Frontier Culture: The Roots and Persistence of “Rugged Individualism” in the United States
Samuel Bazzi, Martin Fiszbein, and Mesay Gebresilasse
NBER Working Paper No. 23997
November 2017, Revised August 2020
JEL No. D72,H2,N31,N91,P16

ABSTRACT

The presence of a westward-moving frontier of settlement shaped early U.S. history. In 1893, the historian Frederick Jackson Turner famously argued that the American frontier fostered individualism. We investigate the Frontier Thesis and identify its long-run implications for culture and politics. We track the frontier throughout the 1790–1890 period and construct a novel, county-level measure of total frontier experience (TFE). Historically, frontier locations had distinctive demographics and greater individualism. Long after the closing of the frontier, counties with greater TFE exhibit more pervasive individualism and opposition to redistribution. This pattern cuts across known divides in the U.S., including urban–rural and north–south. We provide evidence on the roots of frontier culture, identifying both selective migration and a causal effect of frontier exposure on individualism. Overall, our findings shed new light on the frontier’s persistent legacy of rugged individualism.

Samuel Bazzi
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and CEPR
and also NBER
sbazzi@bu.edu

Martin Fiszbein
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
fiszbein@bu.edu

Mesay Gebresilasse
Amherst College
301 Converse Hall
Amherst, MA 01002
mgebresilasse@amherst.edu

Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States*

Samuel Bazzi†
Boston University
NBER and CEPR

Martin Fiszbein‡
Boston University
and NBER

Mesay Gebresilasse§
Amherst College

July 2020

Abstract

The presence of a westward-moving frontier of settlement shaped early U.S. history. In 1893, the historian Frederick Jackson Turner famously argued that the American frontier fostered individualism. We investigate the Frontier Thesis and identify its long-run implications for culture and politics. We track the frontier throughout the 1790–1890 period and construct a novel, county-level measure of total frontier experience (TFE). Historically, frontier locations had distinctive demographics and greater individualism. Long after the closing of the frontier, counties with greater TFE exhibit more pervasive individualism and opposition to redistribution. This pattern cuts across known divides in the U.S., including urban–rural and north–south. We provide evidence on the roots of frontier culture, identifying both selective migration and a causal effect of frontier exposure on individualism. Overall, our findings shed new light on the frontier's persistent legacy of rugged individualism.

Keywords: Culture, Individualism, Preferences for Redistribution, American Frontier, Persistence

JEL Codes: O15, O43, D72, H2, N31, N91, P16

*We thank Alberto Alesina, Quamrul Ashraf, Jeremy Atack, Michael Clemens, William Collins, Klaus Desmet, Benjamin Enke, Marcel Fafchamps, James Feigenbaum, Ray Fisman, Oded Galor, Camilo Garcia-Jimeno, Paola Giuliano, Bob Margo, Nathan Nunn, Ömer Özak, Daniele Paserman, Nico Voigtlaender, Romain Wacziarg, John Wallis, and David Weil, as well as numerous seminar and conference audiences for helpful comments. Yeonha Jung, Max McDevitt, Hanna Schwank, and Huiren Tan provided excellent research assistance. All errors are our own.

†Department of Economics. 270 Bay State Rd., Boston, MA 02215. Email: sbazzi@bu.edu.

‡Department of Economics. 270 Bay State Rd., Boston, MA 02215. Email: fiszbein@bu.edu.

§Department of Economics. 301 Converse Hall, Amherst, MA 01002. Email: mgebresilasse@amherst.edu.

1 Introduction

Rapid westward expansion marked the early history of the United States. According to the influential historian Frederick Jackson Turner, the presence of "a continually advancing frontier line" at the "edge of free land" profoundly shaped American culture (Turner, 1893). The frontier cultivated individualism and antipathy to government intervention. These two traits are encapsulated in the notion of "rugged individualism," popularized by Republican Herbert Hoover in his 1928 presidential campaign.

This paper shows that the American frontier gave rise to a persistent culture of rugged individualism. We combine Census data spanning 150 years with survey and electoral outcomes to identify this indelible legacy of westward expansion. In the 18th and 19th century, frontier populations were not only distinctive demographically but also more individualistic. Across America today, counties with longer historical exposure to frontier conditions exhibit greater individualism and opposition to redistribution and regulation. This anti-statist culture has its roots in both selective migration to the frontier and a causal effect of frontier life on its residents. Both forces were arguably responses to the differential returns to individualism on the frontier.

We track the frontier over time using historical population data and modern Geographic Information System (GIS) methods. Following Turner's classic essay and the 1890 *Progress of the Nation* report by the Census Bureau, we define the frontier line as the boundary at which population density falls below two people per square mile. The frontier is comprised of counties with low population density in close proximity to the frontier line. This time-varying measure of frontier status, detailed in Section 2, is consistent with Turner's (1896) view of the frontier as "a form of society" rather than a fixed area. We calculate total frontier experience (TFE) as the time spent on the frontier between 1790 and 1890, providing the first granular and comprehensive measure of frontier history.

We provide systematic evidence on the demographic and cultural distinctiveness of frontier locations. In line with historical narratives, Section 3 shows that frontier settlers were disproportionately male, prime-age, and foreign-born. These traits are strongly correlated with the two defining features of isolation on the frontier: low density and remoteness. Moreover, we identify structural breaks in these demographics near the density cutoff defining the frontier line, thus validating a seemingly arbitrary historical definition. Individualism also sharply increases along the frontier, as seen through children's name choices—a primordial act of cultural transmission. We measure individualism through the prevalence of infrequent names, which are more pervasive on the frontier even after accounting for the greater prevalence of immigrants.

A rich social science literature motivates our names-based measure of individualism. The informational content of names has been emphasized in economics (e.g., Abramitzky et al., forthcoming; Bertrand and Mullainathan, 2004) as well as psychology and sociology (e.g., Gerrit and Onland, 2011; Lieberson and Bell, 1992). The measure we use comes from social psychology, a field that portrays individualism as the key dimension of cultural variation across countries (Heine, 2010). The foundational contributions of Hofstede (1991) and Triandis (1995) associate individualism with several related traits: a view of the self as independent rather than interdependent, emphasis on self-reliance, primacy of self-interest, and regulation of behavior by personal attitudes rather than social norms. Consistent with these traits, infrequent names reflect a desire to stand out, as opposed to common names, which reflect a desire to fit in (Twenge et al., 2010).

In Section 4, we use our novel measure of TFE to uncover a persistent culture of rugged individualism. In the mid-20th century, several decades after the closing of the frontier, individualistic children's names are more pervasive in counties with greater TFE. Moreover, such individualism goes hand in hand with opposition to big government. In the late 20th century and beyond, residents of high-TFE counties prefer less redistribution and lower public spending, and they pay lower property tax rates. These findings hold across counties within the same state, even after accounting for geoclimatic features, including weather, distance to waterways, and potential agricultural productivity.

These deep-rooted preferences have important political ramifications. High-TFE counties exhibit stronger and, in fact, increasing support for the Republican Party between 2000 and 2016—a period in which attitudes toward taxation and regulation animate a growing partisan divide. We show that voters in high-TFE counties report greater opposition not only to redistribution but also to social protection, minimum wages, gun control, and environmental protection. Republican political discourse on these issues increasingly resonates with frontier culture, embracing opposition to the welfare state, a strong belief in effort versus luck in reward, the right to self-defense, and "manifest destiny". Throughout the 20th century, the partisan divide did not align so clearly with the anti-statist principles of rugged individualism. However, in the few elections when it did, high-TFE counties offered greater support to the candidate who, like Hoover, directly appealed to frontier culture.

The persistent effects of TFE are not merely a reflection of persistently low population density. Rather, they capture a legacy of frontier settlement that cuts across the urban–rural cultural divide. We rule out confounding effects of density in several ways, the most demanding of which identifies the effects of TFE across matched counties with nearly identical contemporary density. We also account for other cultural confounds related to mining, rainfall risk, railroad access, slavery, and diversity. An instrumental variables strategy further isolates exogenous variation in TFE due to national immigration shocks.

Additional results support the proposed link between frontier experience to contemporary culture. First, African American preferences are unrelated to TFE, consistent with the fact that several mechanisms fostering rugged individualism on the frontier (e.g., selective migration, prospects for upward mobility) were of limited historical relevance for blacks, especially in the antebellum period. Second, preferences over policies tangential to frontier culture exhibit little relation to TFE. Third, TFE has similar long-run effects within different regions of the country, including the West Coast, which experienced its own frontier expansion eastward in the mid-1800s. These results point to a shared culture of rugged individualism despite large regional differences in preferences.

Various mechanisms might explain the persistence of frontier culture long after the abatement of frontier conditions. Culture can be sticky and converge at very slow speeds or not at all. With path dependence, initial conditions affect long-run outcomes. In this sense, the earliest stages of development on the frontier were likely a critical juncture in the formation of local culture. Turner (1893) alluded to this possibility, noting that "traits [of frontier society] have, while softening down, still persisted as survivals in the place of their origin, even when a higher social organization succeeded."[1]

Section 5 shows how the culture of rugged individualism took root historically. The frontier attracted

[1] Zelinsky's (1973) "doctrine of first effective settlement" makes a related point: when "an empty territory undergoes settlement [...] the specific characteristics of the first group able to effect a viable, self-perpetuating society are of crucial significance for the later social and cultural geography of the area, no matter how tiny the initial band of settlers may have been."

individualistic migrants, and frontier life caused people to become even more individualistic over time. Cultural change in response to frontier conditions is consistent with theories of utilitarian intergenerational transmission (e.g., Doepke and Zilibotti, 2008) and with Turner's (1901) suggestion that frontier life fostered "a modification of the original stock." These two forces—selection and exposure effects—may be explained by the unique threats and opportunities of frontier life, which rewarded individualism.

Selective migration can be identified using complete-count Census data from the 1800s. Families that moved from settled areas to the frontier were more individualistic—based on pre-move children's names—than families remaining in settled areas. Similarly, families moving from the frontier to settled areas were less individualistic than those staying on the frontier. In other words, individualists selectively move to the frontier, and non-individualists selectively move away from the frontier. These patterns resonate with the view of frontier migrants as individualists willing and able to give up their social environment to settle in remote and isolated contexts (see Kitayama et al., 2006).

The frontier not only attracted individualists but also deepened individualistic culture among those living there. We identify this causal exposure effect using two complementary sources of variation among migrants to the frontier: (i) within-household changes in children's names pre- and post-move, and (ii) cross-sibling variation in frontier experience based on age-at-move. These strategies capture different exposure across the life-cycle, and both identify a significant effect on individualism that is distinct from selective migration. In approach (i), individualistic name choices increase only after arrival, as evidenced by the lack of pre-trends among kids born prior to moving. In approach (ii), we link Census records from 1850 to 1880 and find that, among brothers brought by their parents to the frontier at different ages, those that spent more of their childhood on the frontier give their own children more individualistic names. Both approaches, like our long-run analysis of TFE, emphasize that the length of frontier exposure determines the scope for cultural change.

These cultural dynamics are consistent with differential returns to individualism on the frontier. We provide suggestive evidence using occupational scores to proxy for economic status. Conditional on county fixed effects, fathers whose children have more individualistic names exhibit higher occupational standing in frontier counties relative to settled counties. These patterns are consistent with historical narratives emphasizing that independence and self-reliance were key to survival and success on the frontier. Frontier settlers faced many challenges with little social infrastructure to turn to (Edwards et al., 2017). As Overmeyer (1944) put it, "life was rough, crude, hard, and dangerous." On the other hand, the abundance of land offered profit opportunities (Stewart, 2006), and in an uncharted environment, individualists' non-conformism and inventiveness made them more resourceful (Shannon, 1977).

While individualists may generally oppose interference in the pursuit of self-interest, frontier conditions likely amplified their opposition to redistribution. Land abundance created expectations of upward mobility through effort. Theory suggests that prospects for upward mobility and the importance of effort (versus luck) in income generation make tax-based redistribution unfair and inefficient (Alesina and Angeletos, 2005; Piketty, 1995). Foreshadowing this view, Turner (1893) observed that on the frontier the "tax-gatherer is viewed as a representative of oppression," since the environment "produces antipathy to control." Billington (1974), a noted follower of Turner, wrote that on the frontier "every man was a self-dependent individual, capable of caring for himself without the fostering care of society," which "seemed just in a land that provided equal opportunity for all to ascend the social ladder."

Our paper contributes to the economics literature on individualism (e.g., Beck-Knudsen, 2019; Gorodnichenko and Roland, 2016; Greif, 1994; Olsson and Paik, 2016) and preferences for redistribution (see Alesina and Giuliano, 2011). We provide the first empirical study of America's culture of rugged individualism. Previous work in economic history examines Turner's ideas, but with a different focus (Ferrie, 1997; García-Jimeno and Robinson, 2011). We also contribute to a growing literature on the roots and persistence of cultural traits. Using a wealth of data and new methods, we identify the striking persistence of frontier culture and also provide causal evidence on how it took root historically.

Differences in rugged individualism across the U.S. have suggestive implications for cultural differences with Europe. The forces we emphasize—selective migration, cultural change, an advantage of individualism, and prospects for upward mobility—were arguably important in differentiating American culture. According to Turner (1893), "the Atlantic coast. . . was the frontier of Europe." Comparing the U.S. and Europe, Alesina et al. (2001) conjecture that "American anti-statism" may be linked to the frontier, which "strengthened individualistic feelings and beliefs in equality of opportunities rather than equality of outcomes." Our findings support this hypothesis. Moreover, the deeply-rooted culture that we identify may shed new light on a puzzle in American political economy: the stability of preferences for redistribution over the last 40 years despite significant increases in inequality (Ashok et al., 2015).

Finally, we advance a large literature in the social sciences inspired by Turner. Historians have produced many case studies of frontier populations and rich descriptions of life on the frontier. We provide systematic evidence on the distinctive features of frontier society and measure the historical prevalence of individualism for the first time. Social psychologists have used state-level data to study variation in contemporary individualism, comparing demographic features (Vandello and Cohen, 1999) or infrequent names (Varnum and Kitayama, 2011) between western and non-western states. We go beyond these broad geographic correlations by (i) tracking the frontier historically and introducing a county-level measure of total frontier experience, (ii) accounting for potential confounders, and (iii) disentangling the effects of selective migration and place-based exposure in shaping frontier culture.

Turner's work has attracted immense attention and vast criticism.[2] His narratives contain departures from the historical record, overblown statements, and ethnocentric biases. They paint an idealized portrait of frontiersmen and leave women and minorities out of the picture. The term "free land" appears often when, in fact, land was violently taken from Native Americans, and, in many areas, westward expansion was more about "conquest" than "settlement" (Limerick, 1988). These features of Turner's work may explain why his influence, while still pervasive in history textbooks and classic narratives, has waned in recent historical research. Our study provides empirical support for some important elements of the Frontier thesis, but it is not a general assessment of Turner's work nor an endorsement of its ideological overtones.

2 Mapping the History of the Frontier

We reconstruct the history of the frontier using Census data and GIS methods. This section explains how we track the frontier over time and create a county-level measure of total frontier experience.

From colonial times until the late 19th century, America underwent rapid population growth and

[2]For summaries and references, see, for example, Cronon (1987) and Larson (1993).

westward expansion. The Census Bureau report on the *Progress of the Nation* from 1890, a source of inspiration for Turner's classic 1893 essay, noted that the Thirteen Colonies were "the sources of supply for a great westward migration," as people "swarmed from the Atlantic coast to the prairies, plains, mountains, and deserts by millions during the last century." The report by Porter, Gannett and Hunt (1890) details the decade-by-decade push westward and includes vivid maps of population density (see Appendix Figure A.1). From 1790 to 1890, as the country's population increased from 3.9 to 62.6 million, its settled area grew from under 240,000 square miles to nearly 2,000,000, and its mean center of population moved from Washington D.C. to Decatur, Indiana—a westward shift of over 500 miles.

