For Online Publication

A Appendix Tables

Table A.1: Coverage of Jurisdictions with Estimated Adoption

	N	% All Jurisdictions	Population-weighted %
Incorporated Cities	4199	47%	92%
Counties With Zoning Power	657	72%	87%
Townships With Zoning Power	2225	41%	74%
Newly Annexed Portions of Cities	1468	—	—
Distinct Jurisdictions	7082		

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 A.2: Shock-level summary statistics for Southern Migration instruments

	Black migration		White migration
	Ratios	Percentiles	Ratios
Mean	6.34	0.52	-0.17
Mean, Urban counties	6.48	0.51	-2.40
Standard deviation	10.3	0.29	10.9
First and third quartiles	[0.030, 9.67]	[0.278, 0.773]	[-3.03, 4.08]
Effective sample size (inverse of exposure weights HHI)			
	369.7	369.7	418.8
Largest exposure weight s_{kt}			
	0.0119	0.0119	0.0120

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

	Δ Black		Δ South. White		Δ Foreigner
	Reduce.	IV	Reduce.	IV	OLS
Panel A: Sample varies by specification					
Percentile of $\Delta CC^{Dem}_{c(j)t}$	0.158*	0.402*	0.008	0.038	-0.003
	(0.0828)	(0.237)	(0.0082)	(0.0448)	(0.0056)
Panel N	3418	3418	3952	3952	561
First-stage F-stat		20.46		12.42	
Anderson-Rubin p-value	0.0853	0.0853	0.341	0.341	
Baseline mean	0.534	0.534	0.534	0.534	
Panel B: Common Black destination sample					
Percentile of $\Delta CC^{Dem}_{c(j)t}$	-0.0522		0.003		-0.012
	(0.121)		(0.0070)		(0.0107)
Panel N	3344	3344	3948	3948	236
Baseline mean	0.534	0.534	0.534	0.534	
Panel Units	Shock	Shock	Shock	Shock	Metro
Region–Decade FE	X	X	X	X	X
Incomplete shares	X	X	X	X	X
Pre-period controls	X	X	X	X	X

Significance levels: * = 10%; ** = 5%; *** = 1%.

Notes: This table presents regressions of central city composition change on restrictive lot size adoption, a binary variable defined in Section 3, across measures of composition change for multiple demographic groups,

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

The notes for Table 4 contain further details on the regression specifications.

Table A.4: Robustness of Migration's Effects on Lot Size Excess Mass

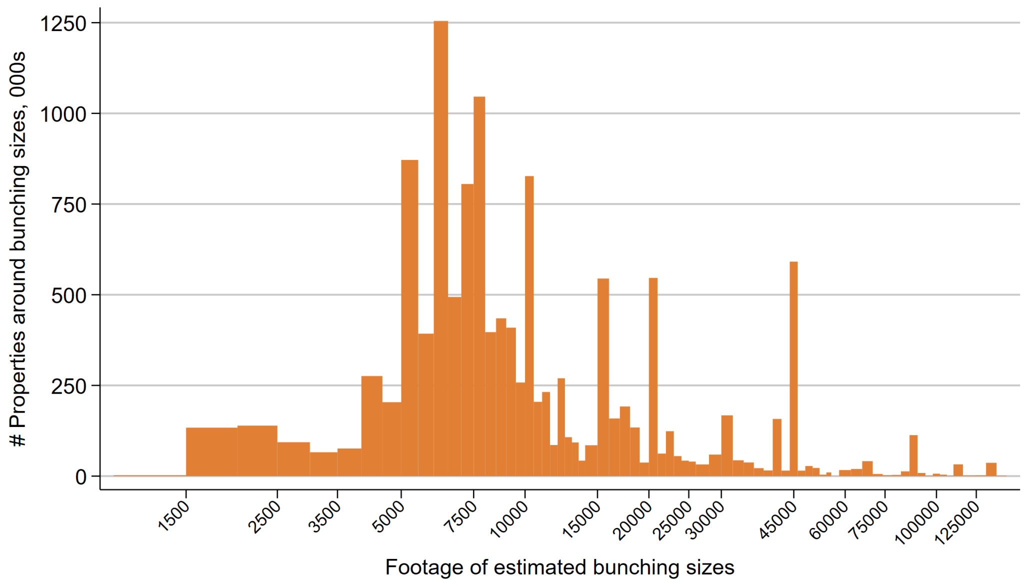
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pctile. $\Delta CC_{c(j)t}^{Black}$ (IV)	10.47***	10.41***	2.346	11.28*	8.520***	8.768*	
	(3.853)	(3.907)	(7.069)	(6.181)	(2.827)	(4.501)	
$\Delta CC_{c(j)t}^{Black}$ (IV)							3.049***
							(1.101)
Panel N	3418	3418	3403	3402	3418	3418	3418
Metros in sample	561	561	468	495	561	561	561
First-stage F-stat	19.51	19.09	8.770	15.76	19.94	19.63	22.56
Anderson-Rubin p-value	0.00755	0.00889	0.746	0.0750	0.001	0.0460	0.00755
Panel N	3418	3418	3403	3402	3418	3418	3418
Metros in sample	561	561	468	495	561	561	561
Main specification controls	X	X	X	X		X	X
Southern white IV control		X					
CC density control			X				
Most anomalous bunching				X			
Jurisdiction FE model					X		
5-year lagged demographics						X	
No pctile transformation							X

Significance levels: * = 10%; ** = 5%; *** = 1%.

Notes: This table presents alternate specifications showing results in Table 3 are robust where lot size adoption is the outcome variable. I show three alternative regression specifications also in Table 5 in which main results remain statistically significant. I exclude misspecification models that do not apply to the binary adoption outcome. The notes for Table 5 contain further details on the regression specifications.

B Appendix Exhibits

Figure B.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 7082 distinct zoning jurisdictions. Histogram bandwidths are defined over the constant partition of lot ranges listed in Appendix Section D.

Figure B.2: Persistence of Postwar Lot Size Restrictiveness

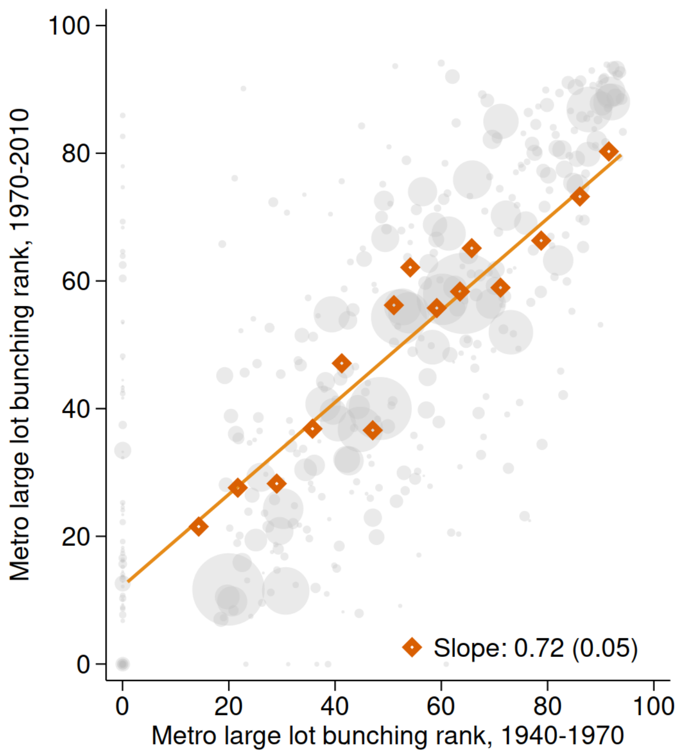
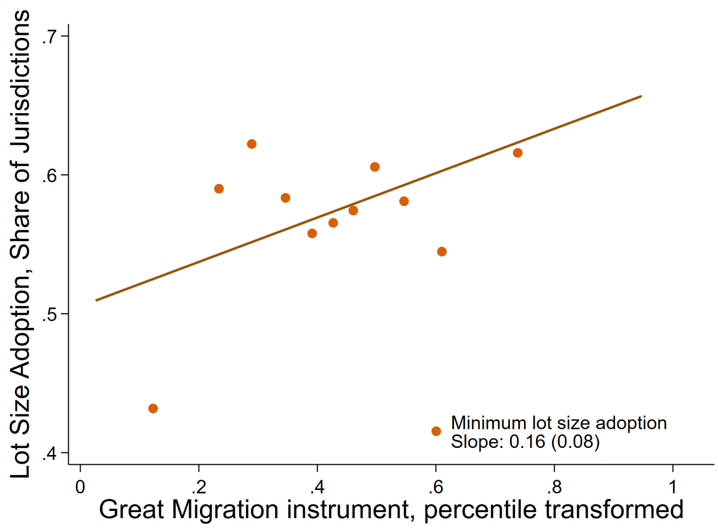
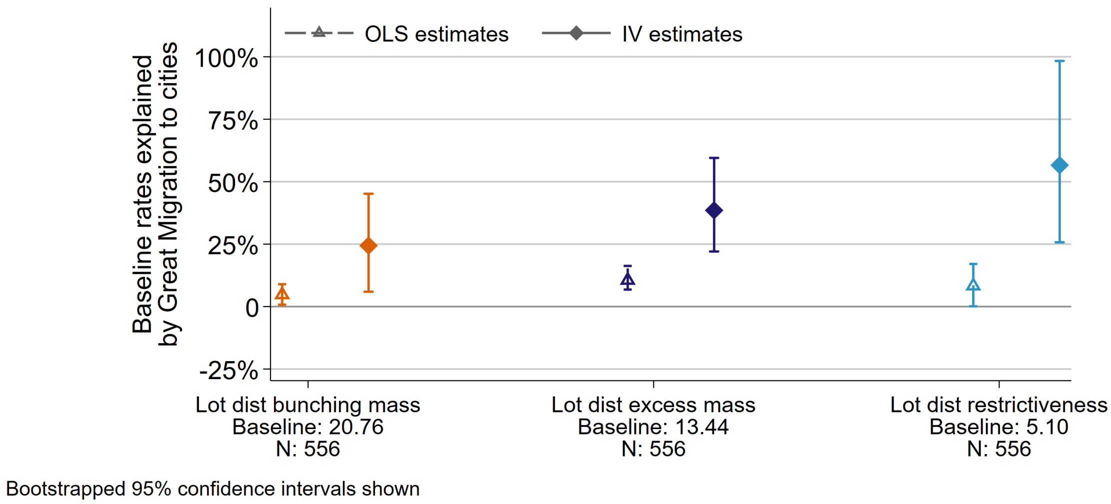


