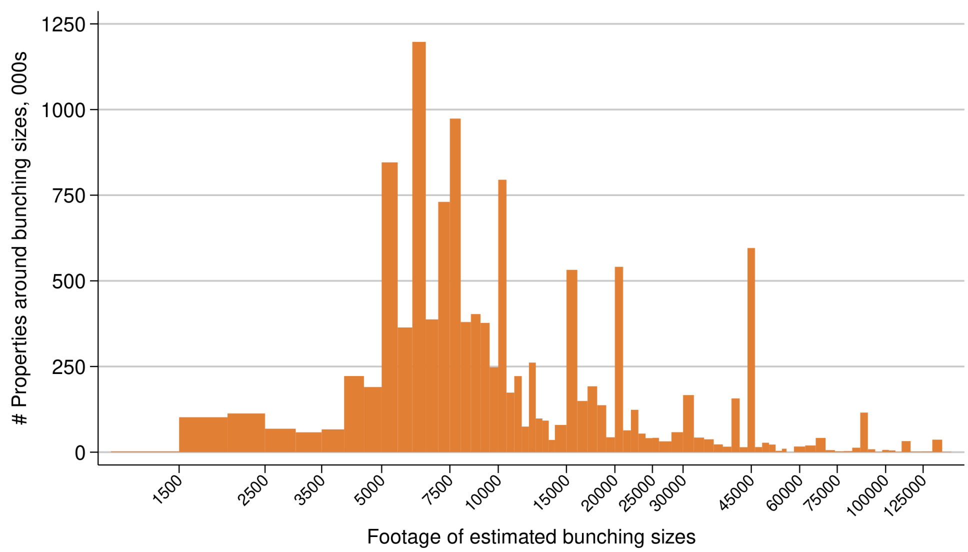
	(1)	(2)	(3)	(4)
Percentile of $\Delta CC^{Black}_{c(j)t}$ (OLS)	5.782***	5.886***	5.694***	5.368***
	(1.059)	(1.067)	(1.042)	(1.081)
Percentile of $\Delta CC^{Black}_{c(j)t}$	3.541*	3.246*	2.954*	3.108
$\times\mathbf{1}$[ADE state] (OLS)	(1.581)	(1.648)	(1.254)	(1.711)
$\mathbf{1}$[ADE state] (OLS)	-2.208*	-1.988	-1.928*	-2.100*
	(1.024)	(1.067)	(0.831)	(1.067)
Panel N	10537	9827	8822	9634
R^2	0.0988	0.0958	0.0980	0.0972
Percentile of $\Delta CC^{Black}_{c(j)t}$ (IV)	4.177*	4.178*	3.500*	4.736**
	(1.629)	(1.633)	(1.622)	(1.625)
Percentile of $\Delta CC^{Black}_{c(j)t}$	6.172**	5.725*	6.134**	6.008*
$\times\mathbf{1}$[ADE state] (IV)	(2.239)	(2.262)	(2.099)	(2.350)
$\mathbf{1}$[ADE state] (IV)	-4.196**	-3.928*	-4.302**	-4.221**
	(1.575)	(1.600)	(1.499)	(1.609)
Panel N	10537	9827	8822	9634
R^2	0.0787	0.0752	0.0739	0.0767
Controls	X	X	X	X
No MA, NJ		X		
No NY			X	
No adoptions in 1920-1940				X

Significance levels: * = 5%; ** = 1%; *** = 0.1%.

Notes: This table presents heterogeneous effects by 1950 state-level public education policies in Table 8, subject to other sample specifications as described in Section 7.3. As in Table 7, I present OLS and IV estimates for each sample specification. The notes for Table 8 contain further details on the regression specifications.

