# There's No Such Thing As Free Land: The Homestead Act and Economic Development

Ross Mattheis
Itzchak Tzachi Raz*

Harvard University

December 31, 2019

**JOB MARKET PAPER**

[Click here for the most updated version]

**Abstract**

The 1862 Homestead Act provided free land conditional on five years of residency and cultivation to settlers of the American West. In total, the Act granted 10% of the land in the United States to 1.6 million individuals. This study examines the impact of the Act on long-run development. Using spatial regression discontinuity and instrumental variable designs, we find that areas with greater historical exposure to homesteading are poorer and more rural today. The impact on development is not only driven through differences in the urban share of the population; cities in homesteading areas are less developed and non-agricultural sectors are less productive. Using newly geo-referenced historical census data, we document the path of divergence starting from the initial settlement. We find that homesteading regions were slower to transition out of agriculture. The historical and empirical evidence is consistent with the hypothesis that the transitory distortions caused by the Act's residency and cultivation requirements induced selection on settlers' comparative advantage in agriculture. This, in turn, inhibited the development of non-agricultural sectors and the subsequent benefits of agglomeration.

*Raz (**Job Market Paper**): Email: iraz@g.harvard.edu. I am grateful to my advisors Alberto Alesina, Melissa Dell, Claudia Goldin, and Nathan Nunn for their support and guidance. We also thank Omar Barbiero, Augustin Bergeron, Moya Chin, Edward Glaeser, Nir Hak, Alejandro Lagomarsino, Maor Milgrom, Jeffrey Miron, Dev Patel, Ljubica Ristovska, Jonathan Roth, Andrei Shleifer, Clemence Tricaud, Chenzi Xu, and David Yang for their valuable comments and suggestions. This project also benefited from comments by numerous participants at the EHA 79th annual meeting, the Harvard Economic History Workshop, the Stanford Economic History Seminar, the Harvard Empirical Legal Studies Series, Clark University, and multiple student workshops at the Department of Economics and the Department of Government at Harvard University. This research was supported by a grant from the Lab for Economic Applications and Policy at Harvard University. Raz also acknowledges financial support from the Economic History Association, the Center for American Political Studies at Harvard University, and the Institute of Humane Studies. All remaining errors are our own.

# 1 Introduction

In the mid-19th-century vast areas of the American West were sparsely settled. Many locations were equally suited for settlement, yet similar locations experienced very different paths of development: some remained mostly rural while others developed into dense and productive urban areas. Federal land policy played an active role in shaping the early patterns of settlement, as the government initially owned the land and its policies determined how it was privatized. This paper focuses on the largest policy in this context—the 1862 Homestead Act. Designed to promote settlement of the West by small family farmers, the Act granted free land to settlers conditional on five years of residency and cultivation. Between 1862 and 1976, the Act granted roughly 10% of the land area of the United States to 1.6 million individuals (Edwards et al., 2017; North, 1974). Critics of the Act have long argued that its provisions distorted settlers' decisions and incentivized them to cultivate land when it was inefficient (Shannon, 1945; Gallman et al., 1972; North, 1974; Anderson and Hill, 2004). Although these distortions only directly affected settlers' decisions during the first five years, they might have altered patterns of settlement and the subsequent path economic development.

This study examines the long-run impact of the 1862 Homestead Act by comparing the development outcomes of locations that experienced different levels of homesteading.[1] Studying the impact of the Act on local development presents a number of challenges. First, settlement under the Act was endogenous to geography and the decisions of settlers, which potentially bias the OLS estimates. Moreover, the Homestead Act was a part of an "incongruous land system" in which many different land policies were in play (Gates, 1936), and the counterfactual land privatization policy varied across regions and the period of settlement. Finally, each land policy was in itself a bundle, affecting who got the land, how much, for what price, and on which terms. Therefore, even holding the counterfactual to homesteading fixed, potential differences in outcomes could be attributed to each of these components, making interpretation difficult.

We address these challenges by focusing on an area where the ability to homestead varies discontinuously across space and the historical context makes the counterfactual mode of settlement clear. Our main strategy exploits the sharp change in exposure to the Act across the historical boundaries of the Osage tribe in southern Kansas in a spatial regression discontinuity (RD) design. Between 1865 and 1870, the Osage ceded their land to the federal government to be sold for their benefit, thereby prohibit-

[1] It is worth highlighting that we cannot test consequences relative to a counterfactual in which the Homestead Act was never enacted. We can only test how differential historical exposure to the Act shaped the spatial distribution of economic development. Specifically, locations with little or no exposure to homesteading may also have been affected by the Act due to the reallocation of economic activity across space. Our results are therefore not indicative of the impact of the Act on the development of the U.S. economy as a whole.

ing homesteading. Outside the cession boundaries, land could be homesteaded and ownership granted for free following five years of residency and cultivation, while inside land had to be purchased, and title was granted immediately. The boundaries of the Osage land were arbitrary straight lines drawn on the map, which allows us to study the impact of the Act on development by comparing outcomes on either side of the boundary. In addition, Osage land was to be sold only to actual settlers, in parcel sizes identical to the homestead plots, implying that the only difference was in the terms of the transfer. Therefore, focusing on the Osage land cessions allows us to isolate one specific component of the Act—the effect of the terms of land privatization—from other potential channels. An additional advantage of this setting is that it provides a natural placebo. Along the eastern part of the boundary there were only negligible differences in homestead assignment, because eastern Kansas was settled before the passing of the Act by settlers who paid for their land. We focus on the western section of the boundary for estimation and use the eastern section as a placebo.

We find that settlement under the Homestead Act reduced local economic development today. To measure long-run development, we use spatially disaggregated 1990 census block-level data on median housing value and rent for our main results, and supplement these measurements using more aggregated 2000 census block-group level data on income per capita, median household income, and educational attainment. Our estimates suggest that in 1990, median housing values and rents in locations just outside the Osage cessions were around 0.5 standard deviations lower relative to locations just inside the cessions. The result is robust to the selection of bandwidth, RD polynomials, different ways of accounting for spatial auto-correlation, and the use of different measures of local long-run development. A fuzzy RD specification suggests that a 10 percentage-point increase in the share of land homesteaded decreases income per capita in the year 2000 by about 0.18 standard deviations. When we focus on a placebo region in the east we find no effect on development.

We also document long-run effects on the industrial composition of the local economy. Historical exposure to the Act decreased year-2000 employment shares in high-human-capital industry sectors, such as information and finance, and increased shares in low-human-capital sectors such as agriculture, construction, and retail. This pattern is consistent with an adverse impact of the Act on the development of cities. However, the impact on development is not simply driven by the share of the population in urban areas; cities in the homesteading region are poorer relative to cities inside the cession.

Although the context of the Osage cession has several advantages, the unique setting may raise questions about external validity. First, the spatial discontinuity in the ability to homestead makes it relatively easy for individuals or firms to relocate in response to differences across the boundary, but the same might not be true in a broader geographical context where homestead-eligibility changes more smoothly. Second, as mentioned, the RD around the historical Osage boundary estimates an

impact relative to a specific counterfactual in which small-scale farmers purchase land, which allows us to isolate the effect of a conditional land transfer relative to land purchase. However, it also begs the question of whether results are similar when compared to a more general counterfactual that also differs along other dimensions—mainly, initial parcel sizes and the ability of absentee landowners to acquire land.

To address these concerns, we broaden the scope of the analysis by using an instrumental variable (IV) design. Our design leverages the component of homesteading availability that is generated by the continuous process of westward expansion and the date of homestead enactment. The intuition behind the instrument is simple. In some counties, a significant share of land was already bought before the passing of the Act. Since land that was already privatized can no longer be homesteaded, prior land privatizations constrained the share of land available to homesteaders. The exact year in which the Act passed was independent of the patterns of settlement in the West. Earlier attempts to pass a Homestead Act go back to 1846 and continue throughout the 1850s, but those were blocked by southern Democrats; the Act only passed in 1862, during the Civil War, after the South seceded from the Union. We use historical land patent data to identify the counties in which significant shares of land were privatized in the decade immediately following Homestead enactment, which we refer to as the "1862 privatization frontier." Then, for counties along the frontier, we use the distance to the 1862 national population center of gravity as an instrument for the fraction of land homesteaded. Intuitively, counties further away had lower shares of land privatized before the Act passed, leaving more land available for homesteading.

The IV estimates are consistent with the RD results. Using county-level data on income per capita for 1969-2000, we find that a 10-percentage-point increase in the share of land homesteaded lowers income per capita by 0.13 standard deviations in the long-run. The IV estimate is not statistically different than the equivalent RD estimate. The result is robust to using different measurements of local long-run development, different definitions of the privatization frontier, and different ways of accommodating spatial auto-correlation for inference. A placebo exercise suggests that effects are not directly driven by the time of settlement or distance to the population center of gravity in themselves, as we do not find a first stage or a reduced-form relationship when we replicate the analysis using counterfactual years of Homestead enactment. When we look at the effect on the composition and productivity of different industries, our findings are consistent with the RD results, and they support the hypothesis that the Act's adverse impact on development is driven by urbanization and the development of non-agricultural industries.

We explore potential mechanisms by studying the impact of the Act starting with the initial period of settlement. The historical and empirical evidence supports the hypothesis that the transitory distortions

of the Act induced selection that set homesteading regions on divergent paths of development. We discuss two channels of selection on relative agricultural skill that may result from the residency and cultivation requirements in the Act. First, settlers with a competitive advantage in agriculture faced a lower opportunity cost to homesteading. Additionally, the frontier was a notoriously challenging environment that had a high level of uncertainty and was subject to adverse shocks (Shannon, 1945; Lee, 1979; Edwards et al., 2017), and more capable farmers were more likely to survive as homesteaders in this environment. Consistent with this pattern of selection, historical evidence suggests that homesteaders who survived the initial five-year period stayed as farmers on their land in greater numbers and for longer periods relative to farmers who purchased land (Lee, 1979). Positive selection on relative agricultural productivity in homesteading regions was likely to affect the path of development as the local economy started to transition out of agriculture. Homesteaders and their descendants would have been less likely to switch into non-agricultural sectors, inhibiting their development and the subsequent benefits of agglomeration.

We provide evidence of divergence in the path of regional development using the RD framework. To this end, we digitized historical maps and geo-referenced individual-level data from the full-count censuses of 1880-1940 into civil townships, towns, and cities around the Osage cessions. Overall, our findings point to diverging paths of development across the Osage boundary starting in 1910, rather than to an initial impact that simply persisted. Consistent with the cultivation requirement of the Act, we find that settlers in the homestead-eligible area were more likely to work in agriculture in 1880. However, after successful homesteaders obtained property rights and the restriction on their occupational choices was removed, the differences in industrial composition across the Osage boundary attenuate and become insignificant by 1900.

At the beginning of the 20th century, the region began to go though a structural transformation. Those with lower relative agriculture productivity were likely to be the first to transition out of agriculture and into manufacturing. Differences in the distribution of skills due to selection during settlement would lead the homestead region to transition more slowly on average. Moreover, entrepreneurs and firms were likely to respond to differences in the composition of local labor markets through their location choices. Consistent with that, we find that the trend in differences in the industrial composition reverses from convergence to divergence. After 1910, estimates of the differences across the boundary grow consistently, generating a U-shaped path. Forces of agglomeration were then likely to exacerbate existing differences. Evidence suggests that as new waves of migrants settled in the region, they were less likely to settle in homesteading areas. By 1940, we again find a positive effect of historical homesteading on the probability that a household head in the region was working in the agricultural sector, and a negative effect on the probability of working in manufacturing, construction, and finance,

insurance and real estate.

Over time, the divergence in the industrial composition of the local economy materialized into divergence in the level of economic development. Using newly digitized data on the assessed valuation of total property from the Kansas State Board of Agriculture, we find a growing gap in wealth across the historical Osage boundary, statistically significant in the decades after 1910. Before the mid-20th century, homesteading locations were already significantly poorer. Using census data, we document that households who reside in homesteading regions reported housing values 0.37 and 0.28 standard deviations lower than households inside the Osage cessions in 1930 and 1940, respectively. Similarly, in 1940, wages in homesteading regions were 0.35 standard deviations lower.

We contribute to several literatures. First, a broad literature across the social sciences studies the development of the American economy (Fogel, 1964; Goldin and Katz, 2008; Donaldson and Hornbeck, 2016; Sequeira et al., 2019), the American West (Bazzi et al., 2018; Leonard and Libecap, 2019), and specifically the Homestead Act (Shannon, 1945; Lee, 1979; Gates, 1960; Shanks, 2005; Edwards et al., 2017). Despite wide acknowledgement that the Homestead Act played a significant role in the development of the U.S., there is little evidence of the effect of the Act. Recent papers study the impact of the Act on state capacity (Poulos, 2019) and land inequality and education (Lillo Bustos, 2018). We provide quantitative evidence documenting its adverse impact on local on economic development.

We also contribute to a broad literature on property rights and land reforms (Banerjee et al., 2002; Besley, 1995; Besley and Ghatak, 2010; Field, 2007; Galiani and Schargrodsky, 2010; Hornbeck, 2010). This literature emphasizes the importance of assigning property rights for the efficiency of economic outcomes. However, in practice, a substantial variation exists in the way different land reforms were designed, and there is little evidence of the advantages and shortcomings of different policy designs in this context. This paper contributes to the literature by studying the implications of a specific governmental policy design, a conditional transfer of land, relative to a specific counterfactual design in which settlers purchase land. Our findings suggest that the transitory distortions that arise from a conditional land transfer policy can have persistent impacts.

This paper also speaks to an important set of studies that demonstrate the impact of historic events on economic development today (La Porta et al., 1998; Acemoglu et al., 2001; Engerman and Sokoloff, 2002).[2] Past work has also gone further by shedding light on specific channels of persistence (Dell, 2010; Nunn and Wantchekon, 2011; Alesina et al., 2013; Lowes and Montero, 2018; Dell et al., 2019; Dell and Olken, 2019). Yet we know less about the way in which effects evolve over time. Beyond showing persistent effects, our results document the path of initial divergence. Our findings suggest that the short-lived distortions placed homesteading regions on a lower path of economic development

[2]See Nunn (2009) for a thorough overview of the literature.

and point to agglomeration and path dependency as important channels of persistence.

This relates to a literature in economic geography and urban economics that studies persistence in the spatial distribution of economic activity. More densely settled areas benefit from agglomerating individuals and firms, reducing costs and accelerating the exchange of ideas (Glaeser and Gottlieb, 2009). As a consequence, impermanent differences in productivity between locations can lead to lasting differences in the density of settlement (Bleakley and Lin, 2012; Hanlon, 2017; Michaels and Rauch, 2017). These differences may be especially important while an area is at early stages of development (Bleakley and Lin, 2012; Jedwab et al., 2017). Our findings highlight that government policy can play an important role in shaping the initial allocation of human capital across space, which can have a persistent impact due to forces of agglomeration.

This paper is organized as follows. Section 2 provides historical background and an overview of the 1862 Homestead Act. Section 3 documents the impact of the Act on local economic development in the long run. Section 4 presents evidence on the path of divergence. Section 5 offers a conceptual framework for interpreting our results, and Section 6 makes concluding remarks.

# 2 Historical Background

> *"The homestead act is now the approved and preferred method of acquiring title to the public lands. It has stood the test of eighteen years, and was the outgrowth of a system extending through nearly eighty years, and now, within the circle of a hundred years since the United States acquired the' first of her public lands, the homestead act stands as the concentrated wisdom of legislation for settlement of the public lands."*
>
> — Public Land Commissioner Thomas Donaldson

## 2.1 Federal Land Policy Before the Homestead Act

During the 19th century, the United States federal government gained control over about 1.5 billion acres of land within the contiguous United States. The administration of these lands was a first-order public policy question for the young nation and subject to vast and contentious political debates.[3] From the founding of the republic, legislators were divided on how to manage public lands. Alexander Hamilton's federalists advocated for the sale of the land in large tracts to raise revenue, while Thomas Jefferson's Democrats favored distributing the land to independent farmers in small amounts.

[3] 3,500 acts of legislation regarding land were passed by the Congress between 1785 and 1880 (Gates, 1970).

Initially, land revenues were imperative to reduce the national debt, and Hamilton's argument held sway for much of the first half of the 19th century. The Land Ordinance of 1785 laid the foundations of U.S. land policy. It set up a land survey system and established the mechanism by which land was to be privatized. After the survey, land would be put on auction and sold to the highest bidder, with a minimum quantity and a minimum price per acre. Importantly, there was no cap on the total amount of land one could acquire, and many absentee owners came to own vast holdings (Libecap and Lueck, 2011; Gates and United States, 1968).

Over the years, many became concerned that the federal land policy was favoring speculators and the wealthy over small farmers. Gradually, the policy started to shift toward favoring actual settlers. Both the minimum quantity and minimum price per acre were reduced, falling to a minimum of 80 acres at $1.25 per acre in 1820.[4] In parallel, other acts allowed squatter-settlers paths to land ownership. Starting in 1830, a series of Preemption Acts were enacted, permitting squatters to purchase public land before it was offered on auctions. Adverse Possession legislation allowed free transfer of land titles from absentee owners to squatters (Raz, 2018).

The shift in federal land policy toward actual settlers peaked with the passing of the Homestead Act in 1862, after persistent advocacy from progressive legislators. The first Homestead bill was introduced in the House by Andrew Johnson in 1846. While the bill did not pass, the homestead movement quickly grew in support. The Free Soil Party (1848-1852) argued for "Free Soil, Free Speech, Free Labor, and Free Men" and included the homestead principle in its political platform.[5] Throughout the 1850s, various homestead acts were introduced in the House but blocked by the Senate. Fearing a loss of population to the West, both the Northeast and Southeast delegates initially objected to the enactment of a Homestead Act. Over time, the political debate surrounding the Act was shifting to center on the slavery controversy. If western land was to be owned by actual settlers farming small tracts, slavery could not expand westward. The Northeast allied with the Northwest, generally supporting the Act, while the Southwest joined the Southeast in opposition to it. In 1860, a Homestead Act was approved in Congress but was vetoed by President Buchanan. After the secession of the South between 1860 and 1861, Northern Republicans were able to pass the Act again. "AN ACT to secure homesteads to actual settlers on the public domain" was finally signed into law by President Abraham Lincoln on May 27,

[4] In 1785, the minimum quantity was 640 acres and the minimum price was $1 per acre. In 1786, the price per acre increased to $2 per acre. The minimum quantity decreased to 320 acres in 1800, and decreased again to 160 acres in 1804. In 1820, both the minimum quantity and price were reduced to 80 acres at $1.25 per acre (Gates, 1970). The Graduation Act of 1854 established a price schedule that declined over the length of time a plot was available for purchase, falling as low as 12.5 cents per acre after thirty years without a buyer.

[5] The party argued "That the public lands of the United States belong to the people, and should not be sold to individuals, nor granted to corporations, but should be held as a sacred trust for the benefit of the people, and should be granted in limited quantities, free of cost, to landless settlers." (Donaldson, 1884).

1862 (Donaldson, 1884; Lee, 1979).

## 2.2 The 1862 Homestead Act

### Provisions and Operation of the Act

The Homestead Act allowed settlers to acquire up to 160 acres of public land for free following five years of residency and cultivation. These requirements were meant to ensure that land is only given for free to actual settlers. Homesteaders were also given the option to buy the land after six months of residency at $1.25 per acre. This provision essentially meant that for the settler, the Homestead Act "contains all of the beneficial features of the Preemption Act" (Donaldson, 1884, p. 350). All adults or heads of family, who are citizens or in the process of becoming one, and did not fight against the U.S. government, were eligible for homesteading.

The process of acquisition under the Homestead Act had three stages (Lee, 1979). First, after choosing a vacant tract on surveyed public land, the homesteader would file their application and an affidavit of qualification at the local land office.[6] Specifically, the settler would have to declare that the claim was "made for his or her exclusive use and benefit, and that said entry is made for the purpose of actual settlement and cultivation." The land register would verify that there were no competing claims for the same tract. Then the homesteader would pay a $10 fee plus a $2 commission to the land agent. Once the receipt was issued, the land was claimed temporarily by the homesteader.

Second, the homesteader would have to "improve" the land through five years of residency and cultivation. Homesteaders were given six months to begin residence on their claims. The dwelling had to be habitable as a place of permanent residence. Improvements such as fences, sheds, or a well were often demanded as evidence of a permanent residency.[7] Continuous residence and cultivation of the land for five years was a necessity. In practice, absence of less than six months was often permitted. Fearing the possibility of having their claim contested, however, homesteaders were cautious to limit their period of absence and to have witnesses that could testify to their return.

The third step was to file for a deed of title. The homesteader had to file an affidavit stating that he had fulfilled the requirements of the Act, supported by a testimony of two witnesses.[8] An additional

[6] If the land office was far from the settlement site, the homesteader could fill out their paperwork in front of a local court clerk and mail it to the land office. If there was high risk of multiple people filing for the same tract, many homesteaders chose to go to the land office in person (Lee, 1979).

[7] Lee (1979) provides two interesting examples that failed to satisfy the requirements: a sod house without a door, window, stove or chimney, and a "partnership tent". Another interesting example is that of a homesteader who failed to move his family to his claim location and was contested by another settler based on the argument that his real residence was elsewhere. He almost lost his claim.

[8] The two witnesses could not be family members or otherwise immediately connected to the homesteader.

fee of $6 was required, bringing the total cost of fees and commissions to $18. If the homesteader abandoned his claim or failed to file for a deed within seven years, the plot would return to the public domain and would become available for others.

It is important to note that the cost of obtaining land was just one component of the total cost of farm-making in the West—and a relatively small one. Capital was needed to finance the travel westward; break the land; buy seeds, farm equipment, livestock, and the building materials needed for a dwelling; and sustain the farmer and his family during the early period in which their farm did not produce sufficient output to support them. Estimates for the minimum required investment for western farm-making vary between about $600 to $1,600 (Danhof, 1941; Deverell, 1988; Lee, 1979). Out of this amount, the homesteader could save $182 relative to a purchase of 160 acres,[9] representing between 30% to 11%. For context, note that the national average daily wage for a laborer in 1860 was $1.11, so Homestead discount was valued at over half of a year of a typical laborer's wage (United States and Edmunds, 1866).

By the time of the repeal of the Act in 1976, about 270 million acres, which is about 10% of the land area of the United States and 27% of the public domain, had been transferred from the government into the hands of about 1.6 million individuals (National Archives, 2016; Edwards et al., 2017; North, 1974). Shanks (2005) estimates that between 46-93 million people living in the U.S. today are descendants of original homesteaders.[10]

While the Homestead Act is particularly well-known and widely perceived as a prominent factor shaping the American West (Gates, 1936; Edwards et al., 2017; North, 1974; Sheldon, 1936), policies of this type are not unique to the United States. Acts of a similar nature were introduced across the New World to settle the frontier, such as in Argentina, Brazil, Canada, Costa Rica, Colombia, New Zealand, and the Philippines.[11] The exact provisions and the historical context varied, but the common element was a free transfer of land conditional on a period of residency and cultivation.

[9] 160 acres × $1.25 = $200 - $18 = $182.

[10] While public opinion regarding the Act is generally positive (Edwards et al., 2017), there does not seem to be a consensus among scholars. While many notable 20th-century historians have referred to the act as a "landmark legislation" (North, 1974, p. 113), "one of the most important laws which have been enacted in the history of this country" (Gates, 1936, p. 652) and even the world (Sheldon, 1936, p. 75), others scholars have argued the Act was not actually significant in the context of the settlement of the West (Shannon, 1945; Edwards et al., 2017). Gates (1936) too argue that "there was built up around the law a halo of political and economic significance which has greatly magnified the importance to be attributed to it and which has misled practically every historian and economist who has dealt with land policies" (p. 653). Similarly, while Allen (1991) claims that "Homesteading gets no respect. Both historians and economists alike find only bad things to say about it," Whaples (1995) finds that in fact, the vast majority of economic historians do not think that the Act had an adverse effect on productivity.

[11] In some countries, these policies were not as effective as the 1862 Homestead Act in providing broad access to frontier land (Acemoglu and Robinson, 2012; García-Jimeno and Robinson, 2008; Engerman and Sokoloff, 2012).

**Homesteading Relative to Land Purchase**

How would land have been privatized in the absence of the Homestead Act? Due to the complexity of the federal land policy, the answer might have been different depending on the location and period of settlement (Gates, 1936). In this paper, we focus on a specific counterfactual scenario in which the frontier is being settled by small-scale farmers who purchased their land from the government. Specifically, our main empirical strategy will focus on a historical context in which an arbitrary historical boundary of Native American tribal land caused a sharp discontinuity in federal land policy. On one side of the boundary settlers were allowed to homestead, while on the other land was only sold to actual settlers in small tracts of up to 160 acres. Focusing on this particular counterfactual has an important advantage. It allows us to isolate one specific economic channel: the terms of the land privatization contract. This rules out other potentially meaningful channels, such as the existence of speculators and absentee landowners or larger initial plot sizes in the region not exposed to homesteading, when interpreting results. Our second empirical strategy extends the scope of the analysis to a historical context in which these potential channels are not blocked, yet the results are nevertheless very similar.

How should we think about the homesteader relative to the farmer-settler who buys land from the government? Both individuals migrated west in the hope of starting a farm on the frontier. However, when seeking to obtain land from the government, they faced different privatization contracts. The key differences in terms of the two contracts are summarized in Table 1. The nature of land privatization in the case of land purchasers was a simple transaction: they would pay a total of $200 for 160 acres, obtain title right away, and face no restrictions on their place of residence or occupation. In contrast, the nature of land privatization for homesteaders was a conditional transfer of land. They were only required to pay a total of $18 for 160 acres; but in return, accepted upon themselves to continuously residing on and cultivating their land for five years. Until these requirements were fulfilled, homesteaders did not get property rights. If these requirements were not fulfilled, homesteaders would lose their claim and any investment they might have made on it.

# 3 Impact on Local Economic Development in the Long Run

Selection is the fundamental challenge in identifying the causal effect of the 1862 Homestead Act on locations is a selection problem. As homesteaders had to reside on and cultivate the land for five consecutive years before obtaining their title, they looked for a hospitable climate, high agricultural productivity, and proximity to markets. These characteristics are likely to be positively correlated with long-run economic development, potentially leading to an upward bias of OLS estimates.

To address this challenge, we use two quasi-experimental designs. A spatial RD design exploits a

discontinuous change in the availability of homesteading for settlers across the historical boundaries of the Osage tribe in southern Kansas (section 3.1). An instrumental variable strategy leverages variation in the availability of homesteading along the 1862 frontier that resulted from the contiguous westward expansion and the timing of Homestead enactment (section 3.2). In both designs, we find that areas that experienced more historical homesteading are significantly poorer today.

## 3.1 Spatial Regression Discontinuity Design

Throughout the process of westward expansion, Native Americans ceded tribal lands to the U.S. federal government. In some cases, treaties specified that the federal government would hold the land in trust, sell it, and use the proceeds to benefit the tribe. Importantly, this prohibited homesteading on the ceded land. Our main strategy focuses on one area in southern Kansas that was ceded by the Osage tribe between 1865-1870 and was subject to such constraints on homesteading.

### 3.1.1 The Setting

In 1825, the Osage tribe ceded most of their land in the great plains to the U.S. government, maintaining a 50-by-276 mile strip of land southern Kansas. In 1865, the tribe signed a treaty with the government to cede two additional tracts of land.[12] The first ("Osage ceded land"), containing about 844 thousand acres, was bought by the government. The second ("Osage trust land"), containing about 3.2 million acres, was ceded in trust for the benefit of the tribe. Land on both tracts was to be sold at a minimum of $1.25 per acre, which specifically barred homesteading. A third tract ("Osage diminished reserve"), containing about 4.8 million acres, was ceded in 1870 after the tribe Osage expressed their desire to leave Kansas and settle in an alternative location in Indian Territory. A bill to remove the Osage from the diminished reserve noted that the Osage trust land and the Osage diminished reserve would be "sold to actual settlers only [...], in quantities not exceeding one hundred and sixty acres [...] at the price of one dollar and twenty-five cents per acre" (Congress Bill, Chap. 296, 1870). This implied that the only difference between the land privatized under the Homestead Act outside these cessions and the land sold inside was that the first was a transfer conditioned on five years of residency and cultivation, while the second was a simple transaction in which land title was purchased.

Our analysis exploits the discontinuous change in the terms of land privatization across the boundaries of these cessions in a spatial regression discontinuity design by comparing the outcomes of locations just outside the cessions relative to locations just inside. The differential assignment of homesteading around the cessions boundaries is presented in Figure 1.

[12] Appendix Figure B.1 shows the historical boundaries of these tracts.

