Population Centers and Coordination: Evidence from County-Seat Wars

Cory Smith & Amrita Kulka

(This paper also appears as CAGE Discussion paper 724)

September 2024

No: 1518

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Population Centers and Coordination: Evidence from County-Seat Wars*

Cory Smith[†] ⓡ[‡] Amrita Kulka[§]

September 19, 2024

Abstract

We study the process of long-run urban growth using a unique setting of close elections that determined "county seats" (capitals) in the frontier United States. Employing a regression discontinuity design, we show that winning towns rapidly became the economic and population centers of their counties as new migrants coordinated on them as destinations. This coordination was largest in the early years of a county's history, but limited in later decades. Using generalized random forests, we show that the economic changes were not zero sum locally: specific choices of county seat could increase long-run county population and income. As county administration was limited in this era, the public sector did not play a substantial role in this growth. Instead, these results illustrate how a political process can select spatial equilibria through a shock that is neither related to locational fundamentals nor confers direct productivity advantages on the location.

*We thank Ryan Schwartz, Fatima Najeeb, Ben Kremer, Roque Bescos, Jake Burgess, and Risa Wagner for their excellent research assistance. We also thank Treb Allen, Dave Donaldson, James Fenske, Hoyt Bleakley, Peter Nencka, Ezra Karger, Enrico Berkes, Randall Walsh, and other seminar participants at Dartmouth College and the University of Warwick for their advice and comments. This research was generously supported by the USDA through an AFRI grant.

[†]University of Maryland (AREC), corybsmith@gmail.com

[‡]Author order was randomly determined via coin toss.

[§]University of Warwick, amrita.kulka@warwick.ac.uk

1 Introduction

The benefits of urban agglomeration mean that individuals and firms prefer to co-locate in more populous communities, even if they believe smaller locations have better fundamental characteristics. This dynamic creates a coordination game with multiple potential equilibria, making the selection of locations a critical question for economic geography.

Several key aspects of this type of coordination remain less understood. While many studies have shown that historical shocks can influence long-term population distributions, some large-scale events have only small or temporary effects (Davis and Weinstein 2002). Evidence on when to expect either outcome remains limited. Further, although multiple equilibria in population distributions are theoretically possible, in most cases the alternative choice to any given location cannot be identified. This means that the details of the selection process remain vague and it complicates the evaluation of counterfactuals.

In this paper, we study the process of coordination using a natural experiment set among towns on the American frontier. In the nineteenth and early twentieth centuries, the selection of the "county seat" (capital) for new counties typically established their long-term population centers. Because county government was of limited size during this period, these population movements were primarily driven by perceptions rather than an allocation of material benefits. We examine cases where the seat was determined by election, using a close election regression discontinuity (RD) design for identification. This setting offers several novel features. First, it illustrates how coordination can be facilitated by a political process. In contrast to other cases in the literature, a close election victory was unrelated to geographical fundamentals and did not directly confer a meaningful productivity advantage. Second, we can fully describe the selection process, as close losers provide specific counterfactual population centers. Finally, the range of election timings allows us to study comparable shocks at many different points in history.

Our first results show that the scope for coordination was large on average. By 2010, winning towns became 1.17 log points (223%) denser than their rivals, representing about an 11% shift of each county's total population. Despite the relatively small size of towns in this setting, winners experience agglomeration benefits associated with larger cities: income measures rise by 8%–18%, white-collar employment by 10 percentage points, and per capita patents by 43%. Historical census data show that the transitions to becoming a population and economic center began in the immediate years after the election, with smaller adjustments continuing for two to three decades. Afterward, relative differences stabilize and we observe few aggregate changes through our end year of 2010. Between 65% and 70% of winning towns are the largest in their counties, both in the initial post-election years and in the long run. The elections thus had a substantial effect on economic geography by establishing counties' primary urban centers in the initial years of their formation.

Several facts establish that election winners grew due to their status and prominence rather than the material components of county government. Election winners very quickly established their own newspapers and appeared more frequently in news articles nationwide. These effects are apparent in the initial post-election years, unlike major public works which would have taken longer to materialize. County seats also disproportionately attracted migrants who did not share the surnames of previous residents, illustrating that county seat status helped coordinate those without preexisting connections to an area. Relative to national or regional capitals, county governments were quite small in size with their administrative employment accounting for less than 1% of jobs even in the capital. Given our large growth estimates, each of these jobs would have had to create 75 others to explain our effects.

Our analysis required the extensive collection of original data. We compiled results for 1117 county-seat election results in 811 counties, individually drawn from newspaper archives, county histories, and administrative records. In 2010, these counties had 71 million residents, or 23% of the nation's population. Building on recent advances in standardizing locations in public microdata (Berkes, Karger and Nencka 2023), we link competing locations to subcounty data from census records, post office locations, geographic characteristics, and tax filings. The result is a detailed political and economic history of our sample locations. The sample spans the population distribution of American communities: On the one hand, a typical place is home to several thousand people in 2010. On the other hand, the very upper tail consists of large cities including Phoenix, Dallas, and Miami. While most elections occurred close to the time of county formation, an important minority occurred in subsequent decades.

For our second set of findings, we identify the conditions that make it easiest to achieve coordination. Using variation in the timing of elections relative to county formation, we examine whether economic geography is shaped more by comparable shocks earlier or later in history (Lin and Rauch 2022). In our setting, initial conditions are most important. The largest economic and population effects occur from elections in the earliest years of county formation, with these effects steadily diminishing to near zero the later the election is held. This is consistent with the idea that coordination is relatively easy in the absence of a well-established spatial equilibrium but difficult otherwise. As urban areas grow, their larger histories, populations, and durable investments make them insensitive to shocks that influence coordination. Additionally, these results provide further evidence that county government per se plays a negligible role in our results. Later elections reliably move the county capital without driving any detectable economic or population changes.

In our last set of results, we evaluate the county-level consequences of coordinating on particular locations. In theory, some towns might function more effectively as county centers than others. We estimate the countywide effects of selecting each election competitor as the seat with the machine learning method of generalized random forests and a set of predetermined characteristics (Athey, Tibshirani and Wager 2019). The algorithm generates out-of-sample predictions for which choice of county seat would maximize long-run county density. We validate these choices with a county-level RD design, showing that, on average, picking this location increased county density by 51%. These same choices also appear to increase modern county income by 5%, although this is just significant at the 10% level. This analysis is possible due to a feature of our setting where close losers provide identifiable counterfactual population centers. Our results show that while the location of the county seat was a zero-sum choice, its effects on broader economic geography were not. Historically, voters' choices were negatively correlated with the machine learning model's picks. Voters tended to pick larger, more established towns at the time of the election, but these characteristics did not significantly influence aggregate, long-run impacts.

We contribute to several strands of literature in urban economics and economic history. First, we add to our understanding of the effects of agglomeration forces. The benefits of geographically concentrating economic activity have been studied both across cities (Rosenthal and Strange 2004; Duranton and Puga 2004; Ellison and Glaeser 1999; Glaeser et al. 1992; Krugman 1991; Kleinman, Liu and Redding 2023; Desmet and Rossi-Hansberg 2013; Redding, Sturm and Wolf 2011; Combes et al. 2012; Michaels and Rauch 2018) and within cities (Ahlfeldt et al. 2015; Heblich, Redding and Sturm 2020). Commonly studied reasons for agglomeration include locations' fundamentals and natural advantages, productivity gains from proximity to dense workers and firms (Ahlfeldt et al. 2015; Heblich, Redding and Sturm 2020; Qian and Tan 2021), and knowledge spillovers (Davis and Dingel 2019; Combes, Duranton and Gobillon 2008; De La Roca and Puga 2017). We are able to study the long-run effects of agglomeration, thereby shedding light on the persistence of historical shocks through population and infrastructure (Hanlon and Miscio 2017). Our setting allows us to examine how a shock with minimal direct economic implications plays out over time by selecting a spatial equilibrium across alternative population centers. Our paper also contributes to knowledge about the less understood effects of agglomeration on the population and occupational structure of smaller, rural counties. Studies of smaller communities are especially limited in many settings in which data on smaller towns are aggregated into larger administrative areas, limiting the ability of researchers to accurately observe them in any detail (Puga 2010). Our findings show that the scope for population coordination can be quite large in these settings and the benefits mirror those in larger cities, including outcomes from income to knowledge output.

While much of the literature on agglomeration recognizes the possibility of multiple equilibria, to our knowledge, there are few papers that study a setting where equilibrium selection from a concrete set of choices can be directly observed. A number of papers have used runner-up designs econometrically similar to ours to examine specific investments (Greenstone, Hornbeck and Moretti 2010; Andrews 2019; Howard, Weinstein and Yang 2024), but their focus lies on understanding the impacts on productivity rather than a way to coordinate population. Allen and Donaldson (2022) model spatial equilibria in the historical United States and define parameter regions in which multiple equilibria do or do not exist but leave the exact mechanisms of coordination unspecified. Empirically, Davis and Weinstein (2008) test for multiple equilibria after World War II bombings in Japan, but conclude that there are none. Other theories predict that "large actors" can facilitate coordination through through public signaling or other forward-looking actions (Behrens and Robert-Nicoud 2015; Henderson and Venables 2009; Henderson 1974; Acemoglu and Jackson 2015). Our paper provides empirical support for these theories as we demonstrate how coordination was facilitated by the public results of an election, organized by the prominent individuals in county government. Our focus on the timing of the shock also connects to work that aims to identify the "critical junctures" in history where change is possible (Callen, Weigel and Yuchtman 2023). In our setting, these appear to be the early years during or just after county formation.

Second, by exploiting historical variation from county-seat elections, we contribute to the literature on factors that shape the growth and location of towns and cities (Bleakley and Lin 2012; Davis and Weinstein 2002; Banzhaf and Walsh 2013; Shertzer, Twinam and Walsh 2018; Ager et al. 2020; Harari 2020; Brown and Cuberes 2020; Dell and Olken 2020; Howard and Ornaghi 2021). County-seat elections represent a novel setting in this literature, but one that is economically significant. We exploit widespread, cross-town variation in large parts of the United States, which broadens our understanding of agglomeration in geographically or industrially concentrated areas; see Hanlon and Heblich (2020) for an overview. This connects to a broader literature on economic development on the American frontier in general (Donaldson and Hornbeck 2016; Hornbeck and Naidu 2014; Acemoglu, Moscona and Robinson 2016; Nagy 2020; Bleakley and Rhode 2023). County-seat elections were a ubiquitous, salient feature of American life outside the original colonies, and the question of their impact on location choices and local economies is important in its own right. Among the close elections in our sample, a back-of-the-envelope calculation suggests that county-seat elections determined the residence of about 5 million people today, making them a large influence on today's population distribution.

Finally, our paper also adds to a subset of this literature on the role of capital cities in economic development (Campante and Do 2014; Ades and Glaeser 1995; Bai and Jia 2021; Chambru, Henry and Marx 2021; Bluhm, Lessmann and Schaudt 2023), which finds that an expanded public sector presence can bring complementary investments in regional and national capital. However, the small scale of local government in our context means our primary mechanisms differ from those researched in this body of work. The rapid growth induced by county-seat selections in the frontier period greatly exceeds the population effects in other contexts, and we find null or negative effects on per-person public-goods spending. Overall, we argue that the long-term county-capital effect is driven by the historical migration effects rather than the political economy aspects discussed in this literature. If our effects were driven only by public employment, it would imply a multiplier of around 75, about two orders of magnitude higher than that found by other papers (Faggio and Overman 2014; Becker, Heblich and Sturm 2021).

Section 2 discusses the historical significance of county-seat elections, Section 3 describes our data collection and assembly, and Section 4 introduces our research design and confirms the relevance (first stage) and exogeneity (balance) of the close-election RD design. Section 5 shows how county seats selected a spatial equilibrium through coordinating population and examines the long-run population and economic impacts and Section 6 details how the results changed over time. Section 7 examines aggregate effects with machine learning, Section 8 tests alternative mechanisms, and Section 9 concludes.

2 Historical Background

2.1 County-Seat Functions

New counties on the American frontier faced the decision of where to locate their seat of government. Tangibly, a county seat required a courthouse, storage of official records, and related employment. County political and administrative officials such as commissioners, judges, or the county clerk would typically live or work in the county seat. The exact functions of county government and court systems depend on which functions individual states have devolved to localities, but some commonalities exist. County governments are usually responsible for property tax assessment, some expenditure on schools, and they often maintain police forces. Some also provide other local public goods such as health facilities, utilities, or parks. No formal rule dictates that county spending focus primarily on the seat as opposed to other towns, though we address questions of bias in Section 8.3. County courts process many nonjury crimes (e.g., traffic violations) and misdemeanor cases.

Today, jobs connected to the legal and administrative functions in county government are rare, representing only 0.24% of national employment. Total county employment accounts for 1.7%, but the majority of these jobs are tied to specific communities rather than the seat (Census 2012); common examples include teachers and police officers. The employment figures are higher within seats themselves, but still quite small; see Appendix Figure C.2.

Intangibly, being named the county seat accorded a large amount of prestige, indicating a town's premier position within a county: "Towns desired county seats ... because the designation brought increased status for the town" (Paher 1969). This status can help explain why, despite few government jobs, county seats are the largest towns in 78% of counties.[1]

2.2 County-Seat Elections

Many counties historically chose seats by holding county-wide elections. While such decisions were typically local, there was substantial state-level variation in this choice. In New England, most states had established counties and seats from the colonial era and desire for change was uncommon. In frontier states that were rapidly establishing new counties, elections were far more common. In states like Kansas they were essentially the default method, though others like Michigan used a mix of elections and non-electoral decisions made by county officials. Most elections involved two locations competing for a majority, but automatic runoff and first past the post systems were also common. Less commonly, pre-election policies at the state level were designed to favor incumbents or other locations and required supermajorities to defeat these locations.

In approximately 80% of counties in our sample, we observe only a single election, usually around the time of a county's creation ("incorporation"). Consequently, election results were usually permanent, though not always. Counties could undo an initial selection with a later election, for example, and our data include elections held decades after incorporation. To avoid issues related to the endogenous calls for elections, our main results focus on only the first election for each county; see Section 4.2.

2.3 Coordinating Role of the County Seat

Despite the small size of county governments, county-seat status was a significant draw for migrants in the frontier period: "Status in the governmental hierarchy can be recognized as an important economic centralizing influence ... the important growth of any frontier village

[1] Authors' calculation for all counties nationwide.

to larger status dates from its becoming a county seat" (Knight 1973). Securing the county seat was a clear signal that a town would become the premier community in its area, meaning the choice of seat served as a coordinating mechanism for early migrants. One way in which information about county seats would spread is through guidebooks written for migrants to western states. These would often highlight county-seat status, indicating its salience; see Bleakley, Rhode et al. (2023) and its underlying sources.

Individual county histories emphasize that population influxes were large and directly followed the election. In Yolo County, California, Woodland gained county seat status and then "entered upon an era of business and social activity. Buildings were erected rapidly, business developed and new people sought a home in the thriving new town" (Gregory 1913). Similarly, in Plainview, Texas, when "the first courthouse was ... completed ... [w]ithin a year the town grew to a population of seventy-five and had a hotel" (Davis 1952). Advocates for the competing communities were cognizant of the high stakes involved, leading to bitter disputes playing out in speeches and newspaper editorials. Some observers thus termed these contests "county seat wars" even though actual violence was quite rare (Schellenberg 1987).

Several further examples highlight that this rapid growth resulted from coordinated movements of people, often organized by prominent citizens. In Dawson County, Texas "Lamesa won the county seat election by five votes. ... A town meeting the next day invited the citizens and merchants of Stemmons to move to the new county seat. ... The offer was accepted and effected within several days" (Texas State Historical Association 2016). In 1871, Phoenix, Arizona's "commissioners quickly followed up their [county seat] electoral success by offering more lots for sale" (Mawn 1977). Here, the commissioners served as prominent organizers who anticipated and supported increased demand for residence in their town. As a final example, Leoti, Kansas triumphed over Coronado in 1888 and an "exodus began immediately ... during the fall and winter of 1888-1889 all of the hundred or more buildings in Coronado were moved to Leoti" (DeArment 2006).

Notably, these histories do not describe a slower process of growth based on jobs or public works from the county government (we discuss this further in section 5). Indeed, such spending was very limited, particularly in this era, and could not have been enacted quickly enough to precede the population movements. In 1913, the average county government had a budget about 2% the size of the average state government and 0.02% of the national government (Sylla, Legler and Wallis 1995). Instead, encouraged by prominent citizens of the election winner, large numbers of citizens relocated to the seat in a coordinated fashion.

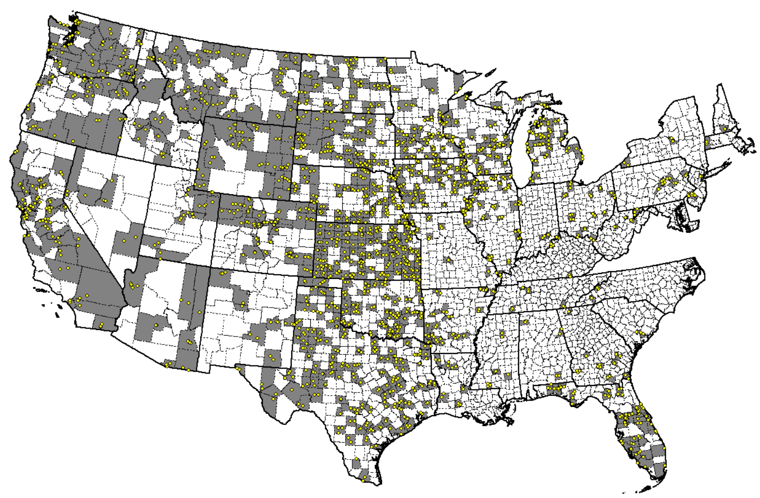
3 Data

Here we provide brief overview of our data sources and variable construction. More details are included in Appendix Section A.

3.1 County-Seat Elections

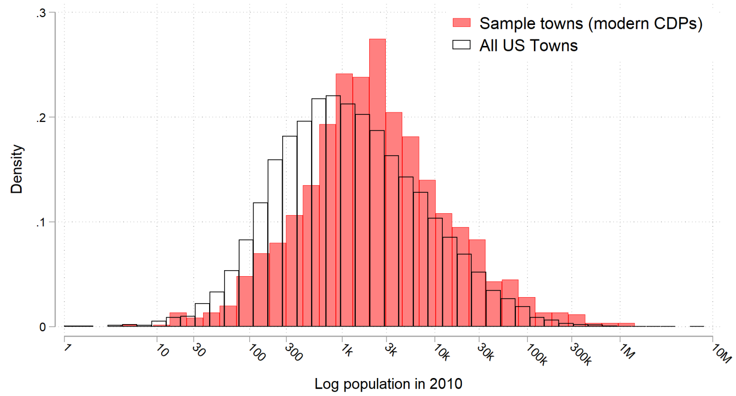
We assemble an original data set of county-seat elections. No comprehensive data set on such elections exists with national or even state-level coverage, meaning an extensive county-by-county search was required. We draw from historical newspapers, state and local archives, county histories, historical societies, and other administrative sources. Our data set consists of 1117 county-seat elections across 811 counties in 42 states mapped in Figure 1. While we cannot formally calculate the fraction of such elections for which we obtain voting results, we assess that a large majority appear in our dataset based on the variety of sources we consult. Many of the counties missing in Figure 1 in fact represent cases where the seat was chosen non-electorally. For example, state legislatures or county officials often selected an initial location which remained unchanged thereafter.

Figure 1: Sample Counties and Towns



Collectively, the counties in our sample were home to 71 million people in 2010, or 23% of the nation's population. Within those counties, the modern (2010) towns covered by our data are fairly close in size to the overall distribution of towns within the United States, albeit modestly larger. Figure 2 plots the 2010 populations of the census places in our sample, compared to all other census places. Appendix Figure C.1 replicates this graph for towns with close (within the main RD bandwidth) elections, and it leaves the pattern substantively unchanged.

Figure 2: 2010 Populations of Sample Census Places



3.2 Other Data

We combine the county-seat election data on historical populations with modern population and economic data from the US Census. We base 2010 population figures on census-block data and historical town-level population data from Schmidt (2018). For modern income, we use Internal Revenue Service (IRS) statistics for 2010, reported at the zip code level. We draw census-block-level data on population and job characteristics from the 2010 LODES data and census figures.

For other historical data, we turn to IPUMS public census microdata from 1940 and earlier and a geocoded panel data set on US post office locations throughout the nation from Blevins (2021). Elevation data are constructed from the SRTM data set, and stream shapefiles are from ESRI. Railroad data are from Atack (2016) and "U.S. National Transportation Atlas Railroads" (2015, distributed by ESRI).

3.3 Linking Data Sources

Some care is required to consistently link the economic data to the towns. In the large majority of cases (82% within close elections), the competing towns are US census-designated places in 2010. In earlier census years, these communities typically corresponded exactly to one or more enumeration districts. In these cases, we link to the corresponding census blocks or enumeration districts, making use of the Census Place Project for historical microdata (Berkes, Karger and Nencka 2023). For the remainder, we match to small subcounty units of the rural areas close to the original site: enumeration districts (census microdata), zip codes (IRS tax data), or 2010 census blocks. See Appendix Section A.14 for further details.

We avoid sample attrition in this procedure by geolocating all competing locations of close elections, regardless of current status. To do so, we extensively review historical county maps and histories and locate the original sites for all cases within the RD bandwidth. The final baseline sample consists of 2130 unique locations, with 1935 geolocated.

3.4 Summary Statistics

Appendix Table C.1 presents summary statistics for the locations within our unrestricted sample. The median election occurs in 1884, and 95% of elections fall between 1840 and 1922. Because we geolocate all towns within our RD bandwidth, it is rare for relevant data to be missing, though it does occur when other data sources (for example, IRS filing data) omit values.

The modern statistics reflect a process of structural transformation as depicted in Appendix Figure C.2. On average, winning towns begin with a primarily agricultural economy that transitions into a range of blue- and white-collar occupations in subsequent decades. In contrast to the economies of most national capitals, local government and lawyer/judge positions do not account for a significant share of jobs. Even among winners in post-election years, these positions account for less than 1% of employment.[2] Along with other evidence, these statistics inform our choice to focus on the effects of density and agglomeration rather than multiplier effects from local government.