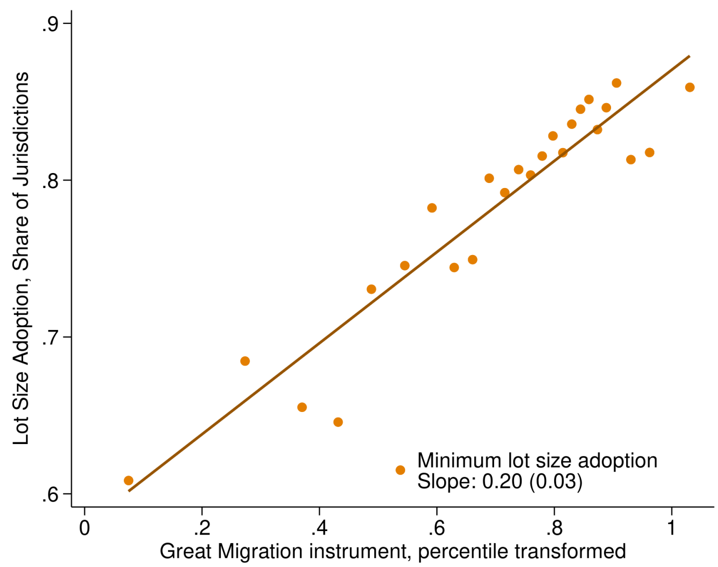
F Appendix Exhibits

Figure F.1: Distribution of Properties Around Estimated Bunching Bins



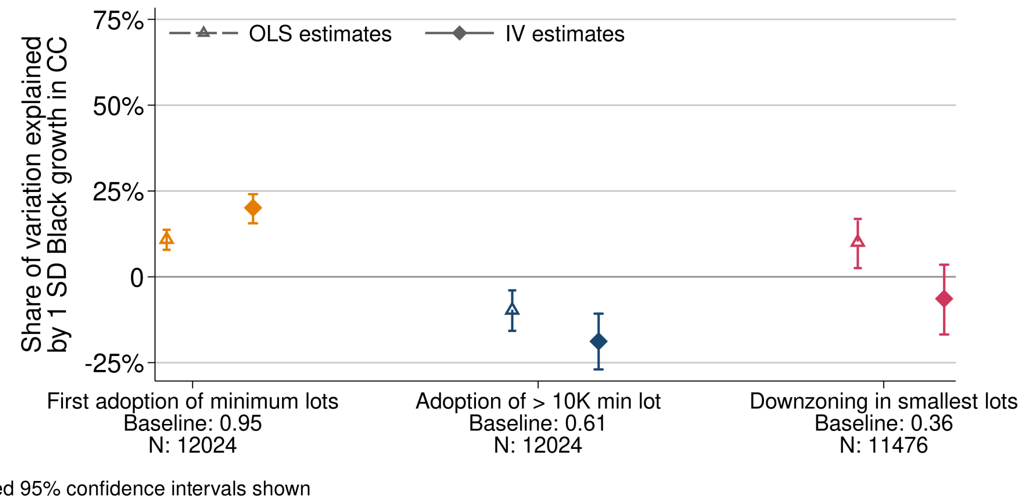
Notes: This figure plots the number of properties whose lot sizes are around the bunching bins detected in Section 2, and which Corelogic tracks as built after the estimated year of adoption. The sample includes properties built from 1925 to the present, covering 7153 distinct zoning jurisdictions. Histogram bandwidths are defined over the constant partition of lot ranges listed in Appendix Section B.

Figure F.2: Lot Size Adoption Linked With Black Composition Change



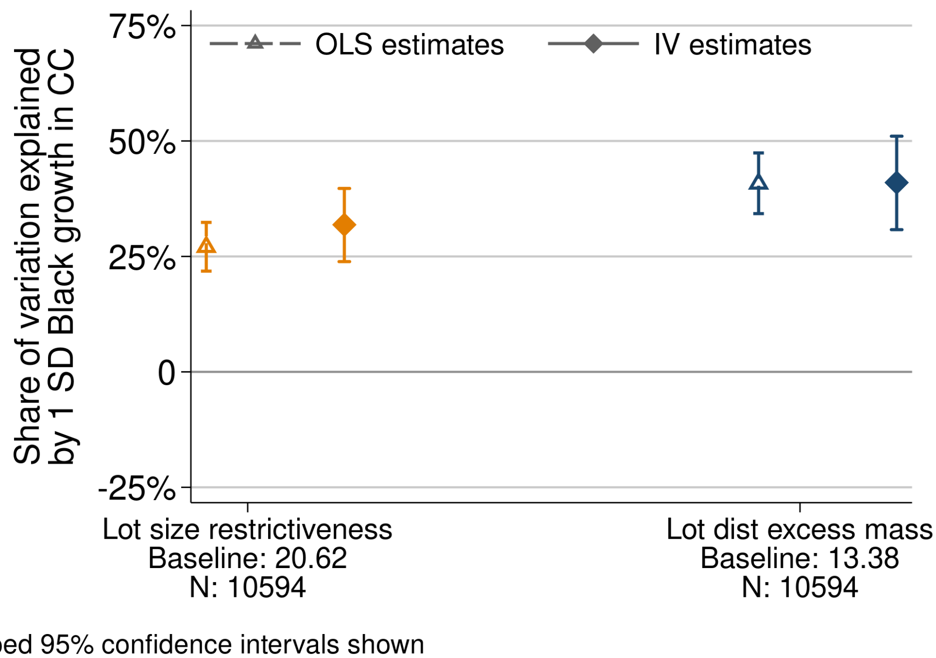
Notes: This figure plots reduced form, nonparametric relationships between the shift-share instrument described in Section 5.3 and a dummy indicating if the jurisdiction adopted lot size controls around the decade's second half. The source data is a panel of non-central city jurisdictions in CBSAs outside of 14 Southern states. The instrument is first transformed from levels to percentiles of each decade's distribution, as in Derenoncourt (2022), then residualized on share exposure variables as described in Section 5.3 and additional controls. Control variables include the CBSA central city's manufacturing share, and analysis sample cities' 1940 black share, homeownership rates and distance to CBD, interacted by period. Reported standard errors are clustered at the CBSA-decade level.
Sources: Calculations from NHGIS Tables (Manson et al. (2021)), Ruggles et al. (2022), CCDB, IPUMS 1940 full count Census (Ruggles et al. (2021)), Boustan (2016), Derenoncourt (2022) and CoreLogic Tax Records.