Figure B.3: Lot Size Adoption Linked With Black Composition Change



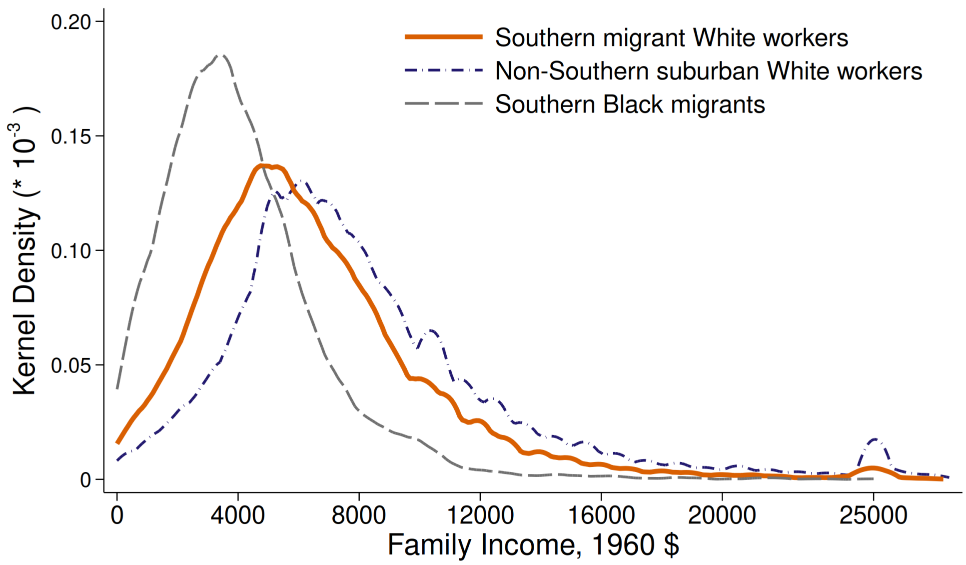
Notes: This figure plots reduced form, nonparametric relationships between the shift-share instrument described in Section 4.2 and a dummy indicating if the jurisdiction adopted lot size controls around the decade's second half. The source data is a panel of non-central city jurisdictions in CBSAs outside 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 4.2 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 B.4: Share of Lot Size Restrictiveness Measures Explained by Great Migration



Notes: This figure presents an aggregation exercise, converting the regression coefficients estimated in Table 3 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 3.1. The specifications used include the full set of controls explained in Table 3. 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.

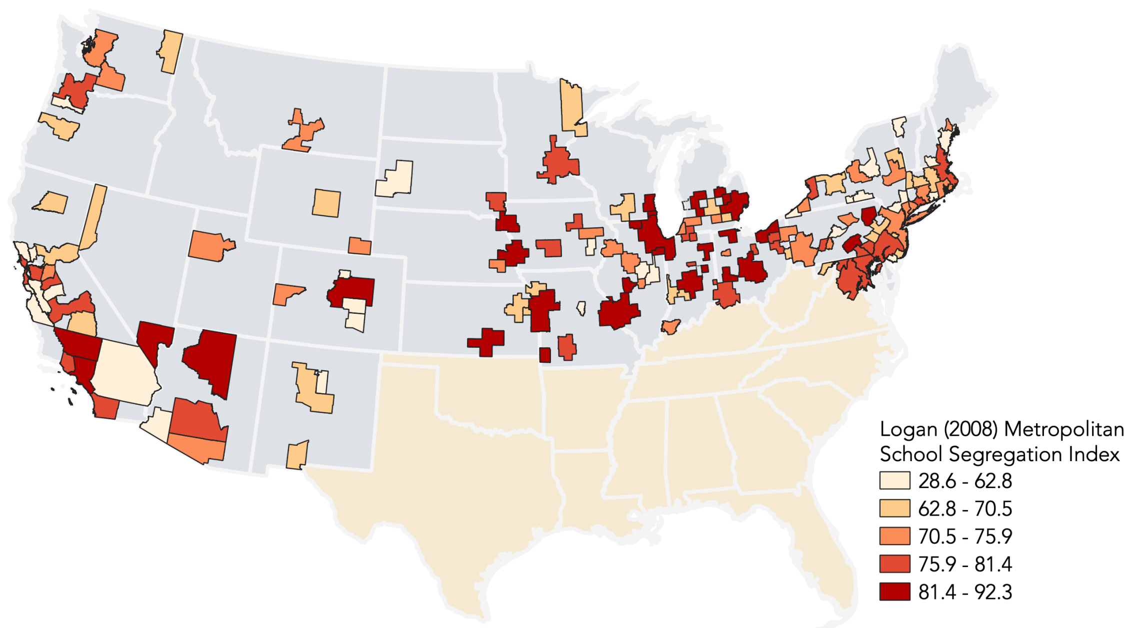
Figure B.5: Household Incomes Within and Between Racial Groups



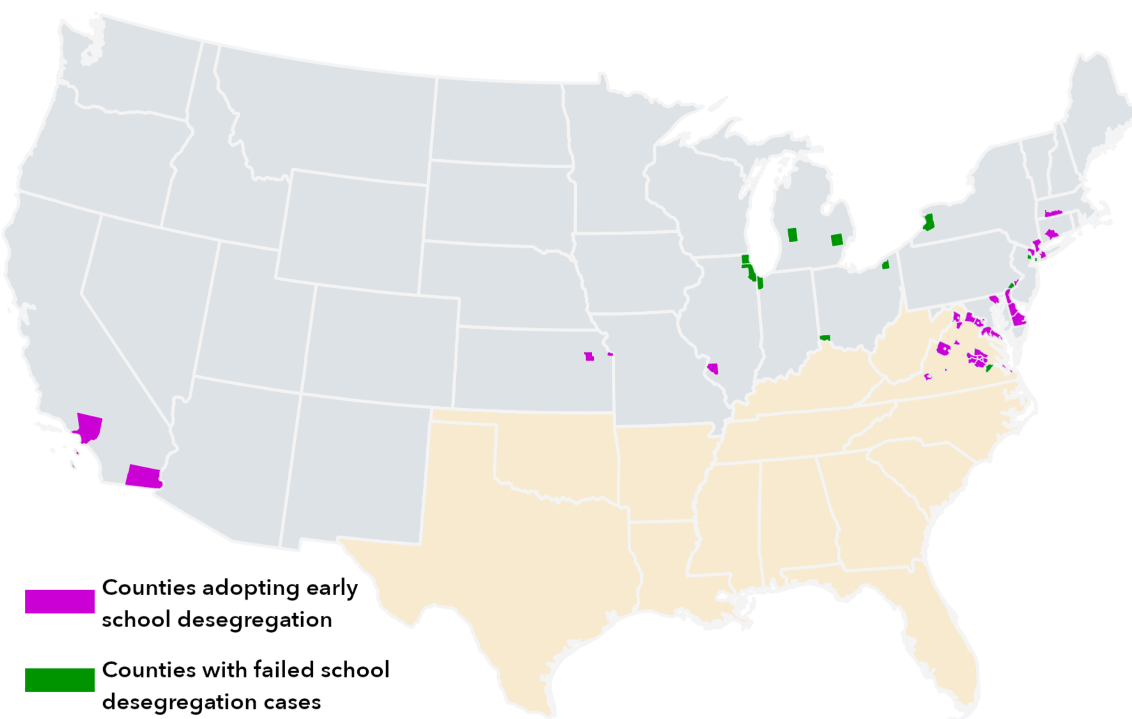
Notes: This figure plots household income distributions in 1960 across three racial and demographic groups, for Northern central cities where the data are available. Household income is defined as self-reported family income in the 1960 Census and is not adjusted for inflation. Southern migrant workers 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).

Figure B.6: The geography of postwar school desegregation

(a) Metro areas by degree of Black school segregation



(b) Counties with early court-ordered desegregation cases



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).

C Details of Data Build

C.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.

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

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

C.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 C.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 Manson et al. (2021) 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 C.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 C.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 D.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]

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.

[2] As an example, the suburb of Mercer Island in King County, WA incorporated in 1960 but did not change its zoning far from the County's zoning in the 1940s. Generically, however, this assumption places a lower bound on how many lot size controls a newly jurisdictions will have, even if they move to simplify the zoning code.

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.