4 Econometric Approach

4.1 Regression Discontinuity

Our main identification strategy consists of RD analyses. In keeping with the best practices recommended in the literature (Calonico, Cattaneo and Titiunik 2014; Gelman and Imbens 2019), we use the following specification:

$$y_i = \alpha \mathbb{1}\left(\text{Win Election}\right)_i + f(\text{Victory Margin})_i + X_i\beta + \varepsilon_i \tag{1}$$

Here, f is a local linear function in vote percentages separately estimated for each side, X_i are controls, and ε_i are error terms clustered by election. By comparing towns that narrowly won to those that narrowly lost, α captures the causal effect of a county-seat election victory on agglomeration measures. Controls typically include county fixed effects and geographic controls (listed in Appendix Section A.12) that cover land quality, pre-election demographics,

[2] See Appendix Section A.15 for definitions. In the census microdata, these figures are inflated because the data combine county- and municipal-government employment into a "local" government category. The figures for county-government employment would necessarily be lower.

railroads, and other infrastructure. The pre-election restriction is necessary to avoid the use of endogenous or downstream controls.

For comparability, we use a single default bandwidth of 14.5 percentage points (pp) selected via the Calonico, Cattaneo and Titiunik (2014) procedure for the first-stage regression without controls. Appendix Section B.2 shows robustness to bandwidth choice, with bandwidths ranging from 5 to 20pp, none of which substantially change the estimates.

4.2 Sample Selection

Our baseline sample includes only the first election for each county. Although 80% of counties in our sample have just one recorded election, this selection procedure avoids potential endogenous selection for the remainder.[3] When we consider effects based on election timing in Section 6, we include all elections to maximize our sample for later periods.

4.3 First Stage and Balance

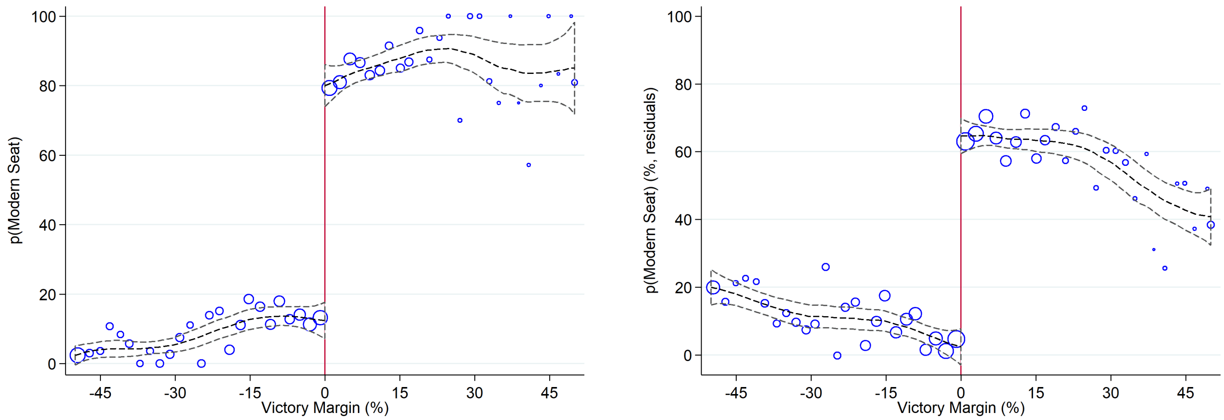
We first confirm that these elections actually determine county-seat choices and that winning is statistically balanced across predetermined characteristics. Figure 3 shows the RD plot and indicates that victory leads to a 67 percentage-point increase in probability of being the modern county seat. We should not expect a 100 percentage-point effect here, as subsequent elections or legislative action could ultimately overturn victories. Still, the estimate indicates that the majority of elections were final; consequently we present subsequent results as reduced-form impacts of victory.

Two categories of issues could have theoretically violated the RD continuity assumption. First, the standard close election RD concerns of fraud or electoral manipulation could apply. Ex-ante, we view this as unlikely since few early frontier settlements would have had the capacity for the "precise control" (Das 2023) necessary to manipulate votes exactly at the cutoff; see Section B.1. Second, bias could stem from elections missing in our sample. Note that such attrition bias would have to come at the county level rather than the from the standard case of attrition in individual observations. Within an election, we obtain data on all close winners and losers through the process described in Section 3.3 and our main results are within-county. For our strategy to be biased, we would have to be differentially more likely to observe elections based on the winner's identity or outcomes.

Balance tests across a range of pre-election and geographic characteristics do not support

[3]Hypothetically, a motivated election loser could call for another election, leading to a sample imbalance in a set of all elections; focusing on a single election removes this possibility. The first election includes runoff rounds where applicable.

Figure 3: Probability of Being Modern County Seat



(a) No Controls **(b)** All Controls

Notes: Binned scatter plot and local linear fit on the probability of being a county seat in the present day. Size of circles indicates number of observations in a bin. Panel (a) ignores controls, and Panel (b) residualizes the y-axis on county fixed effects and geographic controls. Geographic controls are listed in Section A.12. [no controls] <u>RD Estimate</u>: +67.56pp, (z = 16.32)*** [controls] <u>RD Estimate</u>: +66.30pp, (z = 20.08)***

either form of bias. Figure C.3 depicts the results graphically. Across the 9 outcomes, no differences are statistically significant and the highest magnitude among nine z-scores is 1.17, about what would be expected from chance alone. Any form of bias would therefore have to work through unobservables while simultaneously leaving observable characteristics statistically balanced.

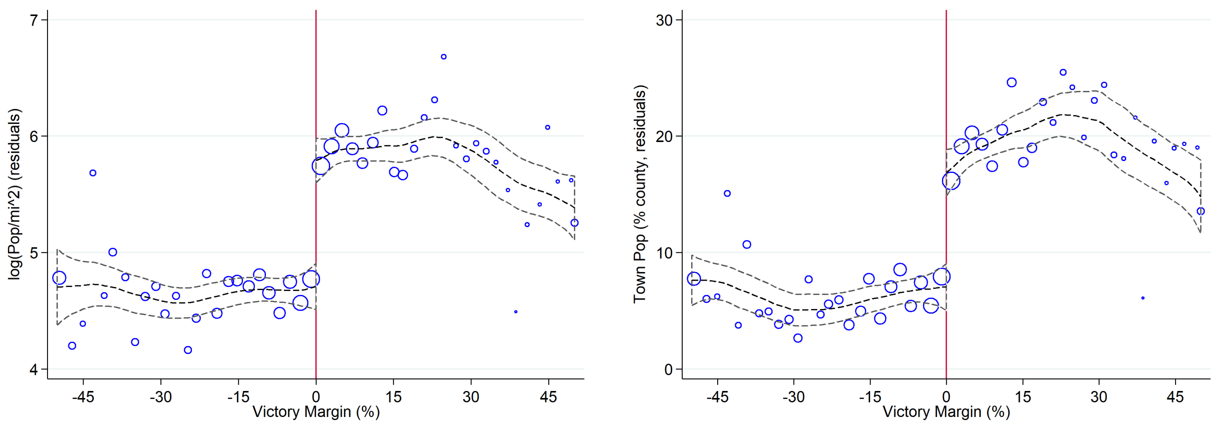
5 Status and Population

In this section, we use estimates of equation (1) to analyze the effects of election victory on towns' population, prominence, and amenities.

5.1 Population Growth

We begin by establishing that the historical accident of a close county-seat election victory had large effects on long-run populations. Figure 4 shows the impact of winning on a town's 2010 population. Our preferred measure is 2010 population density within one mile of the town center (Panel [a]). By this measure, winning towns increase in density by 1.17 log points[4] (223%) over their rivals—slightly more than tripling their density. Since many of the towns in our sample are quite small, a one-mile radius is typical for their extent. While this measure could be influenced by large relative effects on small populations,[5] we also see that a county-seat victory shifts about 11% of the county's population to the winner in panel [b]. This represents a meaningful change in people's location decisions. Other population measures show larger shifts with Appendix Figure C.7 estimating that towns' 2010 census populations increase by 2.61 log points (1263%). We prefer the measures in Figure 4, however, because they rely less on the treatment of zeros.

Figure 4: Effects on 2010 Population



(a) log(Density) < 1 mile

(b) Census Population (% of County's)

Notes: binned scatter plot and local linear fit for population measures. "Density" refers to population per square mile for 2010 census blocks within the specified radius. "Census population" refers to a town's 2010 census population. "Log" refers to $\log(\max(1, x))$. Geographic controls are listed in Section A.12. Panel (a): RD Estimate: +1.17, (z = 9.67)***. Panel (b): RD Estimate: +10.86pp, (z = 8.14)***

The results suggest that the county seats proved attractive to many early migrants and had large impacts on the early distribution of the population. Extrapolating from the result on county-population fraction, 5 million people were shifted simply by the chance occurrence of close elections within our sample and bandwidth. This figure and our previous estimates highlight that we are able to study the impact of relatively large changes in the population

[4] For this and other outcomes that include zero in their support, *log* by default denotes $\log(\max(1, x))$—that is, we bottom-code pre-log values at 1. Among the solutions proposed for log-like transformations by Chen and Roth (2023), this is mathematically equivalent to their suggestion of manually calibrating the effect of a 0/1 change in which we place no weight on this transition. We view this calibration as appropriate given our focus on urban areas. A density of one person per square mile, for example, is extremely sparse and would not represent meaningful progress toward urbanization relative to even lower densities.

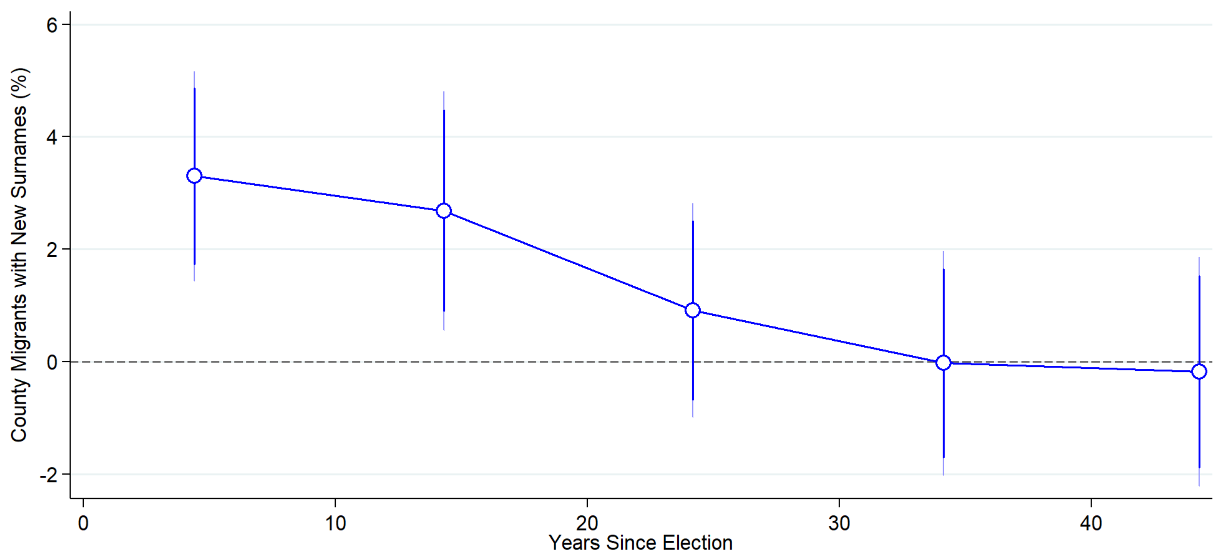
[5] For example, a town that grew from 5 to 50 people would show very large proportional growth, though the growth might not be important in absolute terms.

distribution. The approximate tripling of population density also contrasts with the small number of jobs associated with county seats. As we noted in Section 2, only 0.24% of modern employment is connected to administrative or legal functions of county government. We return to this point in Section 8.

These results can be interpreted through the lens of models of city size and path dependence. Bleakley and Lin (2015) suggest that persistent long-run differences in city size can arise as a consequence of strong local agglomeration forces even if places have similar locational fundamentals. This notion matches our context as winning and losing locations are balanced on geographic fundamentals and other pre-election characteristics but experience different population growth paths as the result of a close election victory that helps centralize a county's population. County-seat elections can then be interpreted as a direct example in which a shock selected one of many population equilibria and thereby shaped the long-run distribution of city size.

5.1.1 Who Moves in Response to Coordination?

Figure 5: Percent of Migrants with New Surnames



Notes: This graph uses linked census microdata 1850-1940 and plots estimates of the RD equation (1) on the fraction of migrants whose surnames do not match any of the county's previous residents, a proxy for lacking familial ties to the area. Migrants are defined as individuals who are linked to the previous census but not to the county in question. Each point represents an RD analysis of equation (1) a specified number of years after the election.

Consistent with a coordination story, we show that the county seat tends to attract migrants who lacked ties to the county. Previous work has emphasized that migration occurs through networks, with people basing their decisions on preexisting family or social ties (Stuart and Taylor 2021; Blumenstock, Chi and Tan 2023). Coordination, by contrast, especially affects those without strong reasons to choose one location over another. One reason for this is that the benefits of agglomeration accrue from general population density. It is therefore beneficial for in-movers to be in the location which promises the amenities related to population density even if they have no further ties to that location. We examine this using the surnames of linked individuals in census microdata. We define new arrivals as individuals who were linked to a different county in the previous census.[6] Among new arrivals to a county, we calculate the fraction in each town whose surname does not match any of those among the county's residents in the previous census – a proxy for lacking familial ties to others in the county.

Figure 5 shows that county seats disproportionately attract migrants with new surnames in the early years of rapid growth, though this effect completely attenuates after several decades alongside the population changes. In the first post-election census, 73% of migrants in the average town have new surnames, indicating that most migrants were initially unconnected to anyone in the area. Appendix Figure C.4 tells a similar story, showing that most new migrants came from outside the state. Overall, the presence of primarily unaffiliated migrants receptive to coordination helps contextualize the large population growth effects we find.

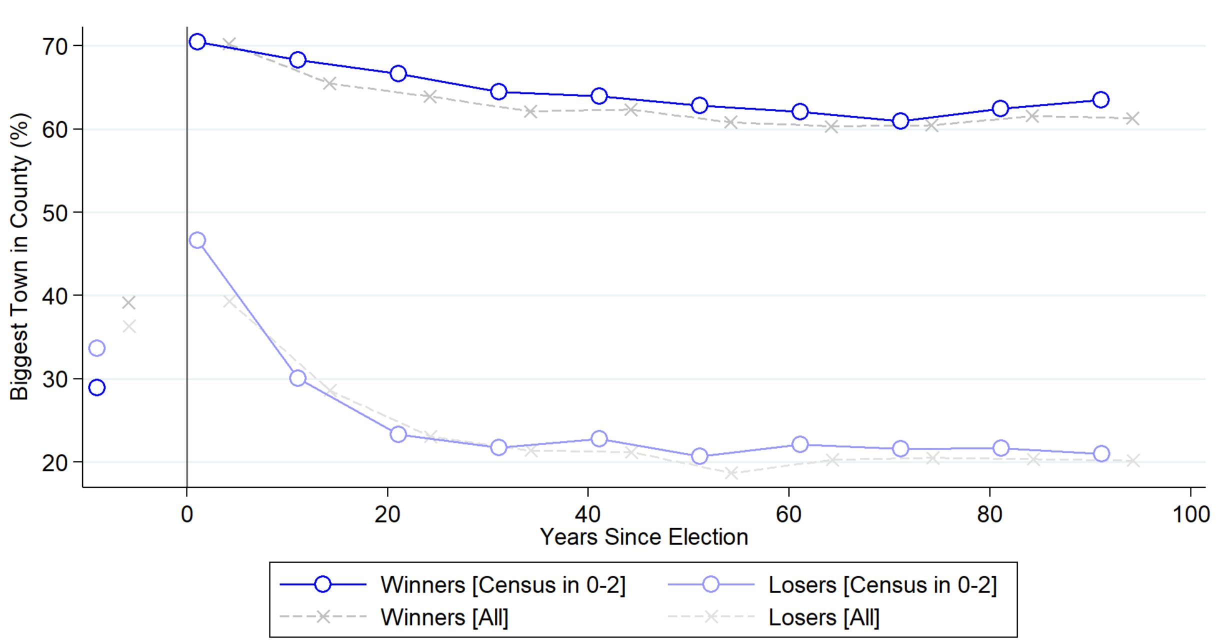
5.2 Status

How were county seats able to broadly coordinate unconnected people? Election victory caused winning towns to quickly become more prominent within their counties according to several different measures. Most apparent to new migrants would have been the fact that winners usually became the largest (or tied for)[7] towns in their counties. Figure 6 charts the evolution of this measure among winning and losing locations. In the first post-election census, winners are (weakly) the largest in about 70% of cases. Since winners remain the seat in about 84% of cases (Figure 3), this value is roughly consistent with the national average of seats being the largest towns in 78% of counties. Losing locations, despite similar pre-election probabilities of being the largest town, experience a loss in relative status over the

[6] Note that this definition mechanically excludes newborns who would not have existed anywhere in the previous census.

[7] In some cases, the census reports no towns within a county, leading to ambiguity on which site could be called the "largest." Given that post offices were allocated to larger communities (Blevins 2021), we resolve these by coding any town with a post office as tied for largest and the rest as not. It is thus possible for multiple locations to be tied for largest within a county. Within a given election, this situation can also occur when a county is split and competitors become the largest within the new counties.

Figure 6: Town Largest in County



Notes: Each point represents an RD analysis of equation (1) a specified number of years after the election. Averages are reported for both the full sample and for the subsample whose election took place 0-2 years prior to a census year. The y-axis measures whether towns are strictly or tied for largest within their counties according to census population. In cases where the census reports no towns within a county, all towns with a post office are considered (tied) for the biggest. The sample consists of the main set of "close" elections within the RD bandwidth.

next several decades, stabilizing at about a 20% chance of being the largest in their counties. As we discuss in Section 8.2, however, this rarely reflects an absolute loss of population.

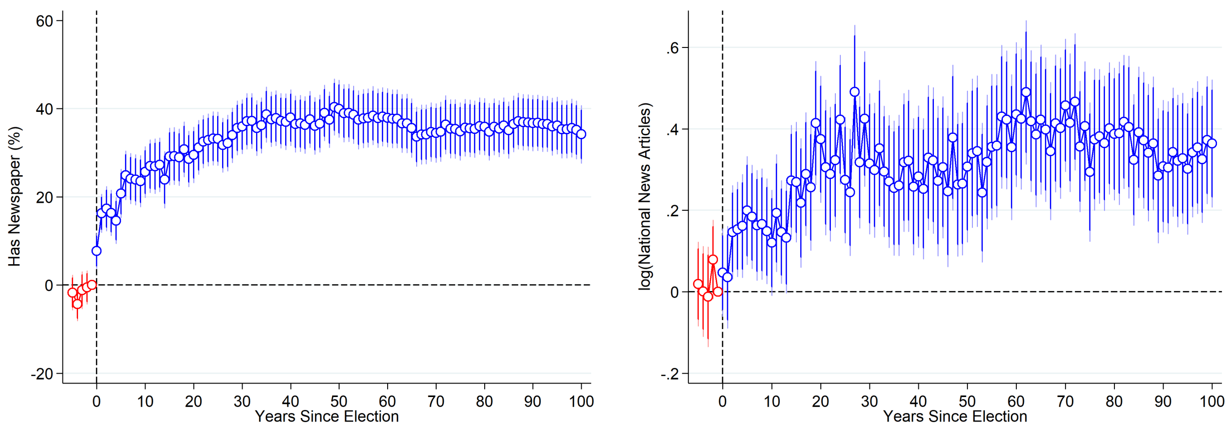
Winners typically achieved this status of "largest" town almost immediately after the election. We highlight this point by analyzing a subgroup of elections that took place just before the census years. One limitation of decadal census data is that the first post-election outcome could occur anywhere from zero to nine years after the election itself. To supplement the full sample, we additionally analyze the subgroup of elections that occurred between zero and two years before a census.[8] The pattern of Figure 6 results for this smaller group is similar to the full sample with the addition that we can demonstrate that 70% became the largest towns within one or two years of the election.

A similarly fast gain of status can be seen when measuring news coverage. For local coverage, panel [a] shows that winning towns quickly establish their own newspapers, with effects discernible even in the election year itself. We show a similar pattern for national

[8]For example, elections in 1898, 1899, and 1900 occurred zero to two years before the 1900 census.

coverage in panel [b], using the Dell et al. (2023) database of news articles. For each location and available year, we code the (log) number of times an article in the database mentions the town's name along with the county name. This measure of status also rises post-election, with winners listed in more articles within just a few years. For both variables, the relative difference between winners and losers plateaus after three to four decades.

Figure 7: News Presence (annual data)



(a) Has Newspaper (%)

(b) log(Articles), national news

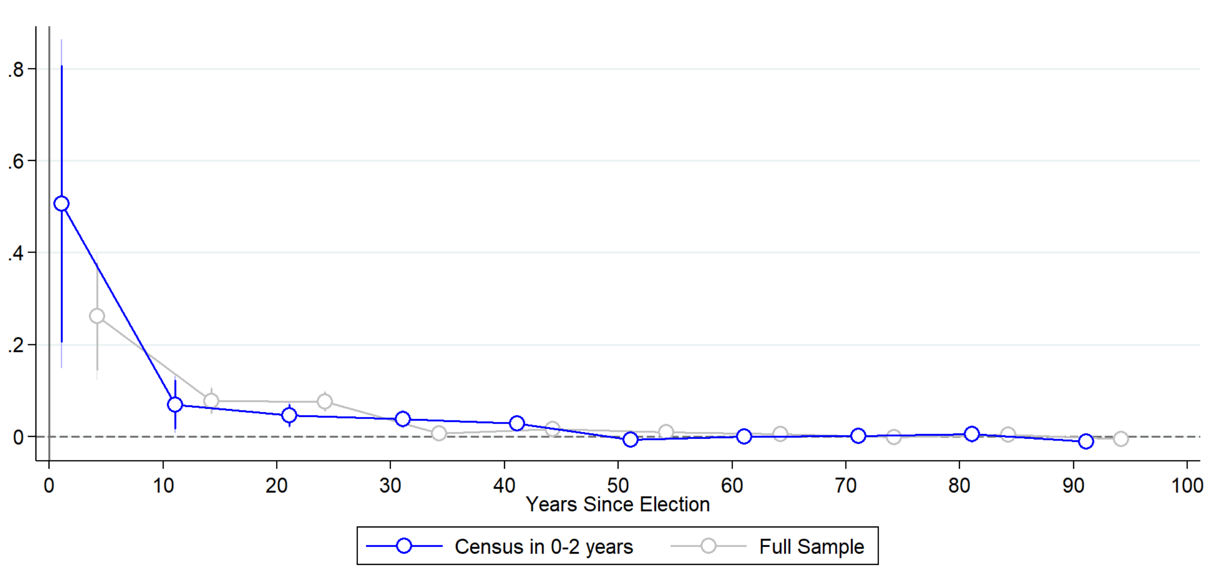
Notes: Effects over time on news coverage. Each point summarizes an RD analysis of equation (1) a specified number of years after the election. Regressions include county fixed effects and controls listed in Section A.12. (a) measures whether a town has a newspaper. (b) measures the log of national news articles in Dell et al. (2023) that mention both the town name and county name. Because the distribution is fat-tailed but the 0/1 difference is meaningful, we use the transformation $\log(\max(0.5, x))$, using the Chen and Roth (2023) suggestion of manually choosing the value of the 0-to-1 and setting it equal to the 1-to-2 changes.