We focus on the western portion of the boundary, since the area outside the cessions along the eastern boundary was settled before the passing of the Homestead Act, leading to negligible differences in homestead assignment across the boundary (Figure 1).[13] Moreover, the differences in timing of settlement in the east are a concern for identification, which requires that all factors beside treatment will vary smoothly at the boundary. We use the eastern portion of the boundary, along with the boundaries of the Osage ceded land and a tract of land ceded by the Cherokee in 1866 (the "Cherokee neutral land," Appendix Figure B.2),[14] as a placebo.

Wichita, the largest city in Kansas, started as a small trading post along the Chisholm Trail inside the Osage cessions. It grew considerably in the following decades, eventually crossing over the historical Osage boundary. Historical evidence suggests that the city had no special advantages in terms of geographical characteristics or exact location. Instead, the key to its success laid in the people who inhabited it during the early years and the subsequent attraction of heavy industry (Miner, 1982). It is possible that the development of Wichita was positively affected by the fact that all the land around it was bought rather than homesteaded.[15] Yet the centrality of Wichita, along with the fact that it had spread over the historical boundary, complicates both analysis and inference. Our baseline analysis flexibly controls for Wichita, but we also document a robustness to dropping it from the sample.

One regional center of economic activity in the proximity of the historical Osage boundary - Dodge city, owes its location and early development to the near-by Fort Dodge. The original town site was located "as close as possible to the fort's compound" (Dykstra and Manfra, 2017, p. 9), and eleven out of the twelve initial town shareholders had economic connections to the fort. This raises a concern that the spatial allocation of economic activity in this region of the boundary was determined due to another federal policy - the location and operation of military forts. Frontier forts provided protection and local economic opportunities, and have been shown to have persistent effects on local population density (Carter, 2019). We will thus document robustness to dropping Dodge City from the sample.

### 3.1.2 The Estimation Framework

The boundaries of the Osage land cessions form a discontinuity in the availability of homesteading in a longitude-latitude space. We use a spatial regression discontinuity to estimate the effect of the

[13] Specifically, our analysis is limited to location west of the 96.88 meridian. We determine this cutoff using the spatial distribution of land privatized prior to the passing of the Act, as the 95th percentile of the spatial distribution on an east-west axis. This cutoff lines up well with historical accounts, which note that "The area in Kansas open to free homestead lay for the most part beyond the 97th meridian" (Gates, 1954, p. 231). See also (Lee, 1979, p. 45).

[14] This tract was ceded in trust by the Cherokee under a treaty signed in 1866. It was to be appraised and sold to the highest bidder, thus prohibiting homesteading. The northern boundary of this tract is aligned with northern boundary of the Osage cessions, forming a straight line on the map from east to west.

[15] Indeed, this is an important part of our interpretation.

Homestead Act on local long-run development. Regressions take the form:

$$Outcome_l = \beta OutsideCession_l + f(Location_l) + Seg_l + \gamma DistWichita_l + Wichita_l + \epsilon_l \quad (1)$$

where $Outcome_l$ is the outcome of interest in location $l$ and $OutsideCession_l$ is the fraction of location $l$ that is homestead eligible. $\beta$ is the coefficient of interest, representing the relationship between the fraction of land that was eligible for homesteading and the outcome of interest. $f(Location_l)$ is the RD polynomial, which smoothly controls for the characteristics of geographic location. $Seg_l$ is a closest boundary segment fixed effect, which ensures that the analysis is comparing locations across the same segment of the boundary. $DistWichita_l$ controls for distance between location $l$ and Wichita. Finally, $Wichita_l$ is a flexible control for 2.5-, 5-, 10-, 15-, and > 15-mile bends around Wichita. These fixed effects are included in the regression in order to explicitly control for the spatial distribution of economic development around Wichita. Flexibly controlling for the location around Wichita is meant to ensure that estimates are not biased by, for example, comparing Wichita's outer neighborhoods to the suburbs. Results are robust to dropping Wichita from the sample.

Our baseline specification uses a local linear RD polynomial, as recommended by Gelman and Imbens (2014), a rectangular kernel, and a 10-mile bandwidth.[16] These specification choices are not essential for our main results. We follow the practice in the literature and document high robustness to a wide range of bandwidths and different RD polynomials (Dell et al., 2019; Lowes and Montero, 2018).

To account for spatial auto-correlation, we follow Bazzi et al. (2018) and cluster observations at arbitrary grid-cells, as proposed by Bester et al. (2011). This approach is considerably less computationally demanding in large samples compared to Conley (1999) spatial standard errors. We choose a grid size of 15 square miles in our baseline specification to allow for maximal spatial correlation without encountering the problem of having too few clusters for our main result (Cameron et al., 2008). However, in some samples and levels of spatial aggregation, the number of spatial clusters drops below 30, raising a concern of a bias as a result of too few clusters. Thus, in these cases we also report Wild bootstrap p-values. In all tables, we also report Conley spatial standard errors with a 50-mile window that are not subject to this concern. Our results are robust to using grid-cells of difference sizes and

[16] To the best of our knowledge, there is no accepted theory on optimal bandwidth selection for multi-dimensional RD studying a wide array of outcomes at different levels of aggregation. For reference, the Calonico et al. (2014) optimal bandwidth for our measurements of long-run development vary from 5.8 to 15.73, depending on the level of spatial aggregation (i.e., block or block group) and whether Wichita is included in the sample. Historical data is only available at a higher level of spatial aggregation.

Conley spatial standard errors with windows of different sizes.

The key identifying assumption is that all relevant factors besides treatment vary smoothly at the Osage cessions boundary. This assumption is required to ensure that locations just inside the cessions, in which land had to be paid for, are an appropriate control group for locations just outside the cessions, where homesteading was possible. Note that in this setting, treatment is the contract under which land was to be privatized. Importantly, treatment was assigned to areas that would be settled *before* the regions were populated by settlers, rather than to an already present population.[17] Therefore, manipulation of the treatment assignment is not a concern in this setting; selective sorting of settlers to different locations is a potential mechanism, not an identification concern.

To assess the plausibility of the identification assumption, Table 2 presents estimates of equation (1) using three different local linear RD polynomials: in latitude and longitude, in distance to the boundary, and both. Column (1) examines the date of privatization, while columns (2) to (10) examine the geo-climatic characteristics: land temperature, precipitation, slope, altitude, and potential yield of wheat, maize, soybean, and sorghum. The dependent variables are all standardized into z-scores. For some combinations of characteristics and RD polynomials, the estimate of $\beta$ is significantly different than zero; however, that is not robust to the choice of the RD polynomial form. Moreover, for most variables and RD polynomials, the estimate is statistically insignificant, suggesting that locations inside and outside the Osage cession are comparable. Appendix Figure B.3 plots the RD boundary over the spatial distribution of central geo-climatic characteristics. It is evident from the figure that the boundary does not coincide with any break in geographical characteristics.

### 3.1.3 RD Results: Long-Run Development

We measure local long-run economic development using data from the 1990 and 2000 decennial censuses of population and housing (Manson et al., 2019). Our main results draw data on median owner-occupied housing value and median renter-occupied rent from the 1990 census. This data is available at the census-block level, which is the smallest geographic area for which the Bureau of the Census collects and tabulates 100% decennial census data (LaMacchia et al., 1994). We supplement these measurements using data on income per capita, median household income and educational attainment from the 2000 census, available at the more aggregate level of census block-group.

[17] To ensure that effects are not driven by difference in the timing of settlement, we focus on the western portion of the Osage cessions.

**Main Results**

Table 3 reports estimates from equation (1) for our main results. The dependent variable is median housing value.[18] Appendix Table A.1 reports results when the dependent variable is median rent. We find that homestead-eligible locations outside the Osage cessions have significantly lower property values compared to locations in which homesteading was prohibited. Our baseline specification, which uses a local linear polynomial in latitude and longitude, suggests that median housing value and median residential rent are about 0.5 and 0.56 standard deviations lower in homestead-eligible areas, respectively (column 1). This result is robust to different specifications to address spatial auto-correlation (Appendix A.1.4).

Column (2) uses a local linear polynomial in distance to the boundary, while column (3) uses a local linear polynomial in both latitude and longitude and distance to the boundary. The results are generally similar across these different linear polynomials. Column (4) drops all blocks within 15 miles from Wichita's center. The point estimate slightly drops in magnitude, but remains economically and statistically significant, suggesting that the effect is not entirely driven by the development of Wichita. Column (5) documents robustness to dropping Dodge City from the sample.

Figures 3-4 present the results graphically in a standard RD plot, with distance to the boundary as the running variable and a linear trend along each side of the boundary up to 35 miles. The dependent variables are residualized on the covariants in the baseline specification of equation (1). Panel (a) uses all of the data, while Panel (b) drops Wichita from the sample. In both outcomes and samples there is a discrete drop in the predicted value of real property at the boundary.

A common concern in an RD analysis is that the effect is being driven by just a handful of observations close to the discontinuity. To address this, in column (6) we run a donut-hole specification, dropping all observations within 2.5 miles of the boundary, which yields even stronger results. A related concern in the case of this study is that some current census blocks do not fall entirely within the treatment or control regions; the higher the level of spatial aggregation in the data, the greater this potential problem. For that reason, our baseline specification uses a continuous treatment assignment, measured by the fraction of the area that falls outside the cessions. Column (10) instead assigns a binary treatment according to the location of the centroid of each census block. Results are similar to the baseline.

Finally, columns (7)-(9) show robustness to dropping the baseline controls—the boundary segment fixed effect, allowing comparison of location far away from each other, the distance to closest center of

[18] To allow for clear interpretation and easy comparison across different outcomes and periods, we present results when the outcome of interest is standardized into z-scores. Results are generally qualitatively similar when levels or logs are used instead.

economic activity, which may correlate with location relative to the cession boundary, and the flexible Wichita controls—allowing for effects to vary by the current spatial distribution of economic activity around the city center. The last attenuates the point estimate, and depending on the specification to account for spatial auto-correlation, statistical significance can be substantially lower.

**Placebo Exercise and Further Robustness Checks**

One possible concern is that the effect is driven by the fact that the land was held by Native Americans right up to the period of white settlement. To address this concern, we estimate equation (1) on the placebo region along the eastern portion of the boundary. Like the western part of the Osage land cessions, homesteading on the Indian cessions in southeastern Kansas was prohibited. However, unlike the western portion, land outside the cessions in the east was mostly settled before the passing of the Act, leading to hardly any difference in homestead exposure across the boundary. Appendix Table A.2 presents the placebo results using different RD polynomials and bandwidths. We find no impact on long-run development in the placebo region.[19]

Figure 5 checks the robustness of the RD estimates to different choices of bandwidths, RD polynomials, and samples. The figure plots the point estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1) for every combination of bandwidth values between $5-35$ miles, common specifications for the RD polynomial, and three different samples. Each facet represents a specific combination of the sample and RD polynomial, and the x-axis within each facet denotes the different bandwidths. Each row uses a different RD polynomial: linear in latitude-longitude (row 1), linear in distance to the cession boundary (row 2), linear in both latitude-longitude and distance (row 3), quadratic in latitude-longitude (row 4), and quadratic in distance (row 5). Each column studies a different sample. Column 1 studies the the full sample and shows that estimates are remarkably robust to the choice of RD polynomial and bandwidth. Column 2 excludes all blocks within 15 miles from Wichita's center, and demonstrates that the result is not driven by the development of Wichita. For most polynomials and bandwidths, there is little change in the point estimate and statistical significance is maintained. At the same time, the results in column 2 are generally weaker than in column 1, suggesting that the development of Wichita as an economic center inside the Osage cession is potentially an important part of the story. Column 3 focuses on the placebo sample and finds no effect. In almost all polynomials and bandwidths, the placebo estimate of $\beta$ is statistically zero.

Finally, in Table 4 we report the results for our main measurements of development (columns 1-2) alongside alternative measurements of local long-run economic development at the block-group level

[19] An important caveat for this placebo is that it potentially also picks up the effect of being settled earlier, which may confound the results.

(columns 3-5). The results are qualitatively similar, but the magnitudes of the effects on the block-group level outcomes are generally higher. Income per capita and median household income are about 0.85 standard deviations lower in locations just outside the Osage cession, relative to locations just inside. Similarly, the share of the population over the age of 25 with at least a bachelor's degree is 1.6 standard deviations lower in homestead-eligible locations. In the Appendix, we document robustness of these results to different bandwidths and RD polynomials (Appendix A.1.3). Appendix Table A.4 reports results for the same outcomes from a fuzzy RD specification in which the Osage cession serves as an instrument for the share of land homesteaded.

### 3.1.4 RD Results: Industrial Composition and Agglomeration

#### Employment Shares in Different Industry Sectors

Using 2000 census block-group level data on employment shares in different industry sectors, we document that settlement under the Homestead Act affected the industrial composition of the local economy in the long run. Figure 6 presents the RD estimates of $\beta$ when the dependent variables are the employment shares in different industry sectors, standardized into z-scores. Our findings suggest that the Act decreased employment shares in high-human-capital industry sectors, while the opposite is true for low-human-capital sectors. Specifically, we find a negative effect of the Act on employment shares in the information industry, finance, insurance and real estate, professional, scientific, management, administrative, and waste management services sectors, as well as in the educational, health, and social services. We find positive effects on employment shares in agriculture and mining, construction, wholesale and retail, and transportation, warehousing, and utilities. This pattern is broadly consistent with contemporary differences in economic activity across urban and rural locations.

#### Urban-Rural Composition vs. Agglomeration

One possible explanation for the impact on development is that homestead-eligible locations outside the Osage cession are more rural. Then the impact on development may be entirely driven by a different urban-rural composition; cities outside the cession might be as developed as cities inside, but there are simply fewer of them. We present evidence suggesting otherwise.

We focus exclusively on 1990 blocks classified as urban to test for gaps in the development levels of cities across the historical Osage boundary.[20] Table 5 presents the results.[21] The dependent variable

[20] Note that the 1990 classification of locations into a urban or rural status is in itself a potential outcome. Therefore, these estimates should not be interpreted as causal effects.

[21] When the sample only includes dense urban locations, all observations are clustered into a small number of locations within the bandwidth. This leads to a small number of arbitrary grid-cell clusters for the calculation of standard errors.

in Panel A is median housing value, while in Panel B it is median rent. We present results for three different local linear RD polynomials: in longitude and latitude (column 1), in distance to the boundary (column 2), and both (column 3). In all specifications, we find that cities just outside the cessions are less developed relative to cities just inside. This suggests that the impact on development is not a simple composition effect. Instead, forces of agglomeration and increasing returns to scale are likely to play an important rule in explaining the long-run impact of the Homestead Act. These results join the results on the industrial composition in suggesting that the adverse impact on development is related to differences in the historical process of agglomeration and the allocation of economic activity across space.

## 3.2 Instrumental Variable Strategy

We extend the scope of analysis and use an instrumental variable strategy that addresses external validity concerns that might arise due to the unique Osage cessions setting; First, the fact that the ability to homestead varied sharply across the Osage boundary made it more likely that individuals or firms would relocate in response to homestead eligibility and the subsequent differences in the initial patterns of settlement. In a broader geographical context where homestead-eligibility changes more smoothly the same might not be true. Second, it focuses on a specific counterfactual, in which land is sold only to small-scale farmers instead of being granted to them in return to five years of residency and cultivation. However, in other contexts, the counterfactual to homesteading also potentially induced differences in initial parcel sizes and the ability of absentee landowners to acquire land. The IV strategy leverages a different source of variation, spread over a wider geographical area and using data at a higher level of spatial aggregation. Therefore, it tests effects against a more general counterfactual, and results in this context are unlikely to be driven by highly localized sorting of individual and firms. Moreover, it relies on different identification assumptions. The results are nevertheless consistent with the RD findings.

### 3.2.1 The Estimation Frameworks

The intuition behind our second identification strategy is simple. After the Homestead Act went into effect, land in the public domain was available for homesteaders. However, in certain areas, significant shares of land were already privatized and thus could not be homesteaded; in other words, the patterns of prior land privatizations constrained the share of land locally available for homesteaders. This

We maintain this approach as a baseline for uniformity, but note that these standard errors should be taken with a grain of salt. As such, we also report Wild bootstrap p-values, but highlight that the Conley standard errors are likely to be more reliable for this sample.

relationship is visible in Figures 7 and 8, which plot the fraction of land homesteaded and the fraction of land privatized prior to the passing of the Homestead Act, respectively, in counties across the U.S. Midwest.

The IV strategy leverages the component of prior land privatizations that results from the definitive date of Homestead enactment and the continuous process of westward expansion, to causally identify the effect of the Act on long-run development.[22] Specifically, we use land patent data from the Bureau of Land Management and General Land Office to identify the areas that were intensively privatized during the decade immediately after the passing of the Act, which we refer to as the "1862 privatization frontier." Then, within this sample of counties, we instrument the fraction of land homesteaded with the distance between each county and the 1862 national population center of gravity.

The fundamental logic of this strategy rests on the following fact: Since the process of westward expansion in the Midwest was generally contiguous, at each point in time during the privatization and settlement period, counties located further away from the center had lower shares of land privatized, on average, relative to counties located closer. This implies that all else equal, counties that happened to be located further out along the privatization frontier when the Homestead Act finally passed in 1862 had a higher share of land available for homesteaders. Therefore, a relationship between distance to the 1862 population center of gravity and the share of homesteading can be found along the 1862 privatization frontier. This pattern is observable in Figure 9.

It is important to note that the identification builds on the fact that only along the 1862 privatization frontier—where some of the land, but not all of it, was privatized before the Act passed—should one expect to find a local relationship between distance to the center and the share of homesteaded land. In locations that were on the privatization frontier long before 1862,[23] essentially all of the land was privatized before the passing of the Act, allowing no land to be homesteaded. In such locations, there should be no relationship between counties' location along the frontier and the share of land homesteaded. Similarly, in locations that were on the privatization frontier only in later periods, none of the land was privatized prior to the passing of the Act, thus placing no restrictions on the share of land that could be homesteaded. In other words, a first stage should be found along the privatization frontier at the time of enactment, but not along the frontiers of different periods. In a counterfactual world—in which, for example, the 1846 homestead bill would have passed into law—a gradient in the

[22] The patterns of prior privatization are arguably not exogenous in themselves. Specifically, it is possible that all else equal, better land would be bought earlier.

[23] It is important to remember that the process of settling and privatizing land is rather slow moving, and that a given county may be a part of the frontier for many years. This also implies that for a sufficiently low $\tau$, there is correlation in the sample of counties that are on the frontier in year $t$ and year $t+\tau$.

share of land homesteaded would have been located along the 1846 frontier instead.[24]

We estimate the effect of the Homestead Act on local long-run development along the 1862 privatization frontier using the following standard IV framework:

$$Outcome_c = \beta Homestead_c + \delta_{s(c)} + X_c\Gamma + \epsilon_c \tag{2}$$

$$First\ Stage: \quad Homestead_c = \alpha Dist PCG_c^{1862} + \theta_{s(c)} + X_c\Lambda + \varepsilon_c \tag{3}$$

where $Outcome_c$ is the outcome of interest in county $c$, and $Homestead_c$ is the fraction of land that was privatized under the 1862 Homestead Act. $\beta$ is the coefficient of interest, representing the relationship between the fraction homesteaded and the outcome of interest. $Dist PCG_c^{1862}$ is the distance between county $c$ and the 1862 national population center of gravity. $\delta_{s(c)}$, $\theta_{s(c)}$ are state fixed effects. Finally, $X_c$ is a wide array of geo-climatic controls: temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat.

### 3.2.2 IV Results: Long-Run Development

We use data on income per capita for 1969-2000 from the Bureau of Economic Analysis to measure long-run economic development at the county level.[25] For our baseline specification we aggregate this data into a single county-level measure of long-run income per capita by first standardizing contemporaneous income per capita into z-scores within each given year and then taking the mean over 1969-2000. The result is robust to using different procedures or measurements of development.

The main IV result can be visualized in Figure 10, which presents the partial correlation plots for the first stage (Panel a) and the reduced form relationship (Panel b). A strong first stage relationship is clearly evident: The further a county is from the 1862 population center of gravity, the higher the fraction of land homesteaded. A strong reduced form relationship is also visible: the further a county

[24] Note that while a relationship between distance from the center and share homesteaded should not generally exist along privatization frontiers of different periods, such relationship will exist when the sample is sufficiently large—for example, within the U.S. Midwest as a whole. This is due to the fact that there was no homesteading in eastern states, and essentially no homesteading in Midwest states that were settled prior to 1862. However, much of this variation is across states, rather than within states, and in broad samples of this sort, the instrument is unlikely to be valid.

[25] While post-2000 data do exist, we restrict the analysis to end at year 2000 in order to have results that are comparable with the RD results, for which outcomes are only available for years 1990 and 2000 (see subsection 3.1.3). Moreover, breakdown of BEA income and earning data by industry sectors, which is used in to explore heterogeneous effects in section 3.2.3 below, exist by SIC classification for 1969-2000, but by NAICS classification post 2000. Our result is robust to using 1969-2017 income per capita instead.

is from the center, the lower is its long-run income per capita. The IV estimator is therefore negative.

Table 6 reports these results more formally. Column 1 reports the OLS estimate of $\beta$ from equation (2). The point estimate is 0.2; however, it is insignificant. We then proceed to the IV estimation. Column (2) reports the first stage. The point estimate of $\alpha$ from equation (3) is 0.128, with statistical significance well above the 0.1% level. The interpretation is that along the 1862 privatization frontier, a 100-mile increase in the distance between a county and the 1862 population center of gravity is associated with about a 13 percentage-point increase in the share of land homesteaded. The Montiel Olea and Pflueger (2013) effective first-stage F statistic is 65.5,[26] indicating a strong first stage. Column (3) reports the reduced form relationship between distance to the center and income. A point estimate of $-0.172$, significant at the 0.1% level, indicates that a 100-mile increase in the distance to the 1862 population center of gravity is associated with about a 0.17 standard deviations drop in a county's long-run income per capita.

Finally, column (4) reports the baseline IV estimate of $\beta$. The point estimate is $-1.345$, with an associated p-value of 0.027. The interpretation of this result is that a 10-percentage-point increase in the share of land homesteaded in a county decreased long-run income per capita in the county by about 0.13 standard deviations. To further interpret the magnitude, note that the mean share of land homestead in our sample of counties along the 1862 privatization frontier is about 32%. The IV result is robust to different definitions of the 1862 privatization frontier, different measurements of long-run economic development, and the use of different cutoff windows for the Conley spatial auto-correlation (Appendix A.2).

Comparing column (1) and (4) suggests a significant positive selection of areas experiencing high shares of homesteading along the 1862 privatization frontier.[27] The Upper Peninsula of Michigan, for example, although being further away from the 1862 population center of gravity, has relatively low shares of homesteading (Figure 9). Indeed, the region's climate, soil, densest timber coverage and distance from markets make it unsuitable for settlement and cultivation, but it is rich in mineral deposits and timber, making it valuable for some purchasers. For the same reasons, this region has low population density and little economic activity today. These factors, which may not be fully accounted for by our geo-climatic controls, will not be picked up by the instrument.

[26] Since our design have one endogenous variable and one instrument, the effective first-stage F statistic equals the robust first-stage F statistic. See Andrews et al. (2019)

[27] With a bigger sample, including all of the Midwest, the selection problem seems to be less severe. The OLS estimate of $\beta$ (equation 2) in this broader sample is $-0.523$, significant at the 1% level.

### A Placebo Exercise

A possible concern with this identification strategy is that the instrument might simply pick up effects that are associated with being settled later or located further away from the center. The concern is that these factors have an independent impact on long-run development. That would violate the exclusion restriction, which requires that distance from the population center of gravity will not be associated with long-run development except through its effect on homesteading.

To address this concern, we implement a placebo exercise in which we regress the first stage and reduced form for every privatization frontier and population center of gravity between 1830 and 1889. The logic of this placebo exercise is that if the exclusion restriction holds, one should expect to find a reduced form effect only when there is a relevant first stage, and for the reasons explained above, the second should only be around 1862. It is important to keep in mind that rate of movement of both the privatization frontier and the national population center of gravity is relatively slow (Appendix Figure B.4), implying a strong serial correlation in both sample and instrument from year to year. As such, one should not expect to find a first stage and reduced form effects *only* in 1862. Instead, if the instrument is valid both estimates would peak at 1862 and decay toward zero in a U or inverse-U shape before and after.

Figure 11 presents the results of the placebo exercise. Panel (a) presents the first stage results of the effect of distance to the center of gravity on the fraction of land homesteaded. Panel (b) presents the reduced form effect—i.e,. the relationship between distance to the center and long-run income per capita. We find that for both dependent variables the estimates long before and long after the time of enactment are close to zero. However, the estimates increase in absolute value as the year gets closer to 1862, reaching a peak right around 1862. This suggests that distance to the center of gravity along privatization frontiers of different years does not affect development in general. Instead, our instrument seem to be picking up something unique that occurs only around the year of Homestead enactment. Moreover, the fact that the first stage and reduced form estimates are mirroring each other suggests that the reduced form effect works through the first stage.

A remaining concern is that a first stage and reduced form relationship only exist around the time of enactment, but not because of the Act. For example, the skeptical reader may still be concerned that while it does not seem to be the case that counties settled slightly later are less developed relative to counties that were settled earlier, having been settled later during the Civil War mattered for development. While concerns of this sort seem less plausible, we cannot rule them out.

### 3.2.3 IV Results: Industry Sectors, Urbanization, and Geographical Mobility

#### Long-Run Earnings, Productivity, and Employment Shares Across Different Sectors

Which sectors of the local economy were affected by the Act? To answer this question, we use data on personal earning per capita, personal earnings per worker, and employment shares, by SIC divisions, calculated using data from the Bureau of Economic Analysis for the years 1969-2000.[28]

Table 7 presents the IV estimates of $\beta$ (equation 2) for different dependent variables in different industry sectors. Each column focuses on a different dependent variable: long-run personal earning per capita (column 1), long-run personal earning per worker (column 2), and long-run employment shares (column 3). Panel A focuses on farming, while panel B on all other forms of economic activity. We find that the aggregate effect on long-run development is driven entirely by non-farming activities. (column 1). The magnitudes of the effect on earning per capita derived from non-farming industry sectors (Panel B, column 1) and the aggregate effect on income per capita (Table 6, column 4) are remarkably similar. A 10-percentage-point increase in the share of land homesteaded is associated with a 0.14-standard-deviations drop in non-farm earning per capita. In contrast, we find no effects on earning per capita derived from farms. The point estimate is actually positive, however insignificant (Panel A, column 1). Columns (2) and (3) suggests that the negative effect on earnings is driven by lower productivity, as measured by earning per worker, rather than a smaller employment share.

Appendix Table B.1 presents further results, breaking down the non-farming side of the economy into different industry sectors. We find that effects on earning per capita are concentrated mainly in three industry sectors: manufacturing, construction, and finance, insurance, and real estate. We also find that the coefficients on earning per worker are negative across the board, although insignificant for many sectors. In no industry, including farming, do we see a robust positive association between homesteading and productivity. Finally, effects on employment shares are mostly insignificant.

#### Contemporary Urbanization and Geographical Mobility

We also find evidence suggesting that the Homestead Act affected urbanization and the extent of residents' geographical mobility. Table 8 presents the results. Column (1) examines the effect on population density. We find a strong and significant effect, suggesting that a 10-percentage-point increase in the share of homesteaded land decreases population density by 0.37 standard deviations.

We then proceed to study the effects on the geographical mobility of residents within a county. As expected in more rural locations, we find evidence for lower mobility. Columns (2)-(3) study effects

[28] We aggregate contemporaneous data into a single county level long-run measure using the same procedure as for our baseline measure for long-run economic development, described above.

on migration flows into and out of the county, using data from Chetty and Hendren (2018). We find that a 10-percentage-point increase in the share of homesteaded land decreases migration inflows by about 0.17 standard deviations and migration outflows by about 0.15 standard deviations. Columns (4)-(8) proceed to study effects on the length of residents' tenure. Specifically, we use census data for the years 1970-2000 (Manson et al., 2019) to examine effects on the share of households in the county with different periods of tenure living in their current house. We find no significant effect on the share with 1 year and $11-20$ years of tenure (column 4 and 7, respectively). We do, however, find that a 10-percentage-point increase in the share of land homesteaded decreases the share of households with $2-5$ years of tenure by about 0.25 standard deviations and the share with $6-10$ years by about 0.09 standard deviations, while increasing the share with a tenure of over 20 years by about 0.24 standard deviations.

## 3.3 Discussion: A Comparison Across Strategies

Our two identification strategies are consistent both qualitatively and quantitatively. The fuzzy RD estimate from the Osage cessions suggests that a 10-percentage-point increase in the share of land homestead decrease income per capita in 2000 by about 0.18 standard deviations (Appendix Table A.4, column 4). The IV estimate for the same dependent variable suggests a slightly lower effect, of about a 0.12-standard-deviations decrease (Appendix Table A.6, column 4).