C.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 C.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 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

[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.

reports the “Up to 1940” and “1940–1969” rows in the County column of Table C.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 C.2 conducts a robustness test by looking at CoreLogic units relative to only units still existing in the 2000 Long-form Census. Unlike Table C.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)).

C.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 C.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 C.3 show a table of lot size controls surveyed by New Jersey planners in 1960.

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

C.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:** 1940 shares tabulated from Bazzi et al. (2023); 1950-1960 levels in 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.

C.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).

C.8 Case Studies of Jurisdictions

Zoning in Lower Merion Township. Lower Merion passed its first zoning ordinance in 1927 without minimum lot sizes. According to Lower Merion Township (1937), the suburb's first planning commission recommended major zoning revisions to "maintain community appearance and the stability of property values." Attached row houses would be banned as a residential use as it is not "necessary for suburban housing." Elsewhere, minimum lot sizes should be "raised to equal that maintained by present development or by deed restrictions" on newer homes. The revised ordinance in 1939 deviated little from the commission's report, setting up five different minimum lot sizes up to a maximum of 30,000 square feet.

While the township did add lot sizes of an acre and two acres in the 1960s, that came more than two decades after the planning commission first recommended a one acre minimum lot size in the original report. There is high persistence between the zoning boundaries passed as part of the 1939 revision and the zoning map as of 2014. As of 2020, the township passed a *form-based zoning code* which still applies the same lot size requirements to areas that had been zoned for 30,000 square feet, but loosens regulations for dense development on blocks close to existing commuter rail routes.

Table C.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 C.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

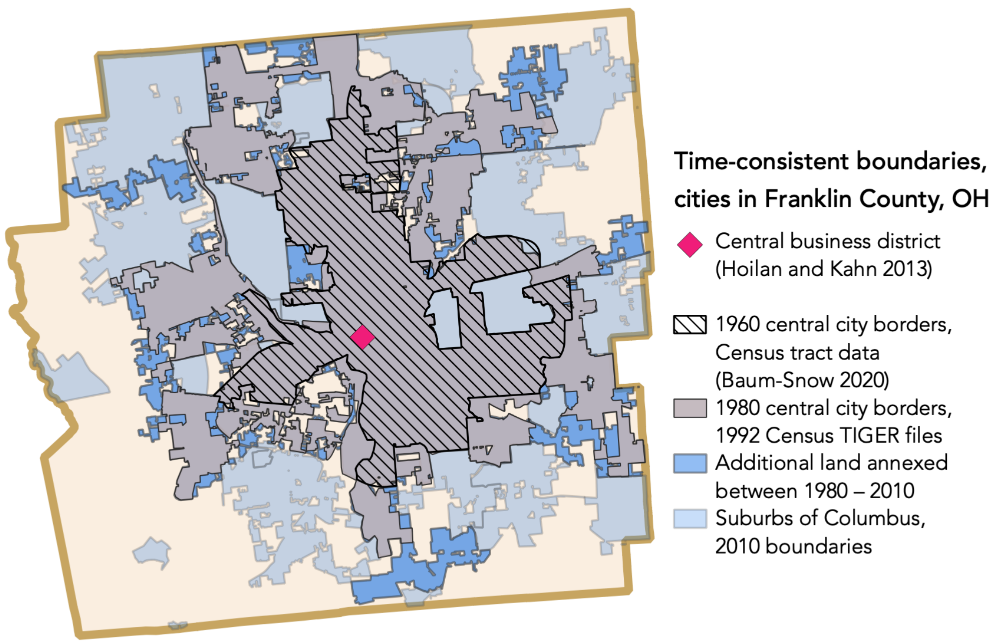
Notes: This table conducts the same matching and averaging procedure as Table C.1, but using one fixed wave of Census estimates. I compare the relative sizes of vintages in the 2000 Census estimates with reported units in CoreLogic built before 2000.

Table C.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

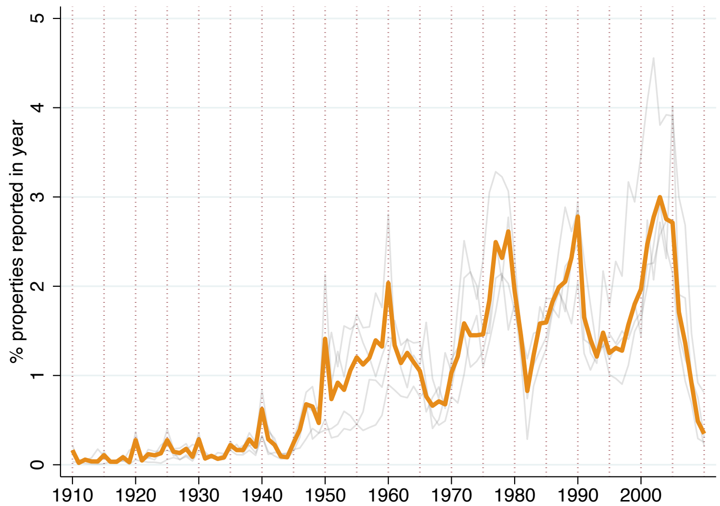
Figure C.1: Sample Jurisdiction Borders in A Metropolitan County



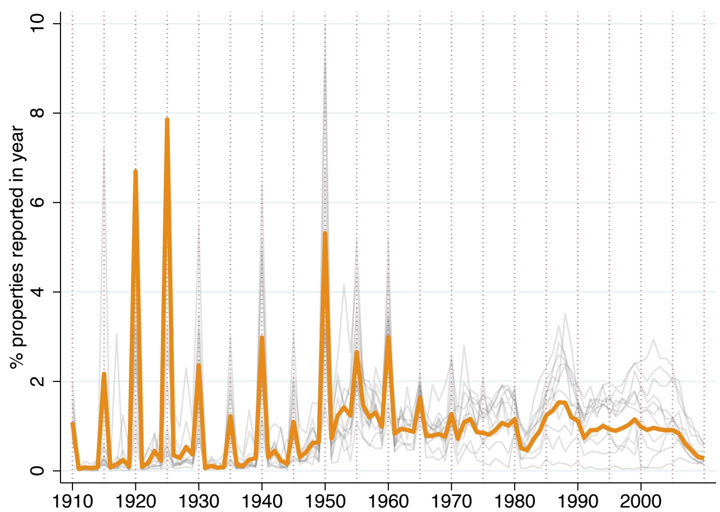
Notes: For Columbus, Ohio and the county it is in — Franklin County, Ohio — I visualize the time-consistent boundaries for the central city that control for postwar annexation. Central city composition change is calculated based on demographics within the tracts making up Columbus's 1960 boundaries. 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 C.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 C.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	1
Maximum Building Area, Percentage of Lot Area..	15%	18%	20%	30%	35%	[illegible]
Front Yard or Setback, in feet	50	40	40	30	25	[illegible]
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

D Details of Adoption Estimation

D.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 D.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 D.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 2, 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 D.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, as later detailed in Section 3.1. 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 D.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 D.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 D.2 and D.3 lists the final vector of parameters and weights. The rightmost column of Table D.3 also shows how each individual structural break specification correlates with the final ensemble estimate reweighing multiple specifications.

Table D.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	≥ 116K
Width of each bin (sq. ft.)	250	500	1000	2000	4000	≥ 8000

Table D.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 D.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 D.1: Visualizing Grid of Lot Bins