Overall, a county seat election victory typically meant that a town could quickly claim premier status within its county, consistent with the historical examples discussed in Section 2.3. In a substantial majority of cases, winners became their counties' largest communities and were increasingly visible in news coverage. In addition to these facts, new migrants to an area would have directly been able to observe the county seat status itself. These rapid changes in a very short period of time suggest a coordination story where both initial and future waves of migrants understood that the county seat offered the best chance of living in a populous community. We explore this in the next section.

5.3 Timing of Population Effects

A key result supporting coordination is that county seat elections led to a very rapid centralization of population in the winning location. Consistent with the historical literature cited in Section 2.3, Figure 8 shows that increases in winning towns' census populations[9] were very rapid and highest in the first few post-election years before tapering off. As in Section 5.2, analyzing the subgroup of elections that occurred within two years of a census allows us to show that the increased growth rates occurred in a very short time frame and tapered off after several decades.

Figure 8: Effects on (log) Census Population Growth Rates



Notes: Estimates of RD equation (1) on growth rates in town census populations between censuses. Formally, we take the differences in log census population from the last census, again bottom-coding population at 1 for this purpose. To calculate growth rates, we divide this by the number of post-election years across censuses with elections in census years being assumed to have one year of treatment. Both the full sample of elections and those that took place zero to two years before the next census are analyzed separately.

Other measures of town size show a similar dynamic. In particular, post offices offer an important measure of community presence in this era. Because of its mission to deliver services to rural communities broadly, post offices were located in many nascent towns too informal to be considered in the census (Blevins 2021). Further, post office presence data

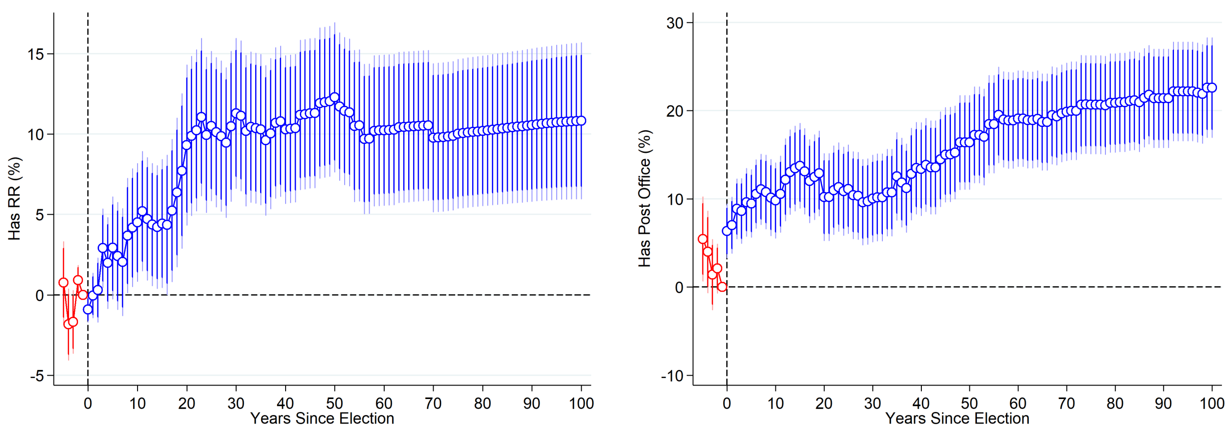
[9] Unfortunately, the density data we use for the modern period are unavailable historically. As noted earlier, we bottom-code populations at 1, but differences may still reflect census recognition of towns. However, the pattern of rapid initial gains with tapering after a few decades is broadly consistent with the other results in this paper.

are available on an annual basis and are not subject to definitional uncertainty like census towns. As suggested by the results of this section, Figure 9's estimates on post offices show that winning towns grow quickly within a single year of the election with growth again slowing after several decades. Both decadal and annual measures of town size thus support the historical literature in stating that populations quickly coordinated around the county seat.

5.4 Amenities

Amenity provision is a key benefit for locating in the larger communities, which county seats quickly became. Other benefits highlighted by previous work on capital cities, such as government employment (Bai and Jia 2021; Chambru, Henry and Marx 2021), are much more limited in the context of the small county governments we study.

Figure 9: Amenity Provision (annual data)



(a) Has Railroad (%) **(b)** Has Post Office (%)

Notes: Effects over time on the presence of a railroad or post office. Each point summarizes an RD analysis of equation (1) a specified number of years after the election. Regressions include county fixed effects and controls listed in Section A.12.

Figure 9 shows that county seats rapidly gained amenities as a result of election victories. We focus on two we can observe on an annual basis: railroad access and post offices. Each point in the figure represents the result of an RD analysis of equation (1) and an outcome measured a specified number of years after the election. In both cases, the gains are fairly rapid, though railroads' arrival lags the population increase; its most rapid rise begins around 20 years after elections.[10] As with the town population effects in Section 5.3, the differences

[10]This lag suggests that railroads did not play an dominant role in coordinating population. We can

stabilize after 30 or so years. The effects on post offices are similar, albeit with a quicker rise, likely reflecting their presence in very small and informal communities.[11] The rapid increase in amenities in the post-election years underlines the causal importance of county seats as influences on early-stage economic geography. Compared with the literature on larger national and regional capitals, the adjustments in our context are much larger in relative terms and occur more quickly.[12]

In the rural counties that represent the bulk of our sample, a number of important amenities were located in a handful of locations. Living in a county's largest town gave migrants the best chance of accessing these amenities. However, as discussed in Section 8.3, election winners do not appear to have more amenities conditional on their population, relative to losers. That is, county seats do not appear to provide more public goods than towns of similar size. The amenities in a county seat thus appear to be a result of its higher population rather than the location of the government per se, highlighting the importance of the agglomeration channel.

5.5 Discussion: the County Seat as Coordination Mechanism

Taking stock, we have shown that the politically organized county-seat shock effectively selected a spatial equilibrium with winners as population centers. Our results suggest that county seats were increasingly visible to new in-movers as attractive locations by rapidly establishing themselves as the largest locations in their counties and receiving increased coverage in newspapers. The population growth effect was broad, but especially large among migrants without existing ties to an area. On average, this equilibrium proved persistent, with transitions completing after several decades and no relative changes afterwards. Our data on the election timing, specific competitors, and their vote shares enables us to precisely

---

also compare our estimates to the literature, though to our knowledge there are no town-level estimates of railroads' impact on population. Donaldson and Hornbeck (2016) find that removing railroads decreases a county's market access by 80% on average and an increase in market access by 1% increases county population by 0.26%. The long-run change in railroad access shown in Figure 9 is about 10 points, translating into about a 2% population growth with those figures. While railroads clearly had a considerable impact on population, their implied effect here is much smaller than the tripling in town size that we find. This provides additional assurance that railroads per se are not driving the effects we find.

[11] Unlike other outcomes, post offices show an additional rise after forty or so years. Our interpretation is that the results reflect the advent of the 1902 "rural free delivery" policy, leading many post offices in smaller towns to be considered redundant (Blevins 2021). With election losers typically smaller, they were more likely to experience closure. This dynamic is discussed in Section 8.2.

[12] For example, Faggio and Overman (2014) finds a small negative impact on private employment. Bluhm, Lessmann and Schaudt (2023) finds about a 20% long-run population shift from a cross-national dataset of regional capitals, with economic changes taking several years to materialize. Figure 4 shows that, in our context, the much smaller county government slightly more than triples population density with Figure 8 showing these gains disproportionately occur in the first years after the election.

describe how and when the selection of spatial equilibria took place.

The size and speed of the transitions we observe underpin our interpretation of this effect as a story of coordination rather than population movements following significant public or private investment in locations. Using data from censuses taken shortly after the elections and annual proxies for population, our results show growth in winners began almost immediately after the election, with the first few years exhibiting the fastest adjustments. Population increases co-moved with or preceded the provision of important amenities such as post offices and railroads in county seats. Moreover, the population movement was large, accounting for a substantial fraction of the county's population and a large relative increase in winner density. This pattern fits the historical examples in Section 2.3 which describe organized movements of people in the immediate aftermath of the county seat votes. Given this, we primarily interpret the choice of a county seat as a shock to early population size through migrants' historical choice of destination.

5.6 Long-run Agglomeration and Economic Transition

In the long run, the substantial growth in early population altered the economies of winning locations. Despite their small size, county-seat election winners transition in ways similar to larger cities: toward higher-income, service-oriented, and innovation-driven economies. Using both census microdata from 1850 to 1940 and 2010 subcounty data, we show that these transformations occurred early in history and persisted.

Across a range of measures, historical county-seat victories increase economic activity in the towns. Table 1 shows this for both 1940 and for modern data, representing effects around 50 and 125 years after the median election in our sample. While outcome definitions and collection procedures differ across the two time periods, all our results show increased economic output per person or per land area. We examine the transition timeline in detail later in this section, with Figure 10 studying outcomes that are consistently measured over time.

The centralizing effect of the county seat increases income, land values, and innovation on a per person or per land area basis. On income, both 1940 census-reported wages and salaries and aggregated 2010 IRS tax returns show increases of 16% and 7.5%. In the former measure, we highlight that this result does not depend on jobs directly connected to the seat by excluding all government workers, lawyers, and judges.[13] On land values, self-assessed home values in 1940 and formally assessed home values per acre increase by substantial amounts. Finally, despite the fairly small size of the towns we study, several measures

[13]The Zip-level IRS data for 2010 underlying the modern measure do not allow us to similarly exclude individuals.

Table 1: Long-run Effects on Economic Measures

	1940				Modern		
	(1) log(Wages) (non-gov)	(2) Home Value	(3) Ed. Years	(4) Patents / 1000	(5) log(Income)	(6) Home Value (per acre)	(7) Bachelor's Worker (%)
Win	0.16***	0.24***	0.45***	0.022***	0.075***	0.65***	1.52***
	(0.019)	(0.036)	(0.044)	(0.0041)	(0.018)	(0.098)	(0.39)
County FEs	Y	Y	Y	Y	Y	Y	Y
Geo	Y	Y	Y	Y	Y	Y	Y
SEs / Clusters	Election	Election	Election	Election	Election	Election	Election
BW (pp)	14.5	14.5	14.5	14.5	14.5	14.5	14.5
N	1044	1041	1044	1044	825	1015	990
N (clusters)	538	538	538	538	528	535	541
$E[y]$	\$721	\$1,920	8.6	.06	\$48,099	\$339,221	16%

Notes: RD estimates of impacts of county-seat victory on income and proxies. Columns (1)–(4) use census microdata from 1940 and respectively examine the log of wages and salary income excluding government and legal workers; home value; years of education; annual per capita patents (per Berkes (2018)) through 1950. Column (5) examines zip-code-level income in 2010 IRS tax data. Column (6) examines the log of assessed home values per acre in 2023 CoreLogic data. Column (7) examines the fraction of jobs with workers with bachelor's degrees in 2010 LODES. Geographic controls are listed in Section A.12. See Section A.14.5 for patent linking and Section A.14 for linking generally.

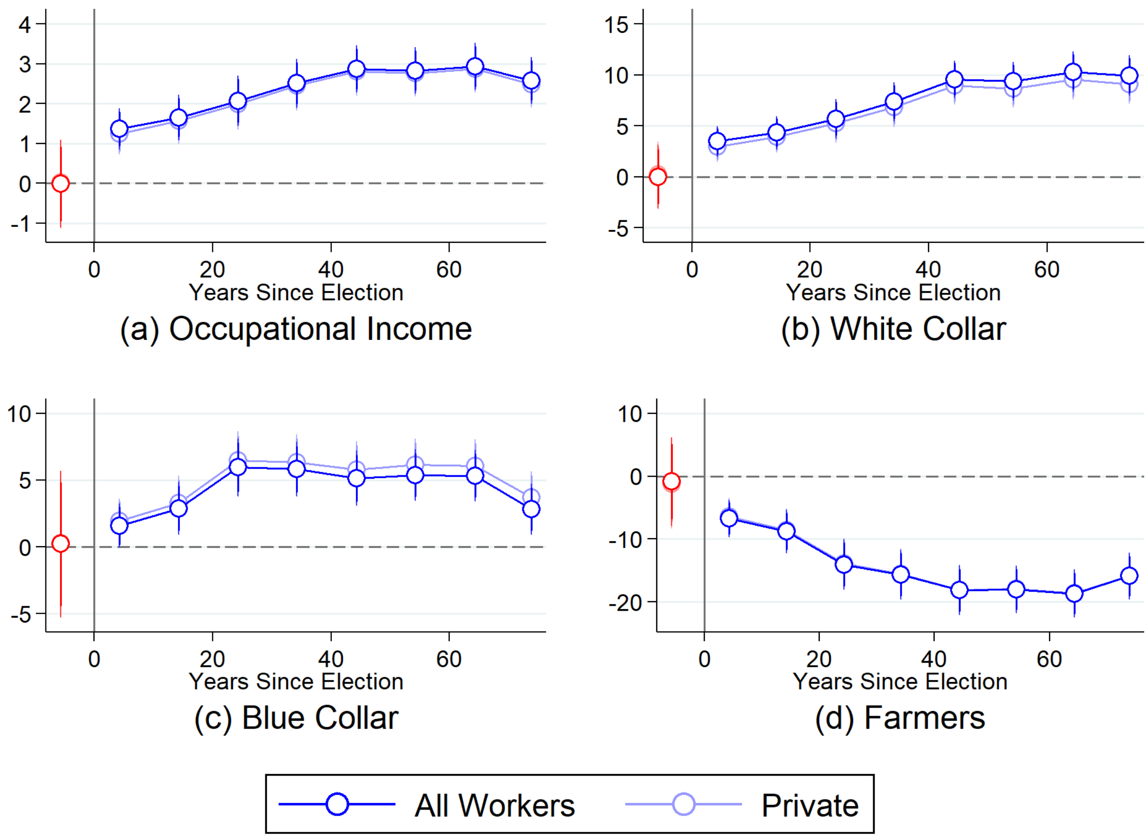
of a knowledge-based economy rise. This includes years of education (1940) and fraction of workers with bachelor's degrees (2010). Most directly, per-person rates of patenting recorded in Berkes (2018) rise by about a third when measured cumulatively from the election to 1950.

The speed of the economic transition closely mirrors those of the population and amenities estimated in Section 5.4. Figure 10 depicts the economic transformations over time according to decadal census data. Each data point represents an RD estimate of equation (1) in years relative to the election, averaged in decadal bins.[14] Winning towns experience a structural transformation, shifting jobs out of agriculture and into a broader range of blue- and white-collar occupations. Occupational-income scores, the best proxy for income in these data, similarly increase by a maximum of 2.5 points. This represents about 23% of the standard deviation in 1850, or an increase of about 13% relative to the towns' pre-election averages, similar in size to the direct income measures presented in Table 1.

Very little of the economic adjustment shown above is tied to the public sector. Indeed, the effects in Figure 10 for all workers essentially overlap with those among private-sector workers only. This similarity is essentially a mechanical result of low rates of government employment. In particular, administrative employment in local government accounts for

[14] For example, the first post-election data point averages results 0–9 years after the election, the second 10–19 years after, and so on.

Figure 10: Occupation and Income Effects (census)



Notes: This figure plots RD estimates for town-level statistics over time. Each data point represents an RD estimate of equation (1) in years relative to the election, averaged in decadal bins. Occupation data including income scores are computed based on census microdata. The income score is created by assigning each occupation in all years a value representing the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950. Each panel contains two sets of estimates: one for all workers and one for all workers except those who work in public administration.

than 1% of jobs in both winning and losing locations. Appendix Figure C.5 demonstrates a similar result by breaking down the 1940 occupational changes by private versus public sector. The largest shifts are toward white-collar workers in the private sector, mainly at the expense of agriculture. As we discuss in Section 8.1, the private sector growth cannot be directly explained by the public employment without multipliers more than an order of magnitude higher than those found in the literature.

In short, historical population-density growth in these primarily small locations precipitated the same changes we associate with density in larger urban areas. These changes stem from adjustments in the first decades after election victory and become permanent, indicating path dependence. Given the importance of this coordination to towns' long-run development, we next turn to the question of under what conditions coordination could occur.

6 Shock Timing

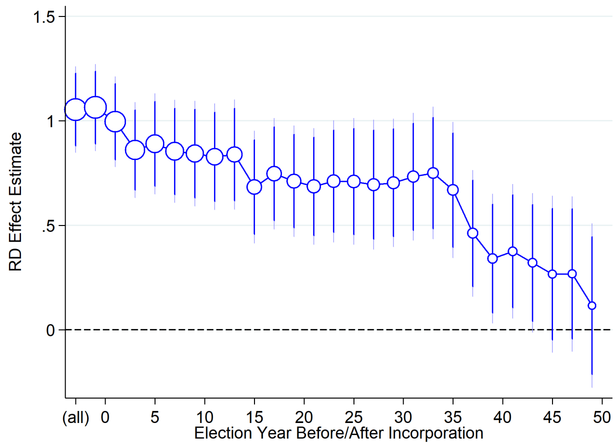
We turn to the question of when the county-seat shocks had the greatest effects on long-run economic geography. Such critical junctures may have occurred early on if initial conditions were most important in shaping long-run outcomes before population distributions became fixed. Alternatively, they may have occurred contemporaneously with later events such as the arrival of railroads or the adoption of certain technologies. To test these theories, we take advantage of the fact that we observe county-seat elections at many points in time. This is another novel feature of our paper. Empirically, we rerun our RD estimates from Section 5 and compare effects across elections that occur increasingly later. To keep time on a consistent scale across areas that were settled in different periods, we categorize elections based on the years elapsed since county creation. An election in the year a county was created occurs in year 0, which can be interpreted as an initial condition.

Figure 11 indicates the importance of initial conditions for the county-seat shock. Each point represents an RD estimate of equation (1), where the sample consists of all elections that occur at least a certain number of years after county creation;[15] moving from left to right, we can compare effects across elections in earlier versus later years. Effects on population (Panel [a]) begin close to our preferred estimates in Section 5 and attenuate to small, statistically insignificant values after about four decades. Estimated effects on income (Panels [c] and [d]) move in a similar pattern, beginning with large increases and attenuating to zero after several decades, albeit with effects on 2010 income estimated more noisily.

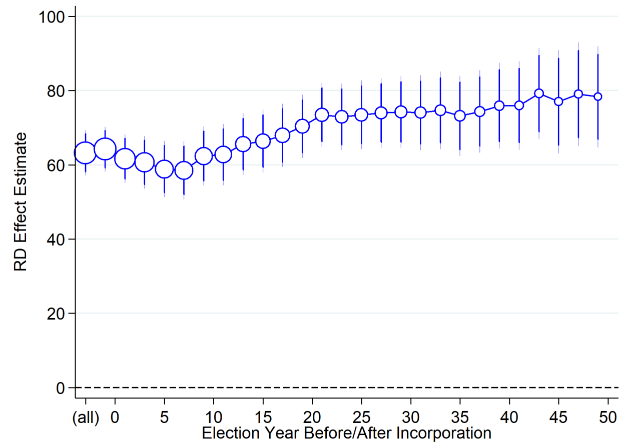
Several aspects of these results reinforce the coordination story. First, Appendix Figure C.9 produces a similar temporal comparison restricted to incumbent county seats. Compared to non-incumbents that get elected, incumbents that get reelected experience much smaller effects that become statistically insignificant after only about five years. This is consistent with the notion that the benefits of county-seat status accrue rapidly and become locked in once a location is chosen. Second, later elections were characterized by migration over shorter distances. Appendix Figure C.4 shows that the initial waves of migrants in early elections mostly came from out of state, but for later elections, migrants were mostly in-state. This is consistent with the idea that later in history, both residents and newcomers

[15] As noted in Section 3, to maximize our sample of non-initial elections, we include all elections in our database for this analysis regardless of whether we observe a previous county-seat election.