The *Progress of the Nation* also deemed the frontier closed by 1890. In a passage quoted in Turner's essay, it stated that "up to and including 1890 the country had a frontier of settlement, but at present the unsettled area has been so broken into by isolated bodies of settlement that there can hardly be said to be a frontier line." As one of the authors put it elsewhere, "the frontier line has disappeared . . . the settled area has become the rule and the unoccupied places the exception" (Gannett, 1893).

2.1 Locating the Frontier and Tracking its Movements

Prior research adopted simplifying definitions of the frontier. In a study of westward migrants in 1850 and 1860, Steckel (1989) identifies the frontier as the states of Minnesota, Iowa, Kansas, Texas, and those farther west. Ferrie (1997) studies migration between 1850 and 1870 and defines 90° west longitude as the frontier's eastern boundary. Kitayama et al. (2010) simply associate the frontier with western states.

We take a different approach. Following Porter et al. (1890) and Turner, we define the frontier as the line dividing settlements with population density of two or more per square mile from those with less.[3] We then define frontier counties as those meeting two criteria: (i) close proximity to the frontier line (100 kilometers in our baseline) so as to capture Turner's notion of the "frontier belt", and (ii) with population density below six people per square mile, a cutoff stipulated by Porter et al. as the beginning of established, post-frontier settlement.

These steps produce a geographically precise, time-varying measure of frontier status. As Turner noted, the frontier was "a form of society rather than an area." Life in such a society was isolating in two ways. Low density meant isolation from other people within a given location. Proximity to the frontier line meant isolation from population centers to the east, and in most cases limited interaction with the federal government. With such isolation came a lack of social infrastructure, making frontier life rough and dangerous. However, isolation also implied relative resource abundance and thus favorable prospects for upward mobility. This attracted pioneering settlers in search of opportunity. It also distinguished low density locations on the frontier. Low density locations in the settled eastern regions were not so isolated from urban centers and were unlikely to be resource-rich.

For each Census year beginning in 1790, we calculate county-level population density per square mile. For intercensal years, we interpolate population density by assuming a constant annual population growth rate that matches the decadal growth rate. We maintain consistent units of observation over time by harmonizing to 2010 county boundaries, but the location of the frontier is very similar when using

[3] Turner (1893) notes, "The most significant thing about the American frontier is, that it lies at the hither edge of free land. In the census reports it is treated as the margin of that settlement which has a density of two or more to the square mile. . . . We shall consider the whole frontier belt including the Indian country and the outer margin of the "settled area' of the census reports."

contemporaneous historical boundaries (see Appendix Figure A.2). The population counts exclude most Native Americans, who were generally not enumerated by the Census prior to 1900.[4]

Using annual county-level population densities, we locate the frontier through contour lines that divide counties with population densities above and below two people per square mile. Figure 1 plots these lines for 1790, 1820, 1850, and 1890. Full details on the underlying GIS procedure can be found in Appendix A. In order to closely approximate historical notions of the frontier described above, we discard all line segments less than 500 km as well as isolated pockets of low density counties within the main area of settled territory (to the east of the main frontier line).[5] Figure 2 shows the evolution of the resulting, main frontier lines in red for 1790–1890.

A second major frontier emerged on the West Coast, starting in California, in the mid-19th century. Spurred by the Gold Rush, this was a large, discontinuous leap in east-to-west expansion. Compared to frontier locations in the heartland, the West Coast frontier had a different type of isolation. It was much farther away from Eastern cities, but proximity to the ocean reduced transportation costs, facilitating flows of goods, people, and ideas. We omit this secondary frontier from the baseline analysis but later show that frontier experience has similar long-run effects in the West as in the heartland.

2.2 Total Frontier Experience

The westward movement of the frontier was fast at times, slow at others. Thus, some locations spent little time under frontier conditions, while others remained on the frontier for decades. This variation is central to identifying the long-run effects of frontier exposure.

To measure the duration of frontier experience for each county, we calculate the number of years spent in the frontier belt from 1790 to 1890. For each year, a binary indicator takes value one if a county is on the frontier as defined by the proximity and low density criteria explained above. The total frontier experience (TFE) for each county is the sum of indicators of frontier status from 1790 to 1890.

We set 1890 as the endpoint for measuring TFE following the *Progress of the Nation* report and Turner. While many places remained sparsely populated long after 1890, the effective isolation of the frontier did not persist with the same intensity. By 1890, transcontinental railroads connected both coasts, and armed conflict with Native Americans had faded. Federal irrigation efforts started soon after. For robustness, we consider a longer time frame for the measurement of TFE, changing the endpoint to 1950.

Figure 3 shows the spatial distribution of TFE for counties in our baseline analysis. Counties to the east of the 1790 frontier line are excluded since it is not possible to measure their TFE without detailed population data before the 1790 Census. TFE ranges from 0 to 63 years with a mean of 18 years and a standard deviation of 11. TFE varies widely both across and within states and is distinct from contemporary population density (see Appendix Figure J.1). For instance, Cass County, Illinois has TFE of 10 years

[4]The Census was not conducted in "Indian Territory" until 1900, and before that time, there were very few individuals enumerated as "Indian", even when some reservations were included in 1870 and 1880. This explains why the frontier remains stuck at the boundary of Oklahoma, which was not fully enumerated until 1900 (see Figure 2). Unfortunately, we did not find a way to circumvent this data limitation. In Section 4.4, we account for exposure to conflict with Native Americans, which was an important part of the frontier history in some regions.

[5]Our results are qualitatively similar when discarding isolated pockets of high density settlement to the west of the main frontier line. The 500 km cutoff discards some contour lines but retains other, large unconnected lines off of the main east-to-west frontier line, e.g., the ones spanning Maine in 1820 and Michigan in 1850. Results are robust to adopting other cutoffs or having no cutoff at all (see Appendix J.2).

and Johnson County, Illinois has TFE of 32 years, but the two counties have nearly identical population density today (see Appendix H for this case study).

3 Distinctive Frontier Society: Demographics and Individualism

This section offers new insight on the distinctive populations living on the frontier. Historians and sociologists have studied frontier demographics (e.g., Eblen, 1965). However, they typically focused on a specific place at a particular time, making it difficult to establish empirical regularities. We offer the systematic look at frontier populations across all censuses from 1790 to 1890.

Historical narratives suggest that frontier settlers were different from those living in settled areas. These accounts often portray young men, immigrants, and the less educated. We explore these traits using historical Census data compiled by IPUMS National Historical Geographic Information System (Manson et al., 2019).[6] Historical narratives also emphasize the individualistic culture of frontier populations. This cultural trait is difficult to measure historically. However, infrequent children's names provide a compelling proxy.[7]

We interpret infrequent names as individualistic. Infrequent names correlate strongly with other proxies for individualism.[8] We define as infrequent those names outside the top 10 within one's Census division; for robustness, we vary both the rank cutoff and the reference group.[9] Appendix K provides a list of common names for selected years (e.g., John and Sarah) as well as a random sample of infrequent names (e.g., Luke and Lucinda).

We take two additional steps to ensure that infrequent names effectively capture individualism. First, we remove variation associated with foreign names, which may be more common on the frontier due to the greater prevalence of immigrants. We restrict attention to children with native-born parents and also compare locations with similarly-sized immigrant populations. The latter accounts for the possibility that native-born adults may choose infrequent names used by immigrants in their community. Second, we exclude spelling variation by first adjusting enumerated names using a phonetic algorithm and then determining whether that name group is infrequent. This ensures that infrequent names are not confounded by misspellings, which may be more common on the frontier but unrelated to individualism.

[6] With the exception of immigrant shares, we measure these traits only for the white population. This maintains consistency across time periods given that non-white populations were not systematically enumerated prior to 1900.

[7] We use Census data with names from the US 100% samples (Ruggles et al., 2019) of the North Atlantic Population Project (NAPP) from the Minnesota Population Center (2019) for 1850 and 1880, and complete-count restricted-access data from IPUMS (Ruggles et al., 2019) for other years.

[8] Varnum and Kitayama (2011) show a positive cross-country correlation between infrequent names and Hofstede's widely used index of individualism. Beck-Knudsen (2019) shows that the names-based measure is strongly correlated with Hofstede's index as well as with the use of first- and second-person singular pronouns across 44 countries (and across regions within five countries). In Japan, Ogihara et al. (2015) shows a strong time-series correlation between the share of common name pronunciations and an index of individualism similar to the one proposed by Vandello and Cohen (1999).

[9] We choose the top 10 cutoff as a baseline following the social psychology literature (Varnum and Kitayama, 2011). With this measure, the majority of children have infrequent names (e.g., 57 percent of boys and 60 percent of girls in 1850). Hence, an "infrequent name" may not be a very unusual name but simply one that is not very common. In any case, the measures for different cutoffs are highly correlated, and results are qualitatively similar across all of them (see Appendix B.3).

3.1 Basic Patterns

We estimate the frontier differential in demographic trait x for county c in Census division d at time t:

$$x_{cdt} = \alpha + \beta \text{ frontier}_{ct} + \theta_d + \theta_t + \varepsilon_{cdt}, \quad (1)$$

where frontier$_{ct}$ is frontier status, and θ_d and θ_t are Census division and year fixed effects, respectively. Panel (a) of Table 1 reports estimates of β, the frontier differential, for each of six outcomes.

Frontier populations have significantly more males, prime-age adults, and foreign-born. Frontier counties have 0.19 additional males for every female relative to non-frontier counties where the average sex ratio is 1.09 (column 1). The population share of prime-age adults (15–49 years old) in the population is 2.6 percentage points (p.p.) higher than in non-frontier counties, for which that share is around 46 percent (column 2). These patterns are consistent with historical accounts of hostile frontier conditions leading to the selective migration of young men. Additionally, frontier counties have 6.3 p.p. higher foreign-born population shares than the average non-frontier county where 7 percent of residents are immigrants (column 3). Meanwhile, literacy rates are not significantly different on the frontier (column 4). While this runs counter to the "safety valve" theory of selective low-skilled migration (see Ferrie, 1997; Turner, 1893), literacy may simply be a noisy measure of skill.

Furthermore, individualistic names are more prevalent on the frontier. In frontier counties, 2.2 p.p. more children age 0–10 have infrequent names relative to the average of 63 percent in non-frontier counties (column 5). This finding is robust to adjusting reported names for their phonetic sound using the Philips (1990) metaphone algorithm (column 6).[10] It is also robust to accounting for the differential presence of immigrants on the frontier using a matching-type exercise that compares frontier and non-frontier counties with nearly identical foreign-born population shares in the given Census year (see even-numbered columns in Appendix Table B.1).[11]

We further characterize the frontier differential by unbundling the two dimensions of isolation: (i) proximity to the frontier line and (ii) low population density. Panel (b) of Table 1 estimates

$$x_{cdt} = \alpha + \beta_1 \text{ near frontier line}_{ct} + \beta_2 \text{ low population density}_{ct} + \theta_d + \theta_t + \varepsilon_{cdt}, \quad (2)$$

where *near frontier line*$_{ct}$ is an indicator for having a centroid within 100 km of the frontier line at time t, and *low population density*$_{ct}$ is an indicator for population density below six people per square mile. The results suggest that both features of the frontier contribute to its distinctive demographics and individualism.[12] As counties transition from frontier conditions to more established settlement, these distinctive traits subside (see Appendix Figure G.1). Yet, as we show later, the duration of exposure to the frontier has persistent implications for culture.

[10] To get a sense of how this algorithm works, consider the common name of John. This metaphone adjustment groups the following variants on "John"—some misspellings and others nicknames—into a single metaphone "JN": Jon, Jno, Johhn, Johnnie, Johnie, Johny, Johnny, Jonnie, Johney, Jone, Johne, and Jonny, among others.

[11] The greater prevalence of infrequent names resonates with a celebrated trilogy about frontier life. *The Awakening Land*, by Pulitzer Prize-winning novelist Conrad Richter, follows the native-born Luckett family that moved from Pennsylvania to the Ohio Valley frontier in the late 1700s. Noted for his painstaking historical research, Richter chose infrequent names for family members: the boys were named Chancey, Worth, and Wyitt, and the girls were named Ascha, Sayward, Sulie.

[12] Column 4 shows that the null for illiteracy in panel (a) is due to offsetting positive effects of low density and negative effects of proximity. This pattern does not arise for other outcomes and suggests scope for further work on the safety valve hypothesis.

3.2 Validating the Population Density Cutoffs

This section further corroborates the distinctiveness of frontier society. The population density cutoffs defining the frontier may seem arbitrary. However, with modern econometric tools, we are able, for the first time, to validate the definition put forward in the landmark *Progress of the Nation* report in 1890.

Each panel in Figure 4 shows a local linear regression function, $g(\cdot)$, and 95 percent confidence band based on the partially linear Robinson (1988) estimator:

$$x_{cdt} = \alpha + g(\text{population density}_{ct}) + \theta_d + \theta_t + \varepsilon_{cdt}. \tag{3}$$

In panel (a), the sex ratio approaches 1.6 in the most sparsely populated counties and declines sharply until population density reaches 3–4 people/mi^2. The slope of $g(\cdot)$ then abruptly flattens out as the sex ratio stabilizes at around 1.05–1.1 males for every female. In panel (b), the prime-age adult share declines sharply as we move towards densities of 2–3 people/mi^2 and levels off thereafter. The foreign-born share (c) and illiteracy rates (d) exhibit more linear, downward-sloping curves. However, panels (e) and (f) show stark nonlinear shapes for both measures of individualistic names.

Together, the graphs in Figure 4 point to structural breaks at population density levels consistent with the frontier definition in the 1890 Census report.[13] Chow (1960) tests soundly reject the null hypothesis of a constant effect of population density above and below 6 people/mi^2 (the upper bound of frontier settlement according to Porter et al., 1890). Zivot and Andrews (2002) tests identify structural breaks in the 2–6 range. In 1850, for example, the sex ratio breaks at 2.7 people/mi^2 and the adult share at 2.0.

Another distinctive feature of frontier society was its limited government presence (see Appendix G.3). Frontier counties had lower taxation and public spending per capita, both of which exhibit sharp structural breaks around 2–6 people/mi^2. This suggests that the institutions of local government structurally change once counties surpass the low levels of frontier population density. There were also fewer post offices, railroads, and canals on the frontier, but these state- and federally-provided public goods vary smoothly with population density.

4 Long-Run Effects of Frontier Experience on Culture

Exposure to frontier conditions laid the foundation for a persistent culture of rugged individualism. This section identifies long-run effects of TFE on individualistic names, policy preferences, and voting behavior. Our analysis is motivated by theories of cultural persistence. While individualism on the historical frontier could have dissipated, the frontier experience may well have shaped the subsequent evolution of culture. With longer exposure to frontier conditions came greater scope for rugged individualism to take root through a set of mechanisms we explore in Section 5. With multiple equilibria and path dependence, the early stages of cultural formation could leave a lasting imprint.

[13] Appendix Figure G.2 provides similar evidence, though with less stark nonlinearities, for proximity to the frontier line.

4.1 Estimating Equation

Our estimating equation relates total frontier experience to modern proxies for rugged individualism:

$$y_c = \alpha + \beta \text{ total frontier experience}_c + \mathbf{x}_c'\boldsymbol{\gamma} + \theta_{s(c)} + \varepsilon_c, \tag{4}$$

where y_c is some cultural trait in county c (e.g., individualism or preferences for redistribution). Total frontier experience (TFE) is the amount of time, scaled *in decades*, that a given county remained on the frontier. Our sample, seen in Figure 3, includes all counties for which the 1790–1890 period contains their entire frontier experience as discussed in Section 2.2. In baseline specifications, $\theta_{s(c)}$ is a state fixed effect, and $\mathbf{x}_c$ includes predetermined or fixed county-level covariates including latitude, longitude, county area, average rainfall and temperature, elevation, potential agricultural yield, and distance to rivers, lakes, and the coast. The coefficient β therefore identifies a local effect of TFE after accounting for geoclimatic factors that may correlate with both TFE and modern culture. Following Bester et al. (2011), standard errors are clustered on 60-square-mile grid cells that cover counties in our sample.[14]

We measure contemporary culture and policy outcomes using numerous data sources, including three nationally representative surveys: the Cooperative Congressional Election Study (CCES), the General Social Survey (GSS), and the American National Election Study (ANES). These surveys are staples in the social science literature, often asking different questions about similar underlying preferences. See Appendix K for details, including a discussion of geographic coverage.