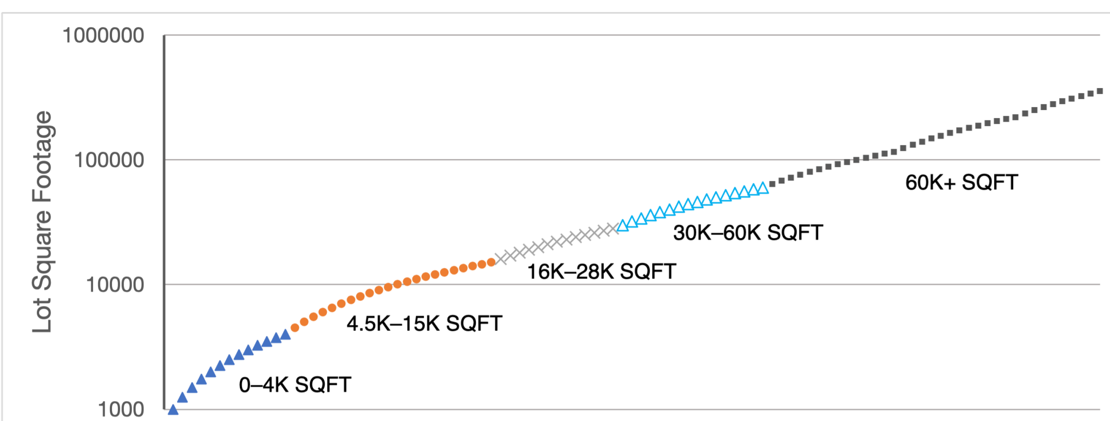
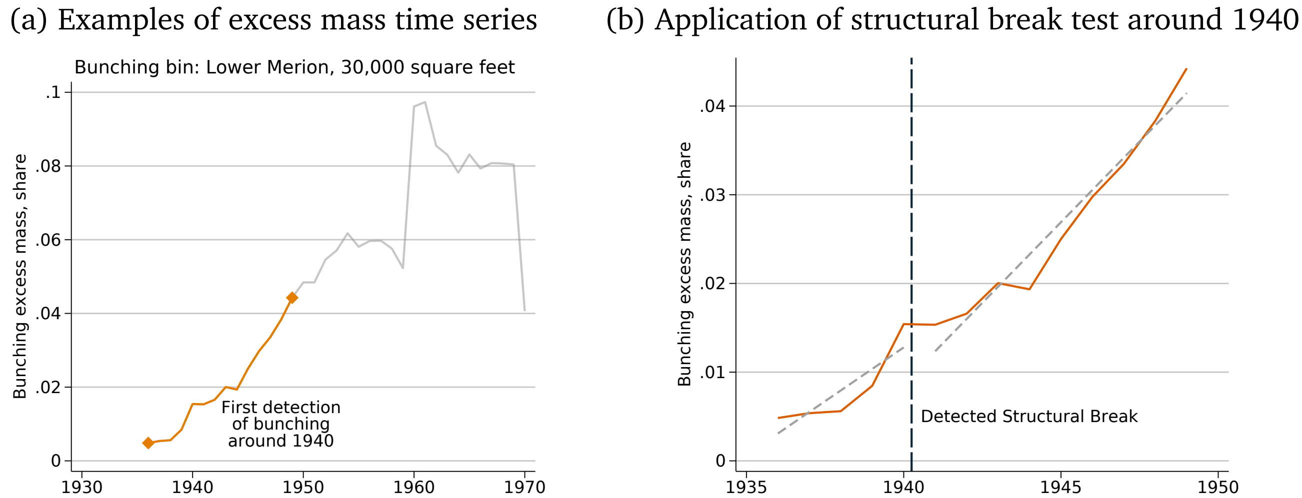
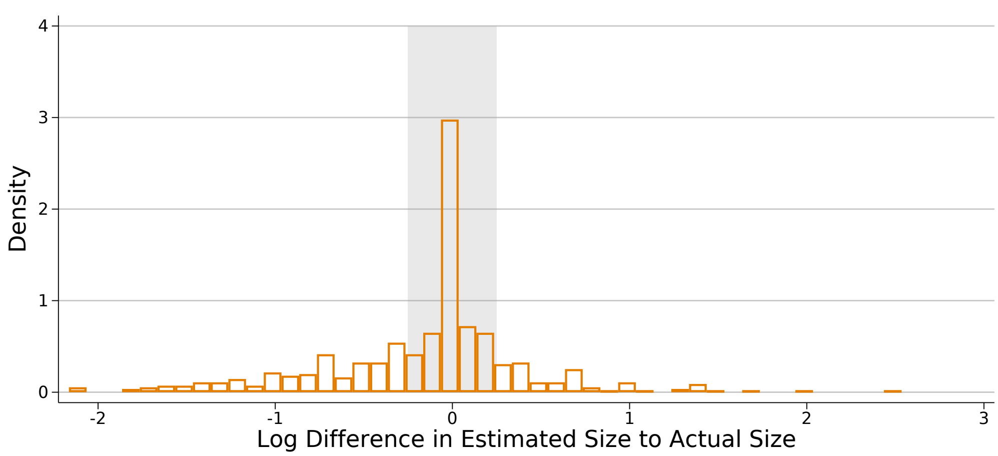


Figure D.2: Visualization of Structural Break Detection



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 D.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.

E Technical Proofs

E.1 Proof of Proposition ID1.

The idea is that, for some lot size distribution $\{\ell\}$, I find threshold values $K, \underline{\delta}$ and construct α within the thresholds such that

$$\mathbb{P}[\underline{\mathbf{L}} \subseteq \underline{\mathbf{B}}] = 1$$
$$\mathbb{P}[\underline{\mathbf{B}} \subseteq \underline{\mathbf{L}}] \geq \underline{\delta};$$

where $\underline{\mathbf{B}}$ is the set of all bunching bins detected by the parametrized gradient statistic classifier.

Step (1) Because every lot size control $\underline{\ell} \in \underline{\mathbf{L}}$ induces marginal bunching after adoption, for the adoption time τ in question and $\tau' < \tau < \tau''$, we have

$$\frac{p^{\tau''}(\underline{\ell})}{p^{\tau''}(\underline{\ell}-\mu)} > \frac{p^{\tau'}(\underline{\ell})}{p^{\tau'}(\underline{\ell}-\mu)}$$
$$\Rightarrow \Delta(p^{\tau''}(\underline{\ell}),\mu) - \Delta(p^{\tau'}(\underline{\ell}),\mu) \equiv G_{\tau'',\tau'}(\underline{\ell}) > 0.$$

The domain of (τ', τ'') where this holds is compact, so for each $\underline{\ell}$ there is a positive maximizer for the gradient statistic at a $(\overline{\tau}', \overline{\tau}'')$ or $\overline{G}(\underline{\ell}) = \Delta(p^{\overline{\tau}''}(\underline{\ell}),\mu) - \Delta(p^{\overline{\tau}'}(\underline{\ell}),\mu)$. Over all such $\overline{G}$, let $\tilde{\alpha}$ to be no greater than the least element.

Then applying the gradient classifier over all possible vintages with rule threshold $\tilde{\alpha}$ will equal 1 for each $\underline{\ell}$, so $\underline{\mathbf{L}} \subseteq \underline{\mathbf{B}}$.

Step (2) Observe that the following inequalities hold for $\tilde{\alpha}$ fixed:

$$\begin{aligned}
\mathbb{P}\big[\underline{\mathbf{b}} \notin \underline{\mathbf{L}} | \underline{\mathbf{b}} \in \underline{\mathbf{B}}\big] &= \mathbb{P}[\Delta(p^{\tau''}(\ell),\mu) - \Delta(p^{\tau'}(\ell),\mu) > \tilde{\alpha} \,\big|\, \ell \notin \underline{\mathbf{L}} \\
&= \mathbb{P}\big[G_{\tau'',\tau'}(\ell) - \mathbb{E}[G_{\tau'',\tau'}(\ell)] > \tilde{\alpha} - \mathbb{E}[G_{\tau'',\tau'}(\ell)] \,\big|\, \ell \notin \underline{\mathbf{L}}\big] \\
&\leq \mathbb{P}\big[G_{\tau'',\tau'}(\ell) - \mathbb{E}[G_{\tau'',\tau'}(\ell)] > \tilde{\alpha} - 2K|\tau'' - \tau'| \,\big|\, \ell \notin \underline{\mathbf{L}}\big] \\
&= \mathbb{P}\left[\frac{G_{\tau'',\tau'}(\ell) - \mathbb{E}[G_{\tau'',\tau'}(\ell)]}{\sigma[G]} > \frac{\tilde{\alpha}}{\sigma[G]} - 2K\frac{|\tau'' - \tau'|}{\sigma[G]} \,\Big|\, \ell \notin \underline{\mathbf{L}}\right] \\
&\leq \frac{1}{1 + (\frac{\tilde{\alpha}}{\sigma[G]} - 2K\frac{|\tau''-\tau'|}{\sigma[G]})^2},
\end{aligned}$$

where the standard deviation $\sigma[G]$ is used to standardize the statistic; the first inequality follows from bounded demand adjustment, as

$$\begin{aligned}
-\mathbb{E}[G_{\tau'',\tau'}(\ell)] = \\
&- \mathbb{E}[\log(p^{\tau''}(\ell)) - \log(p^{\tau'}(\ell))] + \mathbb{E}[\log(p^{\tau''}(\ell-\mu)) - \log(p^{\tau'}(\ell-\mu))] \\
&\geq -2\int K|\tau'' - \tau'| dF(\ell);
\end{aligned}$$

and the second inequality is the one-sided Chebyshev inequality. As $K \to 0$, for all possible vintages the probability $\mathbb{P}[\underline{\mathbf{b}} \notin \underline{\mathbf{L}} | \underline{\mathbf{b}} \in \underline{\mathbf{B}}]$ becomes bounded by $\frac{1}{1+\bar{\alpha}^2} < 1$, and let $\underline{\delta}$ be an intermediate value between the two.

Thus for a sufficiently small K and sufficiently large $\underline{\delta}$, there exists an α for which the gradient classifier set identifies $\underline{\mathbf{L}}$ with the desired probability.

F Details of Identification

F.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 F.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 F.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 F.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 5 normalize all demographic change variables by percentiles across cities in each decade, in an attempt to correct for instrumental variable misspecification.

F.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) $\mathrm{Cov}[g_k(t), g_{k'}(t')\,|\,\overline{\nu}, \tilde{\omega}(k), \mathbf{C}(k,t)] = 0$.

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:

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

F.3 Further validation of instrument

Exposure share controls. The general statement of shift-share identification is that shocks are random with possibly a fixed effect constant across clusters of counties. Adding share controls to control for these fixed effects, as suggested in BHJ, improves estimate precision only if those fixed effects matter in the data. Figure F.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. For both Southern Black and Southern White shocks, overlap of the state shock distributions is substantial: controling for differential exposure to migrants of certain states is unlikely to deliver a more efficient estimator. The only exceptions are larger states that have within-state internal mobility, like Texas or Florida. For that reason, I only control for incomplete share controls in the main specification.

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 F.4.

Further balance tests of shift-share instrument. Figure 6 conducts balance tests of the shift-share instrument, defined over 1950 shocks, on covariates in 1940 suburbs. Figures F.5 and F.6 analyzes covariate balance for just central cities in 1940 and further pre-trends tabulated from the 1930 full count Census, respectively.