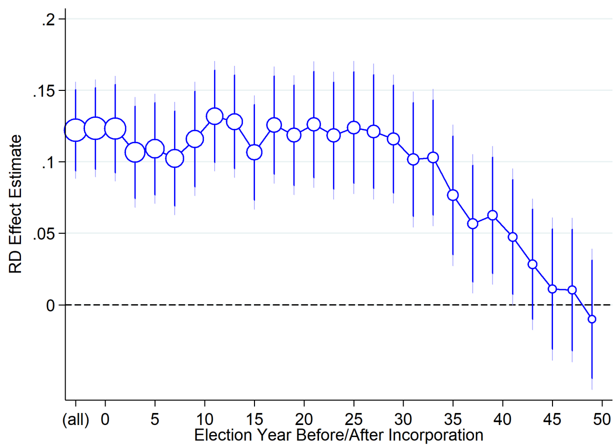
Figure 11: Effect Size Relative to County Creation



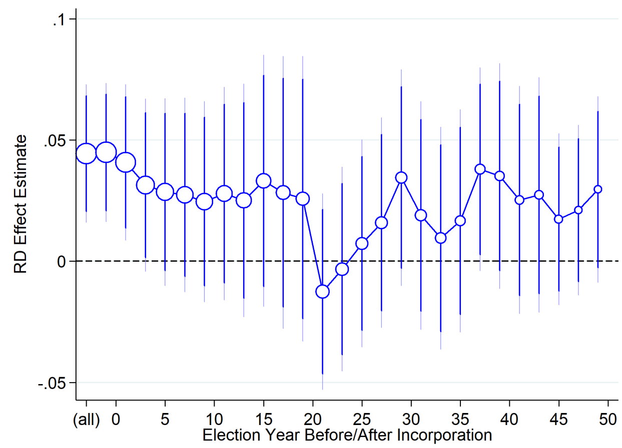
(a) (log) Population Density within 1 Mile



(b) Is Modern Seat (%)



(c) (log) 1940 Wages and Salaries, Private



(d) (log) Adjusted Gross Income (IRS zip code level)

Notes: RD estimates of log population density of census blocks within one mile of town centers (Panel [a]), the identity of the modern seat (Panel [b]), (log) 1940 reported wages and salaries for nongovernmental employees (Panel [c]), and log adjusted gross income at the zip code level (Panel [d]). Each point represents a regression on a subset of elections based on their year relative to county incorporation. The first bin analyzes the full sample, and subsequent ones restrict the sample to elections at least a specified number of years after county creation. Confidence intervals of 90% (thick line) and 95% (thin line) are shown over each estimate. Controls include the geographic characteristics listed in Section A.12 and county fixed effects.

were more familiar with towns in the county and less receptive to coordination.

Figure 11 also indicates that county government per se is relatively unimportant as an explanation for the economic adjustments we observe. Unlike the economic effects of county seat elections that shrink for late-arriving shocks, later elections are actually more likely to determine the modern capital (Panel [b]). If county governments themselves drove economic growth, we would expect to see the strongest economic changes from such late elections as they are more definitive for determining the modern seat. Instead, late elections consistently changed the county seat without many observable economic or population changes.

These results suggest a theory for the heterogeneous literature on long-term effects from historical shocks. Early in an area's history, multiple spatial equilibria are possible, and shocks may have large impacts because they affect equilibrium selection through coordination. In later years, population distributions become increasingly fixed through local returns to scale that lead to agglomeration effects and durable capital investments (see Section 5.4), so coordination is no longer possible. In these later years, towns were sufficiently established that relocating the seat did little more than move a small number of jobs in county administration whose economic impacts were minimal. Other research has explored shocks in a variety of settings, some of which are in initial time periods and others much later; the effect sizes could depend critically on this timing.[16] Our context allows us to explore the role of timing in the effects of a particular shock and this theory fits both the quantitative and qualitative evidence.

Table 2: Effects on 2010 Density by Pre-Election Status

	Railroad		Newspaper		Post Office		Census Pop > 0	
	(1) No	(2) Yes	(3) No	(4) Yes	(5) No	(6) Yes	(7) No	(8) Yes
Win	1.50***	0.54***	1.21***	0.70***	1.12***	0.78***	1.32***	-0.0038
	(0.14)	(0.096)	(0.15)	(0.094)	(0.16)	(0.11)	(0.14)	(0.071)
County FEs	Y	Y	Y	Y	Y	Y	Y	Y
Geo	Y	Y	Y	Y	Y	Y	Y	Y
SEs / Clusters	Election	Election	Election	Election	Election	Election	Election	Election
BW (pp)	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5
N	750	302	581	471	260	792	751	301
N (clusters)	418	198	375	332	196	469	429	201
$E[y]$	953	1,228	842	1,266	1,417	906	933	1,280
E[Years Post Incorporation]	14.3	27.9	14.3	23.1	11.4	20.5	12.5	32.3

Notes: RD estimates of impacts of county-seat victory on log population density within 1 mile, 2010. Each column restricts the main sample of elections to towns based on the presence of absence of a characteristic. Respectively, these consistent of a pre-election railroad, newspaper, post office, and positive population in the most recent pre-election census.

A key reason county seat elections became less consequential over time is that competing

[16] As examples, Bleakley and Lin (2012); Ager et al. (2020) find large roles for geography-based initial conditions in shaping economic geography, but Becker, Heblich and Sturm (2021) finds only "modest" private sector growth from a "substantial" shift in public employment due to the West German capital shifting to Bonn. Davis and Weinstein (2002) finds only temporary impacts from large-scale bombing of Japanese cities in WWII, but suggests that initial conditions may be quite important given the range of densities dating from the Stone Age.

locations were more established, populous, and had more amenities. In these cases, patterns of housing and infrastructure were much more settled and coordination played a more limited role. Table 2 demonstrates this by comparing the effects of election victory in locations with and without particular features. Towns with railroads, newspapers, or post offices all show positive but diminished changes in long-run density as a result of an election victory. More starkly, towns that were established enough to have been previously recognized by the census were not detectably changed. While it is not a primary goal of this paper to disentangle the precise features that solidified a town's existence, the results here point to the formalization and size of communities as the most important component; county seats could still play coordinating roles even after the arrival of economically important investments like railroads. Given this, initial conditions appear to be the most pivotal period for coordination shocks to have long-term effects in this setting.

7 Aggregate Effects

We now explore whether the county-seat decision shaped outcomes at more aggregate levels than the competing towns. Specifically, we consider whether particular choices of seat could influence county-level outcomes. Even given the large benefits to winners, these changes could very plausibly have been zero-sum across competing locations. Alternatively, some locations may have been better suited for the population influx caused by county-seat status, and the selection of county seat would have aggregate effects.

Our analysis hinges on using close election losers as identifiable counterfactual recipients of the county seat shock. This feature of our setting is crucial, as it allows us to compare county-level outcomes across specific choices.

7.1 Machine Learning Approach

We examine the question of aggregate effects through the lens of heterogeneous treatment effects. Using county-level data as our outcome, we consider a modification of equation (1) that allows for location-specific effects:

$$y_c = \alpha_{i,c}\mathbb{1}\left(\text{Win Election}\right)_i + f(\text{Victory Margin})_{i,c} + X_{i,c}\beta + \varepsilon_{i,c} \tag{2}$$

where the equation is unchanged except that treatment effects $\alpha_{i,c}$ can differ for each location i in county c. As described below, we will allow $\alpha_{i,c}$ to vary based on observable characteristics of the competing locations. Since each election can only have a single winner, the true estimates of $\alpha_{i,c}$ should average to zero among the competing towns. Mechanically, ignoring any heterogeneity, methods to estimate equation (2) will produce coefficients of α close to zero since each election will have a location with a positive and negative vote margin but with the same value of y_c.[17] However, this is consistent with both a zero-sum situation where all $\alpha_{i,c}$ are close to zero or a situation with aggregate effects where some $\alpha_{i,c}$ are positive and some are negative. In the latter situation, different choices of county seat would have aggregate effects.

To estimate the $\alpha_{i,c}$, we implement the recently developed machine learning technique of generalized random forests (GRF) (Athey, Tibshirani and Wager 2019). This methodology extends the classic random forest approach (which predicts fixed variables) to the prediction of heterogeneous treatment effects. Given the RD equation (2), the algorithm attempts to predict the $\alpha_{i,c}$ from the town characteristics $X_{i,c}$ in a flexible manner. As before, we use only the predetermined controls listed in Section A.12 as these predictors. To avoid problems with overfitting, all $\alpha_{i,c}$ predictions are made out-of-sample for each location, clustered at the (pre-election) county level. That is, only data from outside a particular county are used to estimate predictions for towns within that county.

7.2 GRF versus Voter Choices

We apply the GRF method to estimate $\alpha_{i,c}$, focusing on locations to maximize long-run county density, defined as the log population density in 2010 within the pre-election county boundaries. For comparability, we restrict the sample to the top two vote-getters within each election and display a histogram of their values in Appendix Figure C.6. The average estimate is sensibly close to 0, but the top and secondary location selected by the algorithm usually diverge significantly. Taken at face value, this spread suggests there is potential for one of the competing locations to impact aggregate outcomes more than the other. We aim to validate the model's predictions below in Section 7.3.

Appendix Table C.4 summarizes and contrasts the determinants of the $\alpha_{i,c}$ predicted by the GRF algorithm and the determinants of choices historically made by voters. We include two sets of predictors: the top five variables most important in choosing the random forest splits and a set of five manually chosen variables that measure town size and status.[18] All regressions include election fixed effects, mechanically centering the average treatment effect

[17] Because some locations have more than two competing locations, the margins need not balance out exactly so estimates will not be exactly 0. In practice, most elections have only two competitors and very few third-place challengers have a substantial vote share, leaving the data almost symmetric around the margin of 0.

[18] For the first category, this is defined as the fraction of trees whose first branch or split is determined by the values of that variable, a standard metric of importance in random forest algorithms. For the second set, we include (log) census population; the presence of a newspaper, railroad, or post office; and the dummy for whether a town has an advantaged status in the election, typically due to incumbency.

on zero and additionally controlling for state and county characteristics.

As shown in Appendix Table C.4, the GRF algorithm tends to favor locations further from the county center with poorer agricultural land, as measured by soil quality and lack of flooding (column 1). In contrast, few of the markers of size or status predict affect the prediction. While we are cautious about making definitive statements based on this predictive exercise, one explanation is that locations best-suited for agriculture were less suited to long-run density in the modern, mostly non-agrarian economy. Similarly, in the large, sparsely populated county boundaries that existed at the time of the election, "central" locations within the county were often more remote from the developed areas that had been previously settled. Given the long-run nature of the outcome we predict, the size of status of the town at the time of the election was not particularly informative.

In contrast, voters tended to select heavily based on contemporary size, status, and convenience (column 2). Larger towns, incumbency, newspapers, and centrality within the county all positively predict vote margins. As such, the GRF-predicted effects on density actually correlate negatively with voters' choices (column 3). The difference between the algorithm's choice and vote margins is notable but perhaps not surprising. Even with perfect foresight, the benefits to any individual voter from increased county density many generations in the future would be unclear at best. As shown earlier in this work, the economic benefits of a county seat to a voter's own town would have been immediate and, qualitatively, most voters tended to choose their hometown.

7.3 Validating the Predictions

We validate the model's predicted choices using a county-level RD. In theory, the range of predicted treatment effects in Figure C.6 could simply have reflected statistical noise rather than true differences. In that case, counties whose voters chose the location preferred by the machine learning model should be similar to those that did not. This motivates the following county-level RD equation

$$y_c = \alpha \mathbb{1}\left(\text{Best Won}\right)_c + f(\text{Best Margin})_c + X_c\beta + \varepsilon_c \tag{3}$$

where, keeping one observation per county, we define the running variable as the vote margin of the town the machine learning model predicts as the density-maximizing option. Given the different sample size, we re-estimate the optimal bandwidth. As controls, we include county-level counterparts X_c to our main control set listed in Appendix Section B.4.

Table 3 shows that the model's density-maximizing choice increased both county density and income. In our preferred specification that includes all controls, counties that narrowly select this seat increase their density by 51% and income by 5.1% . Given that we now have only one observation per election, Table 3 uses robust standard errors which yield statistically significant results for all specifications. However, these do not account for the two-step estimation process. As described in Appendix Section B.4, we address this by bootstrapping our estimates with 500 trials using county-level clustering. This process respectively gives p=0.012 for density and p=0.076 for income; we interpret the latter result as marginally significant.

Table 3: Effects on County-Level Outcomes, 2010

	log(Pop/mi^2)			log(Income), IRS		
	(1)	(2)	(3)	(4)	(5)	(6)
Best Wins	0.84***	0.63***	0.51***	0.083**	0.073***	0.051**
	(0.30)	(0.15)	(0.14)	(0.042)	(0.023)	(0.022)
State FEs		Y	Y		Y	Y
State × x,y		Y	Y		Y	Y
County Geo			Y			Y
SEs / Clusters	Robust	Robust	Robust	Robust	Robust	Robust
BW (pp)	9.25	9.25	9.25	9.25	9.25	9.25
N	387	387	387	387	387	387
$E[y]$	159/mi^2	159/mi^2	159/mi^2	$48,776	$48,776	$48,776

Notes: RD estimates of impacts of the ideal county-seat victory on (pre-election) county outcomes in 2010, as described in Section 7.1. Values are constructed based on 2010 census blocks and zip code polygons, assuming a constant population density within each boundary. Columns (1)–(3) consider (log) population density per square mile. Columns (4)–(6) consider (log) average income, defined as average filer adjusted gross income. County geographic controls are listed in Appendix B.3. Bootstrapped estimates and significance are shown in Appendix Section B.4.

Two things are notable about these results. First, the aggregate effects of the county-seat choice are not zero, particularly for density and plausibly for county income. The density result confirms the relevance of the GRF model's choices and our method of translating a place-level shock into a county-level one. The income result suggests that county-level density remains relevant for economic outcomes, as it did at the town level. An important caveat is that with this methodology, we cannot disentangle whether these impacts are driven by larger spatial spillovers that certain towns have from whether some towns themselves have features that enable them to grow more. Second, the coefficients estimated here are notably smaller than the town-level results shown in Table 1 and Figure 4. The effects of a shock to one particular place are attenuated at a broader scale, even if they do not shrink to zero.

This section highlights that urban coordination was not always zero-sum in our setting. While individual places benefited significantly from receiving the county seat, ex-ante it was unclear whether this simply redistributed economic activity across similar locations. Our machine learning model indicated otherwise, producing a discernible gap in the impacts of competing locations on county-level outcomes. We validated these predictions with a county-level RD whose results show that narrowly selecting the location preferred by the model increased a county's density and plausibly its income. Although we do not aim to conduct a formal welfare analysis here, it is noteworthy that locations preferred by our model diverge from those historically preferred by voters. A planner focused on maximizing long-term population density would have allocated county seats differently.

8 Alternate Mechanisms

The most obvious alternative to our story of long-run agglomeration centers on the direct effects of county government. In this section, we first consider effects of growth directly due to government administrative employment (Bai and Jia 2021) and then consider the effects of a county capital's ability to create higher-quality public goods (Bluhm, Lessmann and Schaudt 2023; Campante and Do 2014). We also look into whether the effects could be driven by negative impacts on losing towns rather than gains to winning towns.

8.1 Government Sector

A priori, the small size of the county-government sector means its growth is unlikely to drive the population and economic changes we document. Figure 4 shows that population density triples as a result of an election victory, yet only about one in 400 job nationally is connected to county administrative and legal functions. Similarly, Figure 10 shows that occupational trends are unchanged by excluding public-sector workers.

If public employment accounted for the job growth we see, the implied multiplier on public hiring would be exceptionally large. In 1940, about 1% of employment in close winners consisted of the sorts of local-government legal or administrative jobs that are potentially connected to the county seat; for close losers, the figure is around 0.4%. Given the 223% increase in density, a multiplier of approximately 75 would be required to explain this change.[19] But Faggio and Overman (2014) find no public multiplier and Becker, Heblich and Sturm (2021) find one of about 0.7. Focusing on the private sector, Moretti (2010) finds a multiplier

[19] $2.23/\left((1+2.23)\times(1.0\%)\ -0.4\%\right)-1\approx 75$. As noted earlier, the 1% figure is an overestimate as it combines county and municipal employment; the multiplier is thus underestimated. A second caveat is that these estimates come from different years. However, we view density as best measured in modern data since area cannot be calculated for census enumeration districts and census microdata offers the most detailed look at occupations in all our data sources. Given that Figure 8 shows that population differences stabilize rapidly, the divergent choice of years should have a small impact.

of 1.6 for the private, nontradable sector. It is unlikely county government has a multiplier of 50 times the largest of these. Instead, the population growth seems to reflect the special role of the county seat in agglomerating early migrants, discussed in Sections 2 and 5.1.

As a final check, we compare our main effects in states where county government plays a larger role to those where it has a smaller role. We measure this by counting the number of major county officers (for example, sheriff or clerk) as listed in Murphy (2009) and dividing states into above and below median. Appendix Table C.3 shows that states with larger roles for county government have similar effect sizes to those with smaller roles, with some point estimates being larger and others smaller. If county-government functions were an important mechanism, we should see distinctly larger effects in states where county government is more important.

8.2 Gains or Losses?

Because the RD estimates relative differences between close election winners and losers, the results in this paper could reflect negative effects on losers rather than any gains to winners. Such a situation would not necessarily invalidate our results. If the economic outcomes we study are primarily a function of density, then we are comparing higher and lower levels of density regardless of gains and losses. However, the interpretation is more complex if population losses have an asymmetric effect compared to gains.

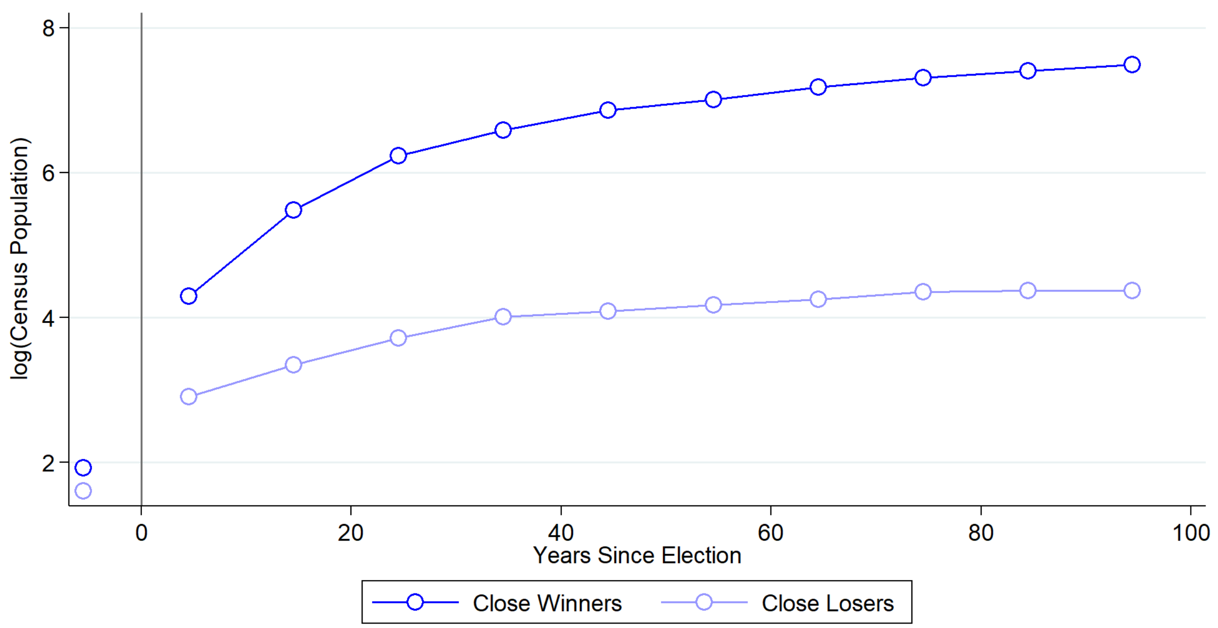
In our context, few or no losing locations lose relative to their pre-election baseline, depending on the outcome. The intuition for this result can be seen most directly in Figure 12 which charts the average (log) census population of close winning and losing towns. Both groups start off at a similar and low level prior to the election and both grow rapidly post-election. However, the winning location grows much faster.

Within the first 25 years post-election, a small minority of losing locations do experience losses of amenities, albeit at a much smaller rate than winners make gains. Appendix Figure C.8 shows the extent to which averted losses from the pre-election baseline explain overall gains in terms of railroads, census population, post offices, and newspapers.[20] Averted losses in railroad access or census population are rare and and small or negligible. For post offices and newspapers, election victory averts a loss in about 5% of locations, accounting for a minority of total gains.

Over longer periods, these losses do expand. 100 years post-election, an election loss results in about 5% of locations losing railroad access and 10% losing a post office or news-

[20]Formally, we simply run regressions of RD equation (1) on whether a particular outcome is lower than baseline, flipping the sign for ease of exposition. So, this variable = 0 if the outcome remains the same or higher than the pre-election value and -100% otherwise. Positive effects thus indicate averted losses.

Figure 12: Town Census Population over Time



Notes: the average of the log census population of towns in the main sample of "close" elections within the RD bandwidth. Log refers to $\log(\max(1, x))$.

paper. In part, these likely reflect the changing relevance of these amenities and public goods over this long duration. In recent years, local news has become less profitable and the US post office began to downsize small, rural offices with the advent of rural free delivery in 1902, effectively changing the relationship between community size and postal presence (Blevins 2015). While important overall, note that these dynamics largely post-date the rapid population growth shown in Section 5.1 and are unlikely to have influenced it directly.

8.3 Public Goods and Service Provision

Another potential explanation for our results is that being the county seat enabled a town to obtain a disproportionate share of public goods, similar to the role of national capitals discussed in Ades and Glaeser (1995). In two analyses, we show that this is unlikely, examining both continuously measurable and lumpy (discrete) public goods. Our results suggest that winning locations appear similar to other towns of similar size.

8.3.1 Continuous Measures of Public Goods

We begin by examining schools—a major focus of most county governments and a case in which a public good scales naturally with population. We turn to Common Core school-

Table 4: School-District Characteristics

	(1) log Students	(2) log Exp / Student	(3) log Rev / Student	(4) Teachers / Student
Win	0.17***	0.024	-0.0049	-0.0026***
	(0.047)	(0.015)	(0.011)	(0.00080)
County FEs	Y	Y	Y	Y
Geo	Y	Y	Y	Y
SEs / Clusters	Election	Election	Election	Election
BW (pp)	14.5	14.5	14.5	14.5
N	1049	1049	1049	1048
N (clusters)	541	541	541	541
$E[y]$	4,698	$14,812	$14,872	.075

Notes: RD estimates of impacts of county-seat victory on locations' school districts. Column (1) reports on log students. Column (2) reports on log expenditure per student. Column (3) reports on log revenue per student from the following sources: total, federal, state, local, and property taxes. Column (4) reports on teachers per student. Columns (2)–(3) are top-coded at the 1% level to remove outliers with few students. Column (4) is bottom-coded at the 1% level to account for districts with zero listed teachers. Geographic controls as listed in Section A.12.

district data for the 2017–18 school year for outcomes.[21] Table 4 shows the results. Unsurprisingly, an election victory increases the number of students in a district, though by less than the population of the winning town (column [1]).[22] Regarding the money available for students, the results suggest there is little difference. The point estimates show that winning results in a 2.4% expenditure increase per student but also 0.5% lower revenue, neither of which is statistically significant. Finally, column (4) suggests a small but statistically significant decrease in teachers per student, with about one fewer teacher per 400 students, or 3% of the sample mean. This last fact may represent a real loss to schools in a district or simply reflect the minimum scale of needing one teacher for very small classrooms.