The main threat to causal identification of β lies in omitted variables correlated with both contemporary culture and TFE. We address this concern in four ways. First, we rule out confounding effects of modern population density. Second, we augment $\mathbf{x}_c$ to remove cultural variation highlighted in prior work. Third, we show that unobservables are unlikely to drive our results. Finally, we use an IV strategy that isolates exogenous variation in TFE due to changes in national immigration flows over time.

4.2 Persistent Individualism

Nearly five decades after the closing of the frontier, individualistic children's names are more pervasive in counties with greater TFE. Table 2 reports the effect of TFE on the share of children age 0–10 in 1940 with infrequent names (panel a) after the metaphone adjustment (panel b).[15] We normalize these outcome variables so that standard deviation effect sizes can be read directly from the coefficients.

Each additional decade of TFE is associated with a significantly higher share of individualistic names by 1940. The baseline specification with geoclimatic controls in column 2 of panel (a) implies 1 p.p. more children with infrequent names when moving across the interquartile range of TFE (11 vs. 24 years). We find similar effect sizes for the metaphone-adjusted measure in panel (b). In both panels, relative to column 1, the geoclimatic controls leave the coefficient unchanged despite a large increase in the R^2. This pattern is consistent with limited selection-on-unobservables according to the parameter δ reported in the table; Oster (2019) suggests $|\delta| > 1$ leaves limited scope for unobservables to explain the results.

[14] As detailed in Appendix F, inference is robust to several alternative approaches to adjusting for spatial correlation, including the Conley (1999) spatial HAC estimator with bandwidths from 100 to 1000 kilometers.

[15] Unfortunately, the 1940 Census is the latest round that provides information on names. Although the Social Security Administration releases baby name counts by state, it does not do so at the county level.

Importantly, the greater prevalence of individualistic names in high-TFE counties is not due to differences in contemporary population density or the prevalence of foreign-born. Columns 3 and 4 of Table 2 bear this out using a matching-type exercise. For each county c within state s, we find the county c' with the most similar population density (column 3) or foreign-born share (column 4) and create matched pairs in ascending order. We then create an indicator for this county pair (c, c'). Finally, we add these 1,018 fixed effects to our baseline specification from column 2. Even in this very demanding specification, the estimated effects of TFE remain statistically and economically significant, with $|\delta|$ well above 1. In other words, the effects of TFE on individualism in the mid-20th century are not merely a reflection of differences in density or immigrant populations that may have persisted from the frontier era.

Furthermore, the results in Table 2 are robust to many alternative measures of individualism inherent in children's names. We demonstrate this in Appendix Table B.3 considering infrequent names with (i) different geographic reference groups (columns 1–4), (ii) a cutoff at the top 100 rather than 10 (column 5), and (iii) the metaphone-adjusted analogues of (i) and (ii) (columns 6–10). We also find similar results using an alternative names-based proxy for individualism that does not depend on the reference group: the absence of inherited names. Patronymic (father-to-son) and matronymic (mother-to-daughter) names may reflect a non-individualistic emphasis on interdependence within the family.[16] High-TFE counties record lower use of patronymic/matronymic names (column 11), and this holds when restricting the outcome to first-born sons and daughters (column 12). The findings across these 12 alternative outcomes are additionally robust to the matching-type exercises with contemporary population density and foreign-born shares (panels b and c).

Moreover, these findings are not artifacts of aggregation bias. Appendix Table B.4 reports individual-level regressions with fixed effects for children's age, birth order and gender. The effect sizes are similar to county-level regressions. At the individual level, we can also control for family surname with nearly 400,000 fixed effects. This recovers the effects of TFE across all households with a given surname, e.g., comparing Smiths in high- versus low-TFE counties. This powerful test leaves our main findings unchanged (panel d) even when coupled with the population density matching-type exercise (panel e).

Together with the findings in Section 3, these results suggest that individualistic names were not only more pervasive in frontier areas historically but also more prevalent in the long run in areas with greater TFE. Indeed, the effect of TFE on infrequent name choices can be seen in the early 1900s with little change thereafter (Appendix Table B.5). This points to the persistence of the early frontier culture of individualism long after frontier conditions abated.

This individualistic culture can be seen in later survey data as well. The 1990 ANES asks whether respondents identify more strongly with self-reliant or cooperative behaviors. Those in high-TFE counties are significantly more likely to favor self-reliance (see Appendix Table B.2). As we show next, frontier culture further manifests in preferences for redistribution and government intervention more broadly.

4.3 Opposition to Redistribution and Regulation

Rugged individualism has profound implications for contemporary politics. This section shows that TFE is associated with preferences for small government and also with lower tax rates in practice. Opposition

[16] Brown et al. (2014) show that the prevalence of patronymic/matronymic names correlates strongly with collectivism across U.S. states. We focus on parental forebears as we only observe a small subset of children with co-resident grandparents.

to specific regulations can be linked to salient aspects of frontier culture historically. We connect these anti-statist preferences to the partisan divide and show that TFE helps explain the growing strength of the Republican Party in the American heartland.

Our analysis spans a wide set of measures that reflect a similar underlying opposition to government intervention. For all outcomes, we report estimates of equation (4) controlling for the geoclimatic characteristics used in column 2 of Table 2 as well as individual demographics (age, age squared, gender, and race dummies) and survey-wave fixed effects where relevant.

Redistribution and Limited Government. Table 3 shows that greater TFE is associated with stronger contemporary opposition to income redistribution. In column 1, we use ANES data from 1992 and 1996, which asks whether respondents would like to see "federal spending on poor people be increased, decreased (or cut entirely) or kept about the same." Around nine percent of individuals would like to see such redistributive spending decreased. Each additional decade of TFE is associated with one additional p.p. increase in support of cuts. Column 2 provides complementary evidence from the CCES measure of support for cutting state spending on welfare. Following Alesina and La Ferrara (2005), column 3 uses the GSS to measure support for redistribution on a scale from 1 to 7 (with 1 being that the government should not be engaged in redistribution and 7 being that the government should reduce income differences through redistribution). Each additional decade of TFE is associated with around 0.02 standard deviations lower support for redistribution. These effect sizes are akin to a 5–10 year age gap in preferences among respondents (with older respondents more in favor of welfare spending cuts).

Areas with greater TFE also display stronger fiscal conservatism. Column 4 uses a CCES question on whether individuals would prefer to cut domestic spending or to raise taxes to balance the federal budget. Column 5 uses an index constructed from several GSS questions about whether the government spends too much on an array of public goods and social transfers. Across both outcomes, individuals are significantly more opposed to high levels of government spending in areas with greater TFE.

Importantly, these reported preferences line up with contemporaneous policy outcomes. In particular, each decade of TFE is associated with a reduction in property tax rates from 2010–2014 by 3.3 percent of the mean (column 6).[17] This is an economically meaningful effect given that much of the variation in tax rates lies across rather than within states. It equals roughly the within-state difference in tax rates between counties that are 10 percent more versus less aligned with the Republican Party, a policy outcome we consider next.

Such strong opposition to redistribution in high-TFE counties also translates into greater support for the Republican Party between 2000 and 2016. While people vote Republican for many reasons, self-reliance and small government have been dominant party themes since the late 1990s (Gentzkow et al., 2019). Column 7 of Table 3 shows that each decade of TFE is associated with a 2 p.p. higher Republican vote share relative to the mean of 60 percent over the five presidential elections from 2000 through 2016.[18]

[17] These rates are estimated by the National Association of Home Builders based on the American Community Survey (ACS) in 2010–14.

[18] This effect size is in line with individual-level regressions using degree of stated support for the Republican Party in the CCES. Using the CCES 2007, 2012, and 2014 survey rounds, we construct an indicator equal to one if the respondent identifies as a "strong Republican" on a seven point scale ranging from "strong Democrat" to "strong Republican" with around 17 percent of individual–years reporting the latter. The estimates imply that an additional decade of TFE is associated with around 4.5 percent greater intensity of strong Republican support.

For perspective, the 2 p.p. effect is roughly the difference in population-weighted, average county-level vote shares in Iowa (48.4 percent) and Wisconsin (46.3 percent) over these five elections.

Voting and Partisan Issues. Several results further elucidate the connection between frontier culture, voting, and partisanship. First, we examine earlier elections when the liberal–conservative divide did not cleanly map into a partisan divide. TFE exhibits little relationship with Republican Party support until around 2000 (see Figure 5). This sharp break coincides with growing polarization after the mid-1990s when Republican leaders launched the "Contract with America" platform for political change, which emphasized tax cuts, balanced budgets, and welfare reform (Gentzkow et al., 2019). After 2000, the effects of TFE trend upward, with significant increases from each election to the next.[19]

The growing electoral imprint of TFE may reflect both supply- and demand-side drivers of political change. Congressional speeches compiled by Gentzkow et al. (2019) point to an increased influence of frontier culture. From the mid-1990s onward, Republican legislators from congressional districts with greater TFE are more likely to discuss topics associated with the "Contract with America" (see Appendix Table C.1). This may be due to an increase in the likelihood of such candidates winning elections or to a change in speech among candidates that would have won otherwise. Either way, this shift against big government among Republican representatives is consistent with strong voter support for such positions in high-TFE regions as seen in Table 3.

During this period of growing partisanship, the Republican party has increasingly opposed not only tax redistribution but also government intervention more broadly. Many salient partisan issues, in fact, resonate with anti-statist principles of rugged individualism. We explore four such issues using the CCES. Beliefs in self-reliance might increase opposition to (1) the Affordable Care Act (ACA) and (2) increases in the minimum wage. Meanwhile, opposition to (3) the ban on assault rifles is connected to views about the right to self-defense, and opposition to (4) Environmental Protection Agency (EPA) regulations on pollution is connected to views about manifest destiny and the pursuit of self-interest. The results in Appendix Table C.2 show that places with greater TFE display significantly stronger opposition to these four hot-button regulations.

Finally, despite limited electoral effects of TFE before 2000, there were four notable exceptions, each with clear links to frontier culture (see Figure 5). The first was in 1928 when Republican Herbert Hoover, who popularized the notion of "rugged individualism", performed relatively better in high-TFE counties. The second was in 1944 when Republican Thomas Dewey campaigned against the inefficiencies and excesses of Democrat Franklin Roosevelt's New Deal programs. The third was in 1972 when Republican Richard Nixon faced Democrat George McGovern, an anomalously progressive liberal candidate who campaigned for a time on giving every American citizen $1,000 per year (akin to a universal basic income). The fourth was in 1976 when high TFE favored the Democratic candidate, Jimmy Carter, who came from a farming family in Georgia descended from the original settlers of Virginia.

Interpretation. Overall, the findings in this section paint a rich picture of how frontier settlement left

[19]Taking a long difference from 2000 to 2016, the average county in our sample exhibits a 9 p.p. shift towards Republican candidates, and each decade of TFE is associated with an additional 1.6 p.p. increase. Alternatively, an interquartile shift in TFE implies an additional 2.2 p.p. Republican Party shift. As a benchmark, Autor et al. (2017) find that an interquartile shift in exposure to import competition from China induces a 1.7 p.p. Republican shift over the same period. Using the original data from Autor et al. (2013), a single regression with both measures puts the TFE effect at around one-quarter as large as the effect of the China shock, with both effects statistically and economically significant.

an indelible mark on America's cultural landscape. As a summary takeaway, we estimate a mean TFE effect of 0.15 standard deviations on the combined index of infrequent names, Republican vote shares, and property tax rates (using the Kling et al., 2007, approach). In combining these outcomes to define a culture of "rugged individualism," we note that individualistic names are strongly associated with higher Republican vote shares and lower property tax rates.[20] In other words, the effects in Table 2 are identified from essentially the same cross-county variation in Table 3, pointing to the close connection between individualism and opposition to redistribution.

There are of course policy preferences for which the individualistic frontier culture does not have clear implications. For illustration, we consider a few foreign policy issues as placebo outcomes: support for U.S. military intervention abroad in the case of genocide or civil war (35 percent in the CCES), opposition to the Iran sanctions regime (20 percent), and opposition to the U.S.-Korea Free Trade Agreement (45 percent). Estimating our baseline specification, we find relatively precise null effects of TFE on these three measures (-0.004, -0.003, and 0.003, respectively).

Importantly, the long-run effects on rugged individualism only materialize for groups able to capitalize on the opportunities afforded by the frontier historically. In Table 5, we find precise null effects of TFE for African American respondents across the six measures of opposition to redistribution and regulation in the CCES. These results support our interpretation of the origins and persistence of frontier culture. A large share of today's black population in the U.S. trace their familial roots to slavery. Slaves of course faced extreme barriers to geographic and socioeconomic mobility, and many of these barriers persisted for blacks in the postbellum period. As a result, the mechanisms linking frontier experience to rugged individualism (e.g., selective migration, upward mobility through effort) would have been irrelevant to blacks living in many high-TFE regions, especially in the South.[21]

The black–white gap in the effects of TFE raises the possibility that racial resentment may explain white opposition to redistribution in high-TFE counties. For example, opposition to affirmative action for African Americans is often linked to beliefs about the role of effort in generating income. Using the CCES, we find a significant association between this type of racial resentment and TFE. However, controlling for contemporary population density undoes this correlation, thus pointing to an urban–rural divide rather than a high–low-TFE divide.[22] In contrast, our key findings in Table 3 cut across the urban–rural divide as we show next.

[20]Conditional on state fixed effects and our baseline controls, a one standard deviation (s.d.) increase in infrequent names is associated with a 0.42 s.d. increase in Republican votes (0.24 s.d. decrease in property tax rates).

[21]The results in Table 5 are driven largely by counties in the South (Census region). Outside of the South, blacks and whites display similar effects of TFE. This may be due in part to the selective migration of blacks out of the South and into frontier areas in the late 1800s. While still subject to greater restrictions on upward mobility than whites, these self-selected black migrants were arguably more exposed to the influence of frontier conditions than those remaining in the postbellum South. See Billington and Hardaway (1998) for a rich exploration of the history of African Americans on the frontier.

[22]The 2010, 2012, and 2014 rounds of the CCES make two statements about racial resentment and ask respondents their degree of agreement on a scale from strongly agree to strongly disagree: (i) "The Irish, Italians, Jews and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors." (64 percent somewhat or strongly agree), and (ii) "Generations of slavery and discrimination have created conditions that make it difficult for Blacks to work their way out of the lower class." (51 percent somewhat or strongly disagree). While TFE exhibits a significant positive association with both measures in our baseline regression (0.010** and 0.012***, respectively), a simple linear control for 2010 population density renders the estimates null and insignificant (0.001 and 0.004, respectively).

4.4 Robustness

This section bolsters our interpretation of the causal pathway from historical frontier experience to contemporary culture. We focus on four outcomes: individualistic names, a simple mean index of preferences over six CCES outcomes, property taxes, and the Republican vote share.

Disentangling Population Density. Differences in population density across locations can be very persistent. Given the well-known cultural divide across rural and urban areas, there is a natural concern that contemporary population density may confound the effects of TFE. Here we disentangle the effects of historical frontier settlement from those of present-day density, showing the robustness of our results.

Table 4 controls for contemporary density in several ways: linearly (column 2), deciles within state (column 3), and county-pair fixed effects (column 4). The latter specification—used in column 3 of Table 2—is very demanding and leaves limited identifying variation. Yet, TFE has a statistically and economically significant effect on individualistic names and Republican vote shares. The effects on mean government preferences in CCES and county-level property taxes are no longer significant, which is not surprising given that these measures exhibit less variation within state.

The remaining columns of Table 4 further establish that the effects of TFE are driven by the history of frontier settlement rather than simply the long-run persistence of low density. TFE has similar effects in urban and rural areas, splitting the sample into counties above and below the 90th percentile of urban population shares (columns 5 and 6). Finally, column 7 separates the history of low density—the number of decades with density below 6 people/mi^2—from TFE. Recall that low density was one of two defining features of frontier locations, proximity to the frontier line being the other. The coefficient on TFE remains significant, indicating that both dimensions of frontier history are important.