The results show there is persistence in sorting of Black migrants towards high manufacturing cities over time, but that differential trends up to 1940 in lot size outcomes are not evident. Several covariates recorded in the 1940 census are unavailable in the 1930 Census, but the 1930 Census also shows suburbs of Black migration destinations have more positive selection in house values and homeownership. I therefore control for these variables in the main specification.

Table F.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.

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

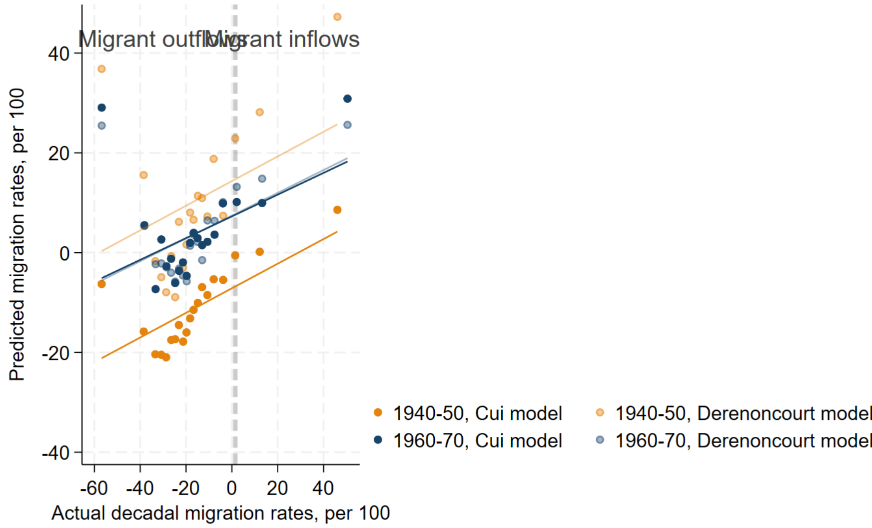
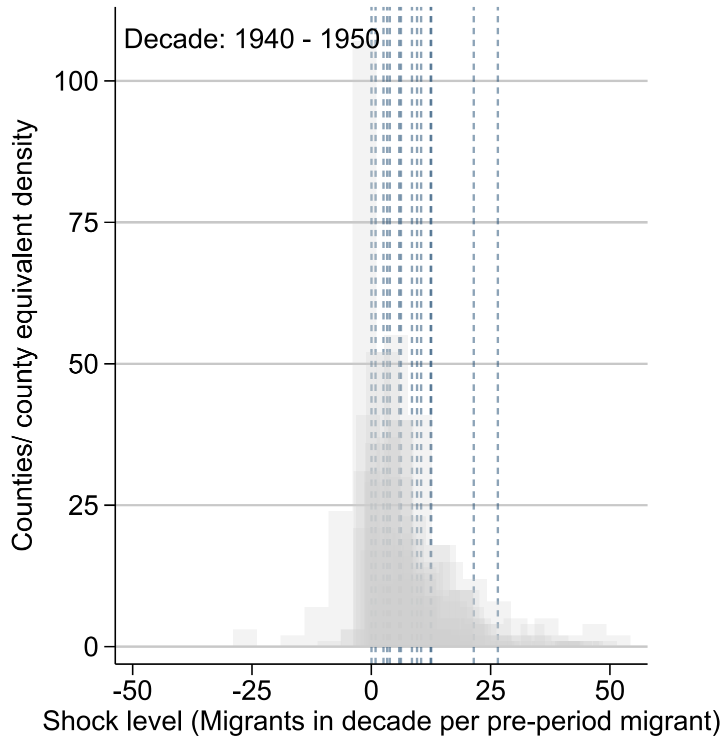


Figure F.2: Black Migration Shock Levels Within and Between States

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



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

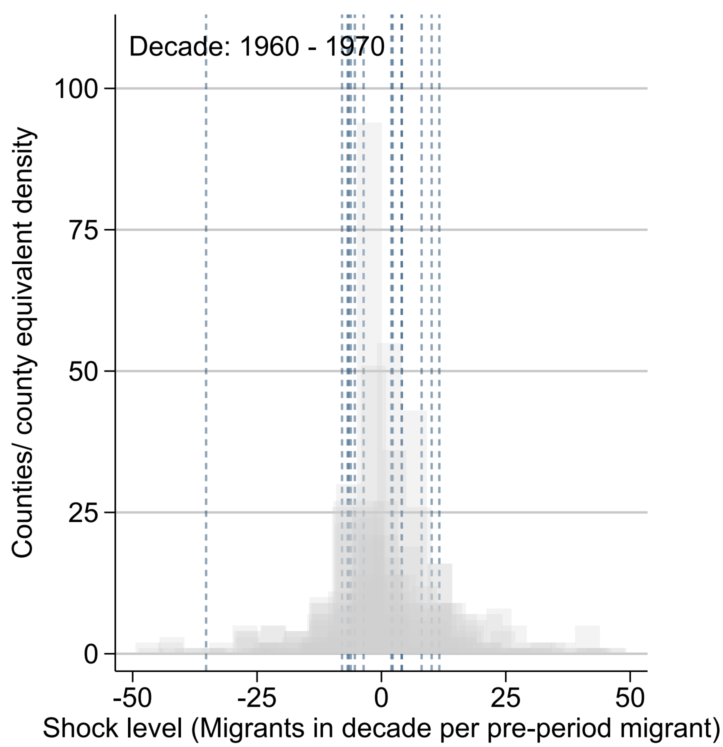
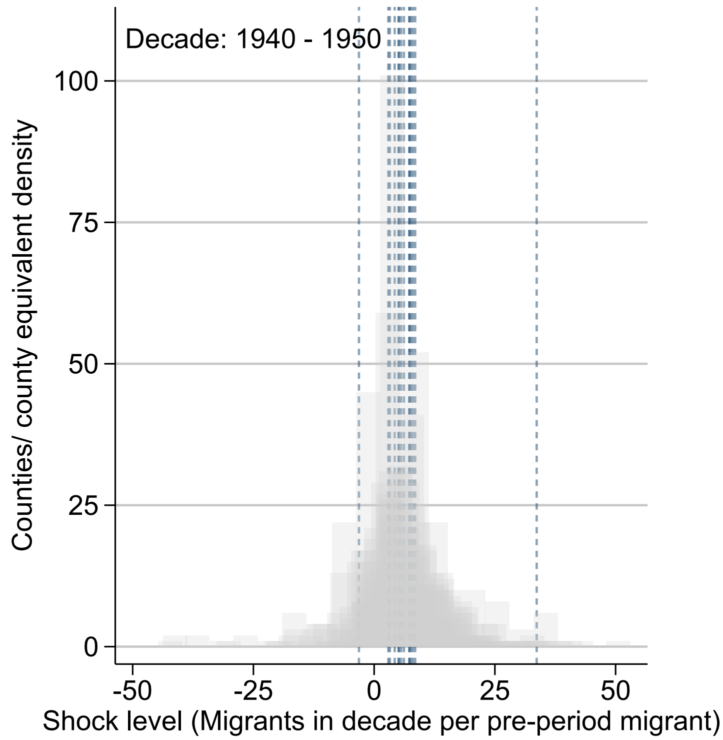