Taken together, the results cut against the notion that migrants were attracted to the county seat to access better public goods. If anything, a small drop in the teacher-to-student ratio suggests a slight decrease in school quality.

8.3.2 Discrete Public Goods

We also argue county seats do not gain disproportionate access to other amenities relative to their population. We examine the presence of railroads and post offices studied in Section 5.4. Unlike school funding, we can trace these amenities back to earlier periods and focus

[21] The earliest for which all key data are available.

[22] This gap likely reflects the fact that low-enrollment school districts were merged to achieve administrative scale.

on the census year 10-19 years after the election and in 2010. These years give us coverage in both short- and long-run adjustments.

Figure C.10 plots the relationship between town population and the presence of the amenity for close[23] election winners and losers in our main sample. The figure displays results both for the first post-election census (short run) and 2010 (long run). Overall, the relationship is similar for both groups. The railroad access curve is slightly higher for close losers while the post office curve is slightly higher for winners. This is consistent with the view that, compared with similarly-sized towns, county seats may have had increased status but did not obtain disproportionate access to material benefits.

Given the limits of this sort of analysis, we consider these results important but suggestive. Small towns are naturally less likely than large towns to have public goods such as railroad connections, meaning that the large population increase from an election victory mechanically produces higher provision in an absolute sense, even absent favoritism. By plotting amenities with respect to population, Figure C.10 essentially conditions on a downstream control, which can bias results. With this caveat, we view the close similarity in the curves as an additional piece of evidence against favoritism in public goods provision.

9 Conclusion

This paper uses the historical politically driven shock of close county-seat elections to study the process coordination in urban growth. County-seat elections gave winning towns a reputational boost in the early phases of settlement when new cities and towns were just forming. Using an RD design, we showed that county-seat selection catalyzed a rapid population coordination and a subsequent structural transformation in winning locations that durably increased their population and income. Population and economic changes were mostly completed within a few decades and remained persistent thereafter. As a matter of economic history, the large effects we find mean that these elections played a substantial role in shaping the population distribution of the modern United States.

We establish several important facts about the temporal and spatial dynamics of this process. First, this comparatively neutral shock, unconnected to geographical or substantial productivity advantages, was able to persistently shape the distribution of towns and economic geography. Second, we show that an increase in status was sufficient to coordinate population movements and select one of several possible spatial equilibria, typically with the winning town as the county's population center. Third, the shock's impact was most pronounced in the early stages of frontier settlement, before a population equilibrium was

[23] Within the default RD bandwidth.

fully set. Shocks occurring several decades after a county's formation have minimal effects, indicating that population distributions became increasingly fixed over time. Finally, using out-of-sample machine learning predictions, we also demonstrate that this process of coordination was not a zero-sum game at the county level. Selecting certain towns over their competitors could boost counties' income and density in the long run.

Our results are consistent with theories in which human behavior plays an important role in selecting among multiple steady states in economic geography. In our setting, the political act of selecting a county seat had a large centralizing effect on the population despite the limited changes it effected in employment and spending. Geographic and other fixed characteristics are balanced in our RD design, showing a case in which these are not required to determine the distribution of town sizes, though they clearly remain important in general. Our results also show that the extent to which such equilibria can be shifted varies greatly across contexts. Most notably, population distributions become less sensitive to the same shock as initial conditions become more fixed over time.

While caution is necessary when considering the external validity of results from any specific setting, our analysis has several implications for economic theories of agglomeration in general. As noted in Section 1, previous literature has found mixed results regarding the long-run effects of historical shocks on economic geography. One plausible explanation is that shocks are more influential more when they come at the right historical moment, particularly during the "initial conditions" when urban centers just beginning to form. Once populations reach an equilibrium, it becomes much harder to transition. Additionally, while multiple equilibria often appear only as theoretical possibilities in studies of economic geography, the case of county seat selection constitutes a historically important choice between different population allocations. Our findings on county-level outcomes suggest that such choices may have consequences not only for individual locations but also for the broader distribution of economic activity across space.

References

Acemoglu, Daron, and Matthew O Jackson. 2015. "History, expectations, and leadership in the evolution of social norms." *The Review of Economic Studies*, 82(2): 423–456.

Acemoglu, Daron, Jacob Moscona, and James A Robinson. 2016. "State capacity and American technology: evidence from the nineteenth century." *American Economic Review*, 106(5): 61–67.

Ades, Alberto F, and Edward L Glaeser. 1995. "Trade and circuses: explaining urban giants." *The Quarterly Journal of Economics*, 110(1): 195–227.

Ager, Philipp, Katherine Eriksson, Casper Worm Hansen, and Lars Lønstrup. 2020. "How the 1906 San Francisco earthquake shaped economic activity in the American West." *Explorations in Economic History*, 77: 101342.

Ahlfeldt, Gabriel M., Stephen R. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica*, 83(6): 2127–2189.

Albright, Keith A. 2011. "Using Subcounty Population Estimates as Controls in Weighting for the American Community Survey."

Allen, Treb, and Dave Donaldson. 2022. "Persistence and path dependence in the spatial economy." National Bureau of Economic Research.

Andrews, Michael J. 2019. "Local Effects of Land Grant Colleges on Agricultural Innovation and Output." National Bureau of Economic Research.

Atack, Jeremy. 2016. "Historical Geographic Information Systems (GIS) database of U.S. Railroads."

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *Forthcoming in the Annals of Statistics*, 47(2).

Bai, Ying, and Ruixue Jia. 2021. "The economic consequences of political hierarchy: evidence from regime changes in China, 1000-2000 CE." *The Review of Economics and Statistics*, 1–45.

Banzhaf, H Spencer, and Randall P Walsh. 2013. "Segregation and Tiebout sorting: The link between place-based investments and neighborhood tipping." *Journal of Urban Economics*, 74: 83–98.

Becker, Sascha O, Stephan Heblich, and Daniel M Sturm. 2021. "The impact of public employment: evidence from Bonn." *Journal of Urban Economics*, 122: 103291.

Behrens, Kristian, and Frédéric Robert-Nicoud. 2015. "Agglomeration theory with heterogeneous agents." *Handbook of regional and urban economics*, 5: 171–245.

Berkes, Enrico. 2018. "Comprehensive universe of US patents (CUSP): data and facts." *UMBC Economics Department Collection.*

Berkes, Enrico, Ezra Karger, and Peter Nencka. 2023. "The census place project: A method for geolocating unstructured place names." *Explorations in Economic History*, 87: 101477.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and path dependence." *The Quarterly Journal of Economics*, 127(2): 587–644.

Bleakley, Hoyt, and Jeffrey Lin. 2015. "History and the Sizes of Cities." *American Economic Review*, 105(5): 558–563.

Bleakley, Hoyt, and Paul Rhode. 2023. "Was Free Soil Magic Dirt? Endowments versus Institutions in the Antebellum United States."

Bleakley, Hoyt, Paul W Rhode, et al. 2023. "De Tocqueville, Population Movements, and Revealed Institutional Preferences." *Journal of Historical Political Economy*, 3(2): 179–210.

Blevins, Cameron. 2015. *The Postal West.* Dissertation.

Blevins, Cameron. 2021. *Paper Trails: The US Post and the Making of the American West.* Oxford University Press.

Bluhm, Richard, Christian Lessmann, and Paul Schaudt. 2023. "The political geography of cities."

Blumenstock, Joshua E, Guanghua Chi, and Xu Tan. 2023. "Migration and the value of social networks." *Review of Economic Studies*, rdad113.

Brown, J, and David Cuberes. 2020. "The Birth and Persistence of Cities: Evidence from Oklahoma's First Fifty Years of Urban Growth."

Callen, Michael, Jonathan L Weigel, and Noam Yuchtman. 2023. "Experiments about institutions." *Annual Review of Economics*, 16.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust data-driven inference in the regression-discontinuity design." *The Stata Journal*, 14(4): 909–946.

Campante, Filipe R, and Quoc-Anh Do. 2014. "Isolated capital cities, accountability, and corruption: Evidence from US states." *American Economic Review*, 104(8): 2456–81.

Census. 2012. "2012 Census of Governments: Finance—State and Local Government."

Chambru, Cédric, Emeric Henry, and Benjamin Marx. 2021. "The dynamic consequences of state-building: evidence from the French Revolution."

Chen, Jiafeng, and Jonathan Roth. 2023. "Logs with zeros? Some problems and solutions." *arXiv preprint arXiv:2212.06080*.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial wage disparities: Sorting matters!" *Journal of urban economics*, 63(2): 723–742.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012. "The productivity advantages of large cities: Distinguishing agglomeration from firm selection." *Econometrica*, 80(6): 2543—-2594.

Das, Sabyasachi. 2023. "Democratic Backsliding in the World's Largest Democracy." *Available at SSRN 4512936*.

Davis, Charles G. 1952. "Plainview, TX (Hale County)." *Handbook of Texas Online*.

Davis, Donald R., and David E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review*, 92(5): 1269–1289.

Davis, Donald R, and David E Weinstein. 2008. "A search for multiple equilibria in urban industrial structure." *Journal of Regional Science*, 48(1): 29–65.

Davis, Donald R, and Jonathan I Dingel. 2019. "A spatial knowledge economy." *American Economic Review*, 109(1): 153–70.

DeArment, Robert K. 2006. *Ballots and bullets: The bloody county seat wars of Kansas.* University of Oklahoma Press.

De La Roca, Jorge, and Diego Puga. 2017. "Learning by working in big cities." *The Review of Economic Studies*, 84(1): 106–142.

Dell, Melissa, and Benjamin A Olken. 2020. "The development effects of the extractive colonial economy: The dutch cultivation system in java." *The Review of Economic Studies*, 87(1): 164–203.

Dell, Melissa, Jacob Carlson, Tom Bryan, Emily Silcock, Abhishek Arora, Zejiang Shen, Luca D'Amico-Wong, Quan Le, Pablo Querubin, and Leander Heldring. 2023. "American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers."

Desmet, Klaus, and Esteban Rossi-Hansberg. 2013. "Urban accounting and welfare." *American Economic Review*, 103(6): 2296–2327.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market access" approach." *The Quarterly Journal of Economics*, 131(2): 799–858.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of urban agglomeration economies." In *Handbook of regional and urban economics*. Vol. 4, 2063–2117. Elsevier.

Ellison, Glenn, and Edward L Glaeser. 1999. "The geographic concentration of industry: does natural advantage explain agglomeration?" *American Economic Review*, 89(2): 311–316.

Faggio, Giulia, and Henry Overman. 2014. "The effect of public sector employment on local labour markets." *Journal of urban economics*, 79: 91–107.

Gelman, Andrew, and Guido Imbens. 2019. "Why high-order polynomials should not be used in regression discontinuity designs." *Journal of Business & Economic Statistics*, 37(3): 447–456.

Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. 1992. "Growth in cities." *Journal of political economy*, 100(6): 1126–1152.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118(3): 536–598.

Gregory, Thomas Jefferson. 1913. *History of Yolo County, California.* Historic Record Company.

Hanlon, W Walker, and Antonio Miscio. 2017. "Agglomeration: A long-run panel data approach." *Journal of Urban Economics*, 99: 1–14.

Hanlon, W Walker, and Stephan Heblich. 2020. "History and Urban Economics." *National Bureau of Economic Research Working Paper Series*, , (w27850).

Harari, Mariaflavia. 2020. "Cities in bad shape: Urban geometry in India." *American Economic Review*, 110(8): 2377–2421.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2020. "The making of the modern metropolis: evidence from London." *The Quarterly Journal of Economics*, 135(4): 2059–2133.

Henderson, J Vernon. 1974. "The sizes and types of cities." *The American Economic Review*, 64(4): 640–656.

Henderson, J Vernon, and Anthony J Venables. 2009. "The dynamics of city formation." *Review of Economic Dynamics*, 12(2): 233–254.

Hornbeck, Richard, and Suresh Naidu. 2014. "When the levee breaks: Black migration and economic development in the American South." *American Economic Review*, 104(3): 963–90.

Howard, Greg, and Arianna Ornaghi. 2021. "Closing Time: The Local Equilibrium Effects of Prohibition." *The Journal of Economic History*, 81(3): 792–830.

Howard, Greg, Russell Weinstein, and Yuhao Yang. 2024. "Do universities improve local economic resilience?" *Review of Economics and Statistics*, 106(4): 1129–1145.

Jeong, Dahyeon, and Ajay Shenoy. 2022. "Can the Party in Power Systematically Win a Majority in Close Legislative Elections? Evidence from US State Assemblies." *The Journal of Politics*, 84(2): 1149–1164.

Kleinman, Benny, Ernest Liu, and Stephen J Redding. 2023. "Dynamic Spatial General Equilibrium." *Econometrica*, 91(2): 385–424.

Knight, Oliver. 1973. "Toward an Understanding of the Western Town." *The Western Historical Quarterly*, 4(1): 27–42.

Krugman, Paul. 1991. "History versus expectations." *The Quarterly Journal of Economics*, 106(2): 651–667.

Lin, Jeffrey, and Ferdinand Rauch. 2022. "What future for history dependence in spatial economics?" *Regional Science and Urban Economics*, 94: 103628.

Mawn, Geoffrey P. 1977. "Promoters, speculators, and the selection of the Phoenix townsite." *Arizona and the West*, 19(3): 207–224.

Michaels, Guy, and Ferdinand Rauch. 2018. "Resetting the urban network: 117–2012." *The Economic Journal*, 128(608): 378–412.

Moretti, Enrico. 2010. “Local multipliers.” *American Economic Review*, 100(2): 373–377.

Murphy, Kathryn. 2009. “County Government Structure: A State by State Report.” *A Publication of the Reseach Division of the National Association of County's County Services Department. Washington, DC March.*

Nagy, Dávid Krisztián. 2020. “Hinterlands, city formation and growth: Evidence from the U.S. westward expansion.” *Working Paper.*

Paher, Stanley W. 1969. “Significant County Seat Controversies in the State of Nevada.” PhD diss. University of Nevada, Reno.

Puga, Diego. 2010. “The Magnitude and Causes of Agglomeration Economies.” *Journal of Regional Science*, 50(1): 203–2019.

Qian, Franklin, and Rose Tan. 2021. “The Effects of High-skilled Firm Entry on Incumbent Residents.”

Redding, Stephen J, Daniel M Sturm, and Nikolaus Wolf. 2011. “History and industry location: evidence from German airports.” *Review of Economics and Statistics*, 93(3): 814–831.

Rosenthal, Stuart S., and William C. Strange. 2004. “Evidence on the nature and sources of agglomeration economies.” *Handbook of urban and regional economics*, 4: 2119—-2172.

Schellenberg, James A. 1987. *Conflict between communities: American county seat wars.* Pwpa Books.

Schmidt, Benjamin. 2018. “Creating Data: The Invention of Information in the nineteenth century American State.” *http://creatingdata.us.*

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

Smith, Cory. 2023. *Land Concentration and Long-Run Development in the Frontier United States.*

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.

Sylla, Richard Eugene, John B Legler, and John Wallis. 1995. *State and Local Government [United States]: Source and Uses of Funds, Census Statistics, Twentieth Century [Through 1982].* Inter-university Consortium for Political and Social Research.

Texas State Historical Association. 2016. "Handbook of Texas: Lamesa, TX."

When is Long-Run Agglomeration Possible?
Evidence from County-Seat Wars

by Cory Smith ⓡ Amrita Kulka

ONLINE APPENDIX

A Data Sources and Variable Construction

A.1 Election Margin of Victory

We compile an original database of county-seat elections from a range of sources including historical newspaper articles, individual county histories, and administrative data. In each case, we record the number of votes received by each location and the required vote percentage for victory, if applicable. We code elections into three categories: automatic runoff/required-majority elections, first past the post elections, and elections with preset super-majority requirements (usually required of a challenger to an incumbent). We code these electorally favored towns as "protected incumbents."[24]

In some cases, our sources only report the absolute vote margin of the winning town, for example "Location A wins over Location B by 112 votes." In cases where exactly two locations compete, we impute the number of total votes based on county population. We log-linearly interpolate county populations based on census decades and, for our sample of elections where total votes are reported, we calculate the median ratio of votes for the top two locations to population, approximately 23%. For elections where only the margin is recorded, we use this ratio to impute the total number of votes. These two pieces of information allow us to construct vote percentages. In the most common case of majority/first past the post[25] elections, we use the margin to mean the gap between winner and loser. In the above example, if we imputed 300 total votes, Town A would have 206 votes and Town B would have 94. In the case of elections with fixed vote percentages, we interpret the margin as the number of votes the largest vote-getter surpassed (or failed to surpass) the required margin. We base these interpretations on examples where we see different sources report both margins and exact vote tallies. In the case where the election In cases where three or

[24]In other parts of the paper, we use incumbent to mean the preexisting county seat. These two can differ: not all incumbents received different treatment in the election. A small number of non-incumbents were also favored when laws required supermajorities to choose a location, for example, farther from the county center. Because these were ex-ante restrictions imposed before votes were cast, they do not bear on the RD's continuity assumption.

[25]For the two-town case, these are mathematically equivalent

more locations competed, we consider the information incomplete and do not enter it into the sample.

In the case where 50% or another fixed percentage of the vote is required, we compute each town's vote percentage minus the applicable threshold as the vote percentage margin. For example, a town receiving 56% of the vote in a majority-required election would have a value of 6%. If it required a 60% supermajority to win, it would have a value of -4%. In a small number of cases, multiple locations ran against an incumbent and a specific vote percentage/supermajority was required; in the case that none of the challengers achieved this threshold, the incumbent would remain the seat. In this instance, we code the challengers' margins as before and the incumbent's margin as the negative of the top challenger's.[26] In the case of automatic runoff elections, we code each round as a separate election where applicable.

In first past the post elections, we initially compute the margin as the location's vote percentage of the location minus the runnerup's vote percentage (for the winner) or the location's vote percentage minus the winner's (for the losers). To achieve consistency with the automatic runoff systems, we then divide these margins by two.[27]

We finally top- and bottom-code margins at +50pp and -50pp respectively, largely for visual clarity of the RD graphs. These reflect the maximum and minimum possible margins for simple majority votes and only affects a small number of observations with other election types. None of these elections would be included within the RD bandwidth under either their new or restricted value.

A.2 Town Population and Location

Our primary source for town location and population is Schmidt (2018) which aggregates data on town population and location deriving from the US census. However, some of our locations do not appear in this data set, either because of an oversight in including census data or because the census never recognized the location as a community. In the former case, we supplement the main Schmidt (2018) data by merging it with data directly from the census including location, modern, and historical populations.

For locations never listed in the census as towns, we attempt to locate them based on historical maps and descriptions of the county. Because linkage failures could be differential

[26] For example, imagine A, B, and C receive 50%, 40%, and 10% of the vote. C is the incumbent and challengers require 60% of the vote. Then the vote margins would be -10%, -20%, and 10% respectively.

[27] Consider the case of a two-location race where majority-required and first past the post elections are equivalent. Say the locations receive 65% and 35% of the vote. Comparing the winner's vote percentage to the required 50% would produce a margin of 15%. Comparing the winner's percentage to the loser's would produce a margin of 30%. Dividing the latter value by two restores the equivalence.

based on election victory, we ensure that we locate every competing location within 15 points of election victory (in either direction).

For locations not recorded in a census decade, we code the town population as 0[28] and interpret this value as the "formal" town population. Because informal and unincorporated communities still exist, we thus supplement this measure with boundary-neutral definitions of population, for example census block population density within 1 mile of the town center.

A.3 Zip-level Income Data

For location income data, we turn to the IRS Zip-level aggregates of income tax returns. Because these derive from administrative data, they have several advantages over their survey-based alternatives like the American Community Survey (ACS). Many of the communities we study are comprised of fewer than 1000 people, meaning that the sample of respondents even in the 5-year ACS would be small. Because some locations we study also no longer exist as census places (CDPs), we would also require comparability between subcounty/minor civil divisions (MCDs) and CDP within the ACS. In general, research on the ACS shows that its accuracy falls for smaller areas, posing a significant issue for its use in our study (Albright 2011). In contrast, administrative sources have a claim to either minimal sampling variation (even a small area with 1000 people is large statistically) or none (because it reflects the universe of an appropriately defined set of people).

We link each location to its Zip code via a polygon file sourced from ESRI. For towns defined as census designated places (CDPs), we use a population-weighted average of all zip codes that intersect the CDP. Our method is equivalent to assuming a constant population density in each census block and then taking a population-weighted average based on the intersected area sizes. For other locations, we use the statistics for the zip code at the latitude and longitude of the location; recall that we geocode all locations within the RD bandwidth even if they no longer exist as towns today. By default, we consider the IRS definition of "adjusted gross income" (AGI) as our measure of income.

A.3.1 Zip-level Income Inequality

IRS data also have the advantage of providing AGI in predefined brackets: under \$25k, \$25-50k, \$50-75k, \$75-100k, \$100-200k, and over \$200k. This differentiation allows us to approximately compute top income shares to measure inequality.

We estimate top income shares via the following procedure by approximating income

[28]Note that, when taking logarithms, this values are <u>not</u> coded to missing as we typically use a $\log(\max(1, x))$ functional form; see Appendix Section A.13.

as constant within each of the six brackets (and equal to the reported average). We then report income shares among specified top percentages of filers. This method offers simplicity, though it likely understates inequality via the approximating assumption.

A.4 LODES

For data on job characteristics, we turn to the 2010 Census LODES data set. These data report information on 2-digit industry codes, approximate wage terciles, and education requirements of jobs at the census block level (with added noise).

For this project, we select the version of LODES based on worker residence as opposed to the job location. There are two primary reasons for this. First, city and town sizes are typically computed based on residence and not the workforce; for consistency, it makes sense to discuss residents' characteristics. Second, smaller communities may have workforces that work in other communities, meaning that their values could be missing — confusing the analysis. Appendix Section A.14.3 discusses how locations are linked to census blocks.

A.5 School District Spending

Data on school district boundaries and revenue/spending come from the National Center for Education Statistics (NCES) for the year 2017, the earliest publicly available for all our relevant variables.