Figure F.3: Share of Extensive Margin Outcomes Explained by Black Composition Change



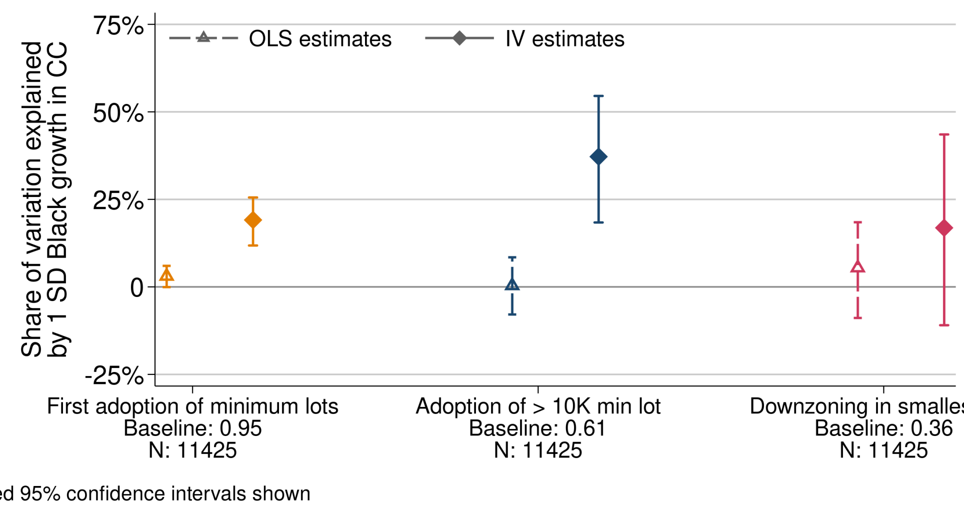
Notes: This figure presents an aggregation exercise, converting the regression coefficients estimated in Table 5 into how much the Second Great Migration explained lot size outcomes in non-Southern metropolitan areas. The aggregation is a three-step procedure detailed in Section 6.2. Outcomes are binary outcomes on the extensive margin of which lot size controls were adopted, as defined in Section 3.1. The specifications used include the full set of controls explained in Table 5. 95% confidence intervals are bootstrapped using random weights on CBSA-decade clusters, following the Bayesian bootstrap of Rubin (1981). For each outcome, I display the sample sizes for the regression specifications and the housing construction-weighted averages for each outcome.