There could be different reasons for these small differences in magnitude, ranging from the different level of spatial aggregation and geographical areas to differences in the historical context. Importantly, while those originally purchasing land inside the Osage cessions had to be actual settlers, the same is not generally true as a counterfactual to settlement under the Homestead Act, and specifically in the context of the IV strategy. The possible existence of large landowners as a counterfactual to homesteading in the IV context, but not in the Osage cessions context, may drive a wedge in the estimated effects.[29] However, in light of all of this, it is quite remarkable to note that each estimate falls within the other estimate's 95% confidence interval. Even though our two strategies rely on entirely different identification assumptions, leverage different sources of historical variation, and use data at a different level of spatial aggregation over different geographical areas, they not only generate similar qualitative results, but also estimates of quite similar magnitudes.

The similarity in results from the two strategies is not limited to the main outcome. Rather, the deeper patterns revealed using both designs are also consistent. The RD estimates suggest that cities just out-

[29] Indeed, recent evidence suggests that an initial privatization to large landowners, as opposed to homesteaders, had adverse consequences for development (Smith, 2019).

side the cession are less developed relative to cities just inside (Table 5). Using the RD strategy, we also document important effects on the industrial composition of the local economy (Figure 6). These effects are broadly consistent with differences in the industrial composition across rural and urban locations in 2000. The IV estimates suggest that counties that experienced higher historical homesteading are more rural today, and their residents are less geographically mobile (Table 8). We also find that contemporary income differences are driven by non-agriculture sectors (Table 7). Both sets of results thus point to the development of non-agricultural sectors, agglomeration, and urbanization as key factors in driving the main outcome.

# 4 The Path of Divergence

We explore potential mechanisms by focusing on the Osage land cessions and providing evidence on the impact of the Act starting from the initial stage of settlement.[30] Our results suggest a path of divergence rather than an initial impact that persists. We find that homesteading regions were slower to transition out of agriculture. Over time, the divergence in the industrial composition of the local economy materialized into a divergence in the level of economic development.

The analysis in this section utilizes micro-level data from the full-count decennial federal censuses for the years 1880, and 1900-1940.[31] We focus on household heads, studying their demographics, occupations, and the industry sector in which they work.[32] In 1930 and 1940, we also study the value of their dwelling or rent cost, and in 1940, their wages. We also use township- and city-level data on the assessed valuation of total property, digitized from the Biennial Reports of the Kansas State Board of Agriculture, for each decade between 1900 and 1940. To facilitate the RD analysis, we use historical maps to geocode the boundaries of civil township and the locations of towns and cities around the Osage cessions boundary in each decade.[33] We then geo-reference the historical data to townships', towns', and cities' centroids.[34]

[30] Note that the IV strategy compares the outcomes of counties that were settled earlier to those of counties that were settled later. Therefore, it is only fitting to study the long-run impact. During the early stages of settlement, counties further away from the population center of gravity will be, almost by definition, less developed, and the nature of economic activity in them is likely to be different regardless of the terms of land privatization.

[31] Data from the 1890 census was lost.

[32] In the early years, this variable is imputed from occupation question, rather than based on data directly recorded.

[33] The boundaries of civil townships have changed quite significantly between 1880 and today.

[34] It is important to note that the level of spatial aggregation in the historical data is significantly higher than the one used to study long-run outcomes (see Section 3.1.3). As such, the RD results in this section of the paper generally have lower power and are less stable across specification. The higher level of aggregation also entails a small change in specification: since all of Wichita is geo-referenced into a single point on the map, we include a single Wichita fixed effect instead of flexible controls for distance bends around the city's center.

## 4.1 Initial Differences, 1880

We begin by studying the impact of the Act during the period of initial settlement. We choose 1880, the first census year following the bill to remove the Osage people from Kansas, as the starting point for the analysis. During this period the western edges of our study region were not yet settled, while the central and eastern parts were still in the early stages of development.

Using the RD framework, we do not find evidence for selective sorting of initial settlers into homesteading based on observable demographics (Figure 12). Settlers in homestead-eligible locations outside the cessions were equally likely to be male, white, single, U.S-born, and born in different regions of the country. They were also of a similar age and had a similar number of children. Of course, it is possible that settlers were selected into homesteading based on unobserved characteristics, such as differences in wealth or skills. Unfortunately, we have no direct data on wealth, and the data on literacy is of poor quality.[35]

We then proceed to examine the initial impact on economic activity. Figure 13 presents the RD estimates on a set of dummies that take a value of 1 if the household head works in a given industry sector, standardized into z-scores. We find that household heads in homestead eligible locations are 0.579 standard deviations more likely to work in agriculture, forestry, and fishing, relative to household heads in ineligible locations just inside the cession (line 1). Homestead eligibility decreased the share of household heads working in manufacturing, construction, transportation, wholesale, or retail by 0.213, 0.437, 0.298, 0.143, and 0.249 standard deviations, respectively (lines 2-6). In all other industry sectors, the estimate is small and insignificant (lines 7-10). These results are consistent with the provisions of the Act, which required homesteaders to cultivate land.

To complement these results, we run a similar analysis, this time using occupational categories rather than industry sectors. The dependent variables here are dummies that take a value of 1 if the household head's occupation is classified into a given category (e.g., "farmer", "laborer", or "craftsman"), standardized into z-scores. The results are presented in Appendix Figure B.5 . We find a similar pattern. Homestead eligibility increased the probability that a household head is a farmer in 1880 by 0.798 standard deviations (line 1). For all other occupations, the point estimates of $\beta$ are negative. We find no significant effect on the probability of working as a farm laborer (line 2), but a strong negative effect on working as a laborer outside the agricultural sector (line 3). Specifically, this probability is about 0.657 standard deviations lower in locations just inside the Osage land cessions relative to locations just inside the cession. Similarly, the probability of working as operatives, craftsmen, clerks, or managers

[35] One possible way to advance this question would be to link individuals across census years, which would allow us to study parental and sibling occupations, as well as the individual's, across periods. These variables may potentially be informative about wealth and talent. We are currently implementing this linkage.

was 0.256, 0.332, 0.049, and 0.203, standard deviations lower, respectively (lines 4-5 and 8-9). The point estimates on occupations belonging to the broad categories of sales, services, and professional are also negative, but statistically insignificant (lines 6-7 and 10). Finally, lines 11-12 examine effects on labor-force participation and occupational income scores and find no effects.

We interpret these results as a direct impact of the Act's cultivation requirement, which shifted the industrial composition of the local economy during the period of initial settlement. Much of the impact was a shift of employment from non-farm laborers to farmers. This, however, had no meaningful effect on average contemporary development, as evident in the null effect on occupational income scores. Moreover, we find no evidence for selective sorting on observable demographics of early settlers.

## 4.2 Divergence, 1880-1940

After examining the initial impact, we proceed to study the path of divergence. Correlational evidence suggests that the homestead-eligible region was slower to go through a structural transformation. Focusing on the causal effect of the Act, we find that the effect of the act on the industrial composition of the local economy first attenuates post-settlement but picks up again over time, generating a U-shaped pattern. We also find suggestive evidence of agglomeration effects, which potentially widened the gap in local economic development. This pattern manifested in a difference in the level of economic development, where areas settled under the Homestead Act had substantially lower wages and property values by the mid-20th century.

### 4.2.1 The Industrial Composition of the Local Economy

A divergence across the boundary can be observed with simple correlational evidence. Figure 14 plots the share of household heads working in four different industry sectors—agriculture, manufacturing, construction, and finance—between 1880 and 1940 within a 35-mile bandwidth on both sides of the historical Osage boundary.[36] Three important facts are noticeable in the figure. First, the area as a whole went through a structural transformation over this period. There is a long secular negative trend in the employment share in agriculture, dropping from an average of about 77% across both regions in 1880 to about 35% in 1940. In all other sectors, the exact opposite is true, as their importance in the economy grows over time. Second, in accordance with the results on the initial impact of the Act on economic activity presented above (Section 4.1), there are initial differences in the industrial composition of the two regions. The employment share in agriculture in the homestead-eligible region is about 79% in 1880, compared to about 74% inside the Osage land cessions. Third, there is a divergence of

[36] Appendix Figure B.6 plots shares in all other industry sectors.

homestead-eligible and non-homestead-eligible regions above the initial gap. By 1940, the employment share in agriculture in the homestead eligible region is about 41%, compared to about 28% inside the cession. The development in some other sectors is the mirror image of that. Specifically, there is a notable divergence in the share working in manufacturing, construction, and finance, insurance and real estate. In other words, the rate at which the homestead-eligible region transitioned out of agriculture and into these sectors was slower. These trends are particularly interesting because, as we observed in Section 3.2.3, these are exactly the sectors in which long-run effects on earning per capita were concentrated. Our two strategies, implemented over different regions, periods, and aggregation levels, again produce consistent results. In all other industry sectors (Appendix Figure B.6) we find parallel trends (e.g. retail and wholesale), and for some there is also no difference in levels (e.g., mining and communication).

Next, we focus on these four industry sectors and utilize the RD framework to provide causal evidence. Figure 15 presents the RD estimates of the effects on the probability that a household head was employed in each of these sectors between 1880 and 1940. Overall, across all four industry sectors, the evolution of the estimate of $\beta$ over time follows a U-shaped (or an inverse U) path; the coefficient first attenuates between 1880 and 1910 and then picks up again (in absolute value) after 1910.

The fact that estimates of the differences in the industrial composition across the Osage boundary attenuate in 1900 relative to 1880 suggests that much of the differences in 1880 were a direct result of the Homestead Act's cultivation requirement. After successfully surviving five years and obtaining property rights, the restrictions on homesteaders' occupational choices were removed. While homesteading in the region did not end by 1900, the rate of homestead entries vastly declined, leading to a very low share of settlers being subject the homesteading constraints in 1900, and even fewer in 1910 (Lee, 1979).

At the beginning of the 20th century, the region as a whole began to go though a structural transformation; however, our findings suggest that this process occurred later and slower in homestead-eligible locations. The RD estimates document that the trend in differences in the industrial composition reverses from convergence to divergence, as the gap across the historical Osage boundary consistently grows after 1910. The divergence path as captured by the RD estimates post-1910 is consistent with the path documented in the raw data (Figure 14). By 1940, we find that household heads just outside the Osage cessions were about 0.62 standard deviations more likely to be working in the agricultural sector, relative to household heads just inside the cessions. In contrast, they were 0.42, 0.2 and 0.09 standard deviations less likely to be working in manufacturing, constriction, and finance, insurance, and real estate, respectively.

Overall, the evidence documents that homestead-eligible locations were slower to transition out of

agriculture, and it points to diverging paths of development across the Osage boundary rather than to an initial impact that simply persisted over time. Our preferred interpretation is that the delay in the local development of non-agricultural sectors meant losing out on the potential benefits resulting from agglomeration, relative to nearby locations that transitioned out of agriculture earlier.

#### 4.2.2 Suggestive Evidence on Agglomeration

We provide suggestive evidence supporting the agglomeration interpretation. Specifically, we provide evidence suggesting that individuals who arrived in the area in later periods settled in greater numbers inside the Osage cessions. Since census data during the early periods contained no information regarding former place of residence, we use alternative variables as proxies. Our first proxy is simply the household head's state of birth. Panel (a) in Figure 16 plots the share of household heads born in Kansas between 1880 and 1940 on both sides of the boundary. Initially, in 1880, all settlers on both sides of the boundary are immigrants to Kansas. Over the years, the share of household heads born locally picks up in both regions; however, a clear divergence is evident. By 1920, the share of local-born household heads is higher in the homestead-eligible region. The RD estimates, presented in Panel (b), paint a similar picture.

A divergence path on this proxy measurement may be driven by different rates of migration outflows as well as different rates of migration inflows. To disentangle these two possibilities, we turn to a second proxy: race. The logic behind this proxy measurement is rooted in the fact that most black residents arrived in the region after the initial period of settlement. Divergence in the share of black households is therefore more likely to represent different rates of migration inflow than different rates of migration outflow. Figure 17 presents the results. Both the correlational evidence in Panel (a) and the RD estimates in Panel (b) suggest that black households migrated to locations inside the historical Osage land cession in greater numbers relative to locations outside. To the extent that these different rates of migration represent a general trend among later migrants, this evidence suggests that agglomeration may have been one of the forces behind the economic divergence.

#### 4.2.3 From Industrial Composition to Economic Development

The differences in industrial composition of the local economies across the historical Osage boundary resulted in differences in economic development before the mid-20th century. To provide empirical evidence we use data on the assessed valuation of total property for years 1900-1940, available at the civil township, town, and city level, which we digitize from Biennial Reports of the Kansas State Board of Agriculture.

Panel (a) in Figure 18 presents the path of the RD estimates of $\beta$ on the log of assessed valuation of total property for the years 1900-1940,[37] while Panel (b) presents estimates for the same period when the dependent variable is standardized into z-scores instead. While we do not find any effects in the earlier years, we do find adverse effects of the Act on total wealth beginning in 1910. By 1940, the assessed valuation of total property of homestead-eligible locations just outside the Osage cession was 0.09 standard deviations lower, or 0.9 log points lower, relative to non-homestead locations just inside the cession. When interpreting these results, it is important to remember that the dependent variable, being the total value of property in a location, measures both development and agglomeration.

More than just documenting a divergence in wealth, these results also demonstrate that the timeline of the divergence in wealth is quite aligned with that of the divergence in the industrial composition of the economy. The RD estimates on the probability of employment in different sectors suggest that divergence in the industrial composition picks up after 1910 (Figure 15). The RD estimates on the assessed valuation of total property only becomes significant at the same time. Therefore, our results suggest that the divergence in the industrial composition, agglomeration, and the accumulation of wealth are related.

To complement this analysis, we use once again individual-level data from the 1930-1940 full count censuses, in which we observe self-reported housing value, rents, and wages.[38] Table 9 presents the results. Overall, the estimates suggest a significant adverse effect of the Homestead Act on economic development before the mid-20th century. Columns (1) and (2) examine effects on housing values. We find that households who reside in locations just outside the Osage cessions reported housing values that were about 0.37 and 0.28 standard deviations lower relative to households residing just inside the cessions, in 1930 and 1940, respectively. The results regarding rents are qualitatively similar, although not statistically significant (columns 3-4). Finally, in column (5) we examine the impact on wages in 1940. We find that historical exposure to homestead settlement reduced the reported wage income of household heads residing in locations just outside the cession by about 0.35 standard deviations relative to household heads just inside.

# 5 Selection, Survival on the Frontier, and Agglomeration

The empirical evidence documents that homesteading regions were slower to go through a structural transformation, and suggests that in turn affected local economic development. But what explains the slower transition out of agriculture in homesteading regions? The effect of the Act on the industrial

[37] We present analysis in logs since the property value data is aggregated to the location level, and more likely to be skewed.
[38] Census data on wages is available only for 1940.

composition of the local economy evolved in a U-shaped pattern, which is hard to reconcile with an explanation based simply on an initial impact and path dependency.

Our preferred interpretation focuses on a process of selection, survival on the frontier, and subsequent agglomeration. Below, we lay out a conceptual framework for thinking about differential selection into homesteading and subsequent economic development. In our framework, the Act can induce two distinct types of selection. The first is that those with lower productivity outside the agricultural sector were more likely to homestead. The second is that settlers with a greater aptitude for agriculture were more likely to survive as homesteaders on the frontier. This resulted in different distributions of skills across space. When firms expanded into the former frontier regions, they were more likely to put down roots in non-homesteading areas, where the non-agricultural labor market was thicker and the geographic concentration of high-productivity laborers higher. Differential selection of settlers and firms across regions would then likely have been exacerbated by forces of agglomeration.

Our conceptual framework is consistent with the empirical patterns we document and a large body of historical narratives. Yet we do not intend to claim that this is the only channel by which the Homestead Act affected development outcomes; other channels may be in play. At the same time, the empirical evidence does not seem to support two central competing hypotheses regarding channels. The first focuses on large landholdings as a result of a faster consolidation in non-homesteading regions, while the second focuses on the Act's effects on local culture and values. We consider these potential channels in Appendix C, but find that neither is supported by the data.

To understand the two selection channels, it is essential to first understand two stylized facts of the frontier environment. First, the frontier was a notoriously difficult environment. Lee (1979) observes that farmers on the frontier faced "the threat of Indian attack, grasshopper plagues, inadequate rainfall, lack of capital, inexperience, [...] low prices for agricultural commodities, etc." (p. 574). Establishing a farm took a considerable investment beyond the cost of acquiring land. As Edwards et al. (2017) note, land obtained through the Act was "hardly free. Successfully creating a farm required years of the entryman's labor, plus investments in a house and fields, livestock, a barn, and fencing. And settlers paid in other ways, too, enduring long freezing winters and hot searing summers, often losing crops to drought or wind or disease or grasshoppers, and suffering deep social isolation" (pp. 11-12).

A second stylized fact is that many prospective farmers were often poorly informed about the quality of land and best cultivation practices for the frontier. Gates (1963) notes that "In the land selection process [...] settlers made many errors that resulted in a high rate of failure on homesteads" (p. 38). According to Shannon (1945), the frontier farmer was in "perpetual conflict with climate he could not conquer and soils that he seldom understood" (p. 4). Frontier land tended to be more arid than the environments to which eastern farmers were accustomed. Soil varied considerably, even within

small areas, but settlers had little knowledge on the subject. "Too often the homeseeker judged the soil largely by its color, but this alone is a faulty basis for selection and often is no criterion at all" (Shannon, 1945, p. 7). The large variety of soil types meant that different agricultural practices were optimal in different locations. This constrained the ability of farmers to learn from their neighbors. As a consequence of all this, settlers were often slow to adopt optimal crops and farming techniques for the frontier.

The combined effect of these two features was a high rate of turnover on the frontier. In particular, the rates of Homestead failure—turnover of homesteaders in the first years of settlement before obtaining property rights—was exceptionally high. "As a common old saw put it, 'The government bets you a quarter-section [160 acres] that you can't survive on the land for five years,' and many who took the bet failed" (Edwards et al., 2017, p. 12). Estimates of the failure rate vary between 37% and 65.5% (Edwards et al., 2017).

Let us now consider the question of selection. One selection channel in our conceptual framework is that settlers with greater aptitude for agriculture are more likely to survive as homesteaders on the frontier. It is intuitive that settlers with a greater aptitude for agriculture were more likely to survive in the challenging frontier environment. However, why should this be more likely to hold for homesteaders relative to settlers who purchased their land?

In our framework, the ability to retain a land title but work outside the agricultural sector acted as insurance against adverse shocks to agricultural production such as prolonged droughts, grasshopper plagues, or low prices for agricultural commodities. During times when the conditions for farming on the frontier were particularly hard, settlers faced two options: leaving the frontier entirely or temporarily seeking work as non-farm laborers in the closest town.[39] Resilient or highly productive farmers may have been able to survive the bad periods while continuing to work their land. But what about the other settlers? Those that owned their land had the option to work in town temporarily while maintaining their title. Selling the land during bad times might have been difficult, and implied losses in terms of a low sale price for the land. Migrating far away without selling or leasing the land implied title risk from squatters (Raz, 2018). This would have incentivized farmers who owned their land to stay in the region. In contrast, for the homesteader, temporarily migrating to town meant abandoning his claim, the investment in breaking the land, and the years of demonstrated residency and cultivation. But if "shaky living" changed into "no living at all," and continuing to farm was not an option, homesteaders with no title to land in the area were more likely to migrate to an unaffected frontier to try their luck

[39] Miner (1982), for example, notes that "the Panic of 1873 changed a shaky living into no living at all for some rural residents [around Wichita]. A good number sold out and left the county [...] Others abandoned their farms and moved into town to eke out what living they could, to beg, or to become wards of public charity" (p. 88).

again or search for employment as a farm tenant back east.[40] This implies that if an adverse shock happened to be realized during the first five years of settlement, settlers with relatively low agricultural productivity who could not survive the shock were more likely to select out of the frontier if they homesteaded than if they purchased land.

A second avenue for selection in our conceptual framework is that settlers with lower productivity as non-farm laborers were more likely to homestead than to purchase land. Land title provided insurance against adverse shocks to agricultural production, but even absent of a shock, there was an option value in owning land outright. Settlers who bought land would have been more readily able to adjust their occupation and the size of their holding after learning about their own capability as farmers on the frontier. Buyers who discovered that they were less capable farmers than they'd hoped were able to immediately sell their land and work in town, while those who tried to secure free land were required to remain working in agriculture for five years or else risk the labor and money they had already invested. In our framework, both the insurance value and the option value are positively dependent on the settler's non-agricultural productivity. Consequently, settlers with a higher relative non-farm productivity were more likely to purchase land relative to settlers with lower productivity as laborers.

The characteristics of settlers in and around the Osage cession observed in the 1880 census provide limited evidence in favor of the notion that settlers selected into homesteading. Figure B.5 shows that individuals in the homesteading region were disproportionately likely to engage in farming and less likely to work as laborers. We interpret this as a direct consequence of the Act's cultivation requirement. But individuals in the Homestead areas were also less likely to work in the relatively high-skilled positions of operatives, craftsmen, and managers. However, these results are only suggestive, as we cannot observe the occupations that individuals in the homesteading region would have chosen absent the cultivation requirement. The results in Figure 12 show that there is no evidence of selection on observable demographics.

Of course, other forces can potentially give rise to selection into homesteading. In particular, credit-constrained settlers may have homesteaded for lack of choice. This may be an important selection channel, since settlers' liquidity was likely associated with many other characteristics. However, historical evidence leads us to believe that capital constraints were unlikely to be a main driver of selection into homesteading, especially in the context of the Osage cessions. First, the price discount of \$182[41] was a relatively small share of the total required investment, which is estimated to be between \$600-\$1,600 (Danhof, 1941; Deverell, 1988; Lee, 1979). Second, settlers on the Osage cessions were

[40] Of course, absent a significant shock, homesteaders will be incentivized to continue farming their claim in order to maintain their title.

[41] This discount refers to the gap between the government base price for 160 acres ($\$1.25 \times 160 = \$200$) and the total cost in fees and commissions (\$18) that a homesteader had to pay to obtain a plot of land of the same size.

allowed one year before payment was due. Prospective settlers could break the land, plant, harvest, and sell crops before paying for their land. Moreover, settlers on the Osage cessions had access to mortgages, which they could use to pay for their land (Gates, 1954; Miner, 1982). Third, more generally, capital markets for land on the frontier were quite developed. Land speculators would often offer settlers options to buy their land on credit, and loans for land purchases were commonly available (Gates, 1960). Of course, accessing credit was associated with a cost, and purchasing land was certainly more costly than homesteading, yet that cost was generally a small portion of total investment. Overall, a lack of personal capital would not prevent settlers from purchasing land.[42]

If agricultural shocks during the settlement process would have been more likely to push homesteaders out of the region, then homesteaders' turnover rates would have been higher relative to that of farmers who bought their land. The high rates of homestead failures are consistent with that. However, due to a positive selection on relative agricultural productivity, surviving homesteaders who managed to obtain titles by remaining on their land for five years would have been less likely to switch out of farming in later periods, as nearby towns grew and the urban wage rose. Consistent with this, Lee (1979) finds that final homesteaders in Kansas were "surprisingly persistent in holding on to their land" (p. 243). They stayed as farmers on their land in greater numbers and for longer periods compared to settlers who purchased land. In interpreting these results, Lee notes that "a selection process had already occurred" (p. 303). The divergence in the industrial composition that we document is also consistent with that.

Differences in initial settlement patterns may have affected the location of industries as the local economy grew. When manufacturing firms began to enter the former frontier region, their location decision was likely affected by conditions in the local labor markets. In our conceptual framework, homesteading regions had a greater proportion of the workforce in agriculture and a lower concentration of high-productivity laborers. Consequently, the first manufacturing firms preferred to establish themselves in locations with lower historical exposure to settlements under the Homestead Act.

The case of Wichita clearly illustrates this process. Historians point out that Wichita did not have clear geographic advantages over other towns in the area. As Miner (1982) puts it: "[Wichita's] precise location within south central Kansas was much less important than the enterprise exhibited by its

[42]Limited information is another potential driver of selection. Specifically, settlers had limited information regarding the location of available land for homesteading in a given region and at a given time. This lack of knowledge was often exploited by land speculators and land agencies. Gates (1954) notes that "Immigrants arriving during the sixties and seventies in Kansas where they hoped to find suitable land available free to them under the Homestead Act were met with posters, flyers, newspapers ads, and runners and agents for numerous land agencies and land-grant railroads, all calling attention to quantities of the "best" Kansas land for sale" (p. 234). While this may be an important factor for the selection of settlers into homesteading versus purchasing, it is difficult to see how this could negatively impact the economic development of homesteading regions overall.

people in competing with equally well-situated places" (p. 172). In the years that followed the period of initial settlement, several firms responded to small differences in the degree of urban concentration in Wichita. Clothing companies, an auto factory, and other small manufacturers set up in Wichita between 1890 and 1910. Differences in the characteristics of the initial settlers might have passed to subsequent generations. One interesting example is that of Clyde V. Cessna. Cessna's parents moved from Iowa to establish a farm inside the Osage cession. As an adult, Cessna became fascinated with aviation and briefly moved to New York to work for a small airplane manufacturer. When he decided to establish his own factory in 1916, he moved to Wichita and rented space in a small auto plant. After struggling through the Great Depression, Cessna's company was contracted to build thousands of planes and gliders for Canada and the United States for use in the Second World War. From this point on, Marshallian forces of agglomeration were evident. A pool of skilled, specialized labor was built up around Cessna's factory in the first half of the 20th century. The intensity of demand during wartime bolstered supplier linkages, and the knowledge of aircraft manufacture concentrated around Cessna spilled across the local economy. Consequently, Wichita became a desirable place for other aircraft manufacturers to locate. Today, aircraft manufacturers are the largest employers in Wichita.[43]

The distortions induced by the Homestead Act were small relative to the growth the region experienced over the subsequent hundred and fifty years. But small differences in the pattern of settlement may have set homesteading regions on diverging paths of development, with forces of agglomeration deepening initial differences in the distribution of human talent across the frontier.

# 6 Conclusion

We examine how differential historical exposure to settlement under the 1862 Homestead Act has shaped the spatial allocation of economic activity in the American West. Between 1862-1976, about 10% of the total land area of the United States was privatized under the Act. However, different regions across the country had different rates of historical exposure to the Act. In some regions significant shares of the land were homesteaded, while other regions experienced no homesteading at all.

This study documents that areas more exposed to historical homesteading have lower property values, income, and levels of education attainment relative to areas less exposed to the Act. To provide causal evidence, we use two empirical strategies. A spatial regression discontinuity design exploits a

[43] Another interesting historical example is that of William Coffin Coleman. Like Cessna, Coleman's patents migrated to southern Kansas during the initial period of settlement in the region. They did not homestead; rather, they settled on a farm in Labette County, located on the Osage ceded land. Like Cessna, Coleman became an important inventor and entrepreneur. He founded the Coleman Company, which grew from a small one-room shop in Wichita in 1902 to be a leading company producing recreation products.

discontinuous change in homesteading eligibility across the historical boundaries of the Osage tribe in southern Kansas, whose land was ceded in trust to the U.S. government to be sold. To broaden the scope of the analysis, we use an instrumental variable strategy which exploits variation in the share of land privatized before the Act passed. These two strategies leverage different sources of variation, over different geographical areas and levels of spatial aggregation, and rely on different identification assumptions. Yet their conclusion is identical: Areas that experienced more historical homesteading are significantly worse off today.

Using newly geo-referenced historical census data, we are able to document the path of regional divergence, starting from the initial stage of settlement. Our findings suggest that locations that were settled under the Act were slower to transition out of agriculture relative to locations in which farmers bought their land. We also find suggestive evidence of agglomeration effects and document an adverse impact on economic development before the mid-20th century. The path of divergence is consistent with the patterns we observe in the long run, which point to the development of non-agricultural sectors, agglomeration, and urbanization as key factors driving differences in development.

Our preferred interpretation for these results focuses on a process of selection and subsequent agglomeration. The empirical and historical evidence are consistent with the hypothesis that the transitory distortions induced by the Act contributed to a selection process of initial settlers. Those with lower non-agricultural productivity selected into homesteading, and those with greater agricultural productivity were more likely to survive as homesteaders on the frontier. The different distribution of skills, in turn, inhibited the development of non-agricultural sectors and the subsequent benefits of agglomeration.

Our findings suggest that even relatively modest transitory distortions in the process of land privatization can have long-lasting impacts. This is generally consistent with recent work by Lowes and Montero (2018), who find that 14 years of exposure to a highly extractive and brutal institution had a substantial impact on development. In contrast, we examine the impact of differences in the terms of an economic contract to privatized land, which only distorted economic incentives of initial settlers for a short period of five years. It is thus perhaps even more surprising that we find a substantial impact on development over a century after the initial land privatization took place.