A.6 Railroads and Other Transportation

Historical railroad data are from Atack (2016) and the "National Transportation Atlas Railroads" (modern, distributed by ESRI). Because the former database only covers years up until 1911 and the latter database only covers the modern period, we adopt the following approach when analyzing railroad construction in a panel setting. In years prior to the election or prior to 1911, we exclusively use the Atack (2016) database and code railroads as existing based on the "in operation by/InOpBy" field and the year of data. For years after 1911 but still prior to the relevant election, we consider only railroads existing in 1911. Although this definition might attenuate results by ignoring track built in the intervening years, we are reluctant to add 21st-century information; this would invalidate the use of this outcome as a pre-election placebo variable. Finally, for years following both 1911 and the election year, we linearly interpolate the existence of a railroad from the 1911 Atack (2016) file and the modern data, using 2010 as the endpoint for the latter.

For modern road data, we use the US Geological Survey (USGS) National Transportation Dataset (NTD) shapefile of major roads.

In the panel setting, we code the existence of railroads based on the distance between the location's coordinates and the nearest active section of track. We primarily consider the existence of railroads within one or two miles of the location's coordinates. For anaysis of modern characteristics, we code a location as having a road or railroad if either (a) it has a segment within 1 mile of its coordinates or (b) the census place (CDP) of which it is part intersects with any segment.

A.7 Post Offices

Geocoded data on US post offices, along with their dates of opening and closing, are available from Blevins (2015, 2021). In our analysis, we typically only count "active" post offices — ignoring any that closed prior to the specified year.

A.8 County Government

Aggregate employment in county government comes from Census (2012). We use these to produce aggregate statistics of county government jobs in particular roles (for example administrative/judicial) for 2012. When comparing to 2010 total jobs, we deflate the figures by the average national employment change from 2010 to 2012.[29]

A.9 Geographic Controls

Raw sources for geographic data are listed here: elevation data are from the SRTM data set, stream shapefiles are via ESRI, terrain slopes from the 2016 LANDFIRE data, and soil quality from the USDA gSSURGO data set. For soil quality, we use versions of the National Commodity Crop Productivity Index (NCCPI) computation to measure soil productivity and the frequency of flooding variables. The latter are broken into discrete categories based on flooding frequency.

For elevation, terrain slopes, flooding frequency variables (by category), and soil quality we compute area-weighted averages within 1 mile of the town center. For the NCCPI soil productivity measures, we code missing values as 0, reflecting the fact that these are usually water bodies or areas unsuitable for farming.[30] For streams, we compute the (log) distance

[29] Aggregated changes in county government employment would have been more ideal but are not available.

[30] Because we consider both the soil properties and the presence of soil types unsuitable for their calculations to be natural features unrelated to election characteristics, the imputation here should not create bias. As with any geographic characteristic in our sample, if the economic development shock caused by our election

to the nearest type of stream. For this purpose, we consider only natural bodies, ignoring human-made channels such as canals.

A.10 PLSS and Settlement Data

For other preperiod election controls we turn to locations' section (approximate square mile) grid square on the Public Lands Survey System (PLSS) and consider pre-election settlement characteristics as reported by the Bureau of Land Management (BLM)'s General Land Office (GLO) records.

From the PLSS itself, we consider the area of the section, whether it functioned as an education section, and whether the section number was even or odd. The latter two pieces of information reflect government land settlement policies of education and railroad land grants; see (Smith 2023) for historical details and relevance. From the GLO, we compute several variables reflecting government-administered settlement up to the year before the election (i.e. preperiod): the log total number of government land distributions ("patents"), the fraction of the section settled under the Homestead Act, and the fraction of the section settled according to "lots" rather than direction-based subsections (for example "quarter sections" of approximately 160 acres). This last variable is useful because agricultural land was typically divided on the basis of gridded subdivisions of sections (for example half, quarter, sixteenth... sections), land for urban areas was typically allotted based on more idiosyncratic "lots."

For areas outside the PLSS system (primarily Texas and New England), we code all these variables to 0.[31]

A.11 Newspaper Presence

The Library of Congress's *Chronicling America* directory includes approximately 157,000 newspapers along with their state and town and years active. We link this list of newspapers to our town sample and define a town as having a newspaper in a year if it has any linked paper active. Note that the directory is a larger sample than the searchable list of digitized newspapers.

caused environmental changes or affected the observability of the characteristic, including it as a control would lead to the bias via inclusion of a "downstream" variable. However, Section 4.3 finds no empirical evidence of imbalance on this and other characteristics.

[31] Because both pre-election settlement characteristics and the location with respect to the PLSS should be irrelevant for winning close elections, this imputation should not create any bias. If locating within or outside the PLSS was connected with election manipulation, then in theory this control could exacerbate bias. However, this is extremely unlikely as almost all counties would have been entirely within or outside of the PLSS; Section 4.3 also finds no empirical evidence of imbalance on this and other characteristics.

A.12 List of Geographic Controls (Town-Level RD)

For the town-level implementation of RD equation (1), we include the following geographic controls: mean elevation, log distance to the (pre-election) county center, average terrain slopes ("ruggedness"), the presence of a (pre-election) post office, log (pre-election) town census population, "protected incumbent" status (a town competitors require a supermajority to beat, usually due to incumbency), miles to the nearest stream, average soil quality ("nccpi3all") and land flooding ("flodfq" categories) in the USDA gSSURGO database, and the presence of a (pre-election) railroad within 1 or 2 miles, and the (pre-election) presence of a newspaper. For locations surveyed in the public lands survey system (PLSS), we include the PLSS section area, even/odd status, status as an education section; see Smith (2023) for a discussion of their relevance. We include the log number of (pre-election) GLO settlement deeds ("patents") and the fraction of these given out under the Homestead Act (or 0% in the case of no patents). We include a binary for being within the PLSS grid, to obviate the importance of the value chosen when the settlement variables cannot be defined. Finally, we include variables defined in the pre-election census data: fraction of workers in agriculture, fraction racially white, fraction female, fraction of people living on farms, fraction of workers in blue collar occupations, average occupational income score. Where unavailable, we use the previous sets of control variables to predict them in an OLS regression.

For terrain and land quality data, we define the value as the average within one mile of the town center. Pre-election is defined as the earliest year available before the election year; in most cases, this is simply the previous year. For census data, it refers to the previous census.

A.13 Logarithms

When presented with fat-tailed variables that have zero in their support, researchers often search for a transformation that roughly matches the properties of the logarithm but gives nonmissing values for zero. By default, we adopt the functional form $\ln(\max(1, x))$ in these cases. This matches the Chen and Roth (2023) suggestion of manually specification and it places as little weight as possible on the 0/1 transition. We view this as appropriate for our variables. For example, a town with exactly 1 resident likely differs little from an area without a town. In most cases, the latter scenario corresponds to an area of sparse but nonzero population.

A.14 Linking

We herein discuss our procedures for linking our locations to different data sources. As detailed in Section 3.3, we have both the name and location centroid of all towns within the bandwidth to use in this process.

A.14.1 Census Microdata 1850-1940

Our linkage here builds off the Census Place Project (CPP) (Berkes, Karger and Nencka 2023) which we supplement with manual adjustments in cases of unclear links. For a large majority of locations, we find a "Census Place" in the above data that has a similar name (per a fuzzy string match) or is very close to the location. The town's characteristics then become those of the people recorded as living in the single or multiple enumeration districts.

In cases where no Census Place matches on either criterion, we adopt a procedure of manual correction. Using the original enumeration district maps in the National Archives,[32] we manually locate the coordinates of the town and link it to the listed district. In cases where these maps are unavailable, as a last resort we average across all enumeration districts that are not part of the CPP. This reflects a decision that in these cases the location would not have formed a sufficiently large town to be classified as a "place" in the CPP.

Because it is uncommon for our locations to exist more than one census year prior to the election, we restrict the final sample to one preperiod of data and subsequently all census years up to 1940.

A.14.2 Census Microdata for 1890

The absence of census microdata for 1890 impose several challenges for our linking exercise. We herein detail our procedures around this year, though we note these choices only affect a small part of the overall sample. Throughout, we maintain the validity of the RD design in pre-election census years. That is, we do not average or otherwise mix pretreatment and post-treatment years.

First, for election years 1891-1900, 1890 is a pretreatment year. We here use 1880 values in place of 1890 values — an earlier preperiod for all variables. For elections 1881-1890, 1890 is the first post-treatment year and we use 1900 values instead — a later post-period. When the election year is 1880 or earlier, all of 1880, 1890, and 1900 are post-treatment years. Here, we linearly interpolate all variables from 1880 to 1900 (i.e., average 1880 and 1900) for the 1890 value. Finally, there are no choices to make for election years after 1900: 1900 serves as the preperiod census and nowhere do we use even earlier preperiod data.

[32] We extend our thanks to Risa Wagner and Jake Burgess for excellent research assistance in this endeavor.

A.14.3 2010 Census Blocks

For the 2010 LODES data, we primarily link locations to census places (CDPs) on a block level using our linkage between our locations and census blocks. We use the residential version of the LODES data set. For the minority of cases where no CDP exists, we use the workplace locations within 1 mile of the central location.

When linking locations to census blocks based on distance (for LODES, population, or other data), we assume a constant distribution of people (or jobs) within the area of each block. We further assume other characteristics (for example ethnicity, job wages...) are constant. Under these assumptions, we intersect a circle of the specified radius with all census blocks and compute the relevant totals (for example total population) and weighted averages from all intersected census blocks, based on the area of intersection.

A.14.4 2010 Zip Codes

We link each location to the 2010 Zip code in which it lies. For locations that are census-designated places (CDPs) in 2010, we use the boundaries set by the CDP polygons. When these intersect with more than one zip code, we use the population-weighted average and calculate population based on census blocks, using a constant density within each one.

A.14.5 CUSP Patent Data to Census Microdata

In this section, we describe our procedure for linking the CUSP patent data to the census Berkes (2018). Our broad approach is to use a fuzzy string match on the first and last name of each inventor to census microdata within each county. We record the year of a patent as the year of issuance and link only within the county recorded by CUSP to the previous census within 10 years. The result is we code, in each year, the fraction of the population that is an "inventor": that has had a patent issued in a particular year. We link inventors to the most recent available census with microdata (i.e., 1850-1940 with 1890 as unavailable).

Several choices are notable here. First, we do not use the "city" information present in the CUSP data. Based on discussions with the authors and our own read of the data, "city" location may in some cases record a mailing address rather than a residential address. Given that not all small locales had post offices, treating the postal address as the true address would result in a mechanically positive effect in favor of the winners – we aimed to avoid inflating our results in this manner. Our read of the data is also that the issuance data of the patent is most complete relative to other years in the data, particularly for the earliest years. Empirically, it produces a higher match rate within the described counties relative to other listed years in the CUSP data.

A.15 Industry Definitions and Classifications

This section describes how we classify industries into "blue collar," "white collar," and farming.

A.15.1 Census Microdata 1850-1940

These definitions are based on the 1950 occupation codes. Farmers constitute codes 100-199 (farm owners, tenants, and managers) and 800-899 (farm laborers and foremen). "Blue collar" constitutes codes 500-699 (craftsmen and operatives) and 900-971 (nonfarm laborers). "White collar" constitutes codes 0-99 (professional and technical services) and 200-499 (managers, clerical, and sales).

We also consider several other specific industries and occupations in these data. Lawyers and judges are defined by occupation code 55. Local government covers industry code 936. Note that that county government specifically is a subset of this code which also would include municipal government positions. Non-local public administration covers codes 906, 916, 926, 946.

A.15.2 LODES Data 2010

These definitions are based on 2-digit NAICS industry codes that are further aggregated by the LODES data set; we do not observe occupation in these data. We define "blue collar" as codes 21-33 (mining, utilities, construction, manufacturing). "White collar" constitutes codes 42-92 (largely administrative, clerical, and service sectors). Informational/tech jobs refers to NAICS code 51 specifically.

B Robustness and Other Results

In this section, we demonstrate robustness to a range of sample definitions and coding decisions. Although Section 4.3 showed no detectable imbalances in geographic or pre-election characteristics, other arbitrary choices in specification could theoretically drive our results, diminishing their meaning.

B.1 Qualitative Discussion of Electoral Fraud and Manipulation

The statistical balance shown above makes it unlikely that fraud or electoral manipulation substantially affects our analyses. Violating the RD design's continuity assumption requires that one party has the ability to "precisely control" the vote share around the win/loss cutoff

(Jeong and Shenoy 2022; Das 2023), either through fraud or exact knowledge of vote shares. For example, a ruling authority could add just enough votes to ensure a victory for their preferred choice. However, our balance checks show that politically protected incumbent towns are slightly less likely to win (not statistically significant).

Qualitative evidence is consistent with the idea that electoral manipulation was unlikely to affect the RD results. Precise knowledge of voting intentions would have been difficult given the period's technology and changing populations. While losing towns sometimes alleged fraud, these complaints typically referred to "unfair" electoral tactics or disputes over voter registration, neither of which is a method of precise control (Schellenberg 1987).[33]

B.2 Bandwidth Selection

Appendix Figure C.11 shows that bandwidth choice does not fundamentally change the results of two of our main variables: population density and (log) income. For the first outcome, decreasing the bandwidth from 20[34] to 5 modestly raises the effect size while leaving all results highly significant. For the second, decreasing the bandwidth again largely leaves the point estimate unchanged, although with the smaller sample size at lower bandwidths, the estimates lose power and do not achieve statistical significance.

B.3 Sample, Optimal Bandwidth, and Controls

Our sample consists of the baseline ("first") elections with at least two competitors.[35] Within these, we focus on the top two competing locations by vote margin.

As with the town-level results in Section 4, we calculate the county-level optimal bandwidth using the no-controls version on county density.

For county-level controls, we adopt equivalents of the town-level controls listed in Appendix Section A.12. Most importantly, we can no longer include county fixed effects as they would essentially subsume the whole sample. In their place, we substitute a linear

[33]For example, this cited work relates a story in which one town's voters were unable to go to the polling place "because they had found a barrel of whiskey that day . . . and many were too drunk" to make the journey. This sort of tactic or allowing unregistered voters to submit ballots could be undemocratic and shift votes, but it could not do so in a discontinuous manner. By analogy, modern candidates may increase their vote shares through campaigns and advertisements without affecting the validity of an RD. Even if candidates violate election regulations in doing so, vote differences will not show up exactly at the threshold of victory.

[34]Note that a small number of communities are not geolocated outside of the 15-point bandwidth, meaning that their outcomes are generally missing. Since this missingness may be endogenous to the election result, we prefer bandwidths within 15 points but include up to 20 points for completeness.

[35]A very small number of elections in our sample were uncontested, meaning that the winning town has a victory margin of 50 percentage points (100% - 50%). While we technically include these in our sample, the margin prohibits them from entering RD estimates of equation (1).

function of county latitude and longitude by state, i.e., state fixed effects and county latitude/longitude interacted with state. For the other controls, we include county-level versions of the town-level variables.

B.4 Bootstrapping and GRF Estimation

To determine the statistical significance of the GRF verification in Section 7, we bootstrap the primary estimates presented in Table 3, columns (3), (6). While we are unaware of theoretical work characterizing the bootstrap's properties for GRF models, to the best of our knowledge this represents the best method for calculating statistical uncertainty given the 2-step procedure in the aforementioned table.

We bootstrap the estimation in the following manner. For each trial, we construct a new sample by bootstrapping in blocks at the county level. We re-run the GRF prediction algorithm, considering each selected county as unique[36] and re-run the RD estimates of equation (1) based on those predictions. In both cases, the bandwidth does not change from the main estimation.[37] As a placebo test, we additionally calculate the RD estimates for randomly selected "ideal" options.

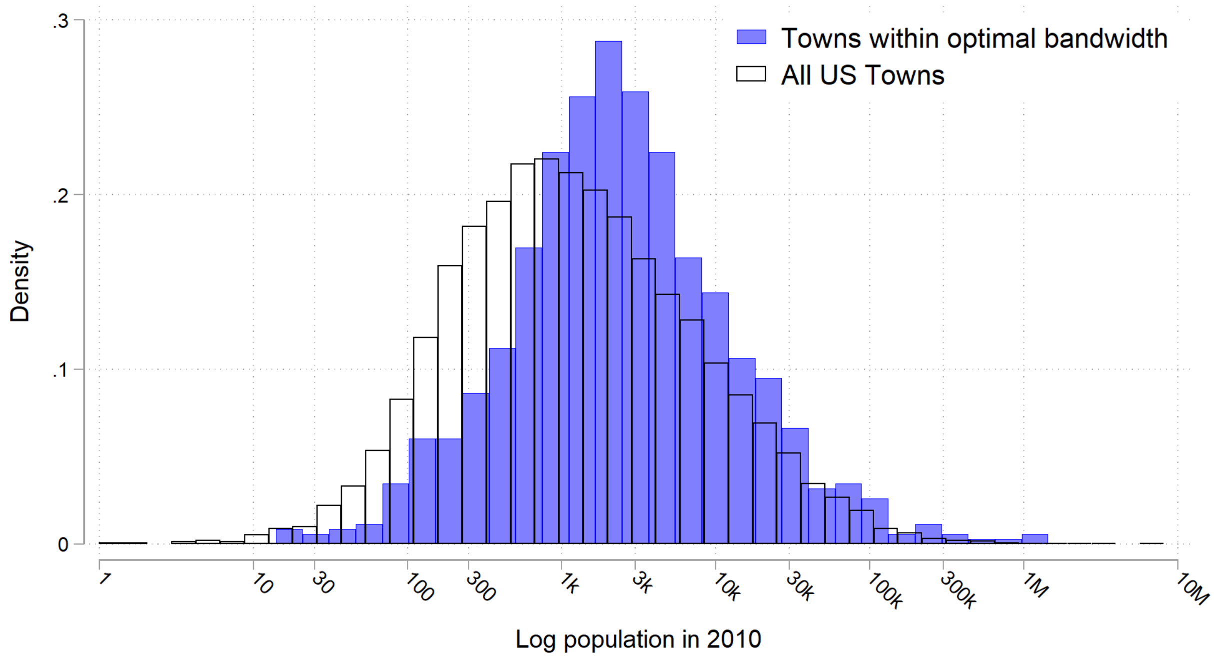
The results from 500 rounds of this procedure are graphed as histograms in Figure C.12. In recording statistical significance, we consider the fraction of estimates that fall below 0. Because we select the seat that the GRF model predicts will maximize density (and presumably income), we treat positive county-level RD estimates as verifications and negative ones as failures. By this metric, we obtain p=.012 for density and p=.076 for income. Unsurprisingly, about half of the placebo estimates are below 0: p=.47 and p=.52 for density and income respectively.

[36]That is, if a single actual county is drawn twice, each selection is treated as though it were a separate county, albeit with the same characteristics.

[37]In a similar way, the standard errors of an RD typically do not change for a given bandwidth regardless of whether that bandwidth is manually or calculated as "optimal."

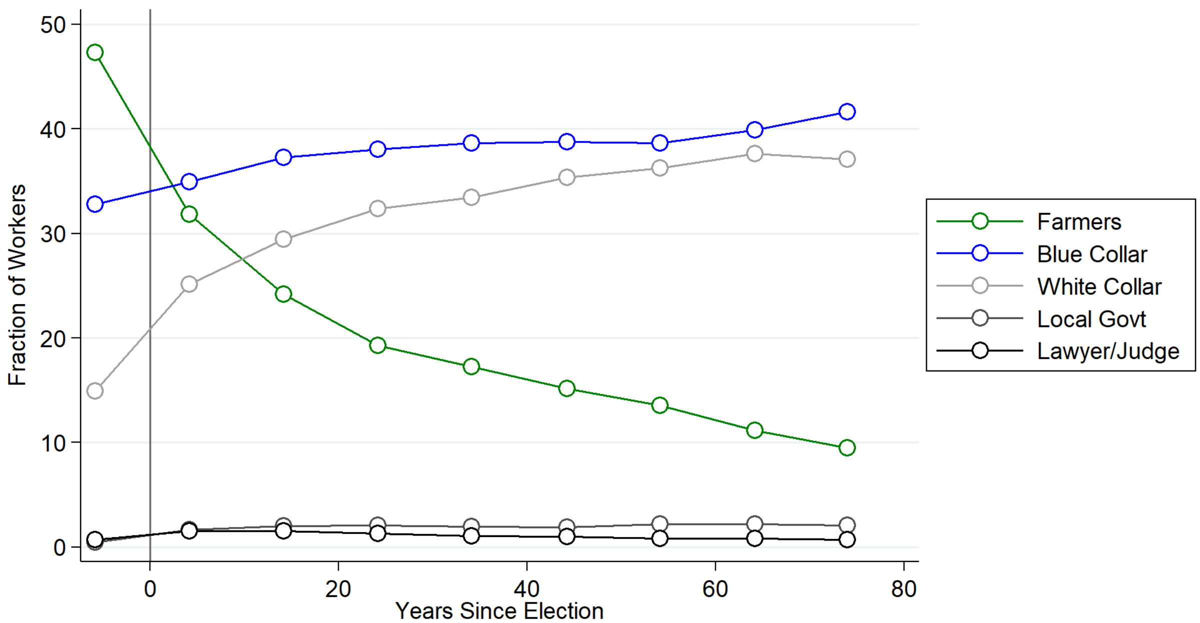
C Appendix Tables and Figures

Figure C.1: Modern Town Populations, Close Elections



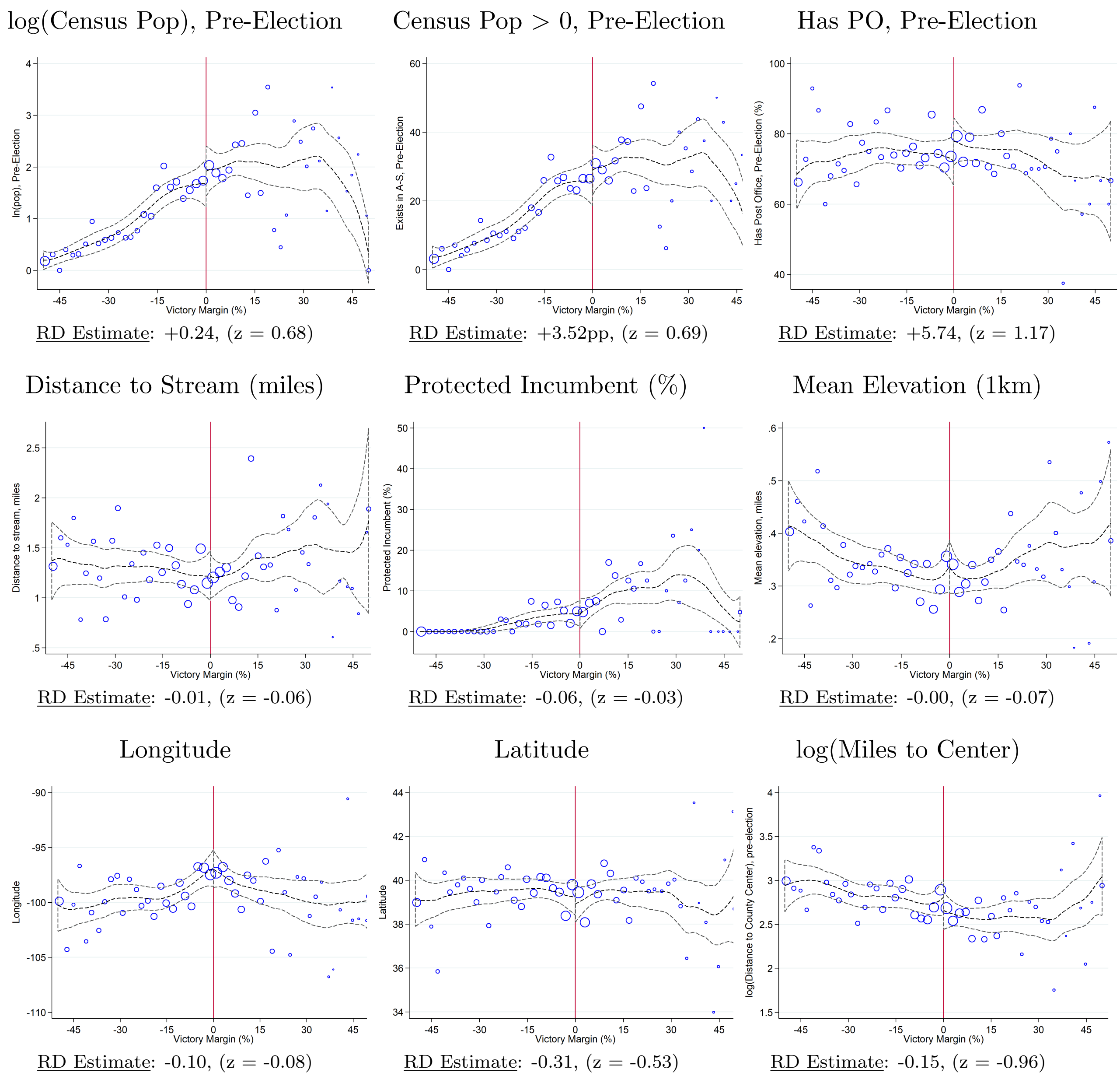
Notes: Distribution of population of census CDPs (a) for all US towns (b) our sample towns, conditional on being within the bandwidth for victory in a county-seat election.