Additional Controls. Beyond population density and our baseline geoclimatic controls, there are of course other factors that may be correlated with both TFE and rugged individualism. Appendix Table B.6 incorporates many such factors: ruggedness (Nunn and Puga, 2012); rainfall risk (Davis, 2016); portage sites (Bleakley and Lin, 2012); mineral resources (Couttenier and Sangnier, 2015); conflict with Native Americans; the prevalence of slavery; the sex ratio (Grosjean and Khattar, forthcoming), immigrant share; Scotch-Irish settlement (Grosjean, 2014); birthplace diversity; the timing of railroad access; and the employment share in manufacturing. These controls add substantial explanatory power but leave the estimated effects of TFE largely unchanged.

Instrumental Variables Strategy. It is of course impossible to control for all plausible correlates of culture that might also have shaped TFE. With the goal of ruling out unobservable location-specific confounders, we introduce an IV strategy that isolates plausibly exogenous variation in TFE.

Our IV is based on historical shocks to the settlement process driven by immigrant inflows to the U.S. Immigrants contributed to westward expansion by exerting population pressure on the eastern seaboard and by going west themselves. The ebb and flow of immigrant arrivals thus determined the time it took for frontier locations in different periods to become established settlements. We can also isolate push factors by predicting migrant outflows from Europe based on climate shocks (following Sequeira et al., 2020). For each county, the IV captures weather-induced emigration flows to the U.S. starting just before the onset of local frontier settlement. These time-varying, national population shocks are unrelated to local conditions of frontier counties and help move us closer to a causal interpretation.[23] These national immigration shocks explain considerable variation in TFE, and, when used as an IV, deliver significant effects of TFE that are slightly larger but statistically indistinguishable from the OLS estimates for our core outcomes. Appendix D describes the IV and the results in full detail.

Regional Variation and the 20th Century Frontier. The effects of TFE on rugged individualism cut across well-known cultural divides in the U.S. Appendix Table B.7 shows that our findings are consistent across three distinct cultural regions of the country: the Midwest, South, and West. Even within the West Coast, high-TFE counties exhibit greater rugged individualism. Such stability across regions is reassuring and points to a specific cultural legacy of settlement history that is shared across an otherwise remarkable diverse country.

When extending the measurement of TFE to 1950, we find somewhat smaller long-run effects on culture.[24] This extension incorporates counties first settled in the early 20th century by which time frontier conditions had changed. Transcontinental railroads and improved communications meant that frontier locations were effectively less isolated than they were historically. According to Lang et al. (1995), "the modern-day [post-1890] frontier is not the nineteenth-century one. It is smaller, more law-abiding and regulated, less isolated, less rugged, and less dangerous," and moreover, "the frontier has not for generations been the dream of those who seek a fortune or a new life." In other words, the 20th century offered relatively less scope for selective migration and treatment effects of frontier life to engender a culture of rugged individualism.

5 The Roots of Frontier Culture

This section explores how "rugged individualism" took root on the American frontier. We use historical Census data to analyze two leading explanations: Section 5.1 examines selective migration, and Section 5.2 identifies causal effects of frontier exposure.[25] Put simply, the frontier attracted individualistic people, and life on the frontier made its residents even more individualistic over time. Both findings are consistent with an advantage of individualism on the frontier, and Section 5.3 provides evidence of differential returns to individualism.[26]

Our analysis below requires individual-level migration data, which we construct in two ways. First, we use information on children's state of birth to infer migration patterns of their parents as in Collins and Zimran (2019). Second, we track individuals over time by linking across Census rounds using an algorithm developed by Feigenbaum (2016) and detailed in Appendix K.

[23] To construct the instrument, we determine the first year in which each county is within 110 km of the frontier line. At this time, the county's local conditions do not affect the contemporaneous process of westward expansion, but the moving frontier is getting close. We then consider the average annual immigrant inflow in the next 30 years. Nearly 85 percent of counties exit the frontier within that time frame. Results are similar for other windows.

[24] See Appendix J.2 for alternative measures of TFE and results excluding 40 counties with zero TFE.

[25] See footnote 7 and Appendix K for details on the data sources used in this section.

[26] Appendix J.4 examines a competing, disease-based explanation for the origins of individualism rooted in biology and known as the parasite-stress theory of values (Fincher and Thornhill, 2012). Using data on disease and illness in the 1880 Census, we do not find evidence in support of this explanation for the frontier differential in individualism.

5.1 Selective Migration

Selective migration increased the prevalence of individualism on the frontier. Using complete-count Census data from 1850–80, we show here that households moving to the frontier had children with more individualistic names than households that remained in settled areas. The opposite holds when looking at movers from the frontier to settled areas.

We estimate the time at which a household moves to the frontier based on the contemporaneous county of residence and differences in the reported birth state of children. Consider a household living in frontier county c in Iowa in 1850 whose first child was born in Virginia in 1842 (a non-frontier state at the time), and their second child was born in Iowa in 1848. We date this household's arrival to the frontier in 1845. If this household did not have a second child, we would date their time-at-move to 1846.[27] An analogous procedure can be used to identify movers from frontier to settled areas.

Table 6 compares the prevalence of infrequent names among children who moved to the frontier relative to children that remain in settled counties. The estimating equation is:

$$\text{infrequent name}_{ict} = \alpha + \beta\, frontier\ migrant_{ict} + \mathbf{FE} + \varepsilon_{ict}, \tag{5}$$

where the binary dependent variable equals one if child i residing in county c in Census year t has a name that falls outside the top 10 nationally in that decade. We consider all children age 0–10 with native-born parents in keeping with the restrictions earlier in the paper. The *frontier migrant* indicator equals one if c is on the frontier in t and i was born in a state with no frontier counties at his/her time of birth. This indicator equals zero for all children of households living in settled, non-frontier counties. Standard errors are clustered by county. The $\mathbf{FE}$ vector includes fixed effects for birth year×gender, birth order, and, in even-numbered columns, child birth state.

Columns 1–2 of Table 6 show that individualists are more likely to move to the frontier. The estimate of β in column 1 is around 3.2 p.p. while the mean for stayers in settled areas is 65 percent. This result holds conditional on child birth state FE (column 2), which captures heterogeneity in individualism across migrants' previous states of residence.

Meanwhile, columns 3–4 show that non-individualists are more likely to leave the frontier. The specification here replaces the indicator for *frontier immigrant* in equation (5) with an indicator for *frontier out-migrant*, which equals one if child i lives in county c that is not on the frontier in t and i was born in a state with at least one frontier county at his/her time of birth. This indicator equals zero for all children of households living in frontier counties. The column 4 estimate of -2.9 p.p. demonstrates significant selective outmigration of non-individualists. Together, the results in Table 6 suggest that selective migration contributed to the greater prevalence of individualism on the frontier historically.

5.2 Frontier Exposure and Cultural Change

The frontier not only attracted individualistic settlers but also made its residents more individualistic. We develop two strategies to identify such a causal effect of frontier exposure on cultural change. Both exploit variation in the length of exposure to frontier conditions, one in adulthood and the other in

[27] This approach misses moves between counties within the same state. In Appendix I, we present complementary results based on a smaller, linked-sample of households where we can identify origin and destination counties in 1870 and 1880.

childhood. With longer exposure comes greater scope for the frontier environment to affect cultural traits. This is the same notion underlying the long-run effects of total frontier experience on culture at the county level. Here, we identify short-run effects of exposure at the individual level.

First, we use an *event-study* approach that exploits within-household variation to show that parents give their children increasingly individualistic names after arrival to the frontier. Second, we use an *age-at-move* approach that exploits cross-sibling variation in the time at which their parents chose to move the family to the frontier. Tracking siblings 30 years later, we find that people with longer childhood exposure to the frontier give their children more individualistic names. Both approaches account for household-specific, time-invariant individualism.

(i) Event Study: Adulthood Exposure. Our first strategy identifies changes in how parents name children born after versus before moving to the frontier. Specifically, we estimate the following equation that relates the name given to child i, born in year $\tilde{t}+j$, to the year $\tilde{t}$ in which his/her household h moved to frontier county c at some time prior to Census year t:

$$\text{infrequent name}_{iht} = \alpha + \sum_{j=-20}^{20} \beta_j 1(born\ in\ \tilde{t}+j)_{ih} + \theta_h + \mathbf{x}_i'\boldsymbol{\eta} + \varepsilon_{iht}. \quad (6)$$

The household fixed effects, θ_h, absorb all time-invariant characteristics that affect h's choice to migrate to the frontier and its individualism. The $\mathbf{x}_i$ vector includes child gender, birth order, and birth cohort trends. We pool across Census years 1850–80 and consider all kids ages 0–20 in 1850 and ages 0–10 in 1860, 1870 and 1880 to avoid double counting. The dependent variable here, and in the age-at-move approach below, is again based on the top 10 gender- and decade-specific names for white children with native-born parents. Standard errors are clustered by county.

The β_j coefficients in equation (6) identify differential individualism across siblings' names with respect to the year $\tilde{t}$ at which h moved to the frontier. The estimates are normalized with respect to children named in the year before arrival on the frontier, such that β_5, for example, identifies how much more likely it is to observe an infrequent name for a child born to family h five years after arrival on the frontier relative to their child born one year before leaving a settled area. The controls $\mathbf{x}_i$ help rule out general trends in infrequent names across time and birth order, thereby isolating within-household variation that is most plausibly related to changes in frontier exposure.[28]

We also estimate an equation with continuous measures of birth years relative to move:

$$\text{infrequent name}_{iht} = \alpha + \beta_{pre}(years\ until\ move)_{ih} + \beta_{post}(years\ after\ move)_{ih} + \theta_h + \mathbf{x}_i'\boldsymbol{\eta} + \varepsilon_{iht}. \quad (7)$$

This specification identifies pre- and post-move trends but is less flexible than (6).

To estimate equations (6) and (7), we require households with at least two children and at least one of them born before the household moved to the frontier. Our sample consists of 57,097 children living in 16,901 households.[29] Consider, for example, a household on the Iowa frontier in 1850 with four children:

[28]This specification is similar to the one in Abramitzky et al. (forthcoming) who relate time spent in the U.S. to the Americanization of names given to native-born children by foreign-born mothers. Whereas their study estimates separate equations for children born pre- and post-move to the U.S., we combine the two in an event-study design centered on the time of move.

[29]The sample is relatively small because the frontier comprised a small share of the entire U.S. population at any given time in the 1800s, and the restriction to frontier migrants with children born prior to moving further reduces the sample size.

John born in 1840, Mary in 1843, Lisa in 1847, and Ruben in 1850. We see John and Mary are born in Virginia and Lisa and Ruben in Iowa. Hence, we impute $\tilde{t} = 1845$ and $j = -5$ for John, -2 for Mary, +2 for Lisa and +5 for Ruben.

The key identifying assumption is that the trend in individualistic names in household h would not have changed had the household not moved to the frontier. While this counterfactual is of course unobservable, the lack of pre-trends in Figure 6 and precise zero on β_{pre} in Table 7 are reassuring. This goes against the concern that parents had already started becoming more individualistic prior to moving and hence for reasons unrelated to frontier exposure. Although frontier migrants are self-selected on prior *level* differences in individualism (see Section 5.1), such migrants are not self-selected on prior *growth* in individualism.

As migrant families reside on the frontier for longer periods of time, their children's names becoming increasingly individualistic. Figure 6(a) reveals a stark trend break in individualistic names within households after moving to the frontier. A child born one decade after their parents moved to the frontier is nearly 8 p.p. more likely to have an infrequent name than their sibling born one year prior to moving. Assuming linearity in Table 7, each additional year of exposure to the frontier increases the likelihood of giving their next child an infrequent name by 0.7 p.p. (column 1).

These estimates suggest that frontier conditions increased parents' own individualism *or* increased their preferences over their children's future individualism. In either case, the prevalence of individualism increased over time within families. Further results below support a causal interpretation.

Robustness. These baseline results are robust to accounting for time trends in individualistic names in several ways: five-yearly birth cohort FE (column 2 of Table 7), three-yearly cohort FE (column 3), child birth order (column 4), and birth order with five-yearly cohort FE (column 5).[30] Additionally, we control for pre-move-state trends in infrequent children's names (Figure 6(b) and column 6 of Table 7). These gender- and cohort-year-specific means account for trends in individualism had the family not left their origin state. Across these checks, we continue to find both a lack of pre-trends and a significant increase in individualistic names as parents spend more time on the frontier.

Moreover, the patterns in Figure 6 help to address a remaining concern about endogeneity. Suppose a household experienced an shock that simultaneously led them to move to the frontier and increased their future individualism irrespective of frontier exposure. For example, they suddenly reap large returns to prior investments with little help from neighbors and local government, inducing them to embrace individualism and also enabling them to move to the frontier in search of opportunities. While such unobservable shocks are impossible to rule out, they seem inconsistent with our findings. In particular, shocks like these would lead to a jump in individualism right around the time of moving to the frontier, whereas Figure 6 points to growing individualism with each additional year of frontier exposure.

While the patterns in Figure 6 are consistent with causal exposure effects, they are identified on a select sample. Households that experience greater returns to individualism early after arrival may be more likely to have more children and also to survive longer on the frontier. With differential fertility, we would see more children born to individualistic households in the later years post-arrival in Figure 6. Differential outmigration (or death) further implies that we are less likely to see households for whom individualism did not increase after arrival to the frontier. Our second approach to causal identification

[30]Corresponding graphical results for each specification can be found in Appendix Figure E.1.

is not subject to these sample selectivity concerns.

(ii) Age-at-Move: Childhood Exposure. Our second strategy exploits variation in age-at-move to the frontier among siblings. We follow these siblings 30 years later to examine differences in individualism revealed in adulthood. In particular, we link brothers aged 0 to 20 in the 1850 Census to the 1880 Census by which time they were 30 to 50 year old household heads with children of their own. With this linked sample of nearly 42,000 individuals spanning 30 years, we ask whether being exposed to frontier conditions from an earlier age makes fathers more likely to give individualistic names to their children.[31] This approach has the advantage of relying on variation in frontier exposure that is not due to the migrants' own choices but rather to their parents' choices.[32]

We estimate the following for child i in the 1880 Census with father f from household h in 1850:

$$\text{infrequent name}_{ifh} = \alpha + \sum_{j=1}^{17} \beta_j 1(f\text{'s age-at-move to frontier}_h = j) + \mathbf{x}'_{if}\boldsymbol{\eta} + \theta_h + \varepsilon_{ifh}. \quad (8)$$

As a baseline, we restrict to children whose fathers moved to the frontier with their parents as children. The key regressors are indicators for those ages $j = 1, \ldots, 17$; we also consider a continuous age-at-move specification. The $\mathbf{x}$ vector includes child i gender and birth order as well as father f birth order fixed effects to absorb variation in individualism unrelated to frontier exposure. Standard errors are two-way clustered on 1850 household and 1880 county.

Given the 1850 household fixed effects (θ_h), the β_j identify differences in the likelihood of individualistic names across cousins in 1880 due to the migration decisions of their paternal grandparents prior to 1850. We normalize age 1 to zero so that each β_j identifies how much less likely we are to observe an infrequent name for cousin i whose father f moved to the frontier at age j compared to cousin i' whose father f' moved to the frontier at age 1.[33]

Differences across β_j identify causal exposure effects under the assumption that the potential individualism of children is orthogonal to the timing of the family's move. This would be violated if families moved to the frontier on the basis of pre-trends in or unobservable shocks to individualism. The earlier, event-study results suggest both are unlikely. Moreover, post-move growth in individualism among parents would not be a source of bias, but rather a channel for the frontier's treatment effect on children.

The core results point to significant effects of frontier exposure. In Figure 7, for example, the likelihood of being given an individualistic name is 10 p.p. higher for children whose fathers moved to the frontier at age 1 compared to their cousins whose father was 10 years old when the family moved to the frontier. Assuming that these age-at-move effects are linear, which seems reasonable given the patterns in Figure 7, column 1 of Table 8 implies that with each additional year of frontier experience as a child, one is 0.7 p.p. more likely as an adult to give their own children individualistic names. This estimate is

[31] Like other studies based on historical linked records, we focus on men as women changed their names upon marriage making it impossible to match them across censuses. The linking generates a sample of 41,975 fathers with 146,845 children in 1880.