The frontier environment in which homesteading was assigned might be central for reconciling the relatively modest historical variation with the persistent and substantial economic impact we document. The Homestead Act distorted decisions during a period in which the spatial allocation of U.S. economic activity was determined. When the slate is clean, even small marks are easily visible. This seems particularly important in the context of the U.S. Midwest, a relatively homogeneous landscape, where no location seems to have a particular special advantage, such that multiple equilibria and path

dependency can arise (Bleakley and Lin, 2012). Our results, therefore, speak to the potential importance of initial conditions in historical development, and more broadly to the existence of critical junctures in history in which even modest variation can have persistent consequences (Acemoglu and Robinson, 2012; Dell, 2012).

# References


41st Congress (1870). "An Act making Appropriations for the current and contingent Expenses of the Indian Department and for fulfilling Treaty Stipulations with various Indian Tribes for the Year ending June thirty, eighteen hundred and seventy-one, and for other Purposes". Chapter 296.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation". *American Economic Review*, 91(5):1369–1401.

Acemoglu, D. and Robinson, J. A. (2012). *Why Nation Fail: The Origin of Power, Prosperity and Poverty*. Crown Publishing Group.

Alesina, A., Giuliano, P., and Nunn, N. (2013). "On the Origins of Gender Roles: Women and the Plough". *Quarterly Journal of Economics*, 128(2):469–530.

Allen, D. W. (1991). "Homesteading and Property Rights; Or, "How the West Was Really Won"". *Journal of Law and Economics*, 34(1):1–23.

Anderson, T. L. and Hill, P. (2004). *The Not So Wild, Wild West: Property Rights on the Frontier*. Stanford University Press, Stanford, California.

Andrews, I., Stock, J. H., and Sun, L. (2019). "Weak Instruments in IV Regression: Theory and Practice". *Annual Review of Economics (Forthcoming)*.

Banerjee, A. V., Gertler, P. J., and Ghatak, M. (2002). "Empowerment and Efficiency: Tenancy Reform in West Bengal". *Journal of Political Economy*, 110(2):239–280.

Bazzi, S., Fiszbein, M., and Gebresilasse, M. (2018). "Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States". Working paper.

Besley, T. (1995). "Property Rights and Investment Incentives: Theory and Evidence from Ghana". *Journal of Political Economy*, 103(5):903–937.

Besley, T. and Ghatak, M. (2010). "Property Rights and Economic Development". In *Handbook of Development Economics*, volume 5, pages 4525–4595.

Bester, A. C., Conley, T. G., and Hansen, C. B. (2011). "Inference with Dependent Data using Cluster Covariance Estimators". *Journal of Econometrics*, 165(2):137–151.

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

Bureau of Land Management and General Land Office. The Official Federal Land Records Site. https://glorecords.blm.gov/.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs". *Econometrica*, 82:2295–2326.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). "Bootstrap-based Improvements for Inference with Clustered Errors". *Review of Economics and Statistics*, 90(3):414–427.

Carter, C. E. (2019). "Forts and the Frontier: The US Army and the Spatial Distribution of Population". Working Paper.

Chetty, R. and Hendren, N. (2018). "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates". *Quarterly Journal of Economics*, 133(3):1163–1228.

Cogswell, S. (1975). *Tenure, Nativity, and Age as Factors in Iowa Agriculture, 1850-1880*. Iowa State University Press, Ames, 1st edition.

Conley, T. (1999). "GMM Estimation with Cross Sectional Dependence". *Journal of Econometrics*, 92:1–45.

Danhof, C. H. (1941). "Farm-Making Costs and the "Safety Valve": 1850-60". *Journal of Political Economy*, 49(3):317–359.

Dell, M. (2010). "The Persistent Effects of Peru's Mining Mita". *Econometrica*, 78(6):1863–1903.

Dell, M. (2012). "Path Dependence in Development: Evidence from the Mexican Revolution". Working Paper.

Dell, M., Lane, N., and Querubin, P. (2019). "The Historical State, Local Collective Action, and Economic Development in Vietnam". *Econometrica (Forthcoming)*.

Dell, M. and Olken, B. A. (2019). "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java". *Review of Economic Studies*.

Deverell, W. F. (1988). "To Loosen the Safety Valve: Eastern Workers and Western Lands". *Western Historical Quarterly*, 19(3):269–285.

Donaldson, D. and Hornbeck, R. (2016). "Railroads and American Economic Growth: A "Market Access" Approach". *Quarterly Journal of Economics*, 131(2):799–858.

Donaldson, T. (1884). *The Public Domain; Its History, with Statistics*. Government Printing Office, Washington, DC, 3 edition.

Dykstra, R. R. and Manfra, J. A. (2017). *Dodge City and the Birth of the Wild West*. University Press of Kansas.

Edwards, R., Friefeld, J. K., and Wingo, R. S. (2017). *Homesteading the Plains: Toward a New History*. University of Nebraska Press, Lincoln, Nebraska.

Engerman, S. L. and Sokoloff, K. L. (2002). "Factor Endowments, Inequality, and Paths of Development among New World Economics. NBER Working Paper No. 9259.

Engerman, S. L. and Sokoloff, K. L. (2012). *Economic Development in the Americas since 1500: Endowments and Institutions*. Cambridge University Press.

Enke, B. (2019). "Moral Values and Voting". Working Paper.

Field, E. (2007). "Entitled to Work: Urban Property Rights and Labor Supply in Peru". *Quarterly Journal of Economics*, 122(4):1561–1602.

Fogel, R. W. (1964). *Railroads and American Economic Growth*. Johns Hopkins Press Baltimore.

Galiani, S. and Schargrodsky, E. (2010). "Property Rights for the Poor: Effects of Land Titling". *Journal of Public Economics*, 94(9-10):700–729.

Gallman, R. E., Davis, L. E., Easterlin, R., and Parker, W. (1972). American Economic Growth: An Economist's History of the United States.

García-Jimeno, C. and Robinson, J. A. (2008). "The Myth of the Frontier". In *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, pages 49–88. University of Chicago Press.

Gates, P. W. (1936). "The Homestead Law in an Incongruous Land System". *American Historical Review*, 41(4):652–681.

Gates, P. W. (1954). *Fifty Million Acres*. Cornell University Press, Ithaca, New York.

Gates, P. W. (1960). *The Farmer's Age: Agriculture 1815-1860*. The Economic History of the United States, Vol. 3. Holt, Rinehart and Winston, New York.

Gates, P. W. (1963). "The Homestead Act: Free Land Policy in Operation, 1862-1935". In Ottoson, H. W., editor, *Land Use Policy and Problems in the United States*. University of Nebraska Press, Lincoln, Nebraska.

Gates, P. W. (1970). Introduction. In Donaldson, T., editor, *The Public Domain : Its History with Statistics*. New York : Johnson Reprint Corp. Reprint of the 1884 ed.

Gates, P. W. (1973). *Landlords and Tenants on the Prairie Frontier*. Cornell University Press, Ithaca and London.

Gates, P. W. and United States (1968). *History of Public Land Law Development*. [For sale by the Supt. of Docs., U.S. Govt. Print. Off.], Washington.

Gelman, A. and Imbens, G. (2014). "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs". NBER Working Paper No. 20405.

Glaeser, E. L. and Gottlieb, J. D. (2009). "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States". *Journal of Economic Literature*, 47(4):983–1028.

Goldin, C. D. and Katz, L. F. (2008). *The Race Between Education and Technology*. Harvard University Press.

Gorodnichenko, Y. and Roland, G. (2011). "Which Dimensions of Culture Matter for Long-Run Growth?". *American Economic Review*, 101(3):492–98.

Hanlon, W. W. (2017). "Temporary Shocks and Persistent Effects in Urban Economies: Evidence from British Cities after the US Civil War". *Review of Economics and Statistics*, 99(1):67–79.

Hornbeck, R. (2010). "Barbed wire: Property Rights and Agricultural Development". *Quarterly Journal of Economics*, 125(2):767–810.

Jedwab, R., Kerby, E., and Moradi, A. (2017). "History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya". *Economic Journal*, 127(603):1467–1494.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1998). "Law and Finance". *Journal of Political Economy*, 106(6):1113–1155.

LaMacchia, R., Marx, R., and Sobel, J. (1994). *Geographic Areas Reference Manual*. Washington.

Lee, L. B. (1979). *Kansas and the Homestead Act, 1862-1905*. Arno Press, New York.

Leonard, B. and Libecap, G. D. (2019). "Collective Action by Contract: Prior Appropriation and the

Development of Irrigation in the Western United States". *Journal of Law and Economics*, 62(1):67–115.

Libecap, G. D. and Lueck, D. (2011). "The Demarcation of Land and the Role of Coordinating Property Institutions". *Journal of Political Economy*, 119(3):426–467.

Lillo Bustos, N. A. (2018). "Land Inequality and Human Capital: Evidence for the United States from the Homestead Act". Working Paper.

Lowes, S. and Montero, E. (2018). "Concessions, Violence, and Indirect Rule: Evidence from the Congo Free State". Working paper.

Manson, S., Schroeder, J., Van Riper, D., and Ruggles, S. (2019). IPUMS National Historical Geographic Information System. Version 14.0 [Database].

Michaels, G. and Rauch, F. (2017). "Resetting the Urban Network: 117–2012". *Economic Journal*, 128(608):378–412.

Miner, H. C. (1982). *Wichita: the Early Years, 1865-80*, volume 1. University of Nebraska Press.

Montiel Olea, J. L. and Pflueger, C. (2013). "A Robust Test for Weak Instruments". *Journal of Business and Economic Statistics*, 31(3):358–368.

National Archives (2016). Teaching with Documents: The Homestead Act of 1862. https://www.archives.gov/education/lessons/homestead-act. Accessed: 2019-09-02.

North, D. C. (1974). *Growth and Welfare in the American Past*. Prentice-Hall, Englewood Cliffs, N.J.

Nunn, N. (2009). "The Importance of History for Economic Development". *Annual Review of Economics*, 1(1):65–92.

Nunn, N. and Wantchekon, L. (2011). "The Slave Trade and the Origins of Mistrust in Africa". *American Economic Review*, 101(7):3221–52.

Poulos, J. (2019). "State-Building through Public Land Disposal? An Application of Matrix Completion for Counterfactual Prediction". Working Paper.

Raz, I. T. (2018). "Use It Or Lose It: Adverse Possession and Economic Development". Working Paper.

Sequeira, S., Nunn, N., and Qian, N. (2019). "Immigrants and the Making of America". *Review of Economic Studies*.

Shanks, T. W. (2005). "The Homestead Act: A Major Asset- Building Policy in American History". In Sherraden, M., editor, *Inclusion in the American Dream: Assets, Poverty, and Public Policy*, chapter 2, pages 20–10. Oxford University Press, New York.

Shannon, F. A. (1945). *The Farmer's Last Frontier: Agriculture, 1860-1897*. J. J. Little and Ives Company, New York.

Sheldon, A. E. (1936). *Land Systems and Land Policies in Nebraska*. Nebraska State Historical Society, Lincoln, Nebraska.

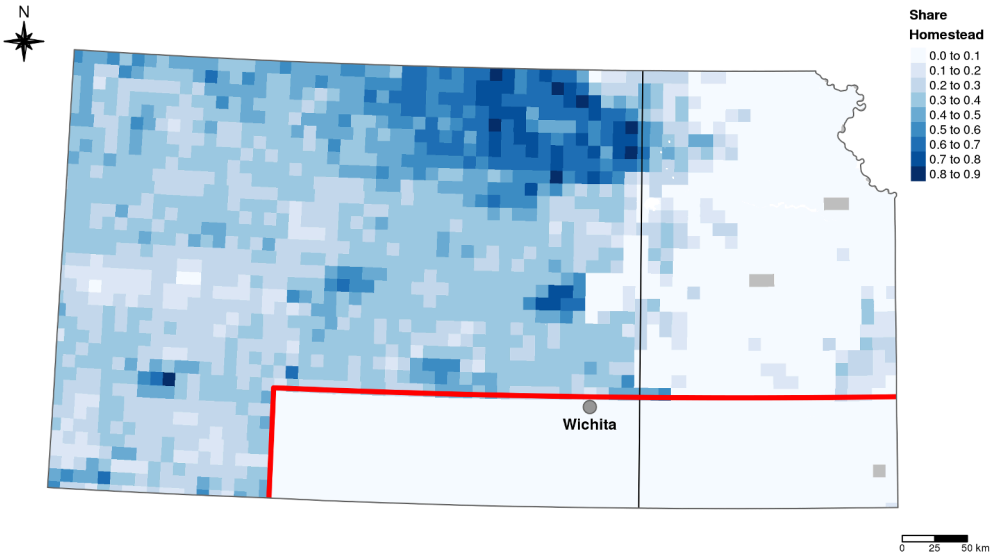
Smith, C. (2019). "Land Concentration and Long-Run Development: Evidence from the Frontier United States". Working Paper.

Swierenga, R. P. (1977). "Land Speculation and Its Impact on American Economic Growth and Welfare: A Historiographical Review". *Western Historical Quarterly*, 8(3):283–302.

United States, C. O. and Edmunds, J. M. (1866). *Statistics of the United States, (including mortality, property, c.,) in 1860: comp. from the original returns and being the final exhibit of the eighth census, under the direction of the Secretary of the Interior*. US Government Printing Office, Washington, DC.

Whaples, R. (1995). "Where Is There Consensus Among American Economic Historians? The Results of a Survey on Forty Propositions". *The Journal of Economic History*, 55(1):139–154.

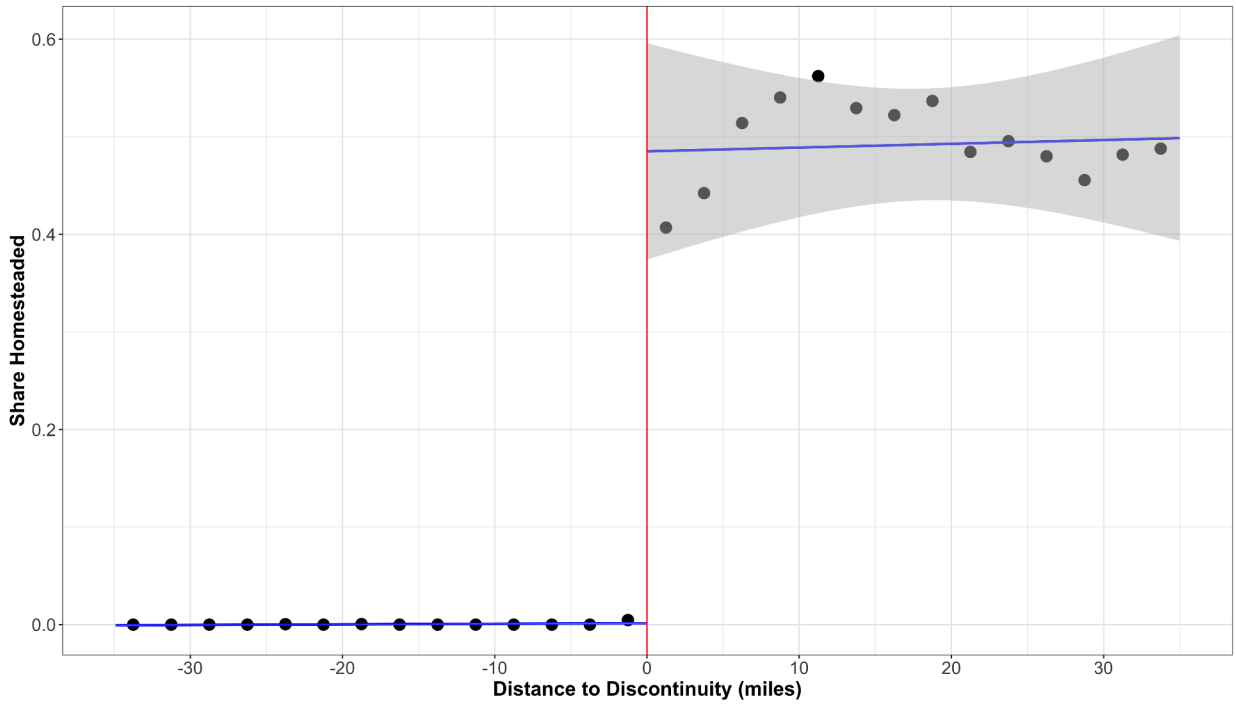
Figure 1: HOMESTEADING ASSIGNMENT AND INDIAN CESSIONS IN SOUTHERN KANSAS




*Note*: This figure plots the fraction of land in PLSS townships in Kansas that was privatized under the 1862 Homestead Act out of the total township land area. Darker color implies a higher share. The red line plots the boundaries of the land cessions of Native Americans in southern Kansas that prohibited homesteading. The black horizontal line is the 96.88 meridian. The gray dot indicates the center of Wichita.

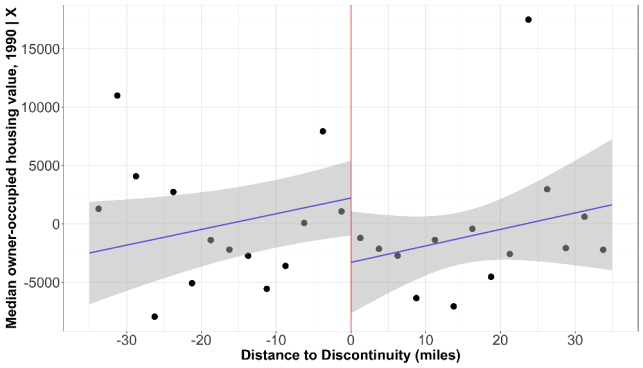
Figure 2: STANDARD RD PLOT FOR SHARE HOMESTEADED

(THE FIRST STAGE)

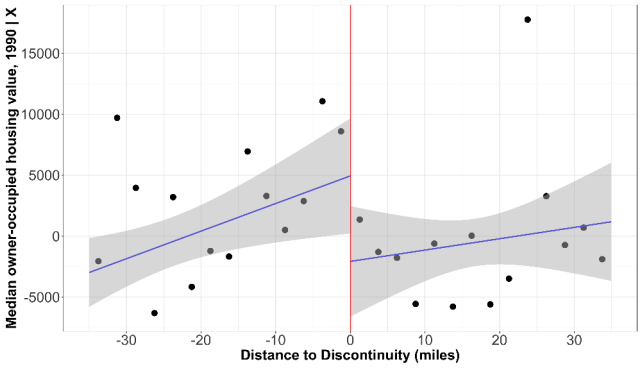



*Note*: This figure plots standard RD plot using linear distance from the boundary as the running variable. Points represent the average value within a 2.5-mile bin. The blue line plots a local linear regression. The gray area shows the 95% confidence interval of the predicted value, with standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011).

Figure 3: STANDARD RD PLOTS FOR MEDIAN HOUSING VALUE



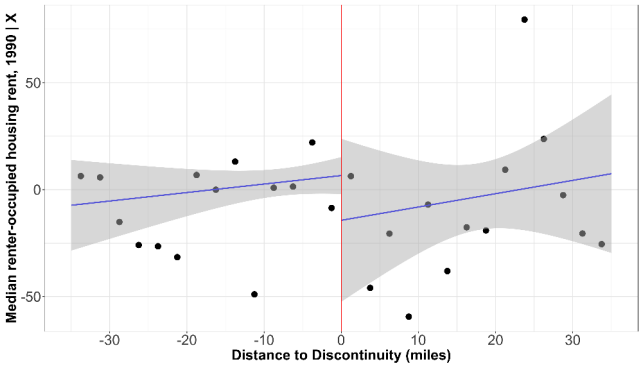

(a) Residual Median Housing Value



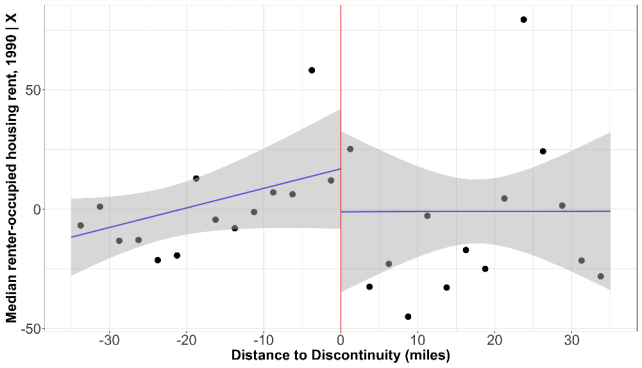

(b) Residual Median Housing Value Drop Wichita

*Note*: This figure plots standard RD plot using linear distance from the boundary as the running variable. The level of observation is 1990 census blocks. Points represent the average value within a 2.5-mile bin. The blue line plots a local linear regression. The gray area shows the 95% confidence interval of the predicted value, with standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011).

Figure 4: STANDARD RD PLOTS FOR MEDIAN RENT



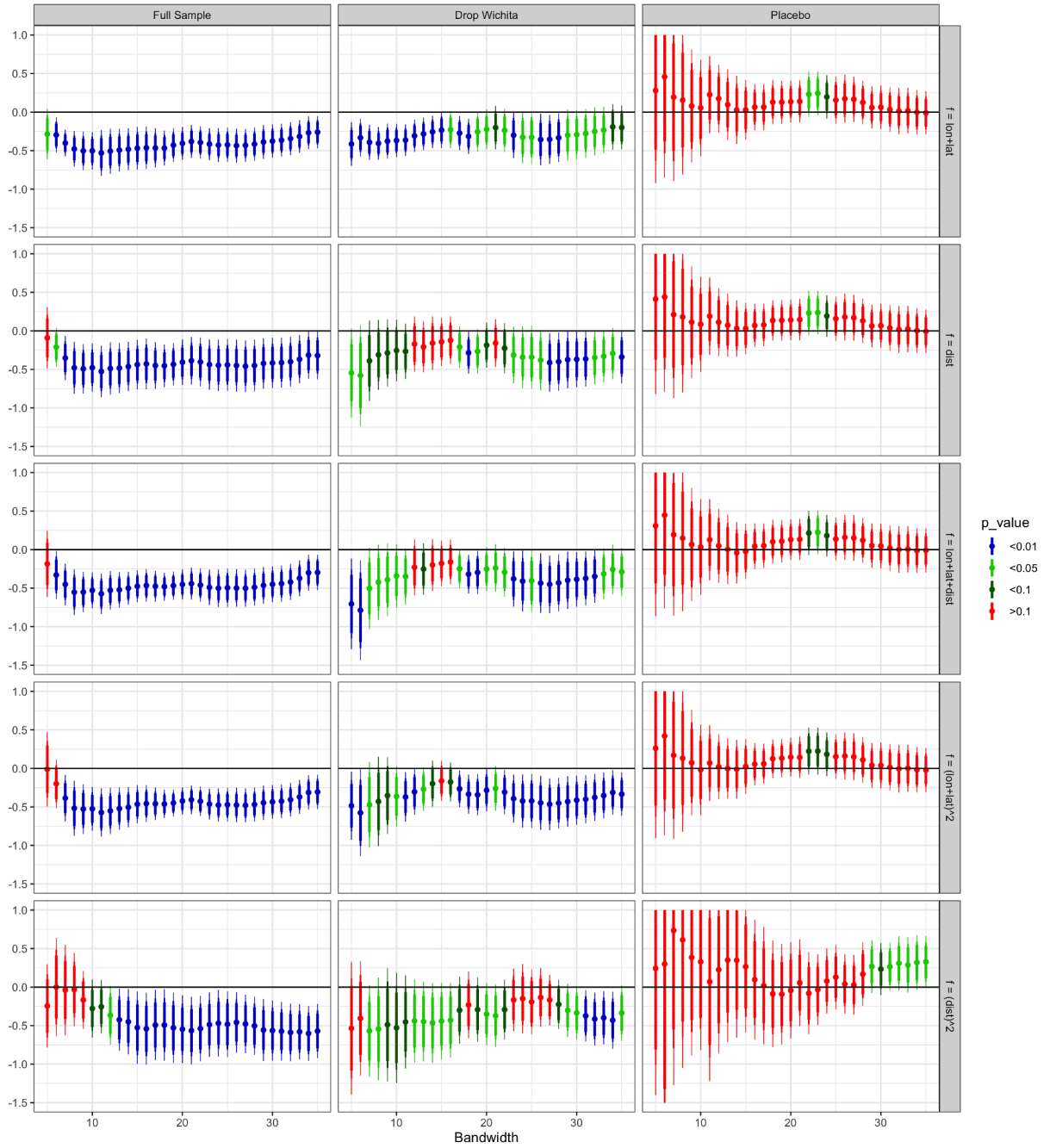

(a) Residual Median Rent




(b) Residual Median Rent Drop Wichita

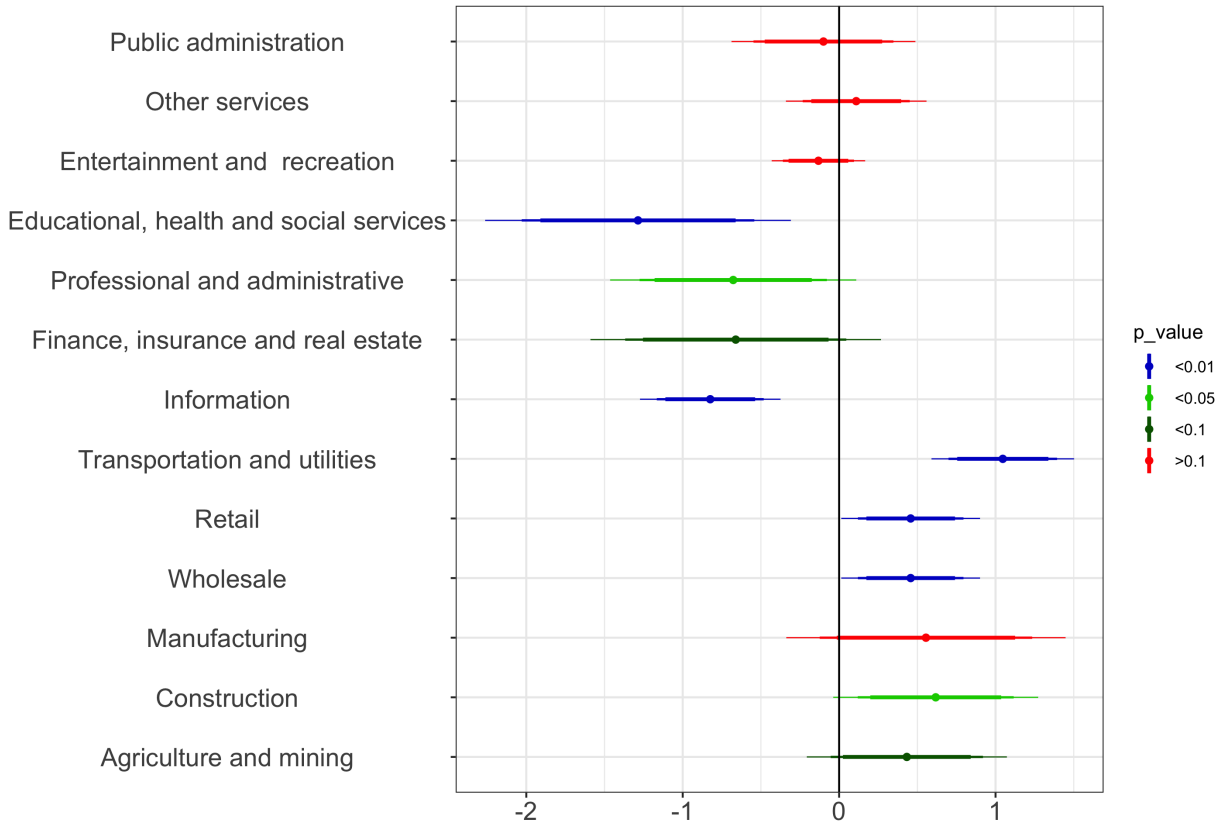
*Note*: This figure plots standard RD plot using linear distance from the boundary as the running variable. The level of observation is 1990 census blocks. Points represent the average value within a 2.5-mile bin. The blue line plots a local linear regression. The gray area shows the 95% confidence interval of the predicted value, with standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011).