Figure C.2: Industry Mix in Census Microdata: Close Winners



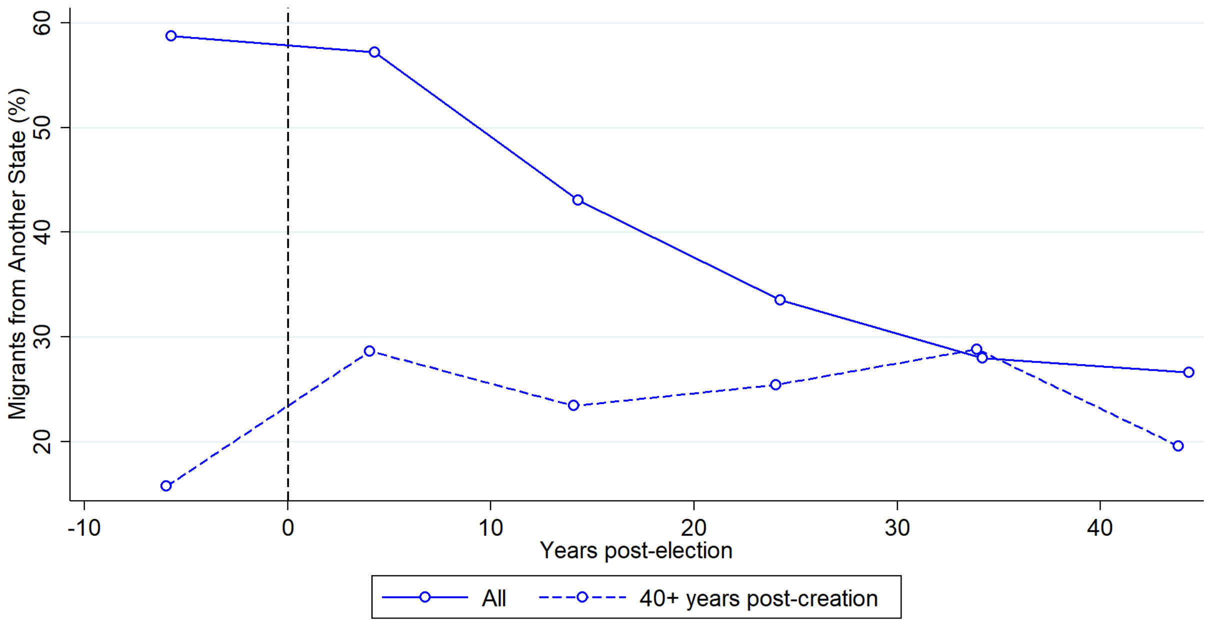
Notes: Average industry statistics in census microdata in years relative to the county-seat election. The sample consists of “close winners”: winning locations within the main RD bandwidth.

Figure C.3: RD Balance Tests — Multiple Outcomes



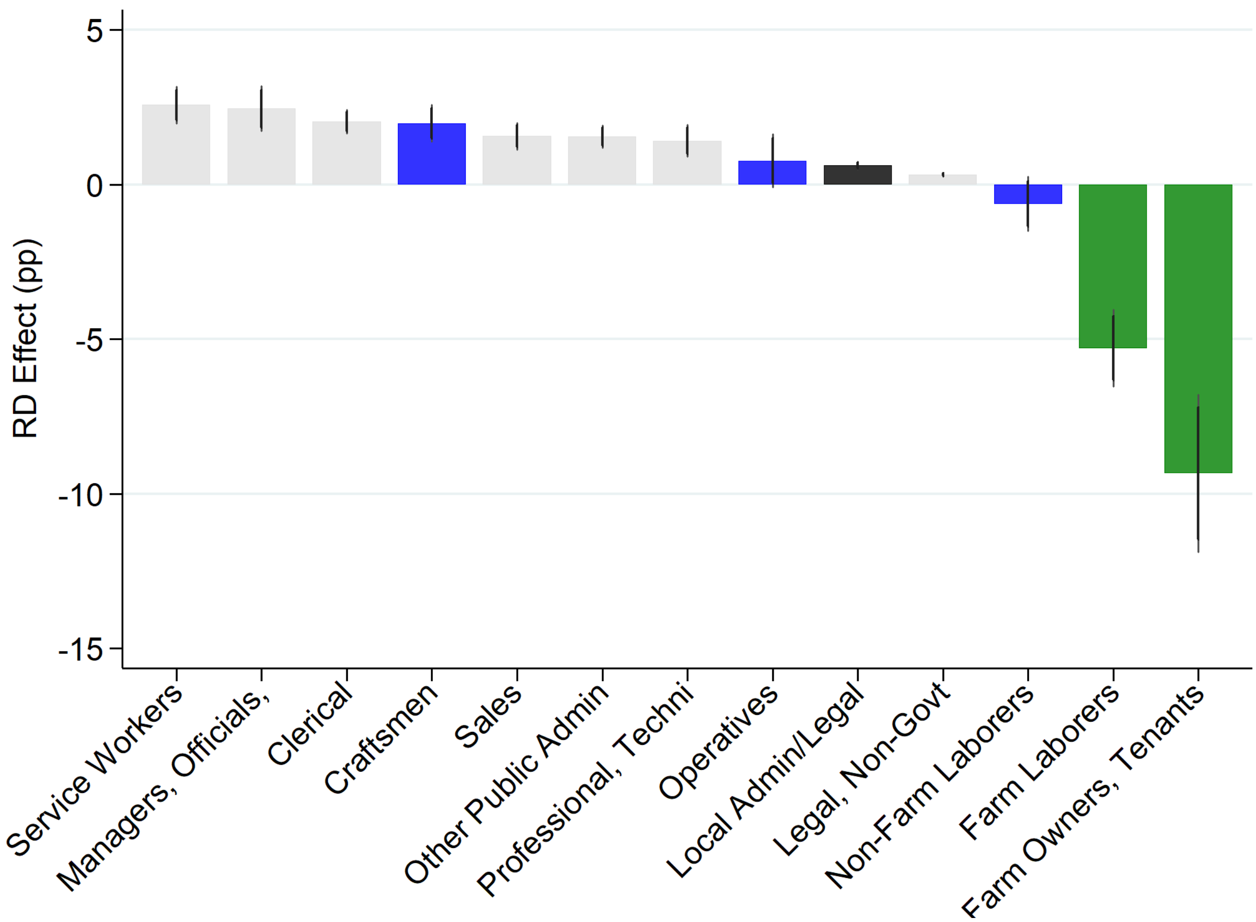
Notes: Binned scatter plot and local polynomial fit for multiple preperiod or geographic characteristics. RD estimates and z-scores are presented below as text.

Figure C.4: Percent of Migrants from Other States



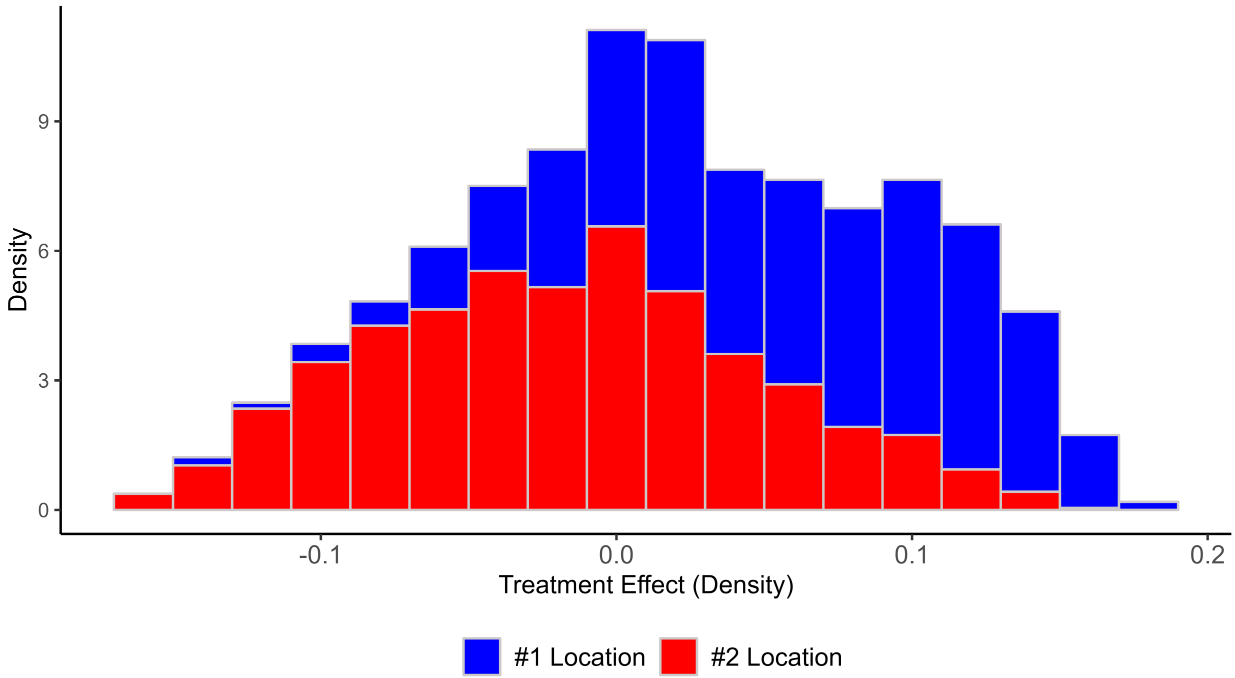
Notes: This charts the fraction of census-linked individuals who were located in another state prior to arriving in a location in our RD sample (close winners and losers in the main election set). The fraction is plotted separately for all elections and those elections occurring forty or more years after the county's creation.

Figure C.5: Occupational Change in 1940



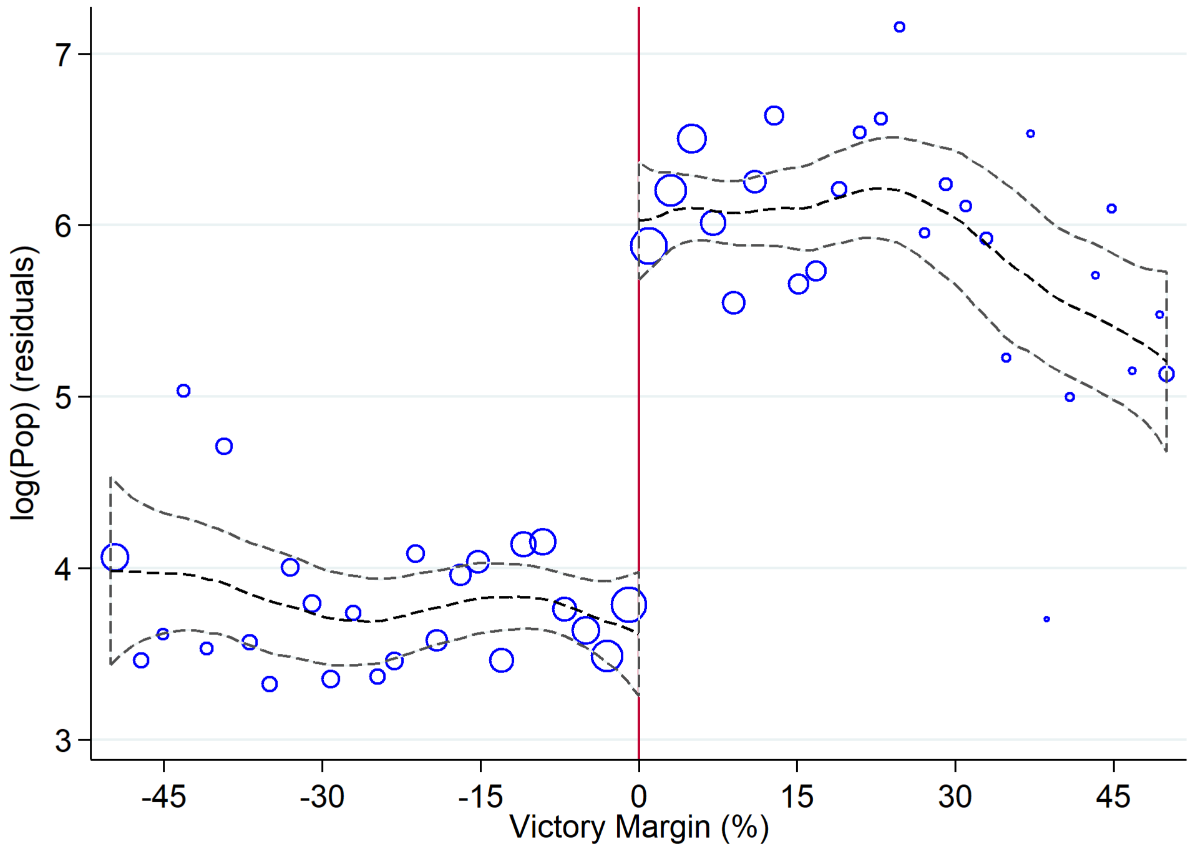
Notes: RD estimates of equation (1) for 1940 job categories for winning towns' enumeration districts compared to losing towns' within optimal bandwidth. Each bar represents a regression for a particular category. Local government and other public administration jobs are defined based on industry. The remaining categories exclude government workers and are coded based on occupation. Colors indicate "white collar" (gray), "blue collar" (blue), agricultural (green), or local government (black) occupations. For occupation definitions see Section A.15. Geographic controls as listed in Section A.12.

Figure C.6: GRF Estimated Treatment Effects: County Density



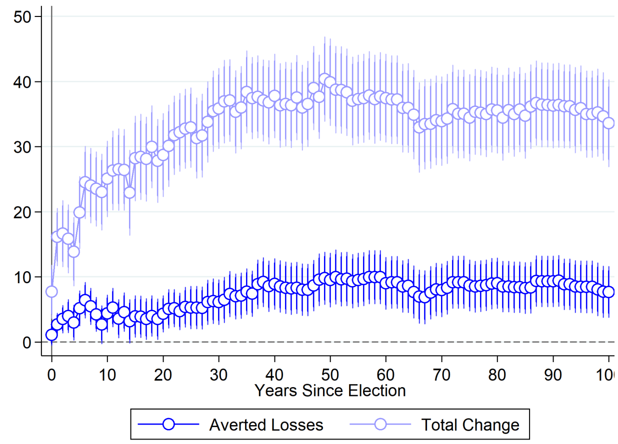
Notes: Histogram of estimated treatment effects on 2010 (log) county density from selecting individual locations as the county seat; see Section 7.1. Within an election, the figure breaks out the location with the highest and lowest effect which are highlighted in separate colors.

Figure C.7: Effects on 2010 Town Census Population

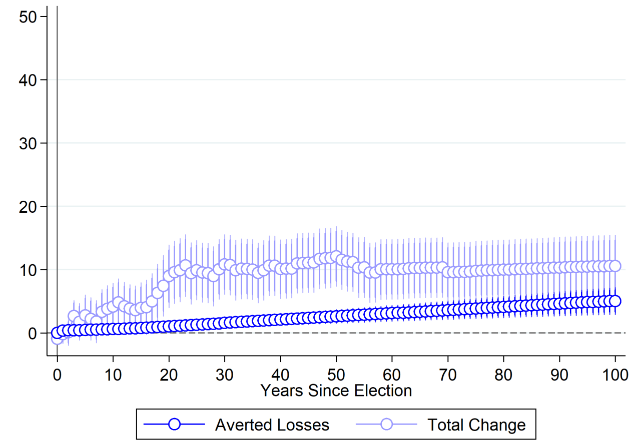


Notes: binned scatter plot and local polynomial fit for the log(Town Census Population) in 2010. RD estimates and z-scores are presented below as text. "Log" refers to $\log(\max(1, x))$. Geographic controls are listed in Section A.12. RD Estimate: +2.61, (z = 11.23)***

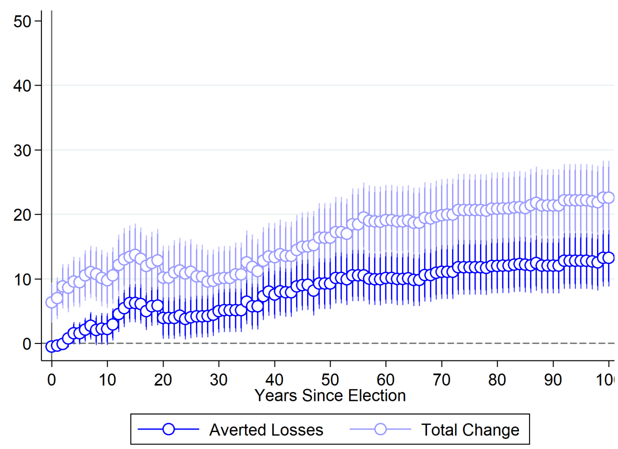
Figure C.8: Gains and Losses of Amenities



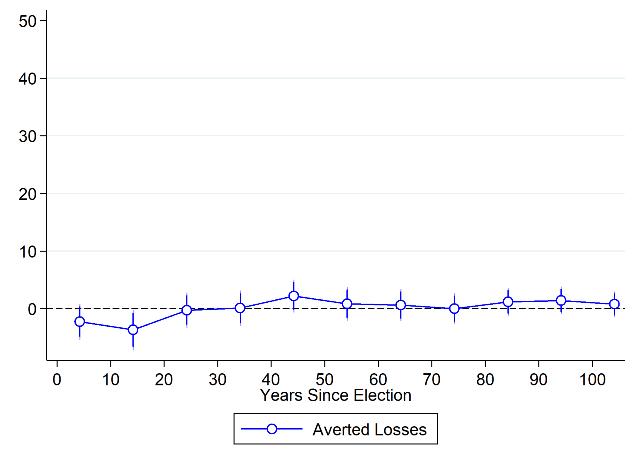
(a) Town Newspaper



(b) Railroad Connection



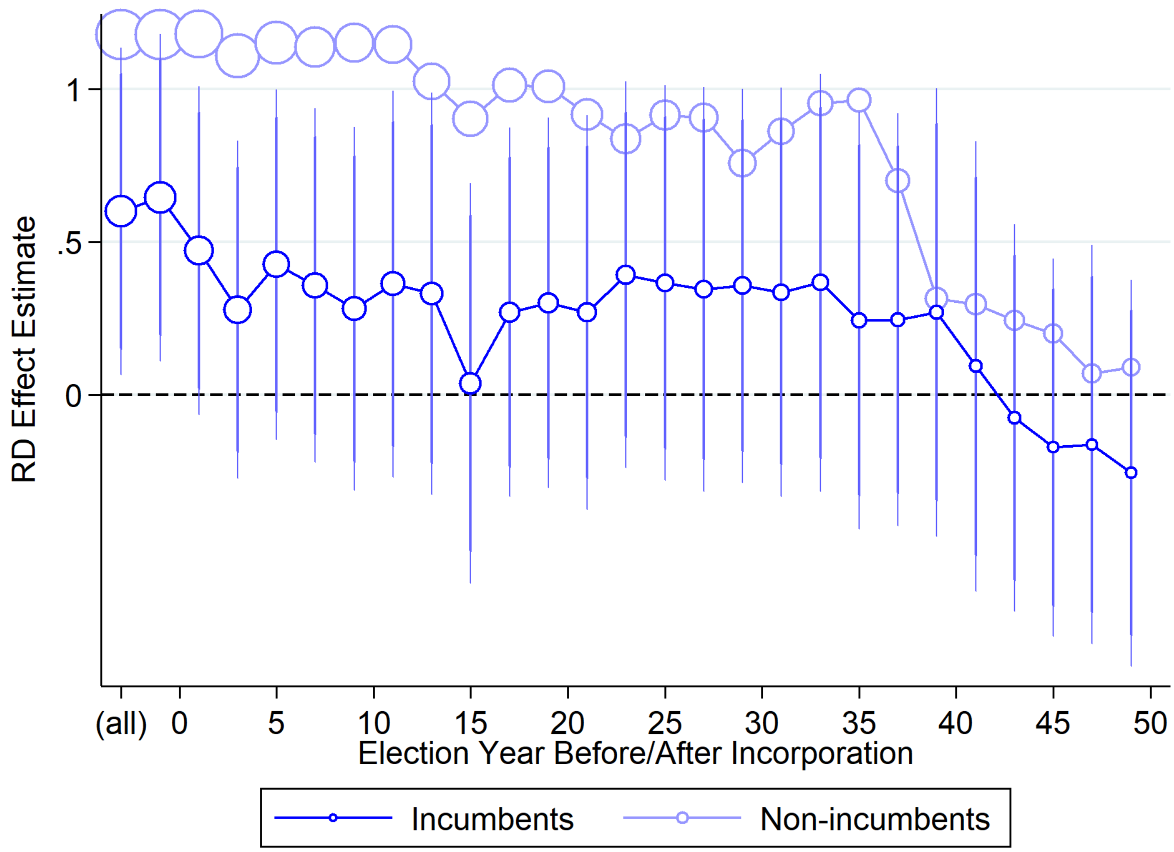
(c) Post Office



(d) Census Population

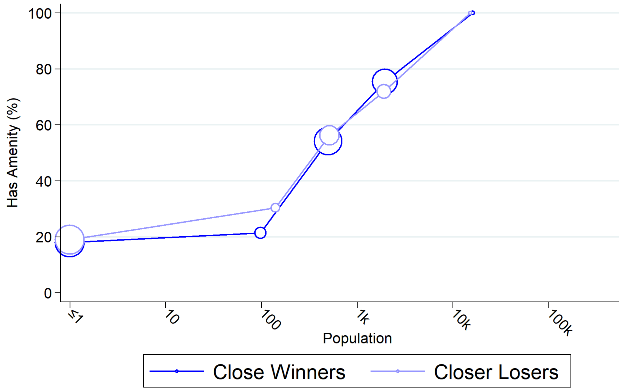
Notes: Effects over time on the total change versus averted loss of a variable. Each point summarizes an RD analysis of equation (1) a specified number of years after the election. y-axes across all panels have been standardized to be consistent. Regressions include county fixed effects and controls listed in Section A.12. "Total change" is the effect on the outcome itself. "Averted Losses" is defined as -1 if a location's outcome is strictly lower in the specified year relative to the year prior to the election and 0 otherwise. It thus measures the contribution of locations that fell below their pre-election values, with the negative sign ensuring that it measures averted losses rather than losses. (a)-(c) study the binary presence of an amenity. (d) compares census population to the most recent pre-election census value.