Figure F.3: Southern White Migration Shock Levels Within and Between States

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



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

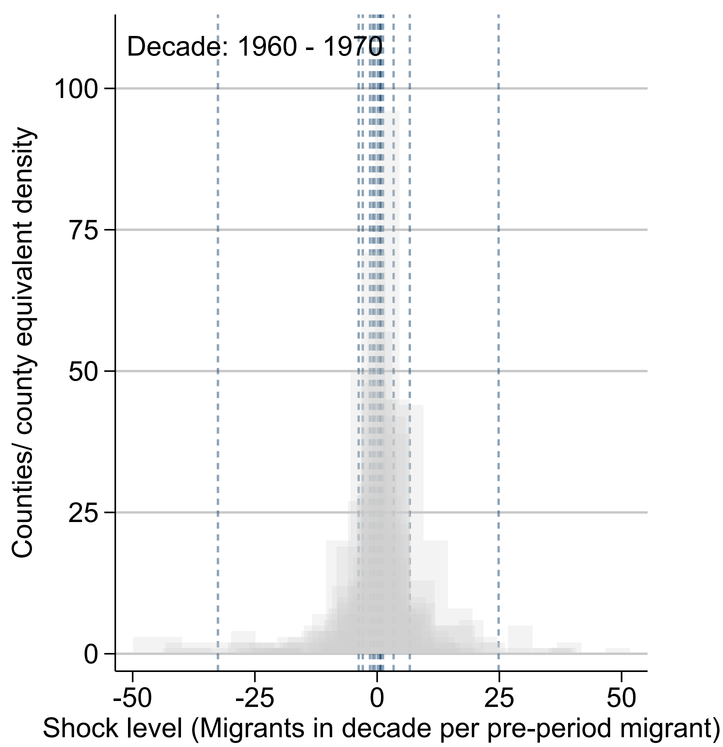


Figure F.4: Central Cities Had Varying 1940 Black Migrant Exposure by State

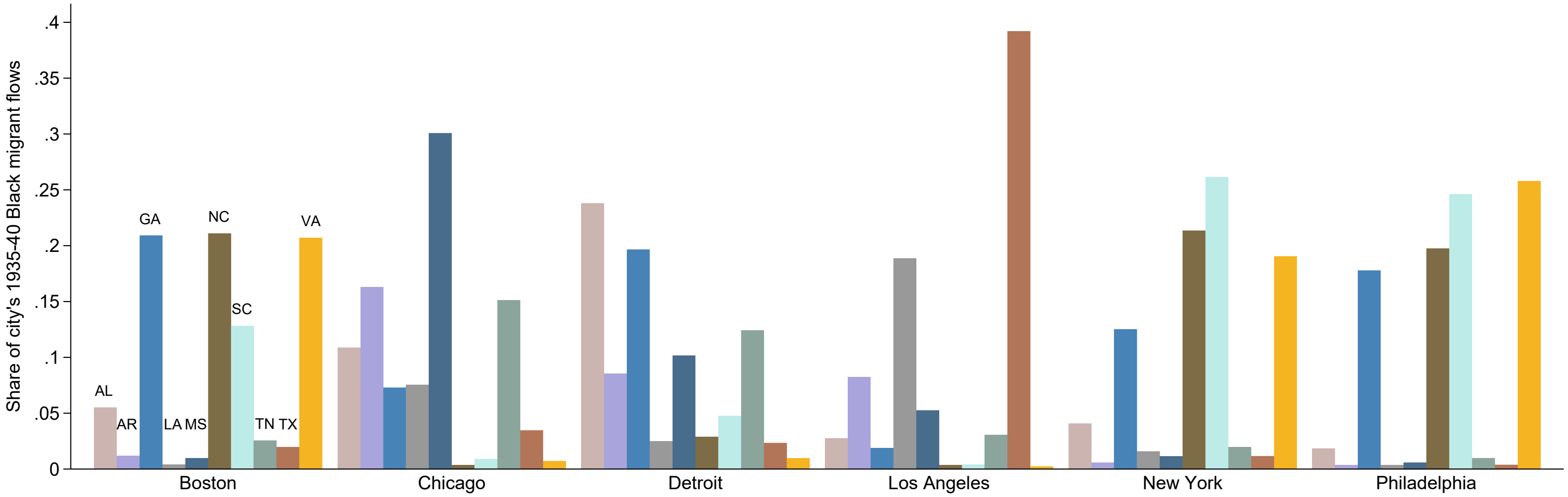


Figure F.5: Balancing Checks for Southern Black Migration Shocks, Central City Covariates

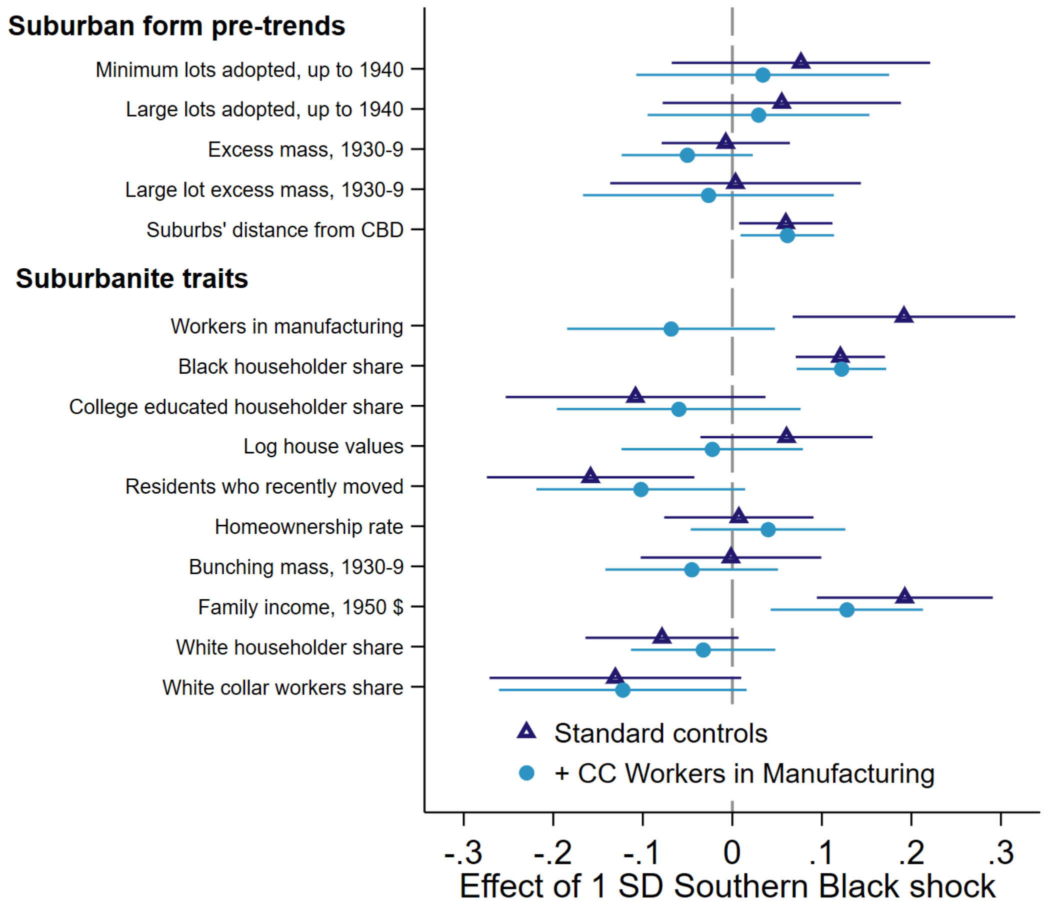
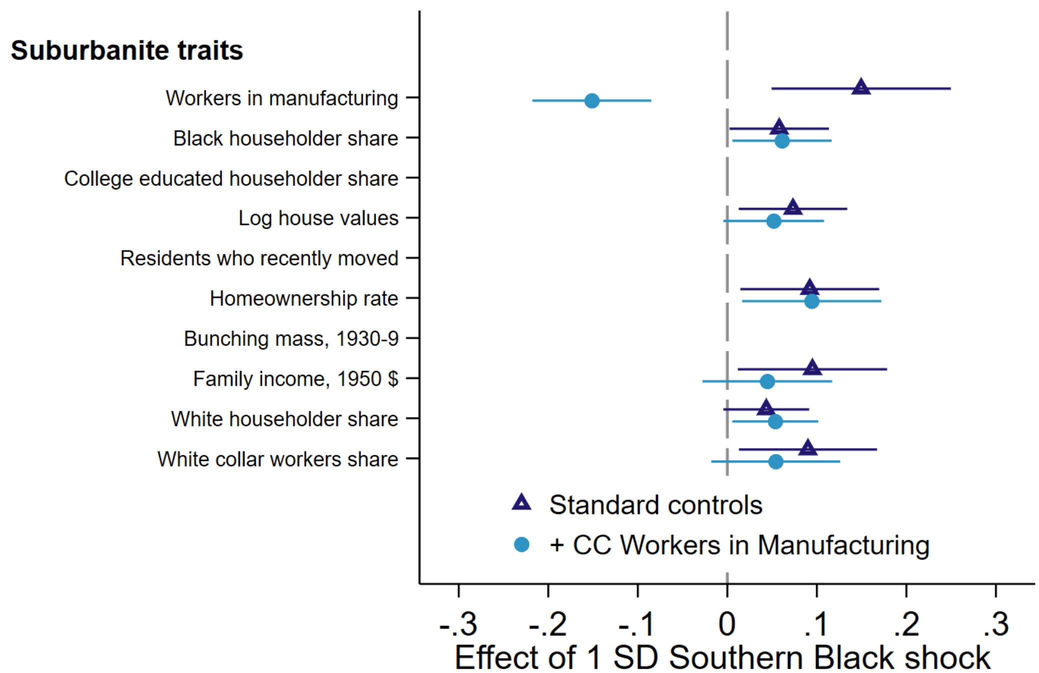


Figure F.6: Balancing Checks for Southern Black Migration Shocks, data as of 1930



G Further Robustness Checks

G.1 Additional checks on main regression specification

Results without outcome grouping. Table G.1 shows results from the regression equation 6, i.e. one that does not group suburban jurisdictions of the same metro together in outcomes. While numerical equivalence is not guaranteed between the specifications, I find results on restrictiveness are similar to my preferred specification but somewhat lower in magnitude. I also run a version of 6 where each jurisdiction has weight $1/|J_m|$, scaled by the number of suburban jurisdictions in a metro J_m. The point estimates from this model are closer numerically to my preferred estimates, but with loss of precision.

Alternative outcomes. If my TSLS estimates correct for differential suburban demand in metropolitan areas with high Black migration, the estimates should have less explanatory power for lot size outcomes more confounded with changing demand. Table G.2 uses the bunching share around bunching bins as an outcome, which Section 3.1 notes is confounded with demand. Point estimates for the bunching share are noisier, and Figure B.4 shows the Great Migration explains baseline bunching share the least out of excess mass bunching and lot size restrictiveness over larger lot sizes. Panel B of Table G.2 also shows my design cannot reject the Great Migration having null effects on the adoption of any minimum lot size.

To examine how sensitive results are to which lot size threshold is used to classify restrictive lot sizes, Figure G.1 reruns the Great Migration and Southern White TSLS specifications for thresholds smaller and larger than 7,500 square feet. Relative to the 7,500 square foot definition highlighted in bold, I find effects of the migrations go in different directions for lower thresholds. For higher thresholds like 10,000 to 20,000 square feet, both specifications yield noisy estimates of increased lot size restrictiveness. Such results are consistent with very large minimum lot sizes adopted in response to general suburban congestion, without a detectable racial exclusionary zoning component.