Figure 5: RD: HOUSING VALUE - BANDWIDTH, RD POLYNOMIAL AND SAMPLE ROBUSTNESS



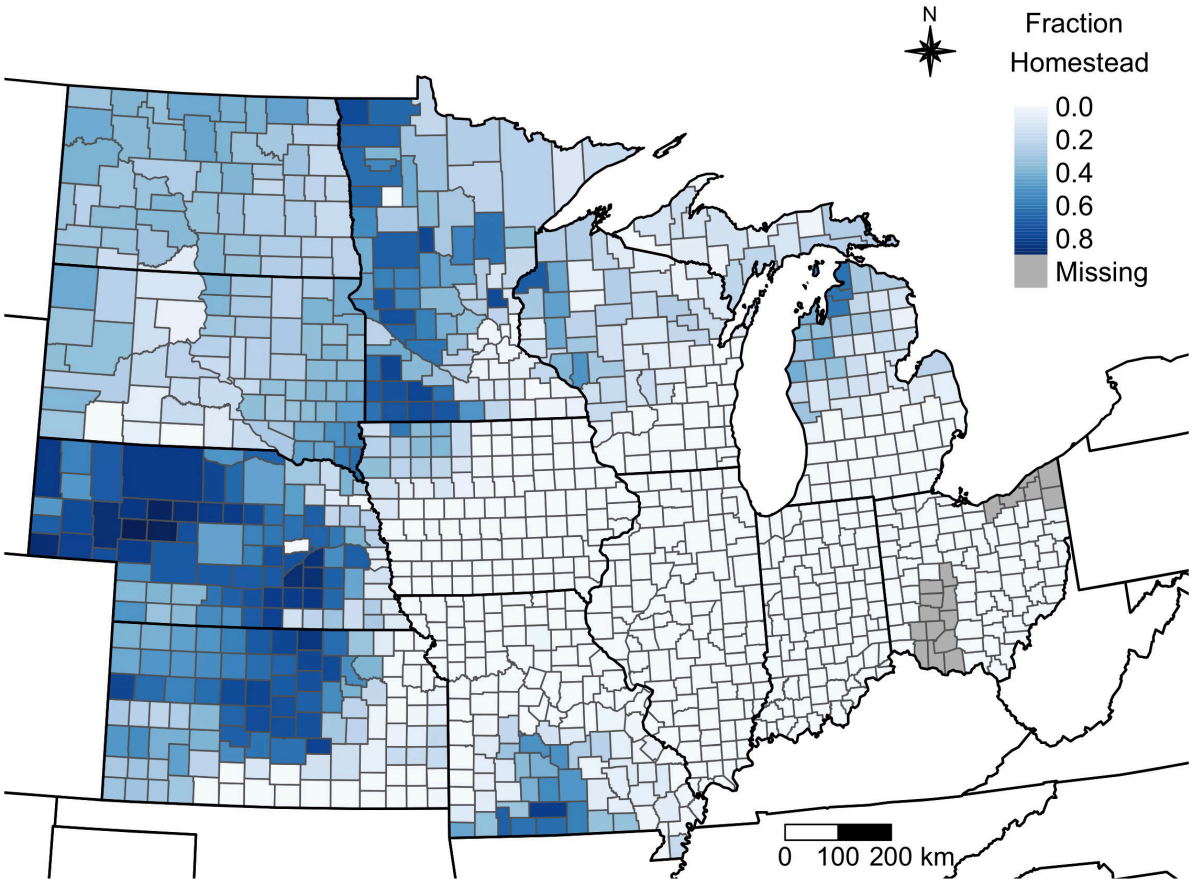

*Note*: This figure plots the RD estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1), for every combination of bandwidth values between $5-35$ miles, common specifications for the RD polynomial, and three different samples. Each facet represents a specific combination of sample and RD polynomial, and the x-axis within each facet denotes the different bandwidths. Each column studies a different sample: the full sample (column 1), excluding all blocks within 15 miles from Wichita's center (column 2), and the placebo sample (column 3). Each row uses a different RD polynomial: linear in latitude-longitude (row 1), linear in distance to the cession boundary (row 2), linear in both latitude-longitude and distance (row 3), quadratic in latitude-longitude (row 4) and quadratic in distance (row 5). Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011).

Figure 6: RD: EMPLOYMENT SHARES IN DIFFERENT INDUSTRY SECTORS



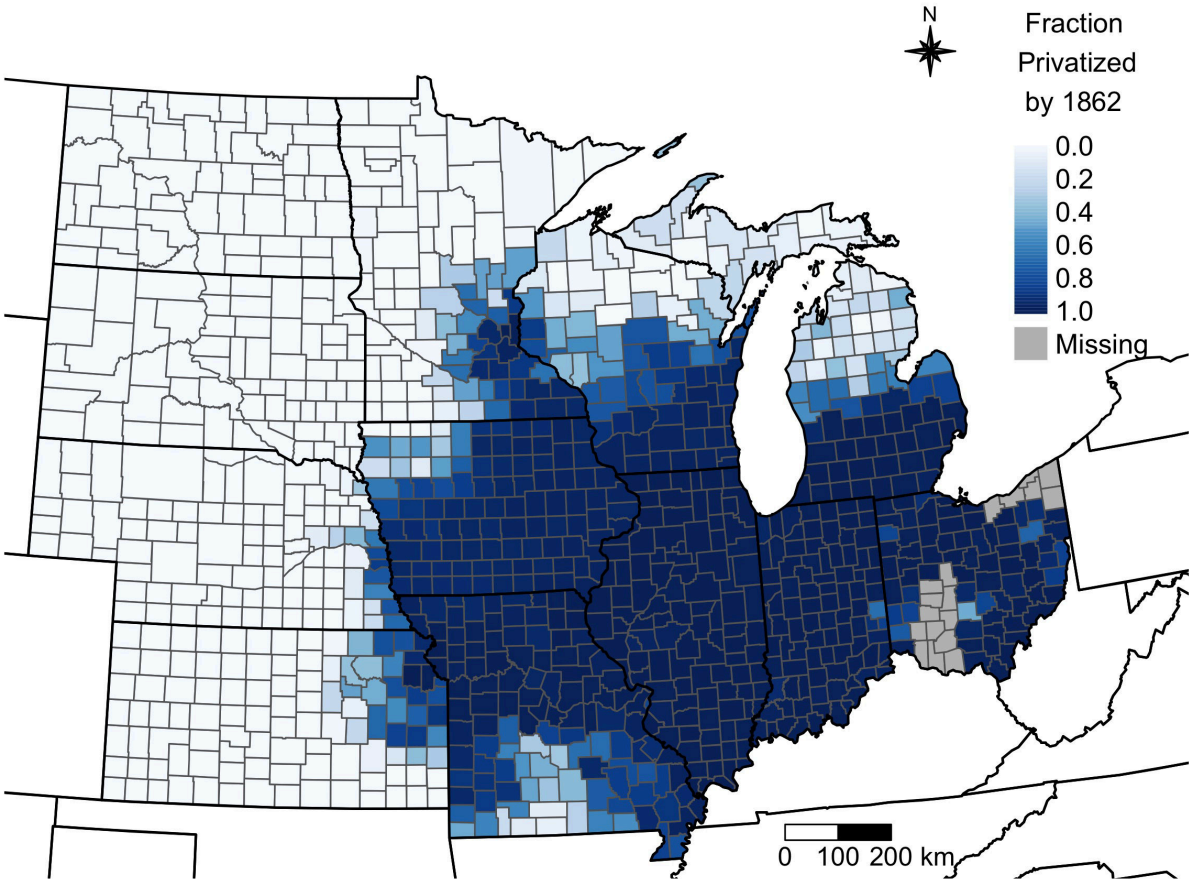

*Note*: This figure plots the RD estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1) when the dependent variable is the employment shares in different industry sectors, standardized into z-scores. The level of observation is 2000 census block-groups. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls.

Figure 7: FRACTION PRIVATIZED UNDER THE 1862 HOMESTEAD ACT



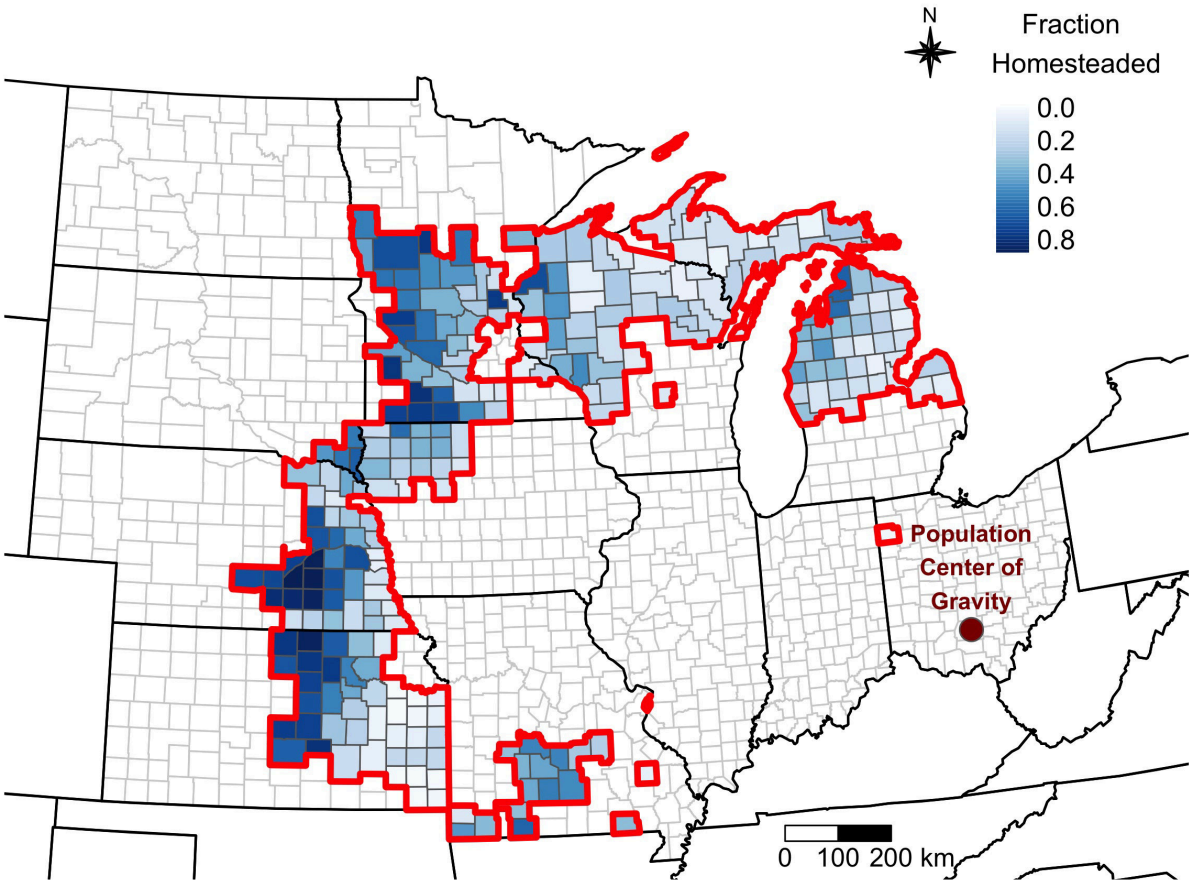

*Note*: This figure plots the fraction of land in a county that was patented under the 1862 Homestead Act out of total land patented. The observation level is counties. Darker color implies a higher share.

Figure 8: FRACTION PRIVATIZED PRIOR TO 1862



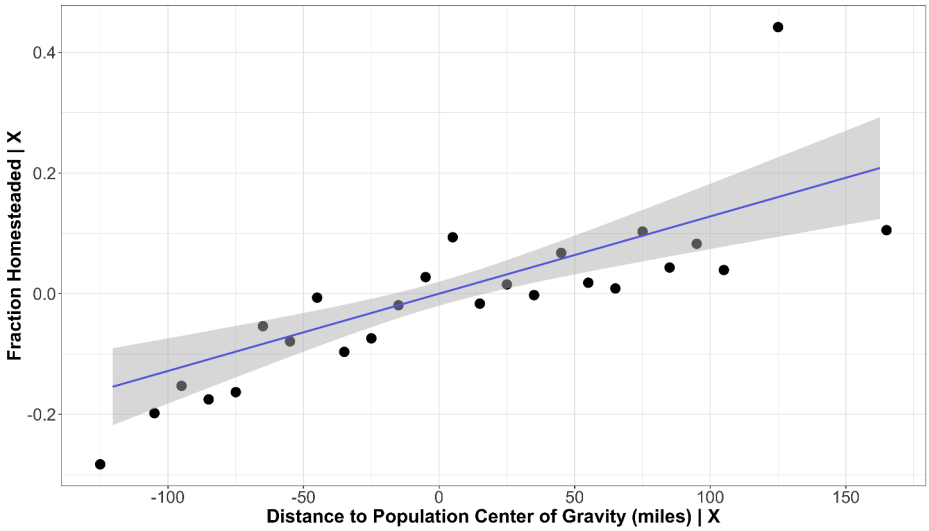

*Note*: This figure plots the fraction of land in a county that was patented prior to the passing of the 1862 Homestead Act out of total land patented. The observation level is counties. Darker color implies a higher share.

Figure 9: FRACTION PRIVATIZED UNDER THE 1862 HOMESTEAD ACT ALONG THE 1862 PRIVATIZATION FRONTIER

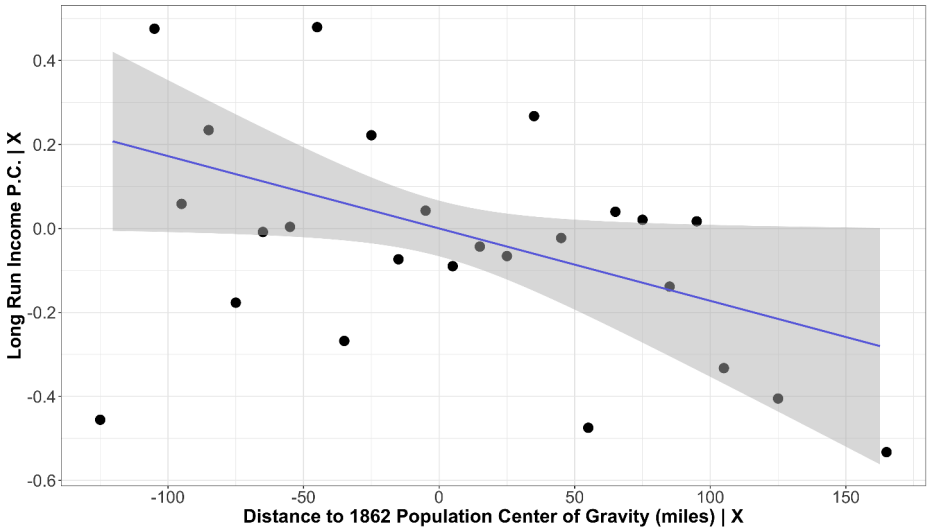



*Note*: This figure plots the fraction of land in a county that was distributed under the 1862 Homestead Act out of total land patented for counties along the 1862 privatization frontier. The observation level is counties. Darker color implies a higher share. The red dot indicates the location of 1862 the national population center of gravity.

Figure 10: IV: First Stage and Reduced Form Partial Correlation Plots



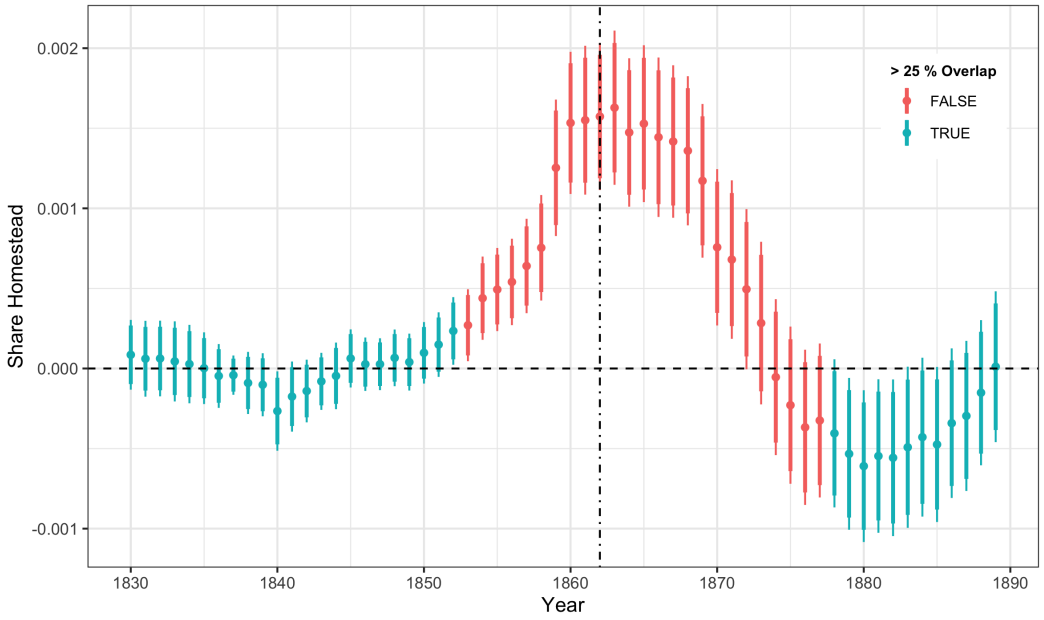

(a) First Stage



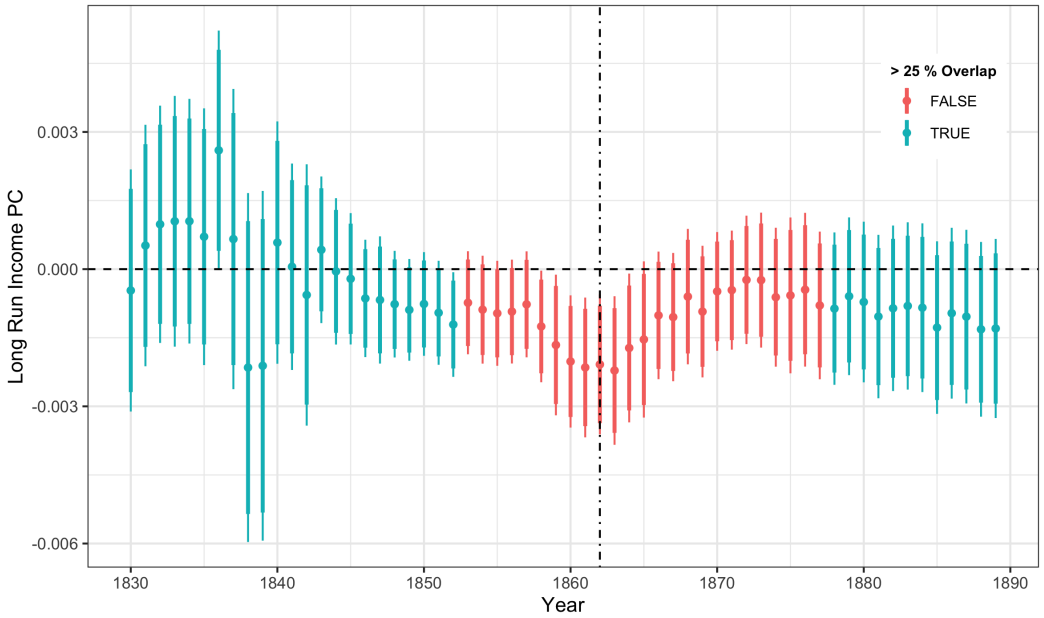

(b) Reduced form

*Note*: This figure plots partial correlation plots for the first stage (Panel a) and the reduced form relationship (Panel b). State fixed effects and baseline geo-climatic controls are partialled out. Observations are at the county level. The sample is the baseline 1862 privatization frontier. Points represent the average value within a 10 miles bin. The blue line plots a linear regression. The gray area shows the 95% confidence interval of the predicted value, using Conley (1999) standard errors with a 500-mile window.

Figure 11: IV: ESTIMATES FOR COUNTERFACTUAL YEARS OF ENACTMENT



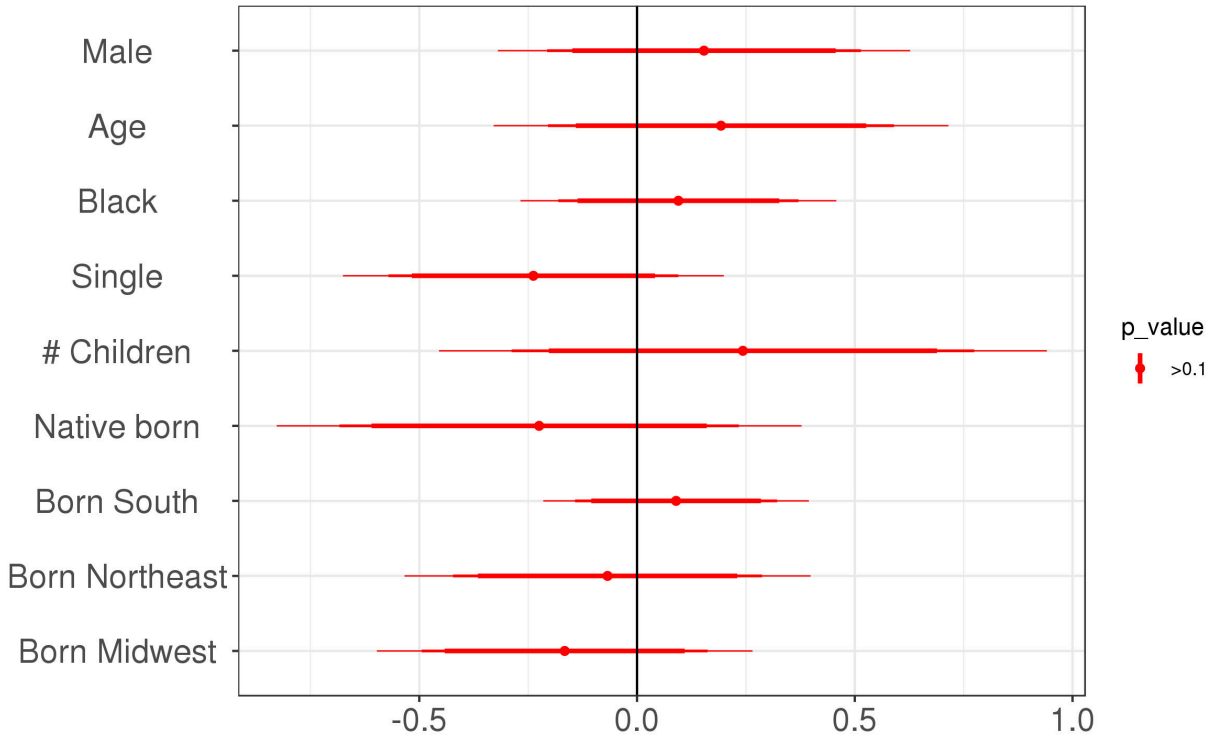

(a) First Stage




(b) Reduced form

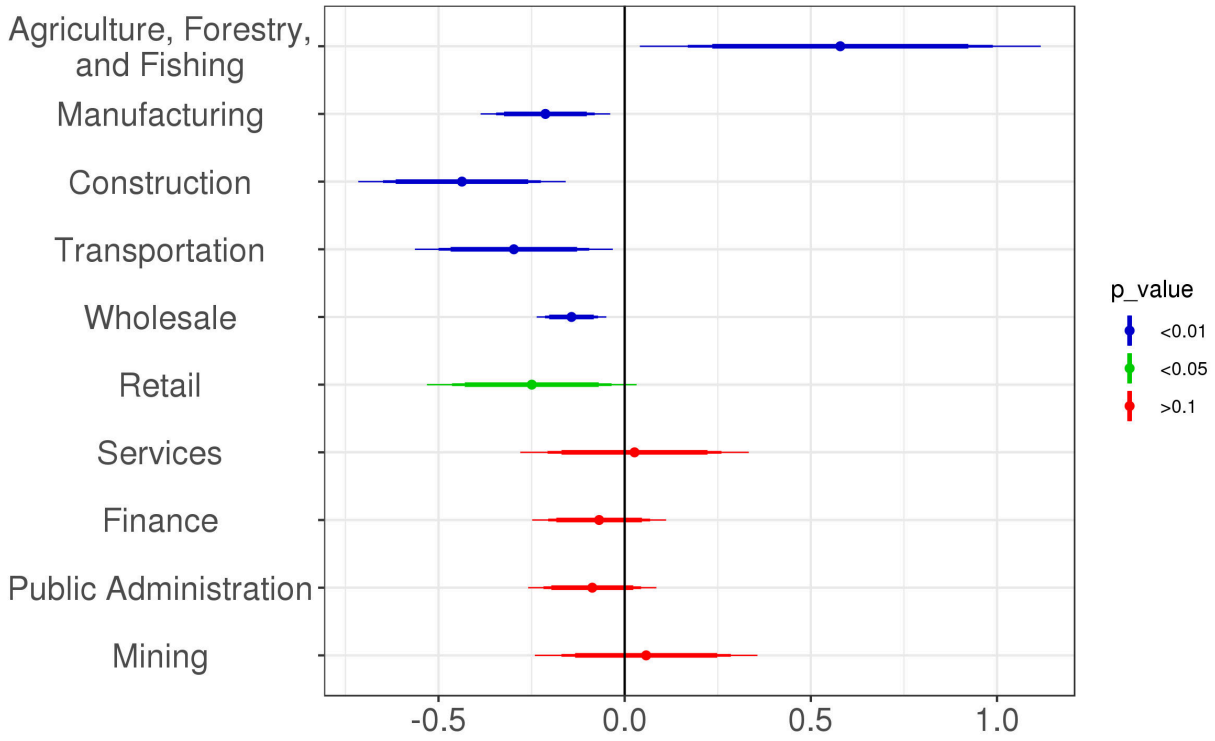
*Note*: This figure plots the point estimate and 99%, 95% and 90% confidence intervals of the first stage (Panel a) and reduced form (Panel b) from the placebo exercise which pools together all privatization years between 1830 and 1889. Estimates are partial effects of the distance from the contemporary population center of gravity, within sample counties on the contemporary privatization frontier. Regressions also control for state × year fixed effects, temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Standard errors are clustered at the county.

Figure 12: RD: OBSERVABLE DEMOGRAPHIC DURING EARLY SETTLEMENT



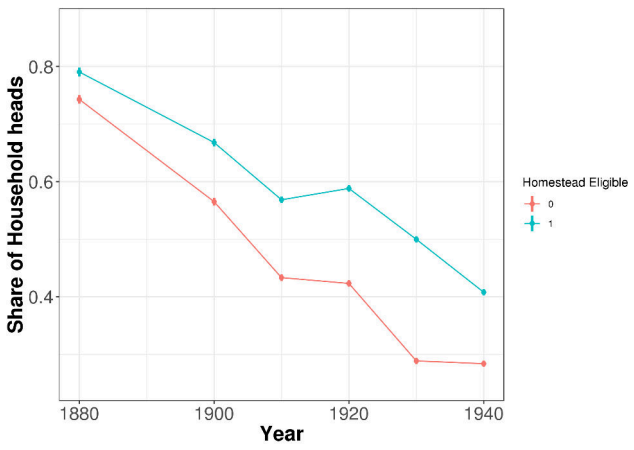

*Note*: This figure plots the RD estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1) when the dependent variables are observable demographics, standardized into z-scores. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from the 1880 full-count census.

Figure 13: RD: INDUSTRIES DURING EARLY SETTLEMENT

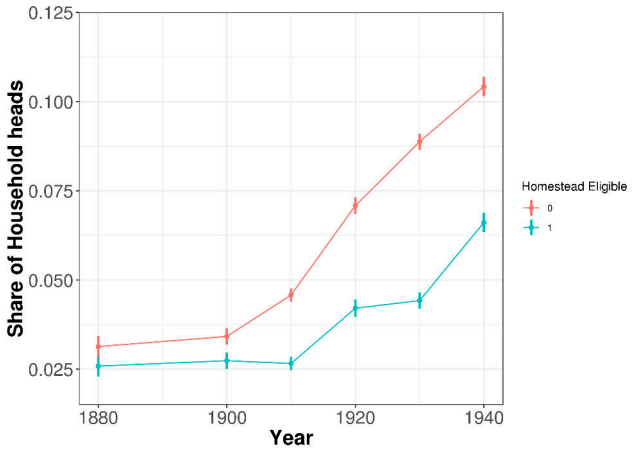



*Note*: This figure plots the RD estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1) when the dependent variables are dummy variables that take the value of 1 if the household head works in a given industry sector, standardized into z-scores. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from the 1880 full-count census.