Figure C.9: Density Effect Size and Timing: Incumbent Towns

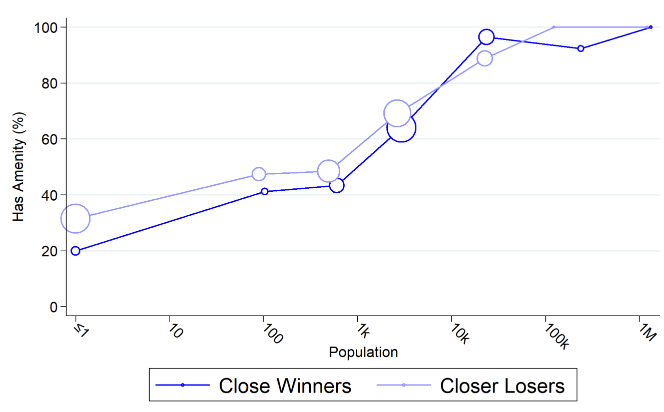


Notes: This figure replicates the analyses of Figure 11a. It splits the sample by towns listed in our archival sources as incumbents and the remaining sample ("non-incumbents"). Standard errors on the latter are removed for visual clarity. Because the incumbents group has, at most, one observation per election, we necessarily remove county fixed effects from the control set as their inclusion would subsume almost the full sample. As a historical note, we code all seats as incumbents regardless of origin. For example, county commissioners in some cases named a "temporary" seat prior to the election. We would code this location as an "incumbent" even if it never won an election in our data.

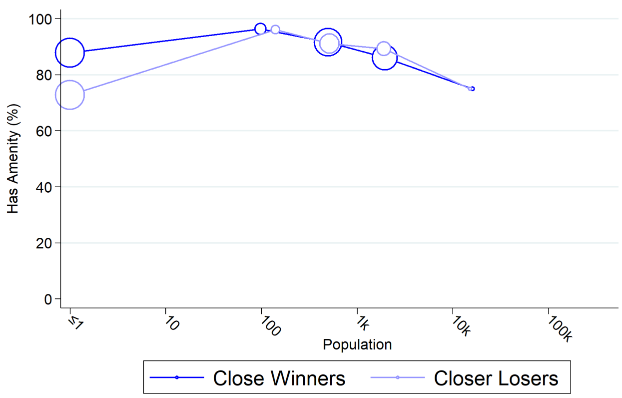
Figure C.10: Public Goods versus Population, Election Winners and Losers



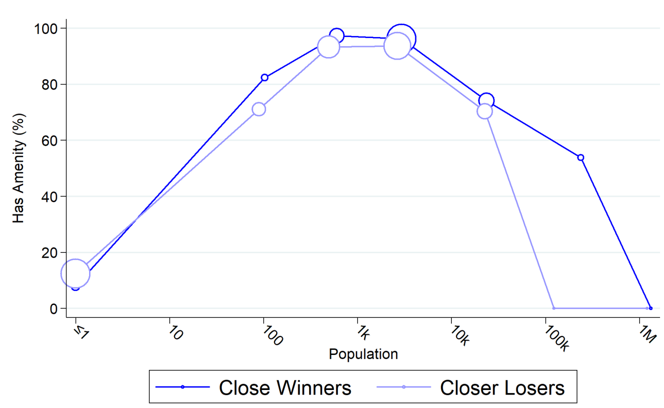
(a) Railroad (next census)



(b) Railroad (2010)



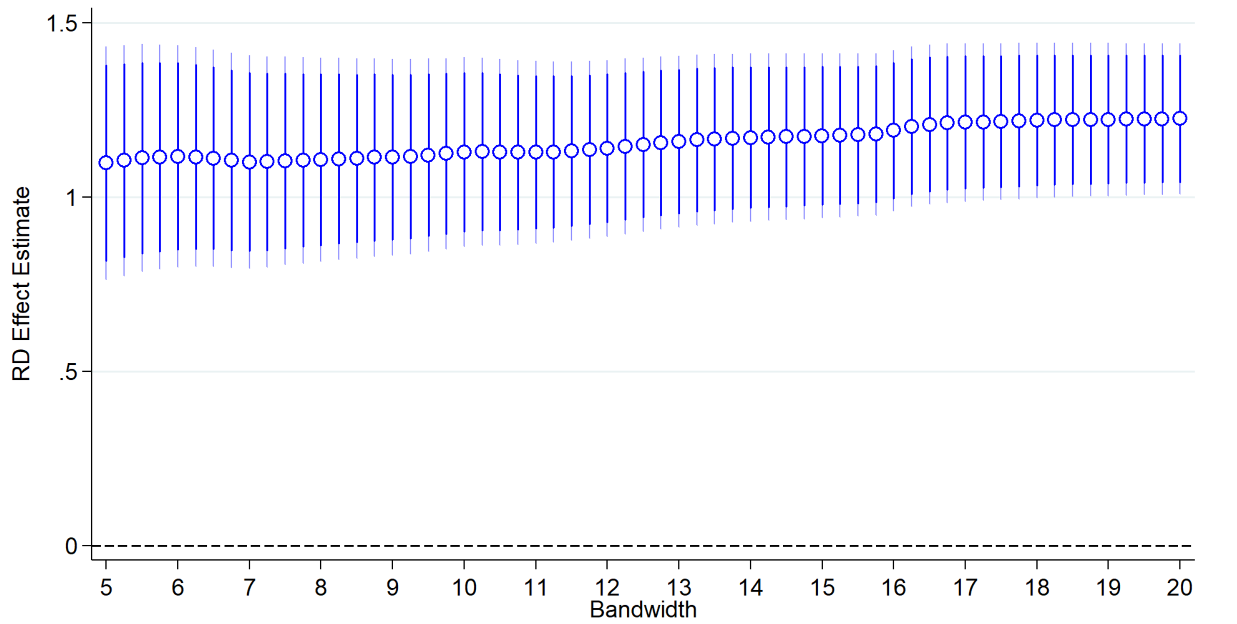
(c) Post Office (next census)



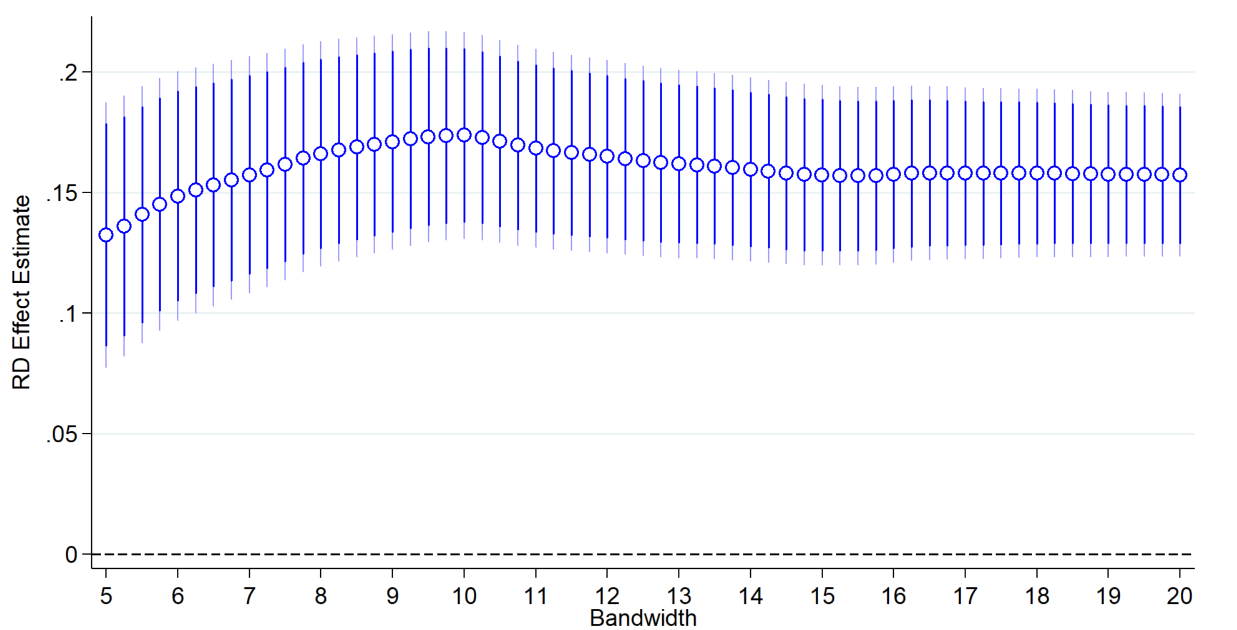
(d) Post Office (2010)

Notes: Relationship between amenities and census population for close (within the default RD bandwidth) election winners and losers. The left column uses data from the first post-election census (0-9 years after), the right column uses 2010 data. The outcomes considered are respectively the presence of a rail line and the presence of a post office.

Figure C.11: Bandwidth Selection



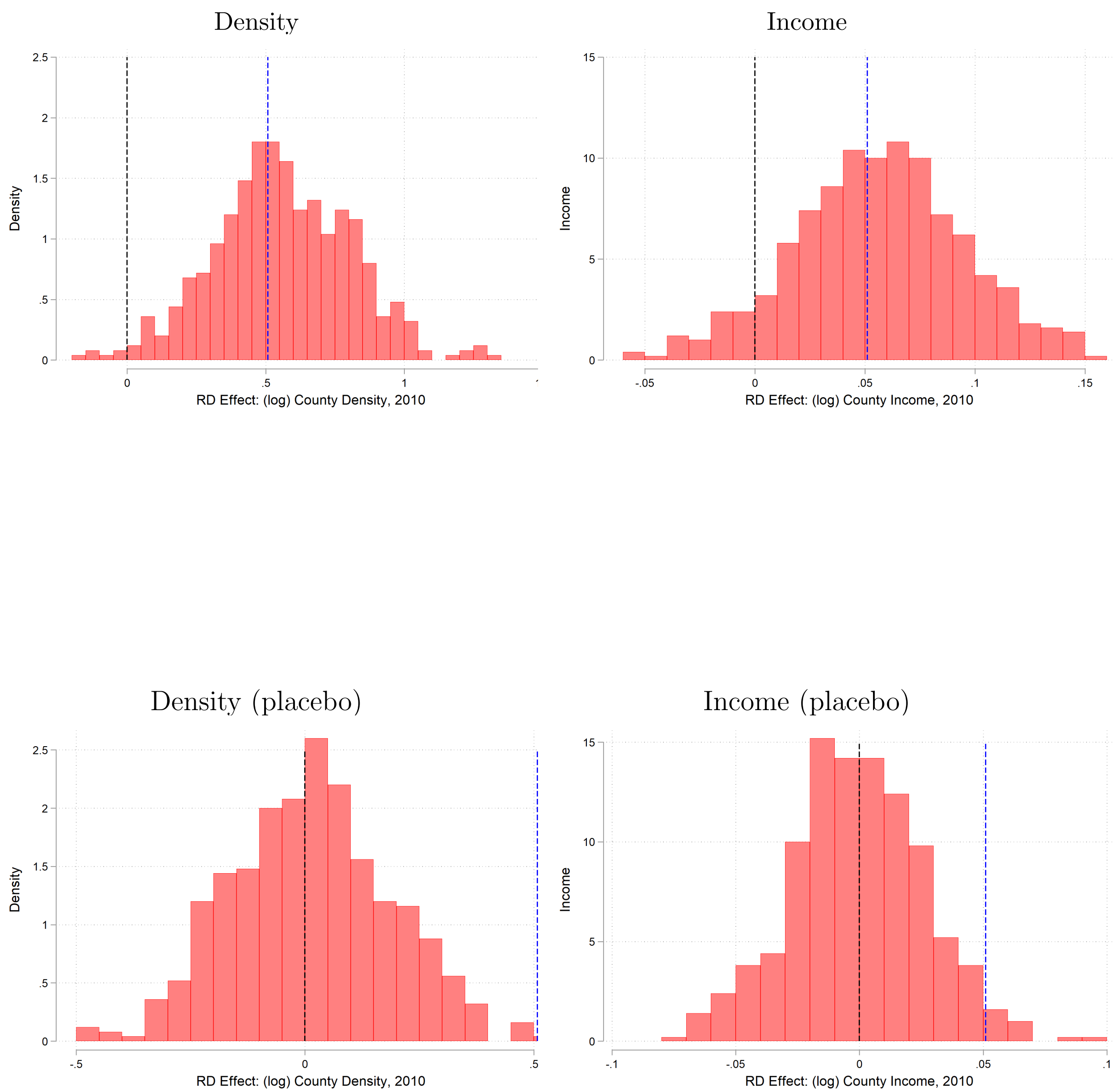
(a) (log) Population Density within 1 mile



(b) (log) 1940 wages, private, nonlegal

Notes: RD estimates on log population density of census blocks within 1 mile of town centers (panel (a)) and 1940 wages excluding government and legal workers (panel (b)). Each point represents a regression at the specified bandwidth. Confidence intervals of 90% (thick line) and 95% (thin line) shown over each estimate. Controls include the geographic characteristics listed in Section 4 and county fixed effects.

Figure C.12: Bootstrapped County-level RD Estimates



Notes: histograms of bootstrapped GRF estimations as described in Appendix Section B.4 and Section 7. Blue vertical lines show the main, full-sample estimate and black vertical lines are marked at 0.

Table C.1: Location-Level Summary Statistics

	N	Mean	SD	p(10)	p(25)	p(50)	p(75)	p(90)
Modern Seat (%)	2,642	38.6	48.7	0	0	0	100	100
Town Pop 2010	2,642	9679.9	57340.7	0	0	927	3560	13542
Pop/mi^2 < 1 mile	2,443	924.3	1475.2	7.51	70.7	416.1	1120.0	2462.3
Election Year	2,642	1884.6	23.5	1856	1870	1884	1903	1914
Election: Years Post Creation	2,642	19.0	22.1	0	2	11	30	49
Election Number	2,642	1.34	0.75	1	1	1	1	2
Vote Margin (%)	2,642	-7.88	21.9	-42.2	-21.8	-4.08	4.94	16.5
Longitude	2,443	-98.7	10.5	-117.0	-102.6	-97.2	-92.7	-85.5
Latitude	2,443	39.6	5.15	31.7	36.4	40.1	43.7	45.9
Zip Income ($)	2,334	48196.5	19879.6	36322.0	39993.6	45216.3	51209.5	59481.6
Top 5% Share	2,334	21.1	6.39	14.4	17.5	20.5	24.3	27.8
White Collar, 1940 (%)	2,448	32.4	14.7	8.39	20.9	36.6	43.8	48.7
White Collar, 2010 (%)	2,259	76.6	18.5	64.6	74.1	80.1	85.6	90.8
Miles to County Center	2,443	35.4	84.9	2.74	5.61	12.0	29.3	64.8

Notes: Summary statistics on location-level data. Shown respectively are the nonmissing count, the arithmetic mean, standard deviation, and 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentile values.

Table C.2: Effects on Income and Proxies

	1940				2010		
	(1) log(Wages) Above Avg.	(2) log(Home Value)	(3) Occscore	(4) White Collar (%)	(5) Top 5% Share	(6) Top $\frac{1}{3}$ Wage	(7) White Collar (%)
Win	0.077***	0.24***	2.48***	9.71***	2.88***	4.14***	4.73***
	(0.013)	(0.036)	(0.29)	(1.00)	(0.43)	(0.61)	(1.14)
County FEs	Y	Y	Y	Y	Y	Y	Y
Geo	Y	Y	Y	Y	Y	Y	Y
SEs / Clusters	Election	Election	Election	Election	Election	Election	Election
BW (pp)	14.5	14.5	14.5	14.5	14.5	14.5	14.5
N	1044	1041	1044	1044	1020	990	990
N (clusters)	538	538	538	538	537	541	541
$E[y]$	$95	$1,920	23	34%	21%	27%	77%

Notes: RD estimates of impacts of county seat victory on various long-run outcomes. (1) shows the impact on log(wages) minus the log of the national average of wages for all workers in the same occupation. (2) shows the impact on the log of home value in the 1940 micro census. (3) shows the impact on the 1940 census occupational income score ("occscore"). (4) shows the impact on white-collar jobs. (5) shows the impact on the share of income going to the top 5% in zip-level IRS data. (6) shows the impact on the fraction of LODES jobs in the highest wage tercile. (7) shows the impact on white-collar jobs in the LODES data. For sector definitions see Section A.15. Geographic controls as listed in Section A.12.

Table C.3: Main Effects by Scope of County Government

	Officers Below Median			Officers Above Median		
	(1) ln(Pop) [< 1 mile]	(2) log(Wages) 1940	(3) log(Income) 2010	(4) ln(Pop) [< 1 mile]	(5) log(Wages) 1940	(6) log(Income) 2010
Win	1.08***	0.16***	0.066***	1.16***	0.12***	0.12***
	(0.15)	(0.023)	(0.023)	(0.17)	(0.029)	(0.030)
County FEs	Y	Y	Y	Y	Y	Y
Geo	Y	Y	Y	Y	Y	Y
SEs / Clusters	Election	Election	Election	Election	Election	Election
BW (pp)	14.5	14.5	14.5	14.5	14.5	14.5
N	667	661	653	385	383	367
N (clusters)	339	337	336	202	201	201
$E[y]$	1,108	$712	$48,642	900	$735	$47,901

Notes: RD estimates of impacts of county seat victory on various long-run outcomes separated by high/low scope of county government. For each state, we count the number of major county offices as listed in Murphy (2009) and split the sample into above and below median states. The outcomes are respectively population density within 1 mile; log(wages) in 1940 for nongovernmental, non-lawyer/judge workers; and 2010 Zip-level income per IRS data. Full-sample estimates are given in Figure 4 and Table 1. Geographic controls as listed in Section A.12.

Table C.4: Determinants of Town Vote Margins

	(1) Predicted TE	(2) Vote Margin (%)	(3) Vote Margin (%)
(log) Distance Center	0.058***	-0.074***	
	(0.0026)	(0.012)	
PLSS Sec. Area	-0.20***	0.16*	
	(0.030)	(0.089)	
Terrain Slope	-0.00025	-0.0014**	
	(0.00019)	(0.00060)	
Non-swamp	-0.086***	-0.028	
	(0.011)	(0.049)	
Soil Quality	-0.045**	0.028	
	(0.018)	(0.074)	
TE: Density			-0.34**
			(0.13)
Protected Incumbent	0.0011	0.16***	
	(0.0061)	(0.028)	
Has Newspaper	0.0054	0.11***	
	(0.0044)	(0.020)	
log(Town Pop)	-0.00091	0.019***	
	(0.0010)	(0.0042)	
Has PO	0.010**	-0.0070	
	(0.0046)	(0.022)	
RR	0.0052	0.082**	
	(0.0059)	(0.032)	
SEs / Clusters	Election	Election	Election
Election FEs	Y	Y	Y
N	1564	1564	1614
N (clusters)	807	807	807

Notes: OLS determinants of machine learning predictions for county-level density, and town vote margins respectively; see Section 7.1. The sample of elections is the baseline of first elections by county and includes the top two competitors within each election. Determinants were chosen based on (a) the top five most important variables of the GRF model, as coded by the fraction of first splits; PLSS section area is top-coded at the 99^{th} percentile (b) markers of size and status. "Protected Incumbent" refers to cases where election regulations favored a particular location (for example, an incumbent) by requiring a supermajority to unseat it.