Figure F.4: Share of Restrictiveness Outcomes Explained by Black Composition Change



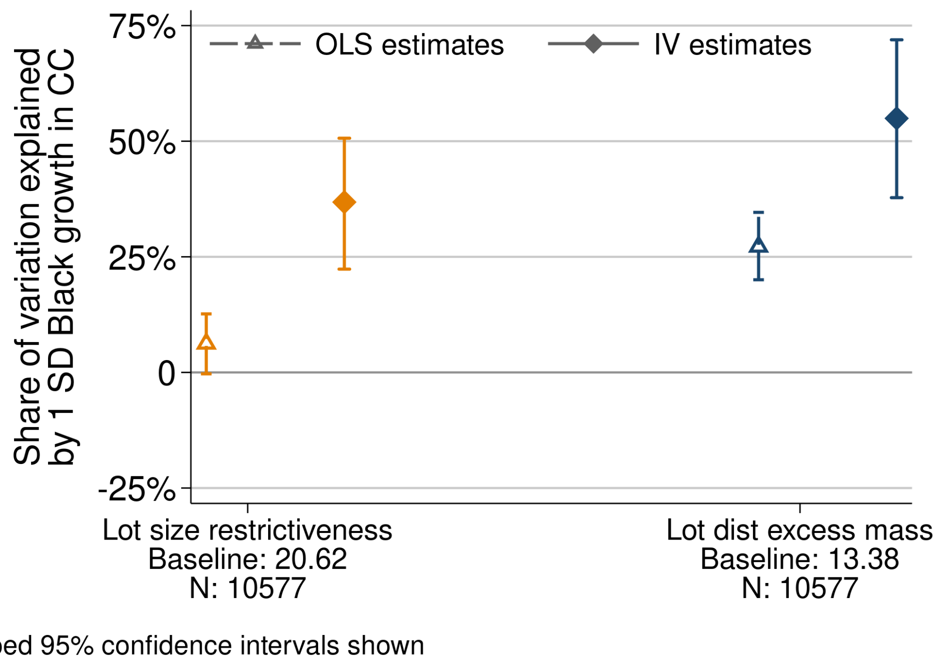
Notes: This figure presents an aggregation exercise, converting the regression coefficients estimated in Table 5 into how much the Second Great Migration explained lot size outcomes in non-Southern metropolitan areas. The outcomes in this figure are continuous measures of lot size restrictiveness as defined in Section 3.2. The notes for Figure F.3 contain further details on aggregated estimates.

Figure F.5: Share of Extensive Margin Outcomes Explained by Black Composition Change, Early Postwar Weights



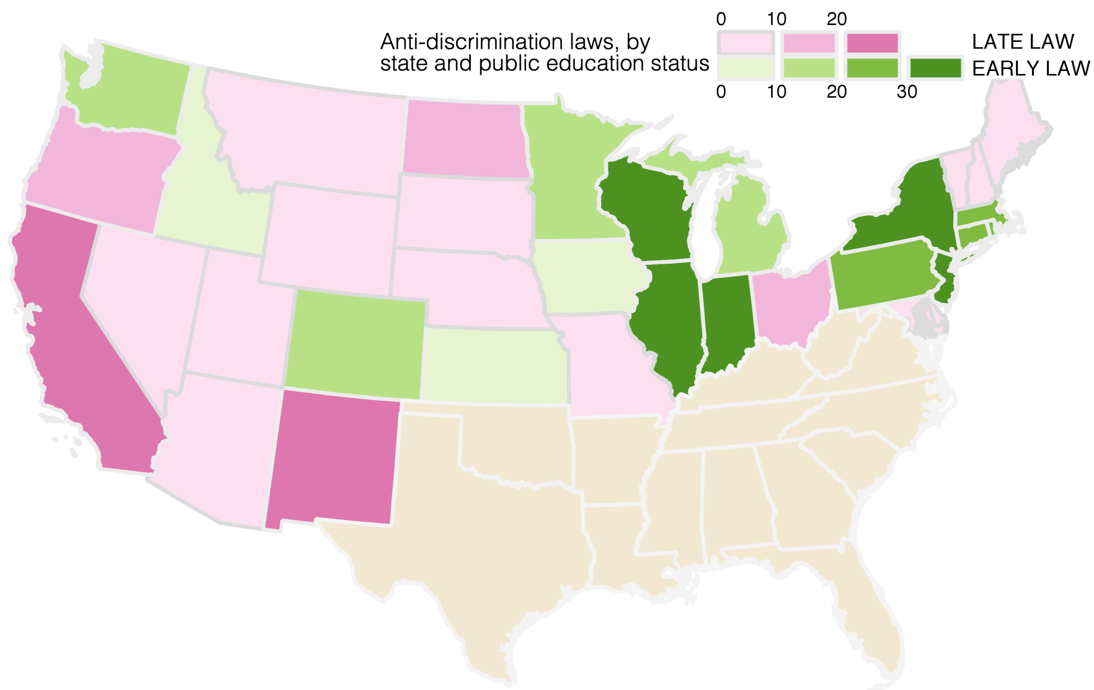
Notes: This figure presents an aggregation exercise, converting the regression coefficients estimated in Table E.4 into how much the Second Great Migration explained lot size outcomes in non-Southern metropolitan areas. Compared to Figure F.3, the regression coefficient changes because it is estimated from a regression using early postwar housing construction weights. The notes for Figure F.3 contain further details on aggregated estimates.