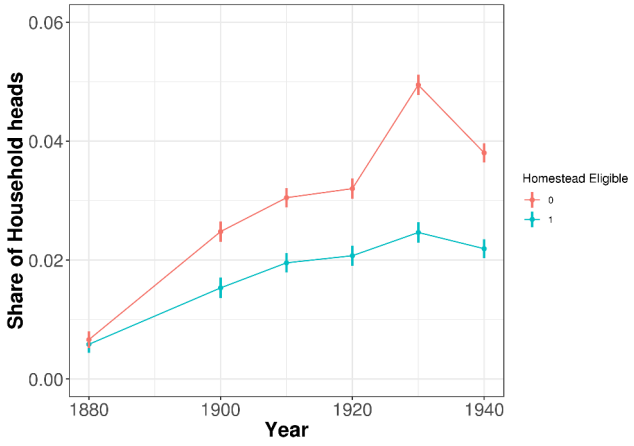
Figure 14: DIVERGENCE IN EMPLOYMENT SHARES ACROSS SECTORS, 1880-1940



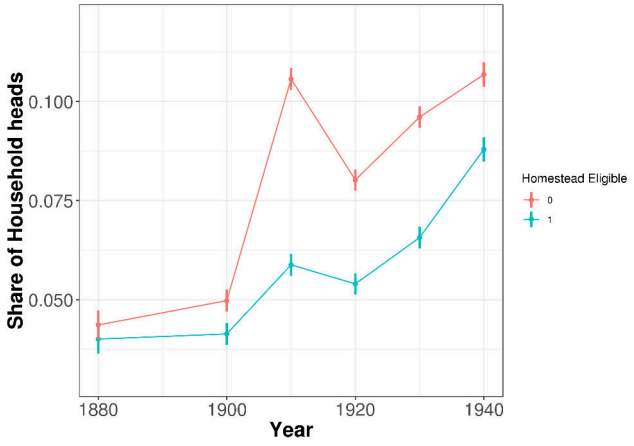

(a) Agricultural, forestry, and fishing




(b) Manufacturing



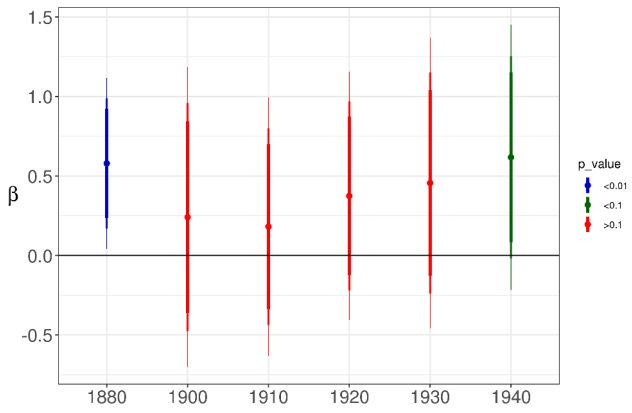

(c) Finance



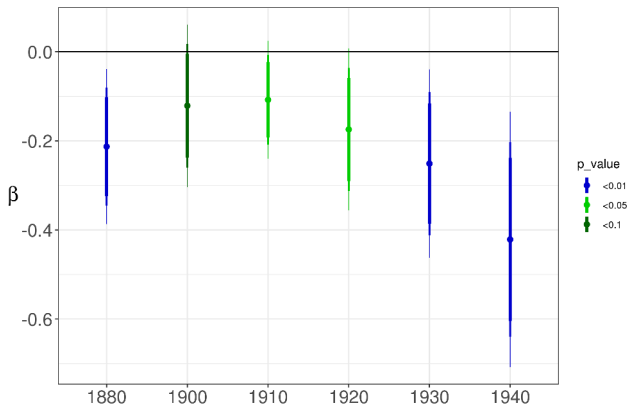

(d) Construction

*Note*: This figure plots the mean and 95% confidence intervals of the share of household heads that were employed in a given industry sector, for both sides of the historical Osage boundary, within a 35 miles bandwidth, for years 1880-1940. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from full count censuses, 1880-1940.

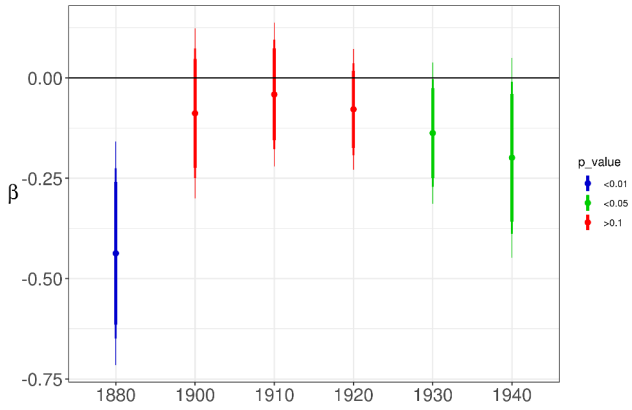
Figure 15: RD: EMPLOYMENT IN SELECTED INDUSTRIES, 1880-1940



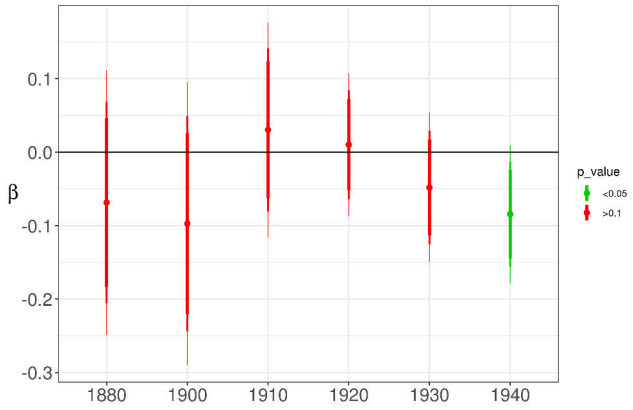

(a) Agricultural, forestry, and fishing




(b) Manufacturing



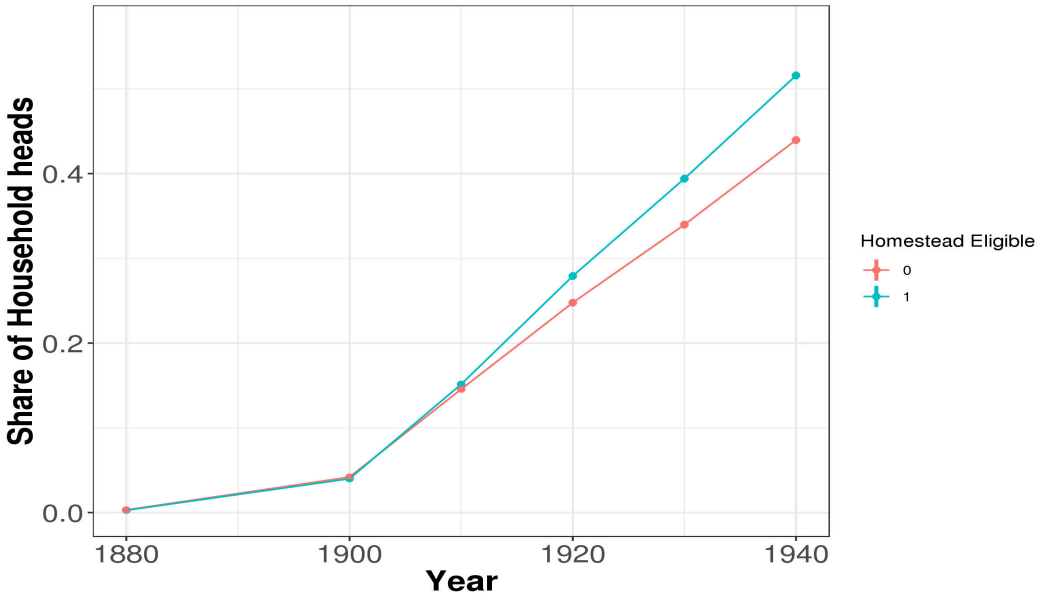

(c) Construction



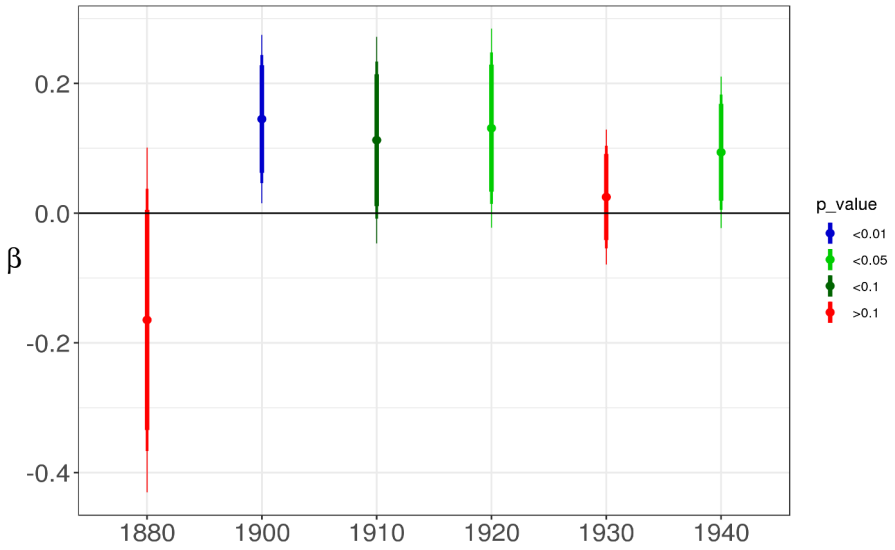

(d) Finance

*Note*: This figure plots the RD estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1) when the dependent variables are dummy variables that take the value of 1 if the household head works in a given industry sector, standardized into z-scores within a year. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from full-count censuses, 1880-1940.

Figure 16: HOUSEHOLD HEAD BORN IN KANSAS, 1880-1940



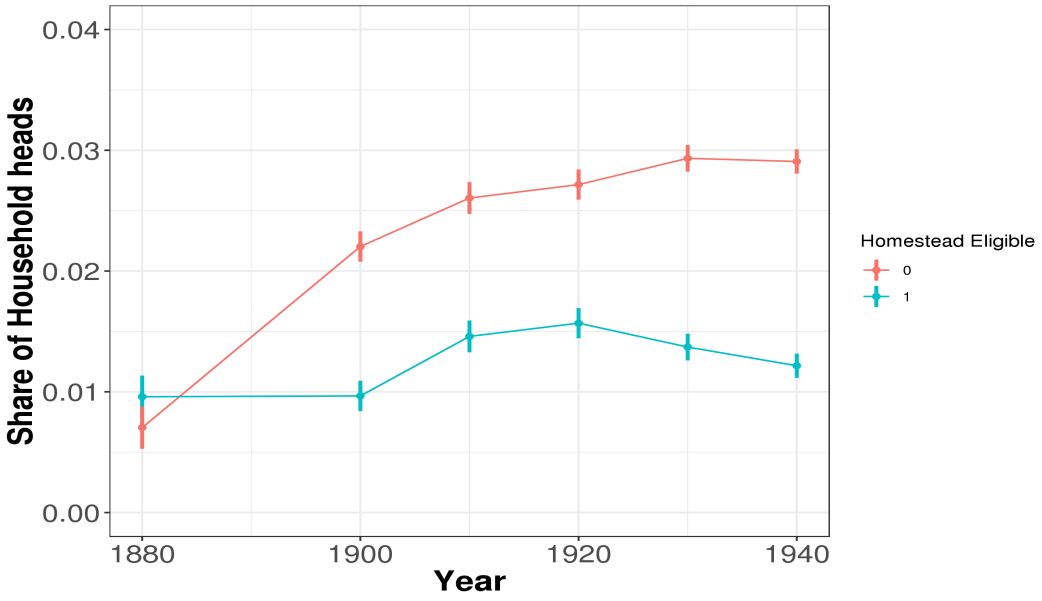

(a) Correlational Evidence



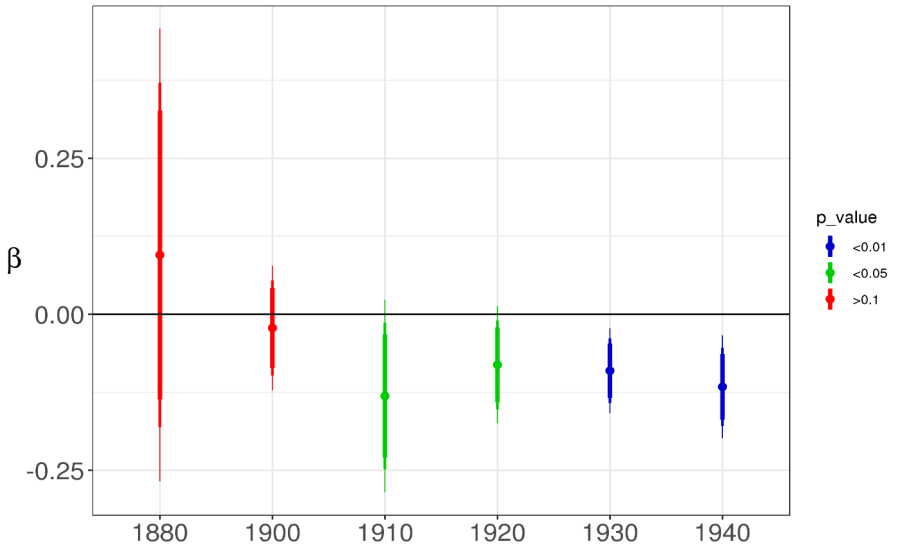

(b) RD Estimates

*Note*: Panel (a) plots the mean and 95% confidence intervals of the share of household heads that were born in Kansas, for both sides of the historical Osage boundary, within a 35-mile bandwidth, for the years 1880-1940. Panel (b) plots the RD estimate and 99%, 95%. and 90% confidence intervals of $\beta$ when the dependent variable is a dummy variable that takes the value of 1 if the household head was born in Kansas, standardized into z-score. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from full-count censuses, 1880-1940.

Figure 17: HOUSEHOLD HEAD BORN IS BLACK, 1880-1940



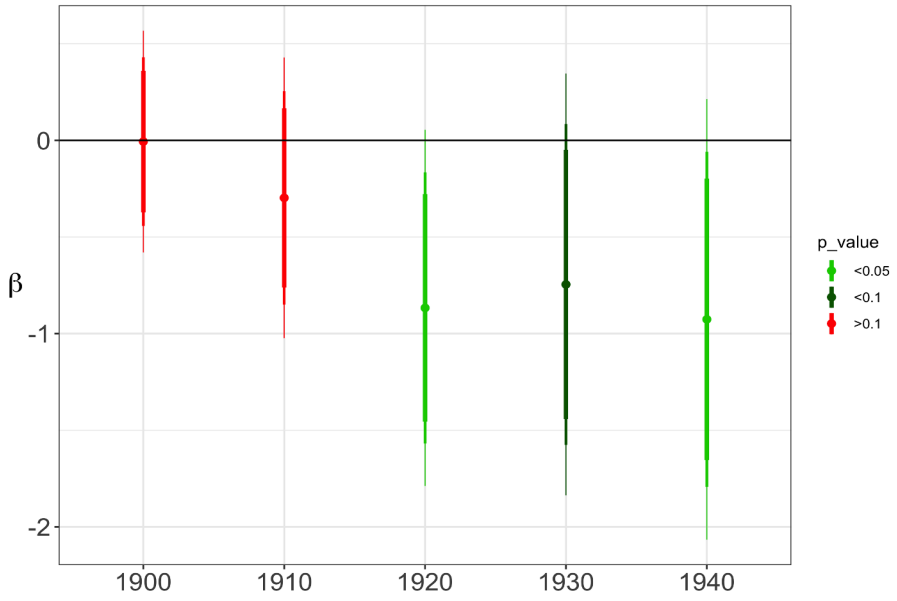

(a) Correlational Evidence



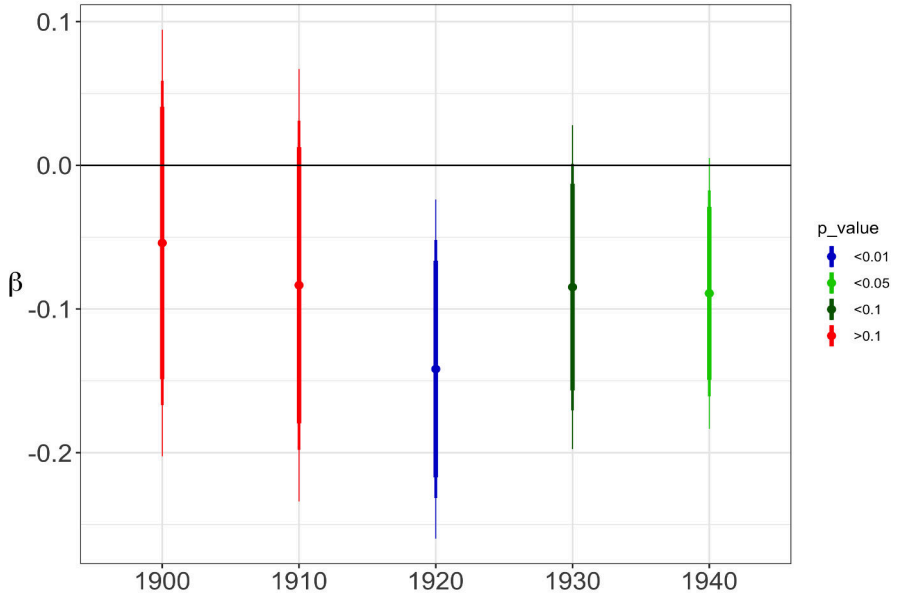

(b) RD Estimates

*Note*: Panel (a) plots the mean and 95% confidence intervals of the share of black household heads, in for both sides of the historical Osage boundary, within a 35-mile bandwidth, for the years 1880-1940. Panel (b) plots the RD estimate and 99%, 95% and 90% confidence intervals of $\beta$ when the dependent variable is dummy variable that take the value of 1 if the household head is black, standardized into z-score. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors are clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from full-count censuses, 1880-1940.

Figure 18: RD: ASSESSED VALUATION OF TOTAL PROPERTY, 1900-1940




(a) Logs




(b) Z-scores

*Note*: This figure plots the RD estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1) on the assessed valuation of total property in 1900-1940. Panel (a) uses a log transformation of the depended variable, and in Panel (b) the dependent variable is standardized into z-scores within a year. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The observation level is a city or a township. Data from biennial reports of the Kansas State Board of Agriculture.

Table 1: THE HOMESTEADER RELATIVE TO A LAND PURCHASER

| | Homesteader | Land Purchaser |
|---|---|---|
| Nature of privatization | A wealth transfer | A transaction |
| Size in Acres | 160 | 160 |
| Cost of Land | $18 | $ 200 |
| Constraints | On location and occupation choices, for five years | None |
| Property Rights | After five years | Right away |

*Note:* This table presents a conceptual comparison between the land privatization contracts that a homesteader and a land purchaser face.

Table 2: RD: BALANCE ON COVARIATES

| | *The dependent variable is:* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Year Titled | Temperature | Precipitation | Slope | Altitude | Flow Accumulation | Wheat | Maize | Soybean | Sorghum |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| | *Panel A: linear in latitude-longitude* | | | | | | | | | |
| **Outside Cession** | 0.154** | −0.070 | 0.016 | −0.043 | 0.112** | 0.053 | −0.039 | −0.005 | −0.037 | −0.119 |
| | (0.064) | (0.061) | (0.012) | (0.110) | (0.045) | (0.050) | (0.064) | (0.035) | (0.073) | (0.131) |
| | [0.066] | [0.067] | [0.017] | [0.107] | [0.035] | [0.074] | [0.069] | [0.046] | [0.083] | [0.082] |
| | *Panel B: linear in distance to the cession boundary* | | | | | | | | | |
| **Outside Cession** | 0.066 | 0.013 | −0.005 | −0.213* | 0.021 | −0.018 | 0.001 | −0.029 | −0.023 | 0.034 |
| | (0.095) | (0.075) | (0.022) | (0.112) | (0.024) | (0.054) | (0.093) | (0.063) | (0.119) | (0.112) |
| | [0.082] | [0.091] | [0.013] | [0.098] | [0.017] | [0.063] | [0.065] | [0.037] | [0.071] | [0.065] |
| | *Panel C: linear in latitude-longitude and distance* | | | | | | | | | |
| **Outside Cession** | 0.089 | −0.010 | 0.004 | −0.225** | 0.026 | −0.009 | 0.014 | −0.016 | −0.004 | 0.036 |
| | (0.088) | (0.070) | (0.011) | (0.109) | (0.020) | (0.053) | (0.082) | (0.052) | (0.106) | (0.101) |
| | [0.086] | [0.098] | [0.008] | [0.107] | [0.021] | [0.063] | [0.065] | [0.036] | [0.068] | [0.063] |
| Observations | 3,650 | 11,990 | 195 | 17,090 | 17,134 | 17,134 | 195 | 195 | 195 | 195 |
| Clusters | 37 | 40 | 32 | 39 | 40 | 40 | 32 | 32 | 32 | 32 |
| DV Mean | 1887 | 23.5 | 641 | 0.005 | 606 | 1,317 | 3.52 | 3.03 | 1.47 | 4 |
| DV SD | 7.48 | 1.42 | 92 | 0.003 | 139 | 12,634 | 0.977 | 1.45 | 0.311 | 0.648 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to Metro | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Wichita Ctrls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. The bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Conley (1999) standard errors with a 50-mile window in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: RD: MEDIAN OWNER-OCCUPIED HOUSING VALUE, 1990

| | RD Polynomial is | | | Sample is | | | Controls dropped | | | Binary Treatment Assignment |
|---|---|---|---|---|---|---|---|---|---|---|
| | Lon Lat | Dist | Lon Lat & Dist | Drop Wichita | Drop Dodge | Donut Hole 2.5 miles | Boundary segment | Dist to Wichita | Wichita FEs | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| **Outside Cession** | −0.502*** | −0.477*** | −0.529*** | −0.369*** | −0.524*** | −0.652*** | −0.357*** | −0.501*** | −0.274 | −0.492*** |
| | (0.094) | (0.121) | (0.090) | (0.083) | (0.097) | (0.227) | (0.114) | (0.092) | (0.173) | (0.092) |
| | [0.115] | [0.123] | [0.094] | [0.058] | [0.106] | [0.245] | [0.139] | [0.114] | [0.077] | [0.115] |
| Observations | 8,233 | 8,233 | 8,233 | 2,737 | 7,684 | 6,488 | 8,233 | 8,233 | 8,233 | 8,233 |
| Clusters | 37 | 37 | 37 | 36 | 37 | 37 | 37 | 37 | 37 | 37 |
| DV Mean | 56,770 | 56,770 | 56,770 | 46,642 | 57,177 | 56,823 | 56,770 | 56,770 | 56,770 | 56,770 |
| DV SD | 36,308 | 36,308 | 36,308 | 27,215 | 36,885 | 37,556 | 36,308 | 36,308 | 36,308 | 36,308 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Dist to Wichita | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes |
| Wichita FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. The bandwidth is 10 miles. The level of observation is 1990 census blocks. The RD polynomial is linear in latitude and longitude in columns (1) and (4)-(10), linear in distance to the boundary in column (2), and linear in latitude and longitude and distance to the boundary in column (3). Baseline specification in columns (1)-(6) and (10) also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. Column (7) drops the boundary segment fixed effects, column (8) drops the linear control for distance to Wichita, and column (9) drops the Wichita fixed effects. Column (4) drops all data within 15 miles of Whichta's center, column (5) drops all data within 5 miles of Dodge City center, and column (6) drops all data within 2.5 miles of the boundary. The specification in column (10) replaces the continuous treatment assignment with a binary one. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Conley (1999) standard errors with a 50-mile window in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: RD: LONG-RUN LOCAL DEVELOPMENT

| Level of observation: | Blocks, 1990 | | Block groups, 2000 | | |
|---|---|---|---|---|---|
| Dependent variable: | Median housing value | Median housing rent | Income per capita | Median household income | Share bachelor degree |
| | (1) | (2) | (3) | (4) | (5) |
| **Outside Cession** | −0.502*** | −0.564*** | −0.845*** | −0.844*** | −1.609*** |
| | (0.094) | (0.126) | (0.324) | (0.175) | (0.355) |
| | [0.115] | [0.093] | [0.232] | [0.151] | [0.314] |
| | | | p < 0.01 | p < 0.01 | p < 0.01 |
| Observations | 8,233 | 6,102 | 383 | 383 | 383 |
| Clusters | 37 | 35 | 21 | 21 | 21 |
| DV Mean | 56,770 | 324 | 19,552 | 41,019 | 0.150 |
| DV SD | 36,308 | 144 | 8,477 | 17,240 | 0.102 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes |
| Dist to Wichita | Yes | Yes | Yes | Yes | Yes |
| Wichita FEs | Yes | Yes | Yes | Yes | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. The level of observation is 1990 census blocks in columns 1-2, and 2000 census block-groups in columns 3-5. The RD polynomial is linear in latitude and longitude. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. The bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Wild bootstrap p-values are also reported in column (3)-(5) due to the smaller number of spatial clusters. Conley (1999) standard errors with a 50-mile window in brackets. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 5: RD: REAL PROPERTY VALUE, URBAN AREAS, 1990

| The RD Polynomial is: | Lon Lat | Dist | Lon Lat & Dist |
|---|---|---|---|
| | (1) | (2) | (3) |
| | *Panel A: Median owner-occupied housing value* | | |
| **Outside Cession** | −0.817** | −0.758** | −0.947* |
| | (0.324) | (0.303) | (0.501) |
| | [0.112] | [0.166] | [0.272] |
| | p < 0.01 | p < 0.01 | p = 0.05 |
| Observations | 5,874 | 5,874 | 5,874 |
| Clusters | 10 | 10 | 10 |
| DV Mean | 56,692 | 56,692 | 56,692 |
| DV SD | 35,816 | 35,816 | 35,816 |
| | *Panel B: Median renter-occupied housing rent* | | |
| **Outside Cession** | −0.866*** | −0.860*** | −0.934*** |
| | (0.185) | (0.126) | (0.238) |
| | [0.122] | [0.151] | [0.258] |
| | p < 0.01 | p < 0.01 | p < 0.01 |
| Observations | 5,020 | 5,020 | 5,020 |
| Clusters | 10 | 10 | 10 |
| DV Mean | 338 | 338 | 338 |
| DV SD | 141 | 141 | 141 |
| Boundary FEs | Yes | Yes | Yes |
| Dist to Wichita | Yes | Yes | Yes |
| Wichita FEs | Yes | Yes | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. The level of observation is 1990 census blocks. The sample only includes census blocks within urban areas. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Wild bootstrap p-values are also reported due to the small number of spatial clusters. Conley (1999) standard errors clustered with a 50-mile window in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: IV: LONG-RUN INCOME PER CAPITA

| | OLS | First Stage | Reduced Form | IV |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Fraction Homestead** | 0.207 | | | −1.345** |
| | (0.311) | | | (0.606) |
| **Dist PCG**$^{1862}$ | | 0.128*** | −0.172*** | |
| | | (0.016) | (0.063) | |
| Observations | 247 | 247 | 247 | 247 |
| DV Mean | 12,125 | 0.322 | 12,125 | 12,125 |
| DV SD | 1,603 | 0.236 | 1,603 | 1,603 |
| Effective F-stat | | 65.503 | | |
| State FEs | Yes | Yes | Yes | Yes |
| Geoclimatic controls | Yes | Yes | Yes | Yes |

*Note:* The dependent variable in columns (1), (3), and (4) aggregates income per capita over 1969-2000 by first standardizing contemporaneous data into z-scores within each given year and then taking the mean. Observations are at the county level. The sample is the baseline 1862 privatization frontier. Geo-climatic controls include temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Conley (1999) standard errors with a 500-mile window in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: IV: LONG-RUN EARNINGS, PRODUCTIVITY, AND EMPLOYMENT SHARES BY SECTORS

| | Earnings Per Capita (1) | Earnings Per Worker (2) | Employment Share (3) |
|---|---|---|---|
| | *Panel A: Farming* | | |
| **Fraction Homestead** | 0.137 | −0.094 | 0.107 |
| | (0.381) | (0.656) | (0.557) |
| DV Mean | 726 | 9,813 | 0.140 |
| DV SD | 783 | 6,391 | 0.091 |
| | *Panel A: Non-Farming* | | |
| **Fraction Homestead** | −1.356** | −1.321*** | −0.107 |
| | (0.579) | (0.450) | (0.557) |
| DV Mean | 6,766 | 15,412 | 0.860 |
| DV SD | 2,440 | 2,355 | 0.091 |
| Observations | 247 | 247 | 247 |
| State FEs | Yes | Yes | Yes |
| Geoclimatic controls | Yes | Yes | Yes |

*Note:* Dependent variables aggregate data over 1969-2000 by first standardizing contemporaneous data into z-scores within each given year and then taking the mean. Observations are at the county level. The sample is the baseline 1862 privatization frontier. Geo-climatic controls include temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Conley (1999) standard errors with a 500-mile window in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: IV: LONG-RUN URBANIZATION AND GEOGRAPHICAL MOBILITY

| | Dependent variable is: | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Population | Migration | | Share with Tenure in Current Unit | | | | |
| | Density | Inflows | Outflows | 0-1 | 2-5 | 6-10 | 11-20 | 20+ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Fraction Homestead** | $-3.742^{***}$ | $-1.712^{***}$ | $-1.493^{***}$ | 0.562 | $-2.482^{***}$ | $-0.932^{*}$ | $-0.665$ | $2.363^{**}$ |
| | (0.705) | (0.658) | (0.405) | (0.417) | (0.854) | (0.492) | (0.599) | (0.923) |
| Observations | 247 | 247 | 247 | 247 | 247 | 247 | 247 | 247 |
| DV Mean | 80 | 0.026 | 0.025 | 0.205 | 0.260 | 0.166 | 0.189 | 0.225 |
| DV SD | 375 | 0.016 | 0.012 | 0.025 | 0.031 | 0.014 | 0.019 | 0.057 |
| State FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Geoclimatic controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

*Note:* Observations are at the county level. The sample is the baseline 1862 privatization frontier. Geo-climatic controls include temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Conley (1999) standard errors with a 500-mile window in parenthesis. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 9: RD: HOUSING VALUES, RENTS, AND WAGES, 1930-1940

| | Housing Value | | Rent | | Wage |
|---|---|---|---|---|---|
| | 1930 | 1940 | 1930 | 1940 | 1940 |
| | (1) | (2) | (3) | (4) | (5) |
| **Outside Cession** | −0.372*** | −0.277** | −0.074 | −0.239 | −0.349*** |
| | (0.056) | (0.112) | (0.063) | (0.201) | (0.126) |
| | [0.059] | [0.123] | [0.058] | [0.181] | [0.124] |
| | p < 0.01 | p = 0.01 | p = 0.22 | p = 0.23 | p < 0.01 |
| Observations | 19,673 | 18,269 | 19,638 | 18,439 | 20,585 |
| Clusters | 26 | 27 | 27 | 27 | 28 |
| Observations | 19,673 | 18,269 | 19,638 | 18,439 | 20,585 |
| Clusters | 26 | 27 | 27 | 27 | 28 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes |
| Distance to Wichita | Yes | Yes | Yes | Yes | Yes |
| Wichita Ctrl | Yes | Yes | Yes | Yes | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. The level of observation is a household head, geo-referenced to a city, town, or civil township. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and a Wichita fixed effect. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Wild bootstrap p-values are also reported due to the small number of spatial clusters. Conley (1999) standard errors with a 50-mile window in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

# Appendix A Robustness Tables and Figures

## A.1 RD Main Result

### A.1.1 RD - Median Renter-Occupied Housing Rent, 1990

Table A.1 reports estimates of $\beta$ from equation (1) when the dependent variable is median rent. The results are similar to the results we find when the dependent variable is median housing value (Table 3), suggesting that homestead-eligible locations outside the Osage cessions have significantly lower property values compared to locations in which homesteading was prohibited.