Measurement error. Misclassification for the binary measure of lot size adoption can be seen with an example: 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 but no misclassification for 1960 nor 1970. 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) is that the bias due to misclassification can be analyzed by comparing misclassified estimates to a validation set, which in my case are the historical records used to train the bunching algorithm. Conditional on a false positive or a false negative, I can calculate expectations of covariates X. 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 the predetermined jurisdiction level controls used in the main specification. I 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 2.4). Table G.4 suggest the reduced form estimate of Z_{ct}^{Black} on restrictive lot adoption is downward biased by -0.594, but this is over a selected sample where effect sizes are larger than the analysis sample. It also suggests that the insignificant results on adoption (Table G.2) may be due to misclassification error downward biasing estimates.

Following Meyer and Mittag (2017), I also simulate a counterfactual where misclassification is missing at random, comparing bias in that scenario with the actual bias. The results show most of the downward bias of the estimator is due to metro areas with greater Black migration having higher rates of misclassification error. This may be because when racial composition changes in central cities, surrounding suburbs try to complement minimum lot sizes with other forms of land use regulation.

Reweighing of jurisdictions. For each metropolitan area in the sample, I group and calculate over all zoning jurisdictions defined as of 1980 that are also defined in modern CBSA borders. However, these jurisdictions could include exurbs developed only in the 1970–80 time period or rural communities disconnected to the central city. Suburbs from which central city residents moved throughout the analysis period, in contrast, should be populated even at the start of the sample.

To correct grouped estimates so they measure outcome variation in "compiler" suburban jurisdictions, I reweigh jurisdictions by their reported 1940 population tabulated using the CPP crosswalk. Results are presented in Table G.3, which shows this adjustment do not greatly alter point estimates. [4]

Robustness to outlier cities. Are the results driven by one large metropolitan area receiving many Black migrants, like New York or Chicago? To efficiently calculate how sensitive results are to collections of outliers, I use the algorithm proposed in Broderick, Giordano and Meager (2021) to calculate the share of observations that must be dropped for estimated effects to switch sign. Results from my preferred specification over all three outcomes in Table 3 do not flip sign unless at least 2% of observations are dropped.

To unpack these findings further, I check which observations must be dropped to flip the sign of estimates, based on their influence in the TSLS specification. The largest metro areas, like New York or Chicago, are not outliers driving results because they have low influence in the specification. One example of a high influence metro area is Decatur, Illinois, a Midwestern metro that was centred on manufacturing and whose central city is 26% Black as of 2020.

G.2 Additional checks on event study specification

Comparing metros of states adopting early civil rights laws. 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 follow (Cook et al., 2022) and tabulate the total number of laws recorded for each

[4] Figure G.2 also convert these reweighted estimates into effects like in Figure 8.

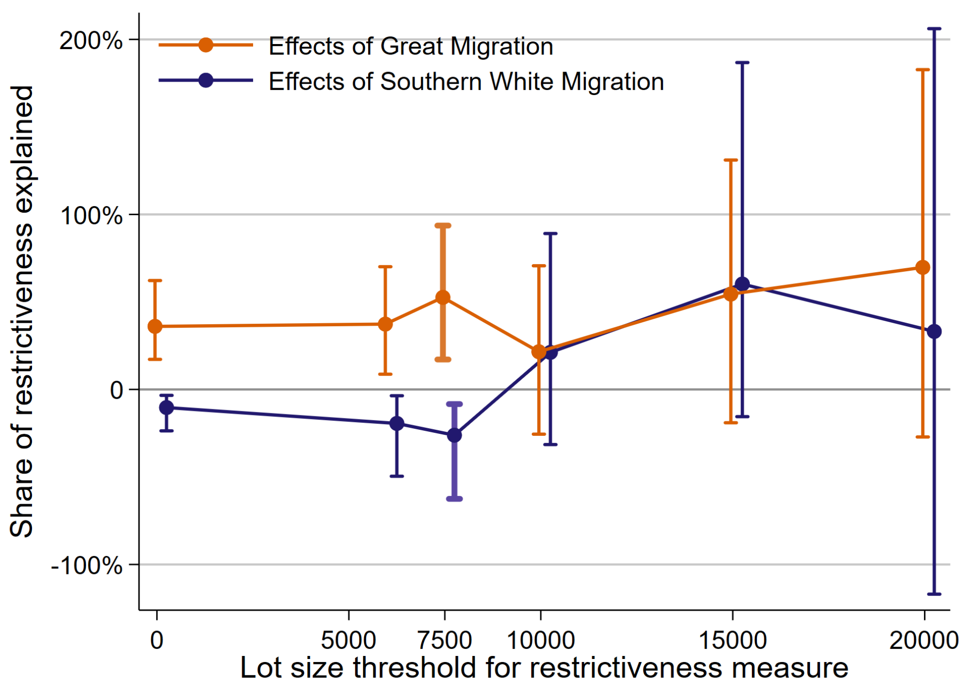
state up to 1950. I divide non-Southern states based on whether they had an above-median number of anti-discrimination laws. The results compared to a partition based on more or less metropolitan school segregation are in Figure G.5. Compared to the results on school segregation, metros in an above-median desegregation law state have insignificant lot size outcome responses.

Effects for counties where lawsuits failed to desegregate schools.

In contrast to counties adopting early school desegregation, there are other counties where unsuccessful lawsuits were launched. In contrast to counties where the presence of a lawsuit corroborates with school desegregation being a contentious political issue, counties with failed cases contain jurisdictions where *de facto* school segregation was in place, but there was no significant coalition of interest groups working to overturn it.

Figure G.6 compares the same event study design ran on counties with failed lawsuits to the estimates in 9, specifically where both designs use the triple difference design and share the same uncontaminated control group. The point estimates using counties with unsuccessful lawsuits are negative and statistically insignificant, which suggests without institutional change, there was no motive to zone for restrictive lot sizes in addition to what is expected from metropolitan Black demographic change.

Figure G.1: Alternative Lot Size Restrictiveness Definitions and Demographic Change



Notes: This figure plots the number…

Table G.1: Results in Table 3 on Ungrouped Data

	OLS		IV	
	(1)	(2)	(1)	(2)
Panel A: Effects on Restrictive Lot Size Adoption				
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.0475**	0.0537*	0.00610	0.0343
	(0.0236)	(0.0284)	(0.0689)	(0.0814)
First-stage F-stat			68.53	54.99
Baseline mean	0.334	0.337	0.334	0.337
R^2	0.00584	0.0122	0.00611	0.0121
Panel B: Effects on Lot Size Restrictiveness				
Percentile of $\Delta CC^{Black}_{c(j)t}$	1.933***	1.943	1.645	2.675**
	(0.432)	(0.577)	(1.176)	(1.324)
First-stage F-stat			69.55	56.23
Baseline mean	2.979	3.080	2.979	3.080
R^2	0.0758	0.0933	0.0760	0.0928
Panel N	16282	12132	15293	12132
Region–Decade FE	X	X	X	X
Incomplete shares	X	X	X	X
Pre-period controls		X		X

Significance levels: * = 10%; ** = 5%; *** = 1%.

Notes: This table presents regressions of exposure to central city Black migration on lot size outcomes, following the specifications in Table 3:

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

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 3 contain further details on the regression specifications.

Table G.2: Migration's Effects on Unconditional Outcomes

	OLS		IV	
	(1)	(2)	(1)	(2)
Panel A: Effects on Lot Size Bunching Share				
Percentile of $\Delta CC^{Black}_{c(j)t}$	1.100	1.976	3.781	10.12
	(1.168)	(1.279)	(4.188)	(6.585)
First-stage F-stat			26.60	19.51
Anderson-Rubin p-value			0.368	0.105
Baseline mean	13.08	13.08	13.08	13.08
Panel B: Effects on Any Minimum Lot Adoption				
Percentile of $\Delta CC^{Black}_{c(j)t}$	-0.002	0.003	0.0578	0.258
	(0.0306)	(0.0359)	(0.158)	(0.247)
First-stage F-stat			26.39	20.46
Anderson-Rubin p-value			0.714	0.295
Baseline mean	0.705	0.705	0.705	0.705
Panel Units	Metro	Metro	Shock	Shock
Panel N	561	561	3418	3418
Region–Decade FE	X	X	X	X
Incomplete shares	X	X	X	X
Pre-period controls		X		X

Significance levels: * = 10%; ** = 5%; *** = 1%.

Notes: This table presents regressions of central city Black composition change, as defined in Section 4.1, to additional continuous measures of lot size restrictiveness defined in Section 3,

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

using both OLS specifications and instrumenting for Black composition change with shift-share Black migration instruments. The notes for Table 3 contain further details on the regression specifications.

Table G.3: Results in Table 3, With Prewar Household Weights

	OLS		IV	
	(1)	(2)	(1)	(2)
Panel A: Effects on Restrictive Lot Size Adoption				
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.0117	-0.0124	0.418*	0.564*
	(0.0458)	(0.0469)	(0.235)	(0.307)
First-stage F-stat			26.39	20.46
Anderson-Rubin p-value			0.0607	0.0579
Baseline mean	0.534	0.534	0.534	0.534
Panel B: Effects on Lot Size Restrictiveness				
Percentile of $\Delta CC^{Black}_{c(j)t}$	0.762	0.326	3.973**	3.958*
	(0.498)	(0.481)	(1.839)	(2.091)
First-stage F-stat			26.39	20.46
Anderson-Rubin p-value			0.0366	0.0658
Baseline mean	2.410	2.410	2.410	2.410
Panel Units	Metro	Metro	Shock	Shock
Panel N	561	561	3418	3418
Region–Decade FE	X	X	X	X
Incomplete shares	X	X	X	X
Pre-period controls		X		X

Significance levels: * = 10%; ** = 5%; *** = 1%.