[32] This approach is similar to Chetty and Hendren (2018) who study childhood exposure to neighborhoods of varying quality. While neighborhood quality varies along a continuum in their setting, county-level frontier status is a binary measure in ours.

[33] To fix ideas, consider brothers John (age 10) and Paul (6), born in Virginia, who we observe on the Iowa frontier in the 1850 Census. Using the procedure above based on children's birth states, we infer that the parents moved the boys to the frontier in 1847 when Paul was 3 and John was 7. Regardless of the precise moving date, Paul could ultimately have as many as four more years of childhood frontier exposure than John. Equation (8) then identifies whether Paul's children observed in the 1880 Census have more individualistic names than John's children.

very similar to the effect of frontier exposure in adulthood seen in Table 7.

Robustness. These results survive key robustness checks. Column 2 of Table 8 includes contemporary state fixed effects, accounting for the possibility that the brothers might reside in different locations in 1880 for reasons unobservable to us but perhaps confounded with their changes in individualism since 1850. We also account for age differences across cousins in 1880 using increasingly stringent fixed effects for birth cohort: decade (column 3), five-yearly (column 4), and three-yearly (column 5). These are in addition to the baseline control for child birth order and help rule out trends in individualistic names that might be correlated with fathers' age-at-move. Like the event-study results in Table 7, these added controls reduce precision but leave significant effect sizes that are statistically indistinguishable from the baseline. Appendix E.2 reports further checks, including robustness to non-classical measurement error in the linking procedure.

Summary: Selection vs. Exposure. Together, these two distinct identification strategies yield evidence consistent with a treatment effect of frontier conditions on individualism. While migrants to the frontier self-select on prior levels of individualism, our findings suggest a causal amplification of this cultural trait after arrival.

To understand the relative magnitude of selection and treatment effects, we ask how many years of frontier exposure it takes to double the differential levels of individualism that self-selected migrants bring with them to the frontier. Comparing estimates of selection from Table 6 with those of exposure effects from Table 7, it takes around 4–8 years depending on which specification one uses. At the upper end, the estimate in column 5 of Table 7 suggests that an additional 8 years of frontier exposure increases the likelihood of giving one's child an individualistic name by 3.2 p.p., which is exactly the differential among frontier migrants compared to those remaining in settled areas as seen in column 1 of Table 6.

Both selective migration and causal exposure effects were arguably responses to an advantage of individualism on the frontier, which we document in the next section. In other words, differential returns attracted individualists to the frontier and also created strong incentives to deepen individualism once they settled there. In addition, the increased prevalence of individualists may have amplified the differential returns to individualism by making it harder for non-individualists to adapt. This would create a feedback loop between selective migration, exposure effects, and the advantage of individualism, amplifying the overall effects of frontier experience.[34]

5.3 Returns to Individualism

This section provides descriptive evidence of an advantage to individualism on the frontier, created by the specific opportunities and threats in this environment. Because people on the frontier primarily had to rely on themselves for protection and material progress, the independent, self-reliant types were likely to fare better (Kitayama et al., 2010).[35] Moreover, frontier settlers often faced unfamiliar agroclimatic

[34] To the extent that more successful settlers also had more children, differential fertility may also have increased the prevalence of individualism on the frontier.

[35] Critics of Turner emphasize the importance of cooperation on the frontier (e.g., Boatright, 1941), but his supporters have argued that cooperation was not inconsistent with individualism. For instance, according to Billington (1974), the frontiersman "spoke for individualism . . . even though he was equally willing to find haven in cooperation when danger threatened or need decreed." While returns to cooperation may have been high at times, maintaining extended reciprocity arrangements would have been difficult in frontier settings with such high population mobility.

conditions in which non-conformism and innovation—two traits associated with individualism—may have been beneficial (see Shannon, 1977).[36]

We estimate the returns to individualism using the following difference-in-difference specification, which relates father i's economic status in county c in Census year t, y_{ict}, to predetermined infrequent names within the household:

$$\begin{aligned} y_{ict} \;=\;& \alpha + \beta \text{ own infrequent name}_{ic} + \eta(\text{own infrequent name}_{ic} \times \text{frontier}_{ct}) \\ &+ \delta \text{ children infrequent names}_{ic} + \zeta(\text{children infrequent names}_{ic} \times \text{frontier}_{ct}) + \theta_{ct} + \varepsilon_{ict}, \end{aligned} \tag{9}$$

where β captures the return to the father's own infrequent name outside the frontier, and η the differential return on the frontier. At the same time, δ captures the association of father's economic status and infrequent children's names outside the frontier, and ζ the frontier differential. We restrict attention to white, native-born fathers with at least one child and define infrequent names as those outside the top 10 nationally. The county×year fixed effects, θ_{ct}, account for all differences in outcomes common across individuals within the same local economy. Standard errors are clustered by county.

We pool data across Census rounds 1850–80 and measure y_{ict} using the occupational score (*occscore*), a widely-used proxy for economic status in the historical literature. This index ranges from 0 to 100 and captures the income returns associated with occupations in the 1950 Census. We use the *occscore* from Ruggles et al. (2019) for 1850 and 1880 and construct the scores directly for 1860 and 1870 using a crosswalk from occupational strings to codes for available years.

Table 9 suggests differential returns to individualistic behavior on the frontier. Focusing on the full specification in column 3, fathers in non-frontier counties that give their children individualistic names exhibit higher occupational scores than those that give their children more common names, and this differential is more than one-third larger on the frontier. The estimate of η around 0.3 is economically meaningful, capturing around one-third the mean difference between the *occscore* for a farmer and a blacksmith. These results are even stronger when restricting the analysis to fathers that are not in farming occupations (column 4). The estimate of $\eta = 0.56$ is around one-half the mean difference between the *occscore* for a blacksmith and a carpenter. Meanwhile, although fathers with own infrequent names perform better in non-frontier counties, there is no significant differential on the frontier.

Beyond greater returns to individualism, the frontier was viewed as a place with favorable prospects for upward mobility and where effort was key to income generation. These views would hone opposition to redistribution. They would also hasten the process of cultural change towards individualism. For example, the greater the returns individualism on the frontier, the more favorable the mobility prospects. Such an environment could lead to lower tax redistribution and, in turn, reinforce the selective migration of individualists. Appendix G.4 provides further background on this complementary mechanism.

[36]The connection between innovation and individualism is discussed at length in Gorodnichenko and Roland (2012). In characterizing the traits of frontier populations, Turner (1893) himself mentions individualism along with the "coarseness and strength combined with acuteness and inquisitiveness" and the "practical, inventive turn of mind, quick to find expedients."

6 Conclusion

This paper shows how frontier settlement shaped culture across the United States. For over a century, the westward-moving frontier attracted sizable swathes of America's young, mobile, and ever-growing population. These settlers created new communities in a context with unique challenges and opportunities. Frederick Jackson Turner famously argued that the frontier fostered a culture of rugged individualism. We provide the first systematic empirical evidence on this prominent theme in American history. The frontier attracted individualistic migrants, and then made them more individualistic over time. This culture persisted over the long run: counties with longer historical frontier experience exhibit more individualistic cultural practices and stronger opposition to government intervention.

Our findings have suggestive implications for the sharp contrast between the U.S. and Europe in terms of redistribution preferences and policies, a recurring topic in the political economy literature. According to Turner (1893), "the advance of the frontier . . . meant a steady movement away from the influence of Europe," as "moving westward, the frontier became more and more American." As settlers of European origin shed their former culture and embraced rugged individualism across the U.S., America as a whole became more and more different from Europe. The frontier roots of opposition to redistribution in the United States may explain why these preferences remain stable despite rising inequality.

In closing, we note that frontier settlement may have had different effects in other countries. For instance, Argentina and Russia also underwent massive territorial expansion in their early history, but were ruled by elites that built more extractive institutions. In their work on the Americas, García-Jimeno and Robinson (2011) argue that frontier settlement hastened the advance of democracy but only in countries with initially equitable institutions. The national institutions of the U.S., which favored relatively high levels of geographic mobility, access to land, and security of property rights, undoubtedly shaped the effects of frontier settlement that we identify. The methods developed in this paper may prove useful in future work to understand the legacy of frontier settlement in the U.S. and elsewhere.

References

Abramitzky, R., L. P. Boustan, and K. Eriksson, "Cultural Assimilation during the Age of Mass Migration," *American Economic Review: Insights*, forthcoming.

Alesina, A. and E. La Ferrara, "Preferences for Redistribution in the Land of Opportunities," *Journal of public Economics*, 2005, *89* (5), 897–931.

_ **and G. Angeletos**, "Fairness and Redistribution," *The American Economic Review*, 2005, *95* (4), 960–980.

_ **and P. Giuliano**, "Preferences for Redistribution," in J. Benhabib, A. Bisin, and M. O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, North Holland, 2011, pp. 93–131.

_ **, E. L. Glaeser, and B. Sacerdote**, "Why Doesn't the United States Have a European-Style Welfare State?," *Brookings Papers on Economic Activity*, 2001, *2001* (2), 187–277.

Ashok, V., I. Kuziemko, and E. Washington, "Support for Redistribution in an Age of Rising Inequality: New Stylized Facts and Some Tentative Explanations," *NBER Working Paper 21529*, 2015.

Autor, D., D. Dorn, and G. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *The American Economic Review*, 2013, *103* (6), 2121–2168.

_ **,** _ **,** _ **, and K. Majlesi**, "A Note on the Effect of Rising Trade Exposure on the 2016 Presidential Election," *Unpublished Mansucript*, 2017. https://economics.mit.edu/files/12418.

Beck-Knudsen, A. S., "Those who stayed: Individualism, self-selection and cultural change during the age of mass migration," *Unpublished Manuscript*, 2019. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3321790.

Bertrand, M. and S. Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bester, C. A., T. G. Conley, and C. B. Hansen, "Inference with dependent data using cluster covariance estimators," *Journal of Econometrics*, 2011, *165* (2), 137–151.

Billington, M. L. and R. D. Hardaway, *African Americans on the Western Frontier*, University Press of Colorado, 1998.

Billington, R. A., *America's Frontier Heritage*, University of New Mexico Press, 1974.

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

Boatright, M. C., "The Myth of Frontier Individualism," *The Southwestern Social Science Quarterly*, 1941, *22* (1), 14–32.

Brown, R. P., M. Carvallo, and M. Imura, "Naming patterns reveal cultural values: Patronyms, matronyms, and the US culture of honor," *Personality and Social Psychology Bulletin*, 2014, *40* (2), 250–262.

Chetty, R. and N. Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

Chow, G. C., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 1960, *28* (3), 591–605.

Collins, William J and Ariell Zimran, "The economic assimilation of Irish Famine migrants to the United States," *Explorations in Economic History*, 2019, *74.*

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

Couttenier, M. and M. Sangnier, "Living in the Garden of Eden: Mineral resources and preferences for redistribution," *Journal of Comparative Economics*, 2015, *43* (2), 243–256.

Cronon, W., "Revisiting the vanishing frontier: The legacy of Frederick Jackson Turner," *The Western Historical Quarterly*, 1987, *18* (2), 157–176.

Davis, L., "Individual responsibility and economic development: Evidence from rainfall data," *Kyklos*, 2016, *69* (3), 426–470.

Doepke, M. and F. Zilibotti, "Occupational choice and the spirit of capitalism," *The Quarterly Journal of Economics*, 2008, *123* (2), 747–793.

Eblen, J. E., "An Analysis of Nineteenth-Century Frontier Populations," *Demography*, 1965, *2* (1), 399–413.

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

Feigenbaum, J. J., "Automated Census Record Linking: A Machine Learning Approach," *Unpublished Manuscript*, 2016. https://jamesfeigenbaum.github.io/research/census-link-ml/.

Ferrie, J. P., "Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner's 'Safety-Valve'," *Department of Economics, Northwestern University Manuscript*, 1997.

Fincher, C. L. and R. Thornhill, "Parasite-Stress Promotes In-Group Assortative Sociality: The Cases of Strong Family Ties and Heightened Religiosity," *Behavioral and Brain Sciences*, 2012, *35* (2), 61–79.

Gannett, H., "The Movements of Our population," *National Geographic Magazine*, 1893, (5).

García-Jimeno, C. and J. A. Robinson, "The Myth of the Frontier," in D. L. Costa and N. R. Lamoreaux, eds., *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, Chicago: University of Chicago Press, 2011, pp. 49–89.

Gentzkow, M., J. M. Shapiro, and M. Taddy, "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 2019, *87* (4), 1307–1340.

Gerrit, B. and D. Onland, "Socioeconomic Determinants of First Names," *Names*, 2011, *59* (1), 25–41.

Gorodnichenko, Y. and G. Roland, "Understanding the Individualism-Collectivism Cleavage and its Effects: Lessons from Cultural Psychology," in M. Aoki, T. Kuran, and G. Roland, eds., *Institutions and Comparative Economic Development*, London: Palgrave Macmillan, 2012, pp. 213–236.

_ and _ , "Culture, Institutions and the Wealth of Nations," *Review of Economics and Statistics*, 2016, *99* (3), 402–416.

Greif, A., "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies," *Journal of Political Economy*, 1994, *102* (5), 912–950.

Grosjean, P., "A History of Violence: The Culture of Honor and Homicide in the US South," *Journal of the European Economic Association*, 2014, *12* (5), 1285–1316.

_ and R. Khattar, "It's Raining Men! Hallelujah? The Long-Run Consequences of Male-Biased Sex Ratios"," *Review of Economic Studies*, forthcoming.

Heine, S. J., "Cultural Psychology," in S. T. Fiske, D. T. Gilbert, and G. Lindzey, eds., *Handbook of Social Psychology*, 2010, pp. 254–266.

Hofstede, G., *Cultures and Organizations: Software of the Mind*, McGraw-Hill, 1991.

Kitayama, S., K Ishii, T. Imada, K. Takemura, and J. Ramaswamy, "Voluntary Settlement and the Spirit of Independence: Evidence from Japan's "Northern Frontier"," *Journal of personality and social psychology*, 2006, *91* (3), 369–384.

_ , L. G. Conway III, P. R. Pietromonaco, H. Park, and V. C. Plaut, "Ethos of independence across regions in the United States: the production-adoption model of cultural change," *American Psychologist*, 2010, *65* (6), 559–574.

Kling, J. R., J. B. Liebman, and L. F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Lang, R. E., D. E. Popper, and F. J. Popper, "'Progress of the Nation': The Settlement History of the Enduring American Frontier," *The Western Historical Quarterly*, 1995, *26* (3), 289–307.

Larson, J. L., "Grasping for the significance of the Turner legacy: An afterword," *Journal of the Early Republic*, 1993, *13* (2), 241–249.

Lieberson, S. and E. O. Bell, "Children's First Names: An Empirical Study Of Social Taste," *American Journal of Sociology*, 1992, *98* (3), 511–554.

Limerick, P. N., *The legacy of conquest: The unbroken past of the American West*, WW Norton & Company, 1988.

Manson, S., J. Schroeder, D. Van Riper, and S. Ruggles, "IPUMS National Historical Geographic Information System: Version 14.0 [Database].," Minneapolis, MN: IPUMS. 2019.

Minnesota Population Center, "Integrated Public Use Microdata Series, International: Version 7.2 [dataset]," Minneapolis: IPUMS, 2019 2019.

Nunn, N. and D. Puga, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

Ogihara, Y., H. Fujita, H. Tominaga, S. Ishigaki, T. Kashimoto, A. Takahashi, K. Toyohara, and Y. Uchida, "Are Common Names Becoming Less Common? The Rise in Uniqueness and Individualism in Japan," *Frontiers in psychology*, 2015, *6.*

Olsson, O. and C. Paik, "Long-run cultural divergence: Evidence from the neolithic revolution," *Journal of Development Economics*, 2016, *122*, 197–213.