### A.1.2 RD - Placebo Exercise using the Eastern Portion of the Boundary

Table A.2 reports estimates of $\beta$ from equation (1) using the placebo region in the east. The dependent variable in Panel A is median housing value, while in Panel B it is median rent. Columns (1)-(2) uses a linear polynomial in latitude and longitude, columns (3)-(4) uses a liner polynomial in distance to the cession boundary, and columns (5)-(6) uses a a combination of both. We report results for both our baseline bandwidth of 10 miles and a larger bandwidth of 15 miles. Across all specification and dependent variables the estimate is insignificant.

### A.1.3 RD - Alternative Measurements of Long-Run Development

Table A.3 reports estimates of $\beta$ from equation (1) when the dependent variable is income per capita (columns 1-2), median household income (columns 3-4) and the share of population over age 25 with a bachelor degree (columns 5-6). Panel A uses a linear polynomial in latitude and longitude, Panel B uses a liner polynomial in distance to the cession boundary, and Panel C uses a combination of both. We report results for both our baseline bandwidth of 10 miles and a larger bandwidth of 15 miles. The results in table A.3 are consistent with the main results presented in table 3. Locations that experienced settlement under the 1862 Homestead Act have significantly lower levels of income and education attainment today.

### A.1.4 RD - Inference

Our baseline RD specification accounts for spatial auto-correlation by clustering observations at arbitrary 15 mile square grid-cells covering the research area as proposed by Bester et al. (2011). A possible concern is that this method and the choice of grid-cell size lead to poor coverage and overly

optimistic standard errors. To address this concern, figure A.1 plots the standard error of $\beta$ and its significance level under different grid-cell size (a) and different windows for Conley (1999) standard errors (b). In all cases, the estimate of $\beta$ maintains high statistical significance.

### A.1.5 Fuzzy RD

Table A.4 reports results from a fuzzy RD specification in which the Osage cession serves as an instrument for the share of land homesteaded.[44] Estimates in these regressions are interpreted as the effect of the share homesteaded on local long-run development. We find that a 10-percentage-point increase in the share of land homesteaded decreases median housing value by about 0.09 standard deviations (column 1), median rent by 0.11 standard deviations (column 2), income per capita and median household income by about 0.18 standard deviations (columns 3-4), and the share with a bachelor's degree by about 0.35 standard deviations (column 5). A disadvantage of the fuzzy RD specification is that it is not clear that a highly local variation in homestead exposure at the block or block-group level, as opposed to differential exposure of larger areas that may constitute different markets, should affect development in the long run. An important advantage of the fuzzy RD estimates is that their broader interpretation allows for a comparison across research designs, and specifically, a comparison with the IV estimates.

[44] At a fine level of spatial aggregation, some observations are lost in this specification due to unavailable BLM data regarding land privatization at that given location.

Table A.1: RD: MEDIAN RENTER-OCCUPIED HOUSING RENT, 1990

| | RD Polynomial is | | | Sample is | | | Controls dropped | | | Binary Treatment Assignment |
|---|---|---|---|---|---|---|---|---|---|---|
| | Lon Lat | Dist | Lon Lat & Dist | Drop Wichita | Drop Dodge | Donut Hole 2.5 miles | Boundary segment | Dist to Wichita | Wichita FEs | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| **Outside Cession** | −0.564*** | −0.384** | −0.438*** | −0.476*** | −0.526*** | −0.708*** | −0.466*** | −0.558*** | −0.221* | −0.540*** |
| | (0.126) | (0.187) | (0.166) | (0.181) | (0.131) | (0.191) | (0.174) | (0.122) | (0.132) | (0.130) |
| | [0.093] | [0.191] | [0.156] | [0.15] | [0.105] | [0.174] | [0.155] | [0.087] | [0.111] | [0.092] |
| Observations | 6,102 | 6,102 | 6,102 | 1,607 | 5,643 | 4,921 | 6,102 | 6,102 | 6,102 | 6,102 |
| Clusters | 35 | 35 | 35 | 34 | 35 | 35 | 35 | 35 | 35 | 35 |
| DV Mean | 324 | 324 | 324 | 249 | 327 | 325 | 324 | 324 | 324 | 324 |
| DV SD | 144 | 144 | 144 | 101 | 147 | 144 | 144 | 144 | 144 | 144 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Dist to Wichita | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes |
| Wichita FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. The bandwidth is 10 miles. The level of observation is 1990 census blocks. The RD polynomial is linear in latitude and longitude in columns (1) and (4)-(10), linear in distance to the boundary in column (2), and linear in latitude and longitude and distance to the boundary in column (3). Baseline specification in columns (1)-(6) and (10) also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. Column (7) drops the boundary segment fixed effects, column (8) drops the linear control for distance to Wichita, and column (9) drops the Wichita fixed effects. Column (4) drops all data within 15 miles of Whichta's center, column (5) drops all data within 5 miles of Dodge City center, and column (6) drops all data within 2.5 miles of the boundary. The specification in column (10) replaces the continuous treatment assignment with a binary one. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Conley (1999) standard errors with a 50-mile window in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## A.2 IV Main Result

### A.2.1 IV - Definition of the Privatization Frontier

We define the privatization frontier as the set of counties that were intensively privatized over the following decade. However, in practice there are multiple possibilities for implementing this definition. Our baseline implementation selects counties over the third quartile of the privatization rate over the next decade. This definition seems to work well, generating a relatively smooth westward expansion of the frontier and selecting frontiers of reasonable sizes. Nevertheless, it is important to make sure that the IV result is robust to different ways of selecting the frontier sample in practice. Columns (2)-(6) in table A.5 documents such robustness, showing that our baseline definition (column 1) does not result in a knife edge case. In column (2), we define the privatization frontier as counties that have above the mean rate of privatization over the next decade. Columns (3)-(4) use counties above the 70 and 80 percentiles, respectively, instead of the 75 percentile baseline. In column (5) the definition starts with a broad set of counties, having above the median privatization rate, and then drop counties in which almost all of the land was already privatized. Column (6) starts with a small set of counties, those with privatization rate above the top decile, and then selects all counties located within a 50 miles buffer from the originally selected counties. Results from all 5 alternative definitions are remarkably similar to the baseline result. Statistical significance is lost when the sample is counties above the 80 percentile of privatization rate (column 4), but this seems to be the result of a noisier estimate due to the smaller sample, rather than a drop in the magnitude of the coefficient.

### A.2.2 IV - Measurements of Long-Run Development

Similarly, there are multiple ways to measure long-run economic development, but these seem to make little difference. Table A.6 reports estimates using different measures and specifications. Column (2) reports the estimate of $\beta$ from a panel regression specification over the years 1969-2000, in which the dependent variable is the contemporary income per capita, standardized into z-scores. This specification also includes year fixed effects. The dependent variable in column (3) is log of the mean income per capita over 1969-2000. The dependent variable in column (4) is income per capita in year 2000, standardized into z-scores. The dependent variable in column (5) is the aggregate household income divided by the number of people aged 16-64 in year 2000, standardized into z-scores. In all cases the estimate is negative and significant. Moreover, when estimates are comparable (columns 1, 2, and 4), they are remarkably similar.

Table A.2: RD PLACEBO: REAL PROPERTY VALUE, 1990

| | The specification is: | | | | | |
|---|---|---|---|---|---|---|
| The RD Polynomial is: | Lon Lat | | Dist | | Lon Lat & Dist | |
| The bandwidth is: | 10 miles | 15 miles | 10 miles | 15 miles | 10 miles | 15 miles |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | *Panel A: Median owner-occupied housing value* | | | | | |
| **Outside Cession** | 0.055 | 0.029 | 0.086 | 0.031 | 0.034 | −0.023 |
| | (0.243) | (0.150) | (0.240) | (0.130) | (0.241) | (0.138) |
| | [0.177] | [0.084] | [0.141] | [0.069] | [0.168] | [0.092] |
| | p = 0.82 | p = 0.84 | p = 0.72 | p = 0.8 | p = 0.88 | p = 0.87 |
| Observations | 2,478 | 3,368 | 2,478 | 3,368 | 2,478 | 3,368 |
| Clusters | 18 | 26 | 18 | 26 | 18 | 26 |
| DV Mean | 32,355 | 31,556 | 32,355 | 31,556 | 32,355 | 31,556 |
| DV SD | 21,868 | 22,356 | 21,868 | 22,356 | 21,868 | 22,356 |
| | *Panel B: Median renter-occupied housing rent* | | | | | |
| **Outside Cession** | 0.233 | 0.324 | 0.265 | 0.280 | 0.224 | 0.213 |
| | (0.337) | (0.227) | (0.315) | (0.210) | (0.316) | (0.229) |
| | [0.291] | [0.308] | [0.239] | [0.273] | [0.266] | [0.297] |
| | p = 0.47 | p = 0.14 | p = 0.4 | p = 0.16 | p = 0.46 | p = 0.32 |
| Observations | 1,529 | 2,018 | 1,529 | 2,018 | 1,529 | 2,018 |
| Clusters | 18 | 26 | 18 | 26 | 18 | 26 |
| DV Mean | 203 | 200 | 203 | 200 | 203 | 200 |
| DV SD | 77.7 | 77.1 | 77.7 | 77.1 | 77.7 | 77.1 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Dist to Wichita | Yes | Yes | Yes | Yes | Yes | Yes |
| Wichita FEs | Yes | Yes | Yes | Yes | Yes | Yes |

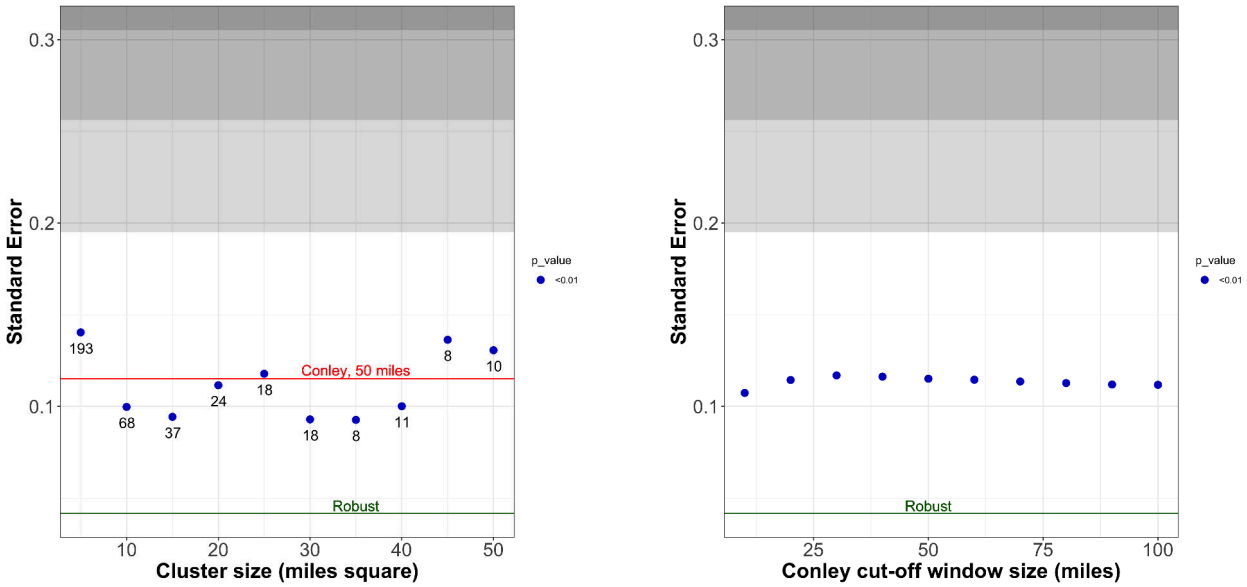
*Note:* The dependent variables are standardized into z-scores within the sample. The level of observation is 1990 census blocks. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Wild bootstrap p-values are also reported due to the small number of spatial clusters. Conley (1999) standard errors with a 50-mile window in brackets. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table A.3: RD ROBUSTNESS: MEASUREMENTS OF LOCAL LONG-RUN DEVELOPMENT

| | The dependent variable is: | | | | | |
|---|---|---|---|---|---|---|
| | Income per capita | | Median household income | | Share Bachelor Degree | |
| Bandwidth | 10 miles | 15 miles | 10 miles | 15 miles | 10 miles | 15 miles |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | *Panel A: linear in latitude-longitude* | | | | | |
| **Outside Cession** | −0.845*** | −1.040*** | −0.844*** | −0.864*** | −1.609*** | −1.712*** |
| | (0.324) | (0.316) | (0.175) | (0.156) | (0.355) | (0.373) |
| | [0.232] | [0.236] | [0.151] | [0.167] | [0.314] | [0.348] |
| | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 |
| | *Panel B: linear in distance to the cession boundary* | | | | | |
| **Outside Cession** | −0.718** | −0.746*** | −0.793*** | −0.663*** | −1.309*** | −1.193*** |
| | (0.338) | (0.212) | (0.197) | (0.157) | (0.355) | (0.270) |
| | [0.357] | [0.222] | [0.145] | [0.118] | [0.373] | [0.314] |
| | p = 0.03 | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 |
| | *Panel C: linear in latitude-longitude and distance* | | | | | |
| **Outside Cession** | −0.829*** | −0.856*** | −0.695*** | −0.566*** | −1.364*** | −1.269*** |
| | (0.270) | (0.191) | (0.243) | (0.204) | (0.284) | (0.229) |
| | [0.272] | [0.181] | [0.223] | [0.164] | [0.305] | [0.284] |
| | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 | p < 0.01 |
| Observations | 383 | 438 | 383 | 438 | 383 | 438 |
| Clusters | 21 | 28 | 21 | 28 | 21 | 28 |
| DV Mean | 19,552 | 19,497 | 41,019 | 41,718 | 0.150 | 0.147 |
| DV SD | 8,477 | 8,017 | 17,240 | 16,909 | 0.102 | 0.098 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to Metro | Yes | Yes | Yes | Yes | Yes | Yes |
| Wichita Ctrls | Yes | Yes | Yes | Yes | Yes | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. The level of observation is 2000 census block-groups. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Wild bootstrap p-values are also reported due to the small number of spatial clusters. Conley (1999) standard errors with a 50-mile window in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.1: RD: MEDIAN HOUSING VALUE - INFERENCE ROBUSTNESS




(a) Grid-cell size

(b) Conley window

*Note*: This figure plots the standard error of $\beta$ from the baseline specification of equation (1) when the dependent variable is median housing value using the approach proposed by Bester et al. (2011) with different grid-cell size (Panel a), and Conley (1999) standard errors with different windows (Panel b). The background color is indicative of the level of statistical significance. The p-value is $< 0.01$ in the white area, and $< 0.05$, $< 0.1$ and $> 0.1$ in the light to dark shades of gray. The y-axis is the standard error of and the x-axis is the size of the grid-cells (Panel a) or Conley window (Panel b). The green horizontal line plots the HC robust standard errors. The numeric label under each point in (Panel a) indicates the number of spatial clusters.

Table A.4: FUZZY RD: LONG-RUN LOCAL DEVELOPMENT

| Level of observation: | Blocks, 1990 | | Block groups, 2000 | | |
|---|---|---|---|---|---|
| Dependent variable: | Median housing value | Median housing rent | Income per capita | Median household income | Share bachelor degree |
| | (1) | (2) | (3) | (4) | (5) |
| **Fraction Homestead** | $-0.879^{***}$ | $-1.111^{***}$ | $-1.827^{***}$ | $-1.830^{***}$ | $-3.501^{***}$ |
| | (0.236) | (0.369) | (0.590) | (0.368) | (0.609) |
| | [0.296] | [0.28] | [0.459] | [0.302] | [0.612] |
| Observations | 7,744 | 5,796 | 382 | 382 | 382 |
| Clusters | 37 | 34 | 21 | 21 | 21 |
| DV Mean | 56,770 | 324 | 19,552 | 41,019 | 0.150 |
| DV SD | 36,308 | 144 | 8,477 | 17,240 | 0.102 |
| Boundary FEs | Yes | Yes | Yes | Yes | Yes |
| Dist to Wichita | Yes | Yes | Yes | Yes | Yes |
| Wichita FEs | Yes | Yes | Yes | Yes | Yes |

*Note:* The dependent variables are standardized into z-scores within the sample. The level of observation is 1990 census blocks in columns 1-2, and 2000 census block-groups in columns 3-5. The RD polynomial is linear in latitude and longitude. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. The bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Conley (1999) standard errors with a 50-mile window in brackets. $^{*}$ $p<0.10$, $^{**}$ $p<0.05$, $^{***}$ $p<0.01$

Table A.5: IV Robustness: Different Frontier Definitions

| | Baseline | Frontier Definition is: | | | | |
|---|---|---|---|---|---|---|
| | | Above mean | Above 70 percentile | Above 80 percentile | Above median & above 20% available | Above top deciles & 50 miles buffer |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Fraction Homestead** | −1.345** | −1.318** | −1.500** | −1.411 | −1.638** | −1.222*** |
| | (0.606) | (0.635) | (0.696) | (0.958) | (0.741) | (0.436) |
| Observations | 247 | 234 | 279 | 201 | 284 | 249 |
| DV Mean | 12,125 | 12,155 | 12,102 | 12,211 | 12,047 | 12,359 |
| DV SD | 1,603 | 1,609 | 1,596 | 1,598 | 1,579 | 1,481 |
| State FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Geoclimatic controls | Yes | Yes | Yes | Yes | Yes | Yes |

*Note:* The dependent variable is the mean z-score income per capita over 1969-2000. Observations are at the county level. Column (1) uses the baseline frontier definition: counties with a 10-year privatization rate above the 75th percentile. Columns (2)-(6) reports estimates using different frontier definitions. Column (2)-(4) uses counties with a 10 years privatization rate above the mean, 70th percentile, and 80th percentile, respectively. Column (5) uses counties with a 10 years privatization rate above the 50th percentile that also have at least 20% of land still in the public domain. Column (6) uses all counties within 50 miles from counties with a 10 years privatization rate above the 90th percentile. Geo-climatic controls include temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Conley (1999) standard errors with a 500-mile window in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

### A.2.3 IV - Inference

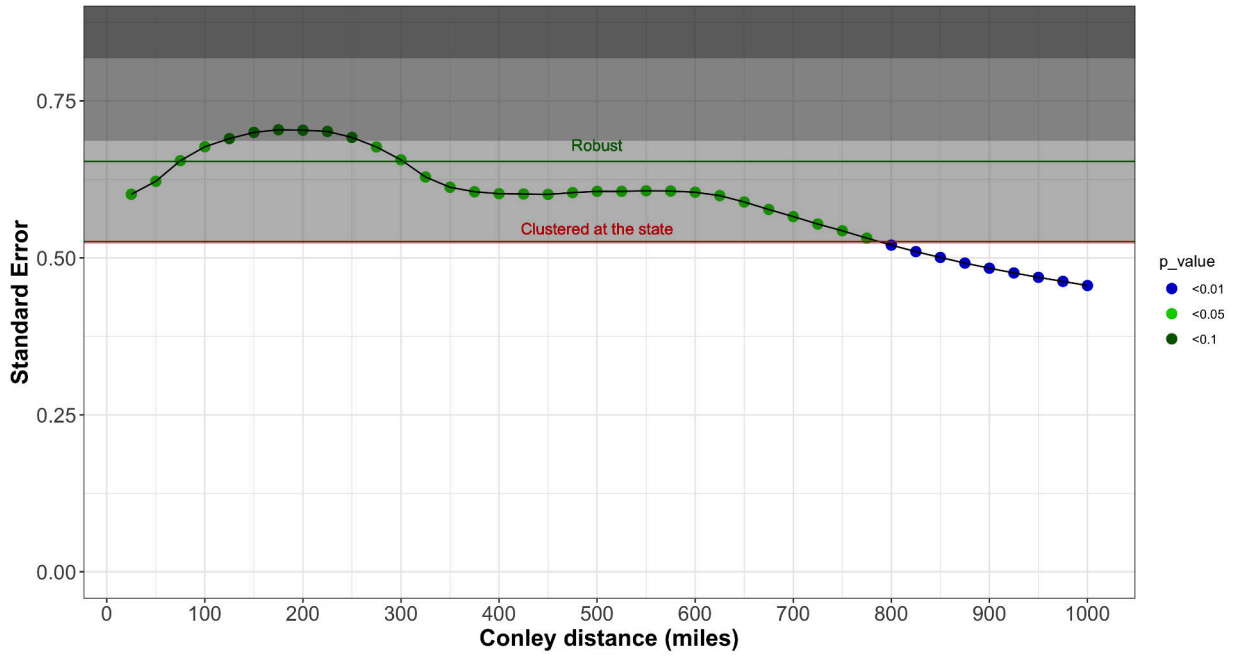
Our baseline IV specification uses Conley (1999) spatial HAC standard errors with a 500-mile cutoff window to accounts for spatial auto-correlation. Figure A.2 documents robustness to different choices of cutoff window size, by plotting the standard errors and significance levels of IV estimate of $\beta$ under different specifications. With cutoff windows between about 125 and 250 miles significance level drops below the 5%, reaching as far as a p-value of 0.057. In all other cases, significance level is below the 5%. The standard error is remarkably stable between cutoff windows of 350 and 600 miles–that is, around our baseline cutoff window of 500 miles.

Table A.6: IV ROBUSTNESS: MEASUREMENTS OF LONG-RUN DEVELOPMENT

| | | Dependent variable is: | | | |
|---|---|---|---|---|---|
| | Baseline | Contemporary income pc (panel) | Log income per capita | Year 2000 income per capita | Year 2000 Household income per capita |
| | (1) | (2) | (3) | (4) | (5) |
| **Fraction Homestead** | −1.345** | −1.325*** | −0.270** | −1.245** | −1.853*** |
| | (0.606) | (0.379) | (0.106) | (0.503) | (0.528) |
| Observations | 247 | 7,638 | 247 | 247 | 247 |
| DV Mean | 12,125 | 12,399 | 9.39 | 24,227 | 34,201 |
| DV SD | 1,603 | 6,588 | 0.13 | 3,613 | 4,263 |
| State FEs | Yes | Yes | Yes | Yes | Yes |
| Year FEs | No | Yes | No | No | No |
| Geoclimatic controls | Yes | Yes | Yes | Yes | Yes |

*Note:* The dependent variable in column (1) is the mean z-score income per capita over 1969-2000. Column (2) reports the estimate of $\beta$ from a panel regression specification with year fixed effects over the years 1969-2000, in which the dependent variable is the contemporary z-scored income per capita. The dependent variable in column (3) is log of the mean income per capita over 1969-2000. The dependent variable in column (4) is the z-scored income per capita in year 2000. The dependent variable in column (5) is the z-scored aggregate household income divided by the number of people aged 16-64 in year 2000. Observations are at the county level. Geo-climatic controls include temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Conley (1999) standard errors with a 500-mile window in parentheses. Inference in column (2) also allows for serial correlation within a 10-year window. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

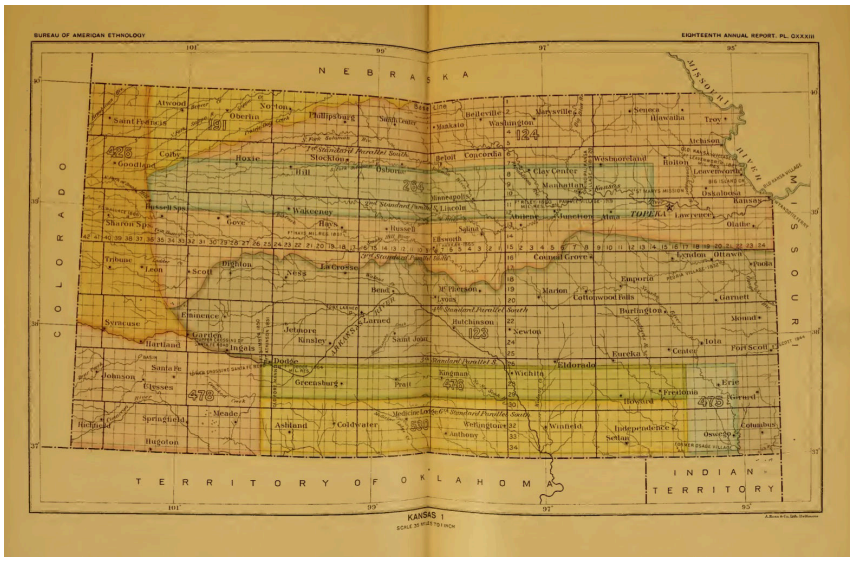
Figure A.2: IV: INFERENCE ROBUSTNESS




*Note*: This figure plots the standard errors and significance levels of IV estimate of $\beta$ for different Conley (1999) cutoff windows. The y-axis is the standard error of and the x-axis is the size of the window. The background color is indicative of the level of statistical significance. The p-value is $< 0.01$ in the white area, and $< 0.05$, $< 0.1$ and $> 0.1$ in the light to dark shades of gray. The green horizontal line plots the HC robust standard errors, while the red horizontal line plots standard errors clustered at the state level.