Notes: This table presents regressions of exposure to central city Black migration on lot size outcomes, following the specifications in Table 3:

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

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 3 contain further details on the regression specifications.

Table G.4: Mismeasurement bias relative to using historical records

	Lot adoption		Restrictive lot adoption	
	Bunch-based	Record-based	Bunch-based	Record-based
heightPercentile of CC^{Black}	1.308*	1.813**	1.072	1.666**
	(0.647)	(0.721)	(0.872)	(0.729)
N	333	333	333	333
Bias with bunching bin outcomes		-0.506		-0.594
Bias assuming random misclass.		0.120		-0.0721
Bias due to correlation, %		-34.52		-31.30

Significance levels: * = 10%; ** = 5%; *** = 1%.

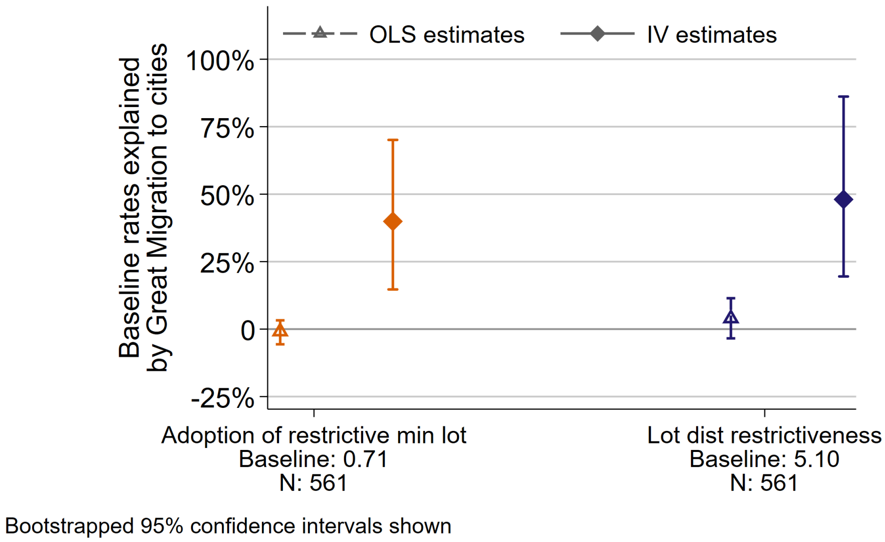
Notes: This table shows results of a reduced form regression on adoption measures, using adoption timing based on bunching on lot sizes versus on historical records. The specification is at the jurisdiction level, adds controls used for Table 3 and in particular uses census region effects:

$$Reg_{jt} = \beta Z^{Black}_{c(j),t} + \delta_j + \mathbf{X}_{j,pre}\Gamma + \varepsilon_{j,c(j)t}.$$

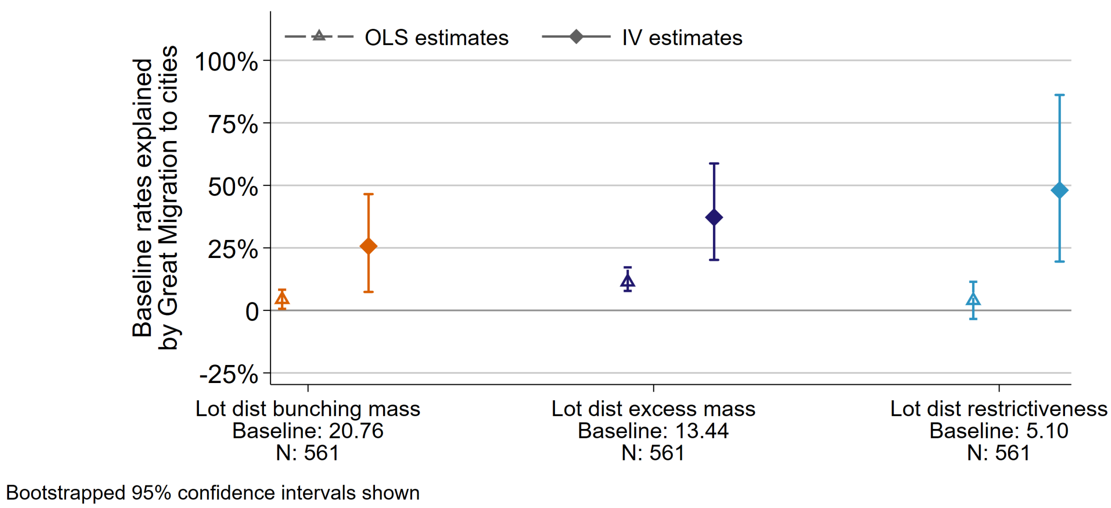
Following the sample size, the bias of the bunching-based measure relative to using records is shown, and decomposed based on correlation of misclassification error and covariates following Meyer and Mittag (2017). Standard errors are clustered at the metropolitan area level.

Figure G.2: Share of Outcomes Explained by Great Migration, Weighted

(a) Main Results on Restrictive Lot Sizes



(b) Main Results on Lot Size Restrictiveness Measures



Notes: This figure presents an aggregation exercise, converting the regression coefficients estimated in Table 3 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 3.1. The notes for Figure B.4 contain further details on aggregated estimates.

Figure G.3: Figure 9 Panel (a), in 5-year bins

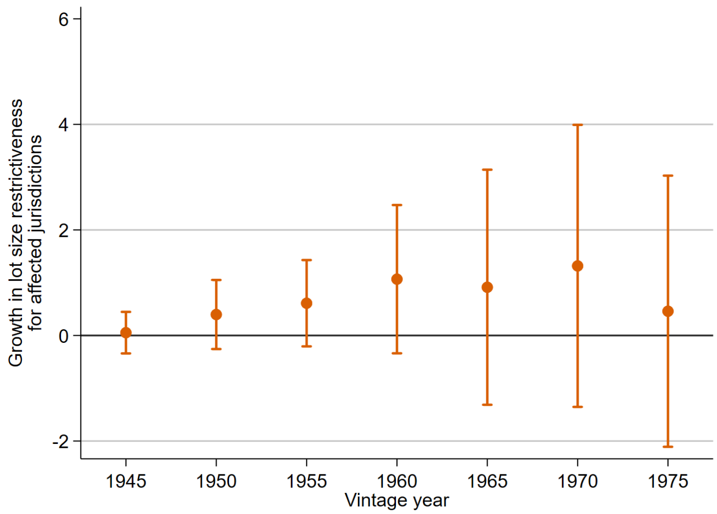


Figure G.4: Figure 9 Panel (b), in 5-year bins

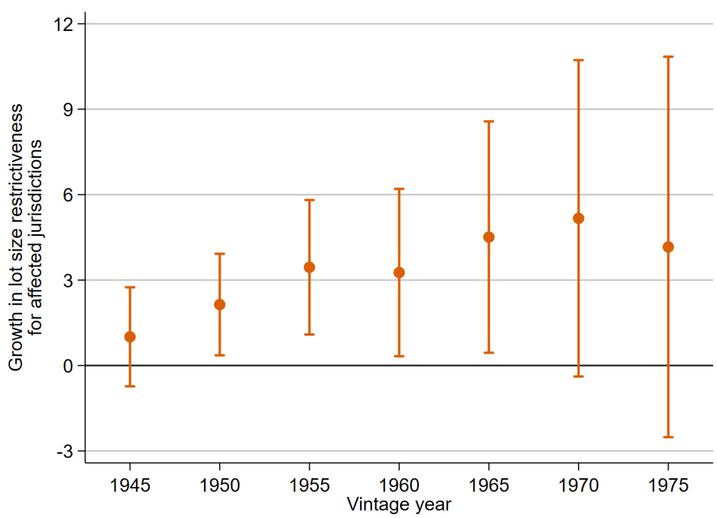


Figure G.5: Figure 9 Panel (a), with early desegregation states

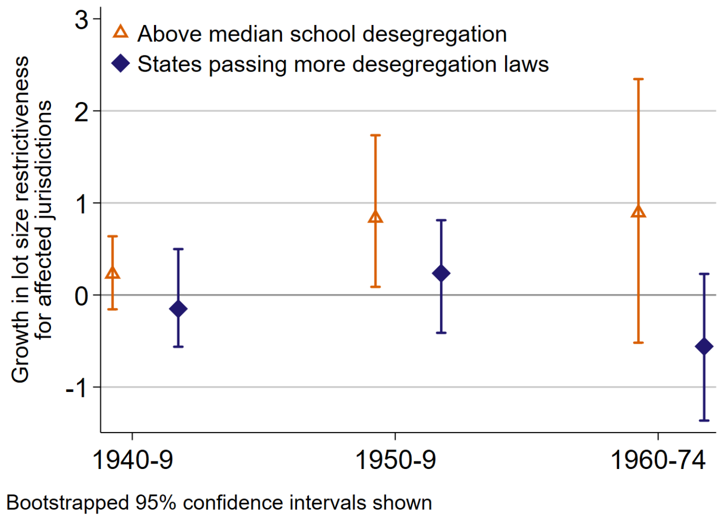


Figure G.6: Figure 9 Panel (b), with failed desegregation cases