Oster, E., "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Overmeyer, P. H., "Westward Expansion Before the Homestead Act," in H. F. Williamson, ed., *The Growth of the American Economy: An Introduction to the Economic History of the United States*, New York: Prentice-Hall, Inc, 1944, pp. 82–112.

Philips, L., "Hanging on the metaphone," *Computer Language*, 1990, *7* (12), 39–43.

Piketty, T., "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, 1995, *110* (3), 551–584.

Porter, R., H. Gannett, and W. Hunt, "Progress of the Nation, including the Map of the Population of 1870," *Report on Population of the United States at the Eleventh Census*, 1890, pp. 13–30.

Robinson, P. M., "Root-N-Consistent Semiparametric Regression," *Econometrica*, 1988, *56* (4), 931–954.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek, "Integrated Public Use Microdata Series: Version 9.0 [dataset]," Minneapolis, MN: IPUMS, 2019. 2019.

Sequeira, S., N. Nunn, and N. Qian, "Immigrants and the Making of America," *The Review of Economic Studies*, 2020, *87* (1), 382–419.

Shannon, F. A., *The Farmer's Last Frontier: Agriculture, 1860-1897*, Vol. 5, ME Sharpe, 1977.

Steckel, R. H., "Household Migration and Rural Settlement in the United States, 1850–1860," *Explorations in Economic History*, 1989, *26* (2), 190–218.

Stewart, J. I., "Migration to the Agricultural Frontier and Wealth Accumulation, 1860–1870," *Explorations in Economic History*, 2006, *43* (4), 547–577.

Triandis, H., *Individualism & Collectivism*, Boulder, CO: Westview Press, 1995.

Turner, F. J., "The Significance of the Frontier in American History," *Proceedings of the State Historical Society of Wisconsin*, 1893.

_ , "The problem of the West," *Atlantic Monthly*, 1896.

Turner, Frederick Jackson, "The Middle West," *International Monthly*, 1901, *IV*, 794–798.

Twenge, J. M., E. M. Abebe, and W. K. Campbell, "Fitting In or Standing Out: Trends in American Parents' Choices for Children's Names, 1880–2007," *Social Psychological and Personality Science*, 2010, *1* (1), 19–25.

Vandello, J. A. and D. Cohen, "Patterns of Individualism and Collectivism Across the United States," *Journal of Personality and Social Psychology*, 1999, *77* (2), 272–292.

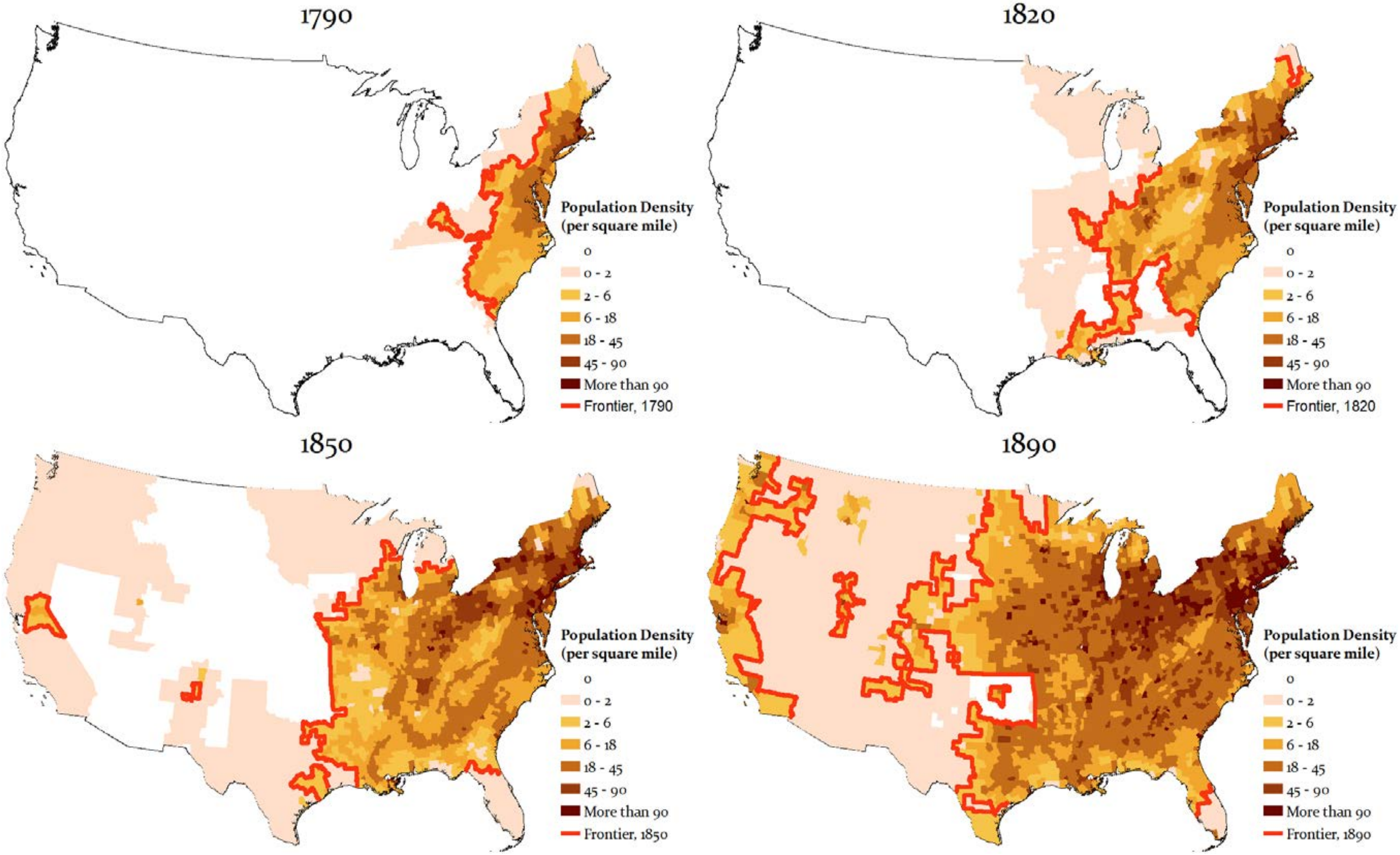
Varnum, M E. W. and S. Kitayama, "What's in a name? Popular Names are Less Common on Frontiers," *Psychological Science*, 2011, *22* (2), 176–183.

Zelinsky, W., *The Cultural Geography of the United States*, Englewood Cliffs, N.J.: Pearson Education, Limited, 1973.

Zivot, E. and D. W. K. Andrews, "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business & Economic Statistics*, 2002, *20* (1), 25–44.

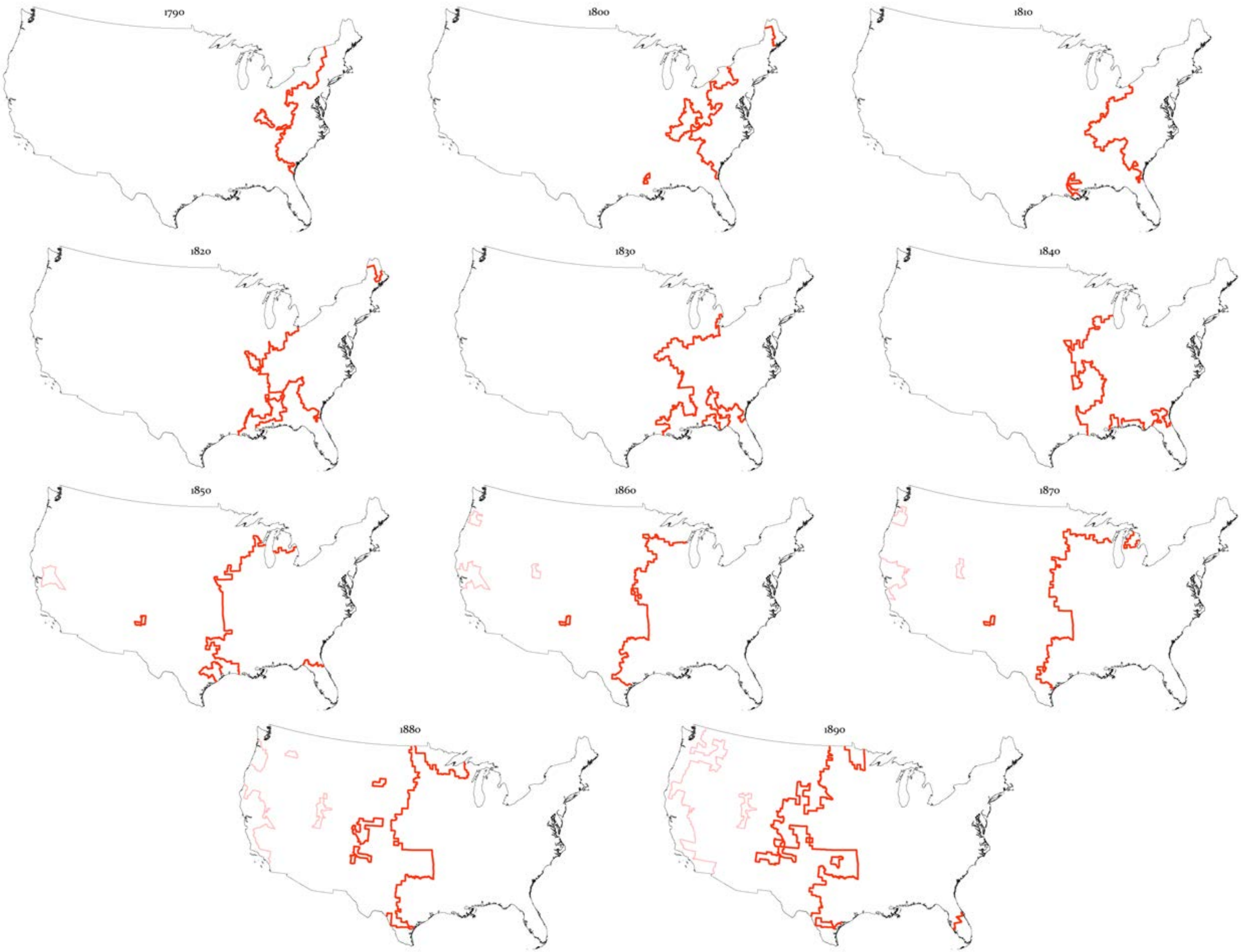
Figures

Figure 1: Population Density and the Frontier for Selected Years



Notes: Based on county-level data from NHGIS (Manson et al., 2019). Population is allocated across years and counties based on the harmonization procedure described in Appendix K. The red frontier line is based on the algorithm described in Section 2.1 and Appendix A. The population density figures exclude most Native Americans, who were generally not enumerated by the Census throughout the frontier era (see footnote 4).

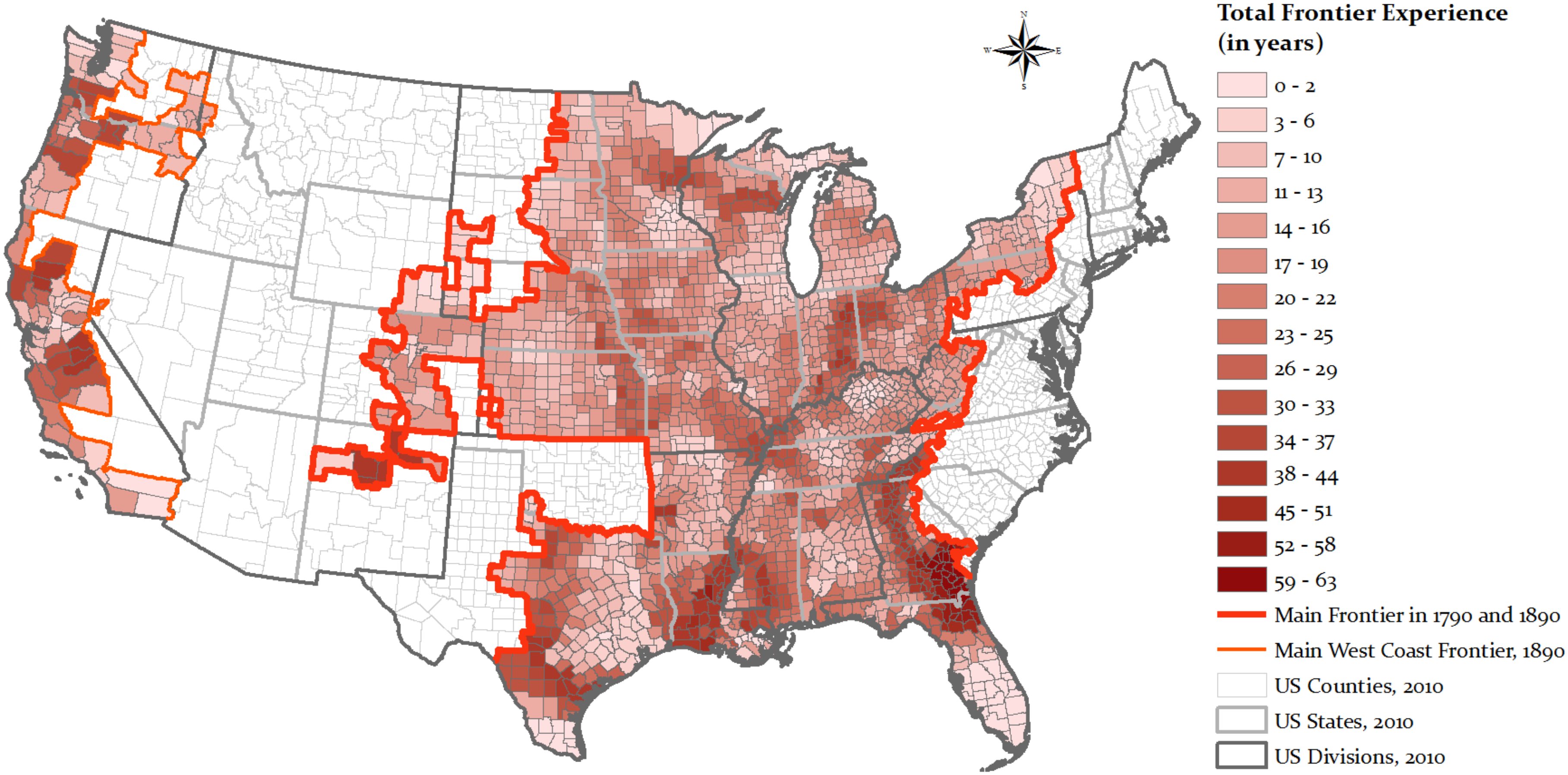
Figure 2: The Evolution of the Frontier, 1790 to 1890



Notes: Based on county-level data from NHGIS (Manson et al., 2019). The frontier lines demarcate the contour of counties with population density below and above 2 people per square mile. The dark red lines correspond to the main frontier lines emerging form east-to-west expansions (our baseline analysis). The light red lines correspond to the frontiers resulting from west-to-east expansions from the West Coast, which we examine for robustness. In both cases, we exclude smaller “island frontiers” in the interior and contour line segments less than 500 km. Full details on the frontier line algorithm can be found in Appendix A.