Figure F.6: Share of Restrictiveness Outcomes Explained by Black Composition Change, Postwar Weights



Notes: This figure presents an aggregation exercise, converting the regression coefficients estimated in Table E.4 into how much the Second Great Migration explained lot size outcomes in non-Southern metropolitan areas. Compared to Figure F.4, the regression coefficient changes because it is estimated from a regression using early postwar housing construction weights. The notes for Figure F.3 contain further details on aggregated estimates.

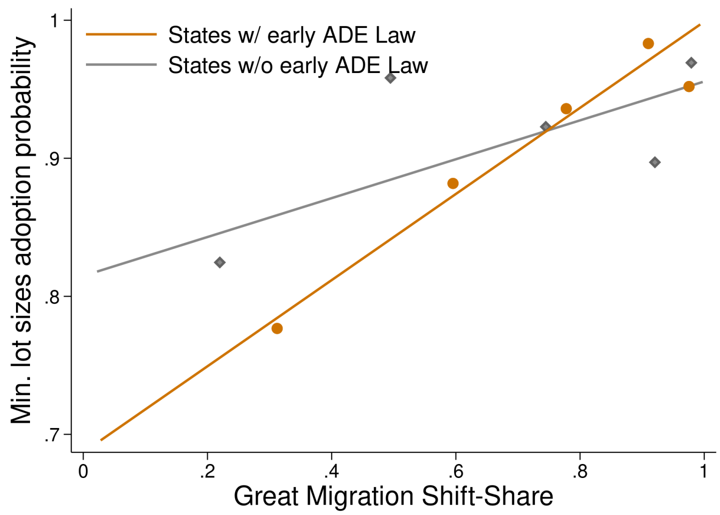
Figure F.7: States by 1950 Anti-Discrimination Legislation



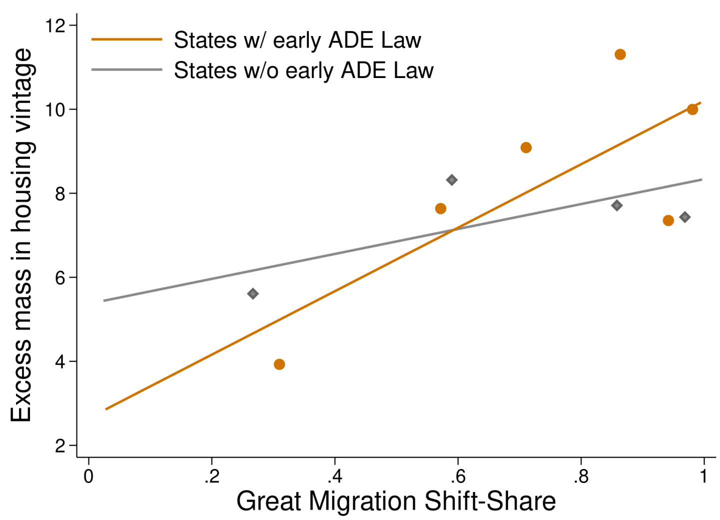
Notes: This bivariate map plots two variables based on data in Murray (1950). Non-Southern states are categorized based on whether they passed a law prohibiting discriminaton in public education by 1950 (pink vs. green), and then on a continuous scale by the number of anti-discrimination laws as coded in Cook et al. (2022).

Figure F.8: Lot Size Outcomes — Black Composition Change Across Legal Regimes

(a) Effects on Lot Size Adoption



(b) Effects on Excess Mass Restrictiveness



Notes: The two panels of this figure plot reduced-form, nonparametric relationships based on the interacted regression specification discussed in Section 7.2. I separately estimate the effect of central city Black composition change for states that adopted an anti-discrimination law in education ("ADE") law by 1950 and states that did not. The X-axis variable is the Black migration instrument defined in Section 5.3. The models for all regressions include the full set of controls as specified in Table 8.