# Appendix B Other Figures and Results

Figure B.1: INDIAN LAND CESSIONS IN KANSAS, 1

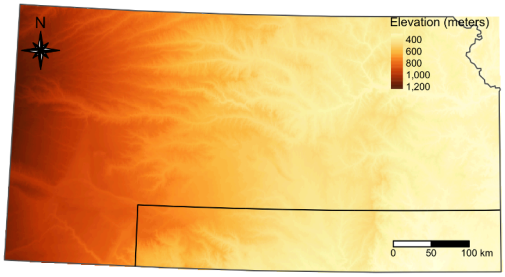



The "Osage ceded land" is marked on the map as area 475, the "Osage trust land" as area 476, and the "Osage diminished reserve as 530. *Source*: Eighteenth Annual Report of the Bureau of American Ethnology to the Secretary of the Smithsonian Institution, 1896–97. By J. W. Powell, Director

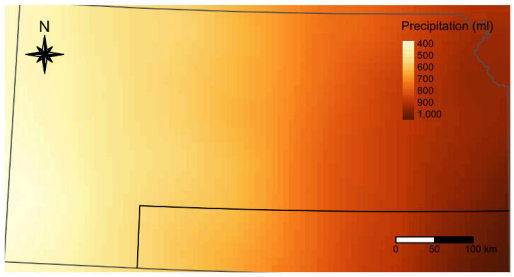
Figure B.2: INDIAN LAND CESSIONS IN KANSAS, 2


The "Cherokee neutral land" is marked on the map as area 490. *Source*: Eighteenth Annual Report of the Bureau of American Ethnology to the Secretary of the Smithsonian Institution, 1896–97. By J. W. Powell, Director

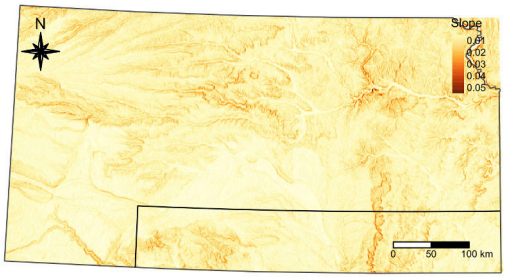
Figure B.3: BALANCE ON GEO-CLIMATIC CHARACTERISTICS



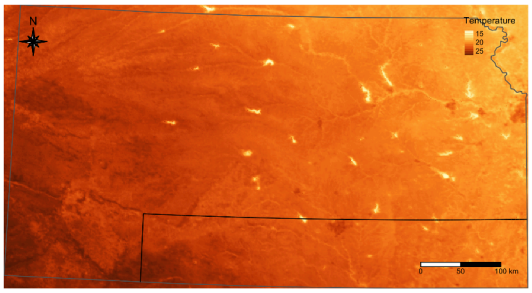

(a) Elevation



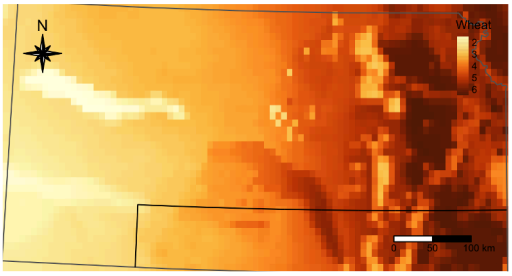

(b) Precipitation



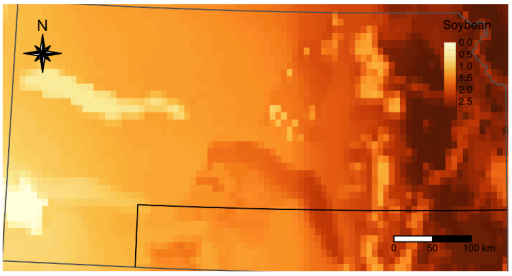

(c) Slope



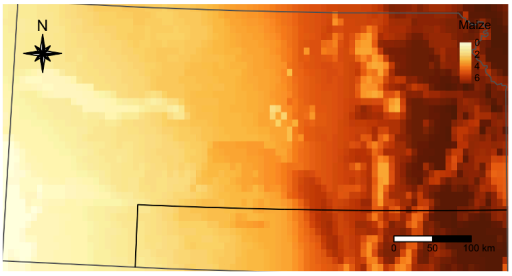

(d) Land Temperature



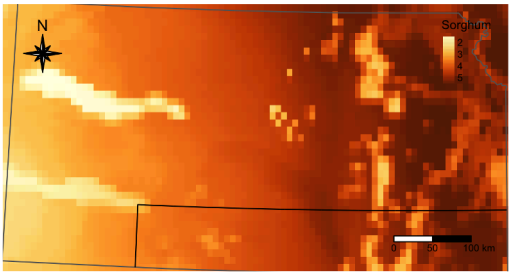

(e) Wheat




(f) Soybean




(g) Maize




(h) Sorghum

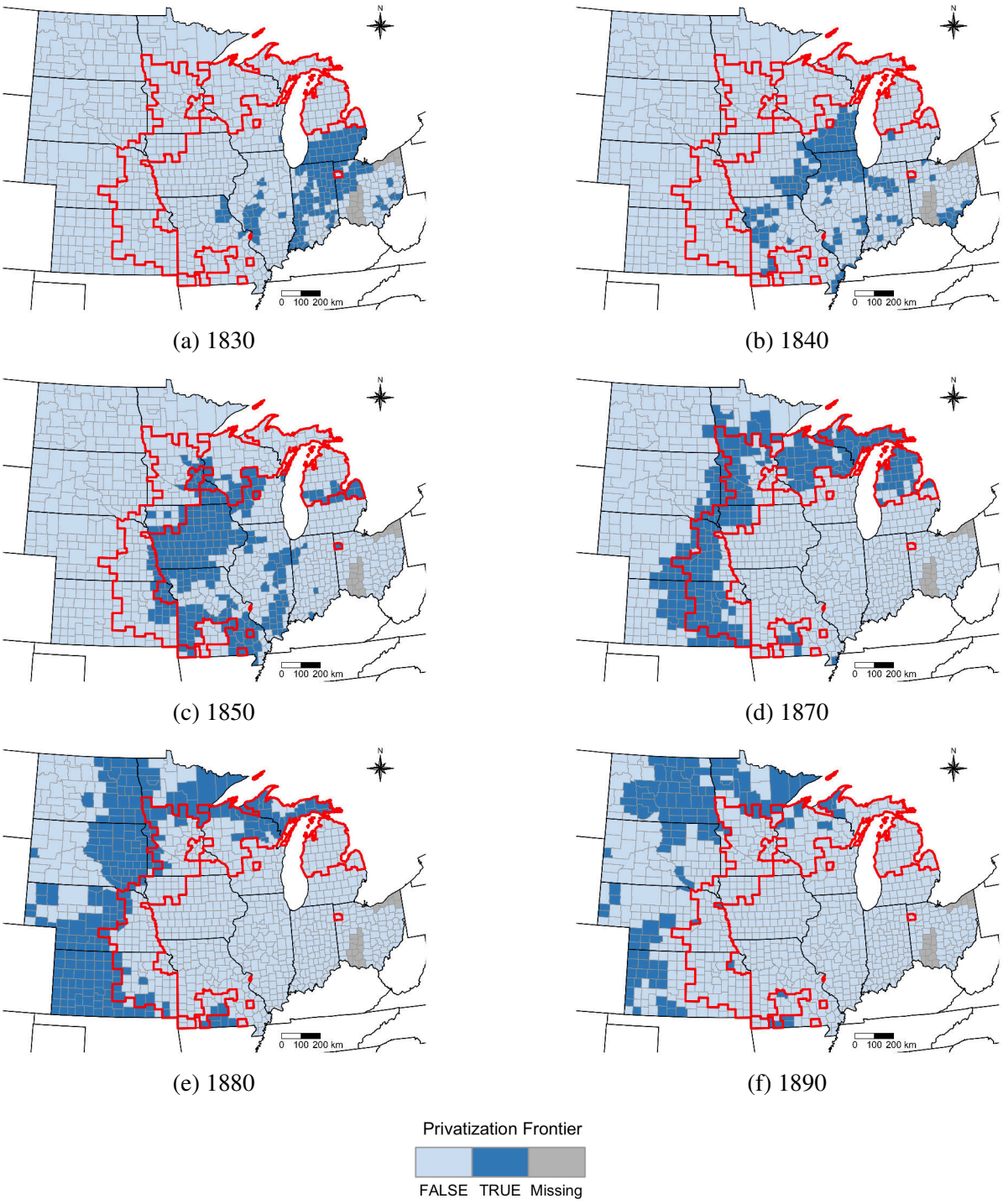
*Note*: This figure plots the RD boundary over the spatial distribution of key geo-climatic characteristics: land temperature, precipitation, slope, altitude, and potential yield of wheat, maize, soybean and sorghum.

Table B.1: IV: EARNINGS, PRODUCTIVITY AND EMPLOYMENT SHARES BY INDUSTRY SECTORS

| Industry | Earnings Per Capita | Earnings Per Worker | Employment Share |
|---|---|---|---|
| Farming | 0.137 | −0.094 | 0.107 |
| | (0.381) | (0.656) | (0.557) |
| Non-Farming | −1.356** | −1.321*** | −0.107 |
| | (0.579) | (0.450) | (0.557) |
| Manufacturing | −1.257** | −1.401*** | −0.701 |
| | (0.507) | (0.412) | (0.442) |
| Finance, insurance, and real estate | −1.509** | −0.524 | −0.672 |
| | (0.649) | (0.553) | (0.559) |
| Construction | −0.783** | −0.771 | −0.698** |
| | (0.326) | (0.601) | (0.354) |
| Government and government enterprises | −0.717* | −1.952*** | 0.289 |
| | (0.373) | (0.470) | (0.325) |
| Services | −1.017 | −0.846** | 0.779 |
| | (0.663) | (0.402) | (0.785) |
| Retail trade | −0.843 | −1.153*** | −0.126 |
| | (0.540) | (0.441) | (0.576) |
| Wholesale trade | −0.904 | −0.415 | −0.806 |
| | (0.725) | (0.347) | (0.637) |
| Mining | 0.044 | −0.138 | 0.182 |
| | (0.290) | (0.294) | (0.432) |
| Transportation and public utilities | 0.004 | −0.881 | 0.500 |
| | (0.614) | (0.584) | (0.592) |
| Agricultural services, forestry, and fishing | −0.128 | −0.511 | 0.133 |
| | (0.353) | (0.446) | (0.537) |

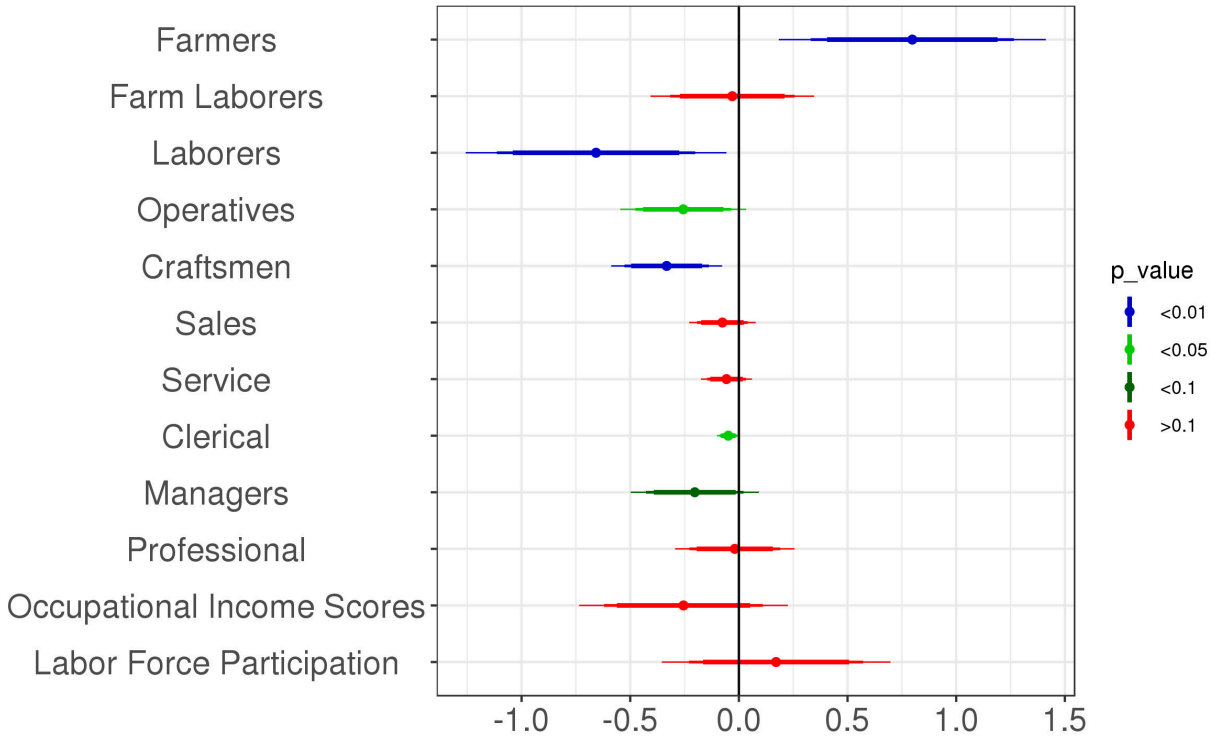
*Note:* This table presents IV estimates of $\beta$ for different dependent variables (columns): personal earning per capita, personal earning per worker, and employment shares, in different sectors of the economy (rows). Observations are at the county level. The sample is the baseline 1862 privatization frontier. Geo-climatic controls include temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Conley (1999) standard errors with a 500-mile window in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.4: THE PRIVATIZATION FRONTIER IN SELECTED YEARS



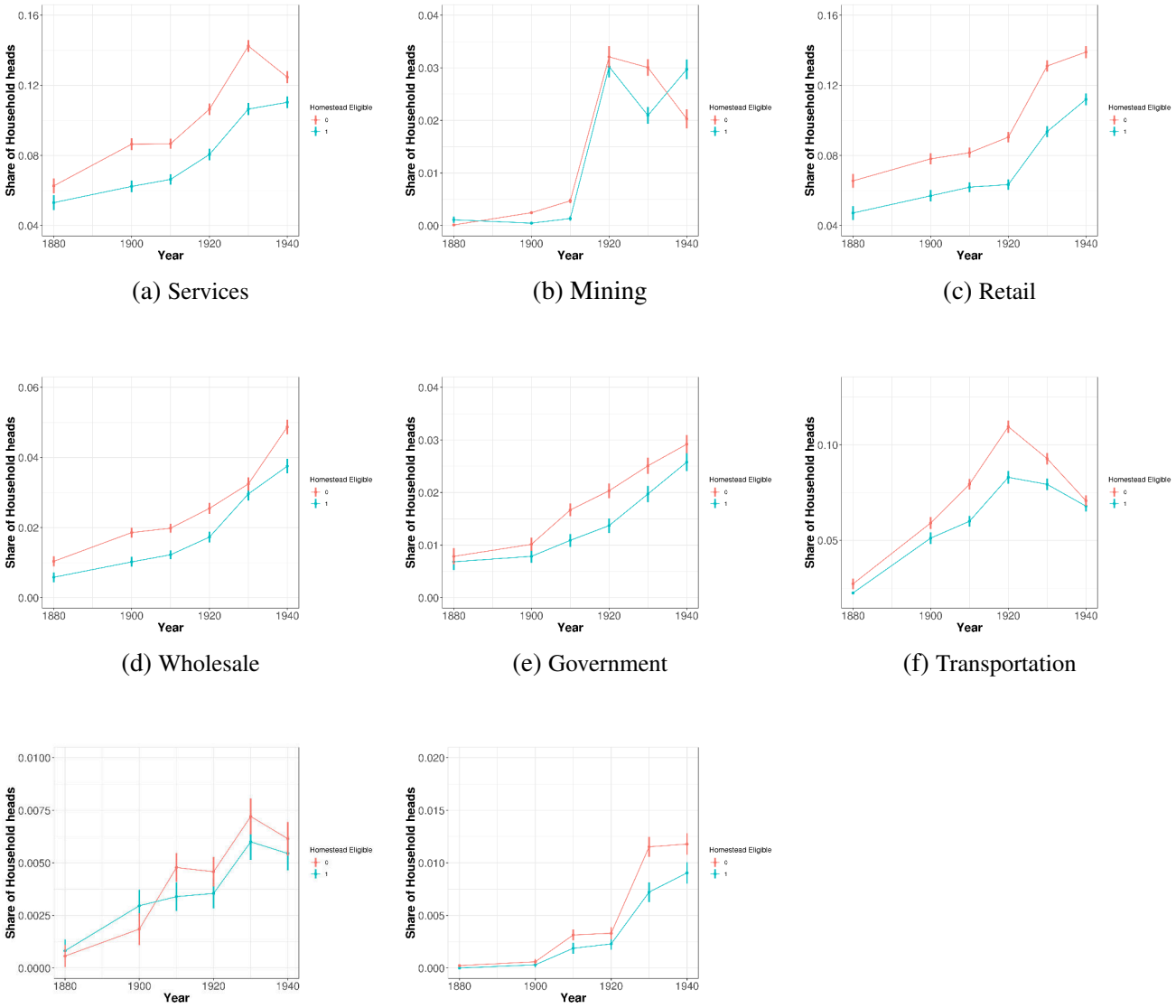

*Note*: This figure plots the set of counties that form the privatization frontiers in selected years (plotted in dark blue) against the counties that form the 1862 privatization frontier (plotted as red boundary). The privatization frontier is defined as counties in which the share of land privatized during the following decade was in the top quartile.

Figure B.5: OCCUPATIONS DURING EARLY SETTLEMENT, 1880




*Note*: This figure plots the RD estimates and 99%, 95% and 90% confidence intervals of $\beta$ from equation (1) when the dependent variables are dummy variables that take the value of 1 if the household head has a given occupation, standardized into z-scores. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from the 1880 full-count census.

Figure B.6: OCCUPATION SHARES ACROSS INDUSTRIES, 1880-1940




(a) Services (b) Mining (c) Retail

(d) Wholesale (e) Government (f) Transportation

(g) Communication (h) Utilities

*Note*: This figure plots the mean and 95% confidence intervals of the share of household heads that were employed in a given industry sector, for both sides of the historical Osage boundary, within a 35 miles bandwidth, for years 1880-1940. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from full-count censuses, 1880-1940.

# Appendix C Alternative Explanations

In this section, we consider two competing hypotheses regarding channels. The first is that a higher prevalence of land consolidation and large landowners in non-homesteading regions had a positive impact on development. The second is that homesteading contributed to the development of cultural traits and values that were unfavourable for long-run economic development. We find that neither hypothesis is supported by the empirical facts.

## C.1 Land Consolidation and Large Landowners

A main competing hypothesis regarding channels focus on the existence of large landowner. Some historians have argued that the large land holdings of absentee owners, often referred to as land speculators, had an adverse effect on economic development (Gates, 1973; Swierenga, 1977). Specifically, it has been argued that absentee owners slowed economic development by keeping land idle and away from settlers, contributed to the raise in farm tenancy and lower investments in agriculture, and reduced local tax collection needed to support public goods. Recent studies in economics provide causal evidence supporting some of these arguments (Raz, 2018; Smith, 2019). On the other hand, other scholars have argued that absentee landownership actually had a positive impact, stressing their important function as land retailers, their risk-bearing and informational roles, their contribution to the development of early cities, and the fact that they helped to attract settlers west (North, 1974; Swierenga, 1977). Moreover, Cogswell (1975) finds a negative correlation between absentee landownership and farm tenancy, and argues that land "speculators may have directly reduced the proportion of tenant farmers by offering land on credit" (p. 27). Can a lower prevalence of large land owners in the homesteading region be responsible for the adverse impact on long-run development?

In the context of the Osage land cessions, this seems less plausible. Similar to privatization under the Homestead Act, land on the Osage cessions was only sold to actual settlers, in tracts not exceeding 160 acres. Nevertheless, this is not the general case for the counterfactual to homesteading,[45] and it is theoretically possible that unlike homestead tracts, for which property rights were only awarded after five years, land on the Osage cessions was quickly consolidated and passed to the hands of large landowners.

To assess the likelihood of this possibility, we utilize census data on the average size of farms, available at the civil townships level in 1930,[46] and use the discontinuity in homestead assignment across the historical Osage boundary to examine effects on farm sizes. The results of the analysis

[45] Specifically, this is note the case in the context of our IV strategy.

[46] To the best of our knowledge, such data was not published by the census bureau for earlier years.

are presented in table C.1. We do not find any statistically significant effects. Not only that, but the point estimates are highly unstable across specification and bandwidth, which suggests that the lack of statistical significance is not likely to be driven by low power alone.

## C.2 Values and Culture

Another competing hypothesis we consider is that settlement under the Homestead Act shaped values and culture, which in turn affected economic activity. We will consider two separate hypotheses.

The first is that the settlement under the Act contributed to the creation of more egalitarian communities with a closely knit social structure. Some historians have argued that the adverse conditions during the first years of settlement draw homesteaders to rely on each other's assistance. Lee (1979) notes that "Homestead neighbors shared labor, provisions and most important of all, each other's company" (p. 552). Moreover, he argues that the relative equality of possessions and social status in homesteaders' communities helped prevent disputes and divisions. The common practices of labor sharing and mutual assistance and the relative social equality shaped the social nature of the newly formed communities, making them more egalitarian and cohesive and strengthening within them the sense of communal values.

Collectivism, or communal values, have been found to be negatively associated with economic growth (Gorodnichenko and Roland, 2011).[47] It is thus possible in theory that homesteading led to stronger communal values relative to communities that were not settled under the Act, which in turn lowered economic development in the long-run.[48]

We use the IV strategy and county level data to examine this hypothesis. Table C.2 presents the results. In column (1), we use survey data from the Moral Foundations Questionnaire[49] to measure the relative importance of universal values over communal values. We follow the procedure in Enke (2019) to aggregate individual level data to the county level. There does not seem to be any association between homesteading and the prevalence of universal values. In columns (2)-(4), we examine effects on the share of religious adherents, the divorce-to-marriage ratio, and the fraction of single female households with children out of total households with children. These measurements were found to be correlated with collectivism or communal values. We do not find any evidence for an impact of the Act on on these outcomes. Moreover, the point estimates across the different outcomes are not aligned. The

[47] Most of the literature had focused on the opposite of collectivism - individualism or universal values.

[48] It is important to note that this hypothesis is not entirely independent from the longevity on settlers' tenure. Indeed, it is possible that communities with lower levels of geographical mobility will become more communal over time. Likewise, the residents of more communal locations are likely to be less geographically mobile.

[49] www.yourmorals.org

Table C.1: RD: AVERAGE SIZE OF FARMS, 1930

| | The bandwidth is: | | |
|---|---|---|---|
| | 10 miles | 15 miles | 20 miles |
| | (1) | (2) | (3) |
| | *Panel A: linear in latitude-longitude* | | |
| **Outside Cession** | 0.062 | 0.092 | −0.587 |
| | (0.417) | (0.308) | (0.372) |
| | [0.252] | [0.203] | [0.385] |
| | *Panel B: linear in distance to the cession boundary* | | |
| **Outside Cession** | 0.949 | 0.440 | 0.047 |
| | (0.792) | (0.342) | (0.154) |
| | [0.79] | [0.294] | [0.139] |
| | *Panel C: linear in latitude-longitude and distance* | | |
| **Outside Cession** | 1.110 | 0.406 | 0.064 |
| | (1.002) | (0.323) | (0.159) |
| | [0.937] | [0.27] | [0.14] |
| Observations | 85 | 135 | 186 |
| Clusters | 25 | 43 | 46 |
| DV Mean | 388 | 377 | 398 |
| DV SD | 251 | 226 | 299 |
| Boundary FEs | Yes | Yes | Yes |
| Dist to Wichita | Yes | Yes | Yes |
| Wichita FEs | Yes | Yes | Yes |

*Note:* The dependent variable is standardized into z-scores within the sample. The level of observation is 1930 civil township. The RD polynomial is linear in latitude and longitude. Regressions also control for closest boundary segment fixed effects, linear distance to Wichita, and Wichita fixed effects. The bandwidth is 10 miles. Standard errors clustered at arbitrary grid cells of 15 miles square in parentheses (Bester et al., 2011). Conley (1999) standard errors with 50-mile window in brackets. $^{*}$ $p<0.10$, $^{**}$ $p<0.05$, $^{***}$ $p<0.01$

Table C.2: IV RESULTS: CULTURE AND VALUES

| | Dependent variable is: | | | |
|---|---|---|---|---|
| | Rel. Imp. of Universal Values | Fraction Religious | Divorced to Married Ratio | Frac. Child w. Single Mothers |
| | (1) | (2) | (3) | (4) |
| **Fraction Homestead** | 0.143 | 0.737 | 0.238 | −0.830 |
| | (0.921) | (1.000) | (1.076) | (0.911) |
| Observations | 205 | 247 | 247 | 247 |
| DV Mean | 10.165 | 0.144 | 0.590 | 0.161 |
| DV SD | 3.290 | 0.036 | 0.182 | 0.041 |
| State FEs | Yes | Yes | Yes | Yes |
| Geoclimatic controls | Yes | Yes | Yes | Yes |

*Note:* The dependent variables are standardized into z-scores. The dependent variable in column (1) is the relative importance of universal values, measured from the Moral Foundations Questionnaire. In column (2), it is the share of religious adherents. In column (3) it is the divorced to married ratio. In column (4), it is the fraction of single female households with children out of total households with children. Observations are at the county level. The sample is the baseline 1862 privatization frontier. Geo-climatic controls include temperature, precipitation, slope, elevation, waterway, and potential output of wheat, maize, alfalfa, barley, buckwheat, soybean, potato, and oat. Conley (1999) standard errors with a 500-mile window reported in parenthesis. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

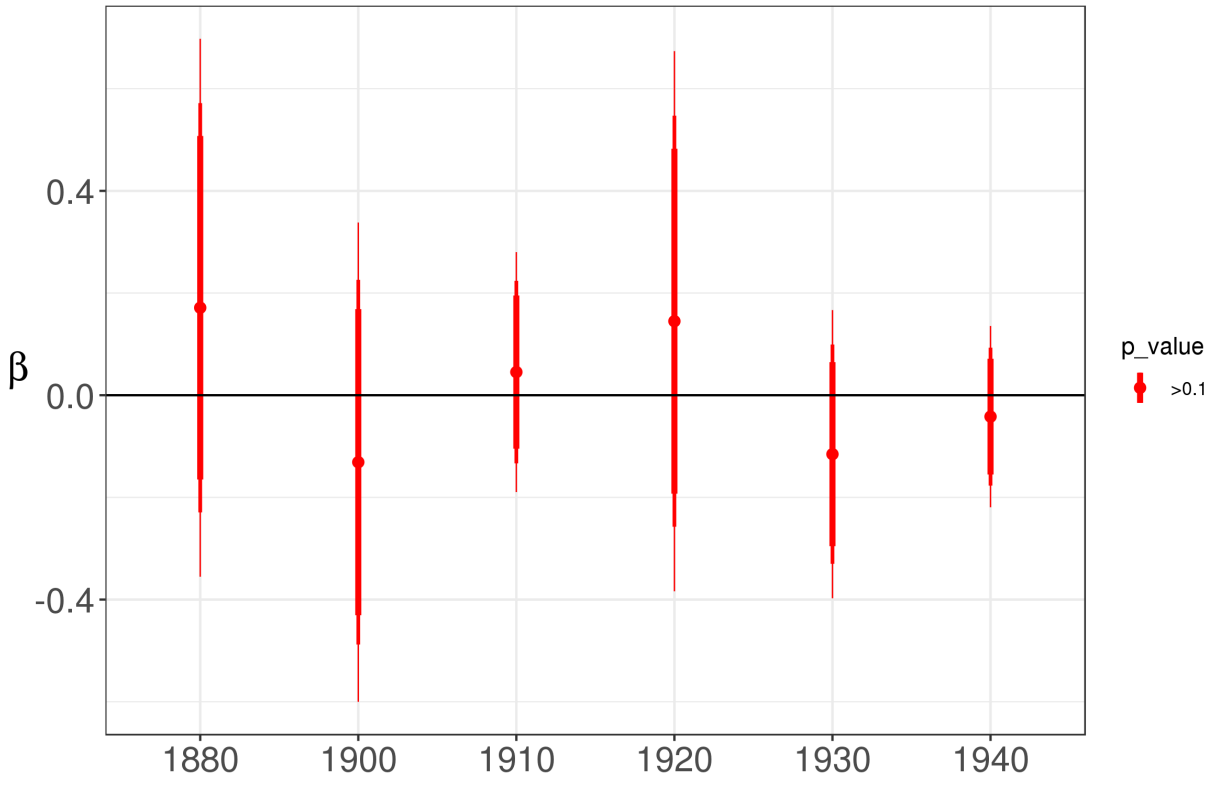
point estimate suggest a positive effect on religiosity, but also on divorces and the relative importance of universal values over communal values, and a negative impact on single parenthood.

Another hypothesis is that acquiring land for "free" affected homesteaders' view regarding the importance of economic success as a mean to achieve personal welfare, and therefore the degree to which they (and latter, their children) were willing to exert effort in the labor market. This in turn contributed to a lower economic development in the long-run.

In order to test the plausibility of this hypothesis, we look for effects of the Act on labor market participation. To do so, we use the RD strategy and full count census data for years 1880, and 1900-1940. The results are presented in Figure C.1. We do not find any association between historical

homesteading and labor force participation. The RD estimate of $\beta$ from equation (1) are insignificant in all years. Moreover, the point estimates are positive in some years, while negative in others.

Figure C.1: LABOR FORCE PARTICIPATION, 1880-1940




*Note*: This figure plots the RD estimate of $\beta$ when the dependent variable is dummy variable that take the value of 1 if the household head is in the labor force, standardized into z-score. The RD polynomial is linear in latitude and longitude, and the bandwidth is 10 miles. Standard errors are clustered at arbitrary grid cells of 15 miles square (Bester et al., 2011). Regressions also include baseline controls. The level of observation is an individual, geo-referenced into a city, town, or civil township. Data from full-count censuses, 1880-1940.