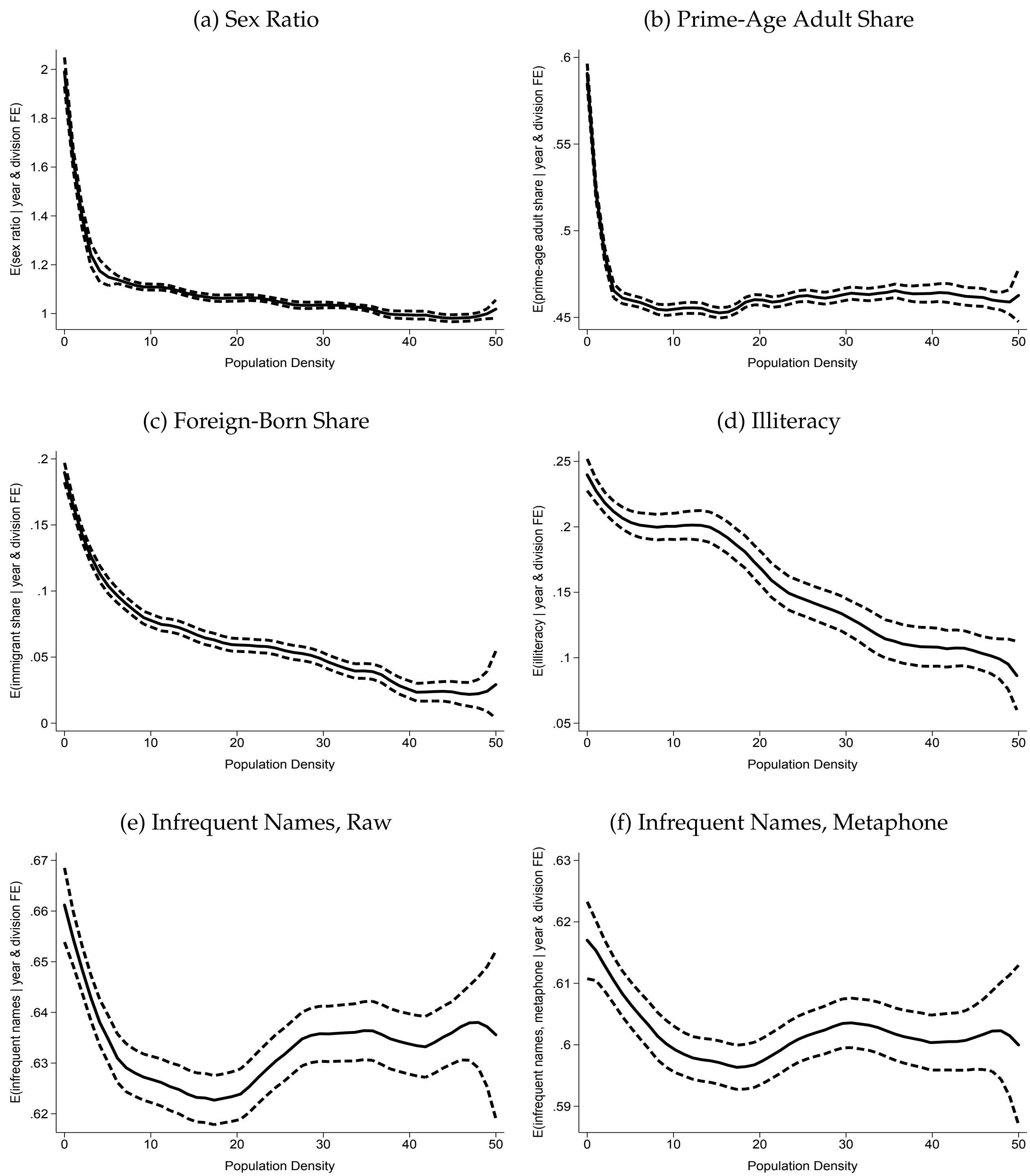
Figure 3: Total Frontier Experience, 1790 to 1890

(baseline sample lies between 1790 and 1890 main frontier lines, see Section 2.2 and notes below)



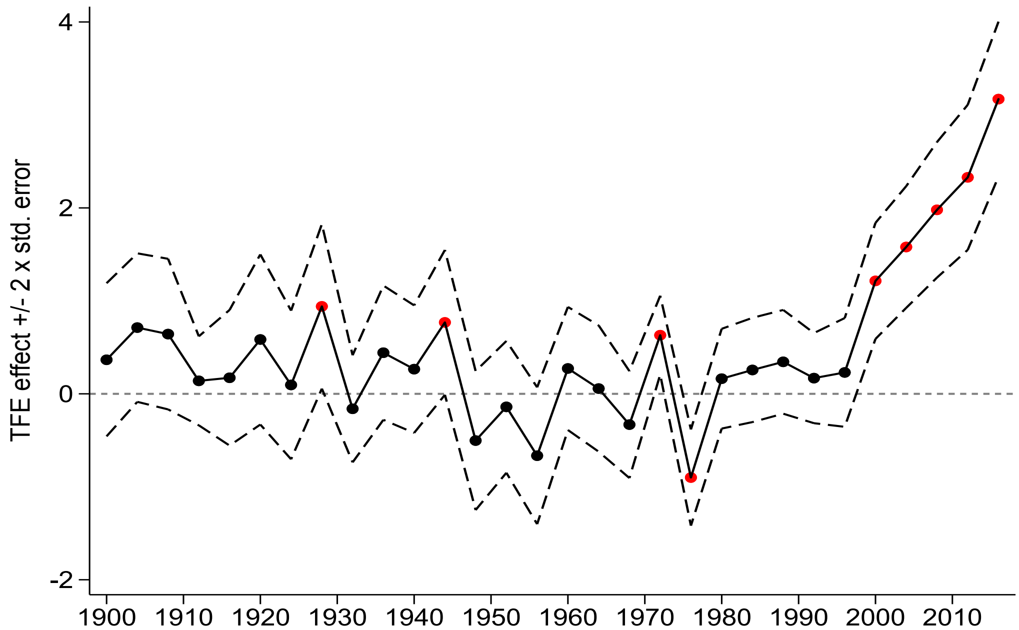
Notes: Based on county-level data from NHGIS (Manson et al., 2019). Total frontier experience is the total number of years the county was within 100 km of the frontier line and its population density was below 6 people per square mile, between 1790–1890. The white areas to the east of the 1790 main frontier line are counties for which we do not know frontier history given the lack of Population Census data before 1790. The white areas to the west are beyond the 1890 frontier line and hence not included in our baseline sample, which is restricted to the frontier era as defined by Porter et al. (1890) in the Census *Progress of the Nation* report. We include many of those counties to the west when extending the frontier era through 1950 for robustness.

Figure 4: Demographics and Individualism by Population Density, 1790 to 1890



Notes: These figures plot semiparametric estimates of equation (3) relating population density to demographic characteristics prominent in historical accounts of the frontier (a-d) and proxies for individualism (e-f). We estimate these curves $g(\cdot)$ based on the Robinson (1988) partially linear approach, pooling across all available years 1790–1890 for each county c. The specification includes Census division and year fixed effects, which are partialled out before estimating these shapes, and are based on an Epanechnikov kernel and rule-of-thumb bandwidth. The dashed lines are 95 percent confidence intervals. The estimates are recovered over all counties, but the figure zooms in on those with less than 50 people/mi^2 for presentational purposes. (a) *Sex Ratio* for whites is the ratio of the number of white males over white females. (b) *Prime-Age Adult Share* is the fraction of whites aged 15–49 over the total number of whites. (c) *Foreign-Born Share* is the ratio of foreign-born persons over total population. (d) *Illiteracy* is the illiteracy rate for whites aged 20 or older. (e) *Infrequent Names* is the share of children with names outside of the top 10 most popular names in their Census division with the sample restricted to children aged 0–10 with native-born parents. (f) adjusts the measure in (e) applying the metaphone procedure to enumerated names prior to computing the infrequency indicator.

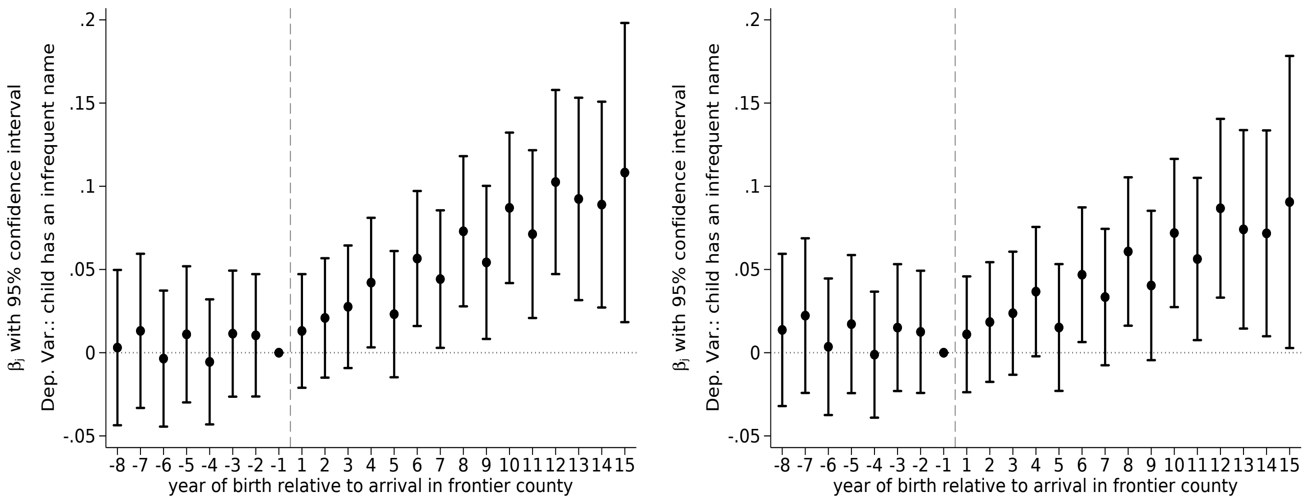
Figure 5: TFE and the Republican Presidential Vote Share, 1900–2016



Notes: This figure reports point estimates and +/- 2×standard error confidence bands on the effects of TFE on the Republican Presidential vote share in each election from 1900 to 2016. The red circles indicate statistical significance at the 95% level.

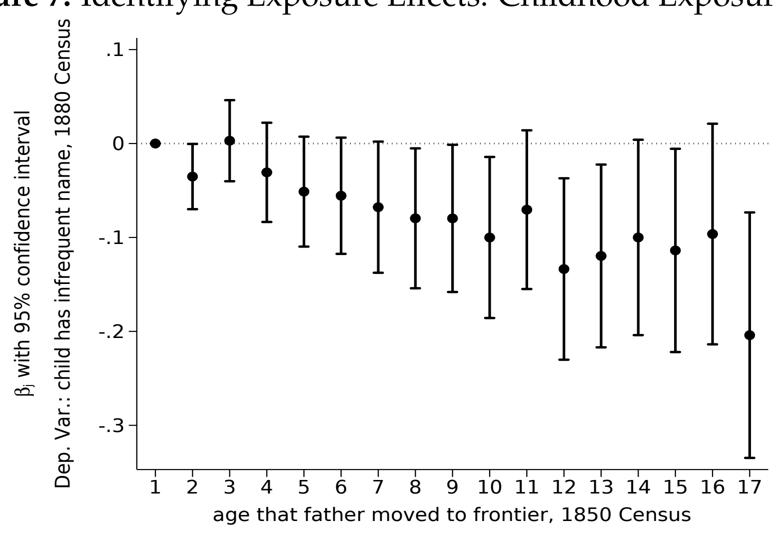
Figure 6: Identifying Exposure Effects: Adulthood Exposure (I)

(a) Baseline: Birth Decade FE (b) Pre-Move State Name Trends



Notes: This figure isolates within-household, cross-child variation in parental exposure to the frontier. Each graph reports estimates of β_j and 95% confidence intervals in equation (6) for $j = -8, \ldots, 15$ (with other j included but suppressed for presentational purposes). Each β_j can be interpreted as the differential likelihood of an infrequent name being given to a child born j years before/after their parents moved to the frontier, relative to the child born one year prior to moving. The sample includes 57,097 children born to 16,901 families headed by white, native-born parents that moved with at least one child to a frontier county as we observe them in the Census records in 1850, 1860, 1870 or 1880. All estimates control for household fixed effects and child gender. Graph (a) additionally includes child birth decade FE, and (b) includes controls for the mean gender-specific infrequent name share in each child birth year in the state from which each family migrated from before arriving on the frontier. Standard errors are clustered by contemporaneous county.

Figure 7: Identifying Exposure Effects: Childhood Exposure (II)



Notes: This figure reports estimates of β_j and 95% confidence intervals in equation (8). Each β_j can be interpreted as the differential likelihood of an infrequent name being given to a child whose father's family moved to the frontier at age j compared to a child born to that father's younger brother who was 1 when the family moved to the frontier. The sample consists of 81,823 children age 0–20 in the 1880 Census with fathers hailing from 17,778 families observed in the 1850 Census and where at least two brothers (one brother) were born before the family moved to the frontier. We link the fathers from 1850 to 1880 using a procedure detailed in Appendix K. There are 16,776 children with fathers that moved at age 1, 8,463 at age 2, ..., 3,164 at age 10, ..., and 487 at age 17. These estimates control for 1850 family fixed effects, father birth order, child gender, child birth order, and an indicator for duplicate matches in the linking process.

Tables

Table 1: Demographics and Individualism on the Frontier

Dependent Variable:	Male/Female Ratio	Prime-Age Adult Share	Foreign-Born Share	Illiterate Share	Infrequent Child Names Raw	Infrequent Child Names Metaphone
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel (a): Baseline Frontier Definition: Low Density and Proximity to Frontier Line					
frontier county	0.190	0.026	0.063	-0.007	0.022	0.018
	(0.021)	(0.004)	(0.008)	(0.011)	(0.007)	(0.006)
Mean Dep. Var. in Non-Frontier Counties	1.09	0.46	0.07	0.18	0.63	0.60
Number of County-Years	11,594	5,508	11,062	2,779	6,907	6,907
R^2	0.09	0.20	0.33	0.17	0.32	0.33
	Panel (b): Distinguishing Low Density and Proximity to Frontier Line					
near frontier line	0.103	0.022	0.058	-0.058	0.022	0.018
	(0.012)	(0.003)	(0.009)	(0.013)	(0.007)	(0.006)
low population density	0.127	0.005	0.033	0.055	0.004	0.006
	(0.014)	(0.003)	(0.008)	(0.011)	(0.006)	(0.005)
Mean Dep. Var. in Non-Frontier Counties	1.09	0.46	0.07	0.18	0.63	0.60
Number of County-Years	11,594	5,508	11,062	2,779	6,907	6,907
R^2	0.10	0.19	0.35	0.19	0.32	0.34

Notes: This table reports OLS estimates of equations (1) and (2) in Panels A and B, respectively. The dependent variables are the same as in Figures 4 (a)–(f). The sample size varies across columns depending on availability in the given Census round. All variables, except foreign-born share, are defined over the white population. Infrequent names capture the share of children with names outside of the top 10 most popular names in their Census division. The measure in column (5) is based on the raw enumerated name and in column (6) on the metaphone-adjusted name. In both cases, the means are restricted to white children aged 0–10 with native-born parents. *Low population density* equals one if the county has density less than 6 people per square mile, and *near frontier line* equals one if the county is within 100 km of the frontier line in the given year. The sample excludes counties to the east of the 1790 frontier line and west of the main 1890 frontier line in keeping with our baseline long-run sample restrictions. All regressions include year and Census division FE. Standard errors are clustered using the grid cell approach of Bester et al. (2011) as described in Section 4.1.

Table 2: Total Frontier Experience and 20th Century Individualism

	(1)	(2)	(3)	(4)
	Dependent Variable:			
	Panel (a): Infrequent Names (standardized share)			
total frontier experience	0.138	0.141	0.096	0.086
	(0.024)	(0.021)	(0.022)	(0.018)
Oster δ for $\beta = 0$		-13.80	2.63	1.96
Number of Counties	2,036	2,036	2,036	2,036
R^2	0.55	0.60	0.85	0.87
	Panel (b): Infrequent Names, Metaphone (standardized share)			
total frontier experience	0.138	0.141	0.088	0.089
	(0.025)	(0.021)	(0.022)	(0.019)
Oster δ for $\beta = 0$		-16.20	2.32	2.41
Number of Counties	2,036	2,036	2,036	2,036
R^2	0.52	0.58	0.85	0.85
State Fixed Effects	✓	✓	✓	✓
Geographic/Agroclimatic Controls		✓	✓	✓
Within-State Nearest Neighbor Matching:				
on Population Density (1,018 FE)			✓	
on Foreign-Born Share (1,018 FE)				✓

Notes: This table reports estimates of equation (4) for our infrequent names as defined in Table 1. Both are defined over white children age 0–10 with native-born parents in the 1940 Census. In the average county, 76.6 percent of children have infrequent names and 72.7 percent have infrequent names after metaphone adjustment with standard deviations of 5.6 and 5.3 percentage points, respectively. Total frontier experience is expressed in decades. The dependent variables are standardized so that the coefficient indicates the standard deviation effect of each additional decade of frontier exposure historically. This baseline sample is based only on counties inside the 1790–1890 east-to-west frontier. Alternative definitions of the dependent variables names are considered in Appendix Table B.3. Column 1 simply includes state fixed effects, and column 2 adds the following controls: county area; county centroid latitude and longitude; distance to oceans, lakes and rivers from county centroid; mean county temperature and rainfall; elevation; and average potential agricultural yield. Column 3 includes fixed effects within-state for pairs of counties that have the most similar population density in 1940. Column 4 includes fixed effects for within-state pairs of counties that have the most similar foreign-born population shares in 1940. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. Columns 3 and 4 additionally cluster (two-way) on the county-pair. Alternative approaches to inference can be found in Appendix Table F.1. The Oster (2019) tests in columns 2–4 are each with reference to the baseline specification in column 1 with only state fixed effects.

Table 3: Total Frontier Experience and Opposition to Government Intervention and Redistribution

Dependent Variable:	Prefers Cut Public Spending on Poor	Prefers Cut Public Spending on Welfare	Believes Gov't Should Redistribute	Prefers Reduce Debt by Spending Cuts	Index of Preferences for Spending Cuts	County Property Tax Rate, 2010–14	Republican Presidential Vote Shr., 2000–16
Scale:	binary	binary	standardized	binary	standardized	[0, 100]	[0, 100]
Data Source:	ANES	CCES	GSS	CCES	GSS	ACS	Leip
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
total frontier experience	0.010	0.007	-0.022	0.014	0.028	-0.034	2.055
	(0.004)	(0.003)	(0.012)	(0.002)	(0.011)	(0.007)	(0.349)
Oster δ for $\beta = 0$	16.01	3.10	97.95	5.89	3.53	-27.45	-8.55
Mean of Dependent Variable	0.09	0.40	0.00	0.41	0.00	1.02	60.04
Number of Individuals	2,322	53,472	9,085	111,853	5,739	2,029	2,036
Number of Counties	95	1,863	255	1,963	253	2,029	2,036
R^2	0.04	0.04	0.06	0.04	0.07	0.82	0.33
Survey Wave Fixed Effects	✓	✓	✓	✓	✓	–	–
Individual Demographic Controls	✓	✓	✓	✓	✓	–	–
State Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports estimates of equation (4) for several measures capturing preferences for redistribution and state spending as well as actual property tax rates and the Republican vote share. Total frontier experience is expressed in decades. Full details on the outcomes can be found in Appendix K. We use all available survey rounds with the given outcome, and in all cases, we restrict to those counties in our baseline sample as described in the notes to Table 2. All columns are based on the specification in column 2 of Table 2 with additional individual-level controls for age, age squared, gender, and race in columns 1–5. The ANES measure in column 1 equals one if the respondent prefers that federal government spending on poor people be cut. The CCES measure in column 2 equals one if the respondent would prefer to cut public spending on welfare programs. The GSS measure in column 3 is a standardized measure of intensity of support on a 7 point scale of the statement that the government should reduce income differences in society through redistribution. The CCES question in column 4 equals one if the household would prefer that the state budget be balanced through spending cuts rather than tax increases. The GSS measure in column 5 is a standardized first principal component analysis (PCA) index based on a series of questions about whether the government spends too much on different public goods and transfer programs. The measure of county-level property tax rates in column 6 is estimated by the National Association of Home Builders using the American Community Survey data from 2010–14. Column 7 captures the mean county-level Republican vote share in the last five presidential elections with data from the Leip Atlas. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. The Oster (2019) tests are with reference to a baseline specification that only includes state fixed effects.

Table 4: Disentangling the Effects of Population Density

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population Density, 1940/2010		✓					
Population Density Decile Within-State, 1940/2010			✓				
Population Density-Neighbor Matching Within-State, 1940/2010				✓			
Sample Restriction	None	None	None	None	$> 90th$ percentile urban pop. share, 1940/2010	$\leq 90th$ percentile urban pop. share, 1940/2010	None
	Panel (a): Infrequent Children's Name Share in 1940 (standardized)						
total frontier experience	0.141	0.130	0.086	0.096	0.139	0.100	0.081
	(0.021)	(0.020)	(0.019)	(0.022)	(0.034)	(0.022)	(0.024)
total low density experience							0.095
							(0.019)
Number of Counties	2,036	2,036	2,021	2,036	242	1,794	2,036
R^2	0.60	0.62	0.68	0.85	0.83	0.60	0.61
	Panel (b): Infrequent Children's Name Share in 1940, Metaphone (standardized)						
total frontier experience	0.141	0.129	0.083	0.088	0.122	0.100	0.080
	(0.021)	(0.021)	(0.020)	(0.022)	(0.033)	(0.023)	(0.024)
total low density experience							0.095
							(0.019)
Number of Counties	2,036	2,036	2,021	2,036	242	1,794	2,036
R^2	0.58	0.59	0.66	0.85	0.82	0.58	0.59
	Panel (c): Mean Government Preferences Outcomes (CCES), 2006–16						
total frontier experience	0.014	0.008	0.009	0.008	0.013	0.004	0.011
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)
total low density experience							0.006
							(0.002)
Number of Individuals	112,759	112,759	111,704	112,759	68,436	44,323	112,759
Mean of Dependent Variable	0.41	0.41	0.41	0.41	0.38	0.46	0.41
R^2	0.05	0.05	0.05	0.07	0.05	0.05	0.05
	Panel (d): County Property Tax Rate in 2010						
total frontier experience	-0.034	-0.020	-0.010	-0.001	-0.022	-0.014	-0.028
	(0.007)	(0.006)	(0.005)	(0.005)	(0.012)	(0.005)	(0.008)
total low density experience							-0.008
							(0.005)
Number of Counties	2,029	2,029	2,014	2,020	223	1,806	2,029
Mean of Dependent Variable	1.02	1.02	1.02	1.02	1.34	0.98	1.02
R^2	0.82	0.85	0.87	0.95	0.90	0.86	0.82
	Panel (e): Republican Vote Share, Average 2000–16						
total frontier experience	2.055	1.532	1.535	1.655	1.280	1.489	1.255
	(0.349)	(0.346)	(0.357)	(0.356)	(0.886)	(0.347)	(0.404)
total low density experience							1.256
							(0.290)
Number of Counties	2,036	2,036	2,021	2,034	223	1,813	2,036
Mean of Dependent Variable	60.04	60.04	60.11	60.04	49.29	61.36	60.04
R^2	0.33	0.40	0.38	0.73	0.27	0.38	0.35
State Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓	✓	✓

Notes: This table disentangles the effects of TFE from the effects of historical and contemporary population density. Those in panel (a), (b), (d) and (e) are from prior tables, with the baseline estimates reproduced in column 1. The outcome in panel (c) is the mean of the six binary indicators from the CCES survey from Tables 3 and C.2. Column 2 and control for contemporaneous population density (i.e., 1940 in panels (a) and (b), 2006 in panel (c), 2000 in panel (d), and 2010 in panel (e)). Column 3 includes indicators for the decile of within-state population density. Column 4 implements the nearest-neighbor matching specification from column 3 of Table 2. Columns 5 and 6 split the sample into counties above and below the 90th percentile of contemporaneous urban population shares. Column 7 controls for the total number of years that the country had population density less than 6 people/mi^2 from 1790–1890. This is one of the aspects of total frontier experience, the other being the total number of years that the county was within 100 km of the frontier line during that period. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. Column 4 additionally clusters (two-way) on the county-pair. The Oster (2019) tests are with reference to a baseline specification that only includes state fixed effects.

Table 5: Racial Differences in the Long-Run Effects of Frontier Experience

Dependent Variable:	Prefers Cut Public Spending on Welfare (1)	Prefers Reduce Debt by Spending Cuts (2)	Opposes Affordable Care Act (3)	Opposes Increasing Minimum Wage (4)	Opposes Banning Assault Rifles (5)	Opposes Regulation of CO_2 Emissions (6)
total frontier experience × white	0.009	0.016	0.027	0.025	0.018	0.018
	(0.003)	(0.003)	(0.004)	(0.008)	(0.004)	(0.004)
total frontier experience × black	-0.008	-0.002	-0.000	0.005	-0.002	-0.008
	(0.006)	(0.005)	(0.007)	(0.014)	(0.008)	(0.007)
total frontier experience × other	0.010	0.015	0.014	0.009	0.019	0.029
	(0.007)	(0.007)	(0.009)	(0.019)	(0.008)	(0.008)
white	0.044	0.065	0.047	-0.064	0.006	0.046
	(0.012)	(0.012)	(0.015)	(0.032)	(0.012)	(0.015)
black	-0.177	-0.065	-0.215	-0.285	-0.148	-0.067
	(0.014)	(0.011)	(0.021)	(0.038)	(0.021)	(0.025)
Number of Individuals	53,472	111,853	29,446	5,134	29,404	29,215
Number of Counties	1,863	1,963	1,728	1,066	1,723	1,718
TFE(black)=TFE(white), p-value	0.010	0.000	0.000	0.125	0.022	0.000
Mean of Dependent Variable, Whites	0.43	0.44	0.58	0.32	0.39	0.35
Share White Respondents	0.79	0.77	0.76	0.86	0.76	0.76
Share Black Respondents	0.11	0.13	0.11	0.09	0.11	0.11
Share Other Respondents	0.10	0.11	0.13	0.05	0.13	0.13
State Fixed Effects	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓	✓

Notes: This table allows the effects of TFE to vary by (self-identified) race of respondents for the six CCES outcomes used in Tables 3 and C.2. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1.

Table 6: Selective Migration and Individualism on the Frontier

	Dep. Var.: Child Has an Infrequent Name (Children Named Prior to Moving)			
	(1)	(2)	(3)	(4)
omitted reference group:	stayers in settled counties		stayers in frontier counties	
household migrated from settled to frontier	0.032	0.028		
	(0.006)	(0.006)		
household migrated from frontier to settled			-0.049	-0.029
			(0.011)	(0.006)
Observations	8,734,740	8,734,740	370,999	370,999
Mean of Dep. Var., Stayers	0.65	0.65	0.64	0.64
R^2	0.02	0.03	0.04	0.07
Birth State Fixed Effects		✓		✓

Notes: This table estimates equation (5) in columns 1–2 and an analogous specification for movers from frontier areas to settled areas in columns 3–4. All columns include fixed effects for birth year×gender as well as birth order. The sample pools across Censuses from 1850–1880 and restricts to white children age 0–10 with native-born parents. The dependent variable is an indicator for whether the child has a non-top-10 name in the Census division and decade in which s/he was born. In columns 1–2, the sample includes all children living in non-frontier counties as well as children who were born in non-frontier counties and are currently living in frontier counties as a result of a family move. In columns 3–4, the sample includes all children currently living in frontier counties as well as all children who were born in frontier areas and are currently living in non-frontier counties as a result of a family move away from the frontier. The non-movers (i.e., stayers) are the omitted group to which the estimate differential refers, with the dependent variable means at the bottom of the table computed over these stayers. These mover households and children are identified using variation across reported child birth states and current county of residence (see Section 5.1 for details). Standard errors are clustered by county.

Table 7: Identifying Exposure Effects: Adulthood Exposure (I)

	Dep. Var.: Child Has an Infrequent Name					
	(1)	(2)	(3)	(4)	(5)	(6)
year of birth relative to move, pre-move	0.002	0.001	0.003	-0.001	0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)
year of birth relative to move, post-move	0.007	0.005	0.007	0.004	0.005	0.005
	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)
pre move = post move, p-value	[0.005]	[0.018]	[0.023]	[0.007]	[0.03]	[0.009]
Observations	57,097	57,097	57,097	57,097	57,097	57,097
Number of Families	16,901	16,901	16,901	16,901	16,901	16,901
Mean of Dependent Variable	0.65	0.65	0.65	0.65	0.65	0.65
R^2	0.36	0.36	0.36	0.36	0.36	0.36
Household Fixed Effect	✓	✓	✓	✓	✓	✓
Birth Year Fixed Effect	decade	5-yearly	3-yearly		3-yearly	
Child Birth Order				✓	✓	
Pre-Move Birth State Yearly Name Trend						✓

Notes: This table reports estimates of equation (7), which estimates a continuous version of the event study specifications in Figure 6. That is, the *year of birth relative to move, pre-move* measures the number of years until the household moves to the frontier, and *year of birth relative to move, post-move* measures years since arrival to the frontier. We also report the p-value for equality across the two. See the notes to Figure 6 for further details on the sample and specification. Standard errors are clustered by county.

Table 8: Identifying Exposure Effects: Childhood Exposure (II)

	Dep. Var.: Child Has an Infrequent Name				
	(1)	(2)	(3)	(4)	(5)
age-at-move to frontier	-0.008	-0.008	-0.007	-0.006	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	81,823	81,823	81,823	81,823	81,823
Number of Families	17,778	17,778	17,778	17,778	17,778
Mean of Dependent Variable	0.69	0.69	0.69	0.69	0.69
R^2	0.26	0.27	0.26	0.27	0.27
Extended Family (1850 Household) FE	✓	✓	✓	✓	✓
State of Residence FE in 1880	–	✓	–	–	–
Child Birth Cohort FE	–	–	decade	5-yearly	3-yearly

Notes: This table reports the continuous analogue to the age-at-move-specific estimates in Figure 7. All estimates control for 1850 family fixed effects, father birth order, child gender, child birth order, and an indicator for duplicate matches in the linking process. Column 1 is the specification used in Figure 7. Column 2 here additionally includes 1880 state fixed effects to allow for the possibility that brothers from 1850 may live in different locations today. Columns 3–5 control increasingly flexibly for child birth cohort. See the notes to that figure for details on the the sample and specifications. Standard errors are two-way clustered by 1850 family and 1880 county.

Table 9: Returns to Individualism on the Frontier, 1850–1880

	Dep. Var.: Father's Occupation Score			
	(1)	(2)	(3)	(4)
infrequent children's names, mean	0.834		0.820	1.452
	(0.033)		(0.032)	(0.053)
frontier x infrequent children's names, mean	0.304		0.303	0.560
	(0.065)		(0.064)	(0.156)
father has infrequent name		0.209	0.173	0.466
		(0.016)	(0.015)	(0.027)
frontier x father has infrequent name		0.014	-0.003	-0.086
		(0.041)	(0.040)	(0.106)
Observations	5,673,688	5,673,688	5,673,688	1,993,201
Mean Dep. Var.	18.1	18.1	18.1	26.1
R^2	0.07	0.07	0.07	0.10
County×Year Fixed Effects	✓	✓	✓	✓
Excluding Farmers				✓

Notes: This table reports estimates of equation (9). The sample includes all white native-born men with a non-missing occupational score and at least one child age 0–10 in the 1850–1880 Censuses. The dependent variable, occupational score, range from 0 to 100 and are provided by Ruggles et al. (2019) for 1850 and 1880. We construct the scores directly for 1860 and 1870 using a crosswalk of occupational descriptions to codes for available years. Column 4 omits all fathers with an occupational description that includes the string "farm". The infrequent children's names are computed over all of the children age 0–10. The infrequent father's name is defined with respect to all other native-born white men born in the same decade as the father. The frontier indicator equals one if the county is within 100 km of the frontier. Standard errors are clustered by county.

Appendix

This appendix includes the following material: Appendix A details the procedure for mapping the frontier; Appendix B presents robustness checks for Sections 3 and 4; Appendix C presents additional results on partisanship; Appendix D describes our instrumental variables results; Appendix E provides robustness checks for Section 5.2; Appendix F presents alternative approaches to inference; Appendix G provides further results characterizing historical frontier demographics and institutions; Appendix H presents a case study to illustrate the long-run effects; Appendix I presents an alternative estimate of selective migration; Appendix J presents additional results; and Appendix K describes data sources and construction.

A Mapping the Frontier

This section provides a step-by-step description of how we construct the frontier lines for each year between 1790–1890.

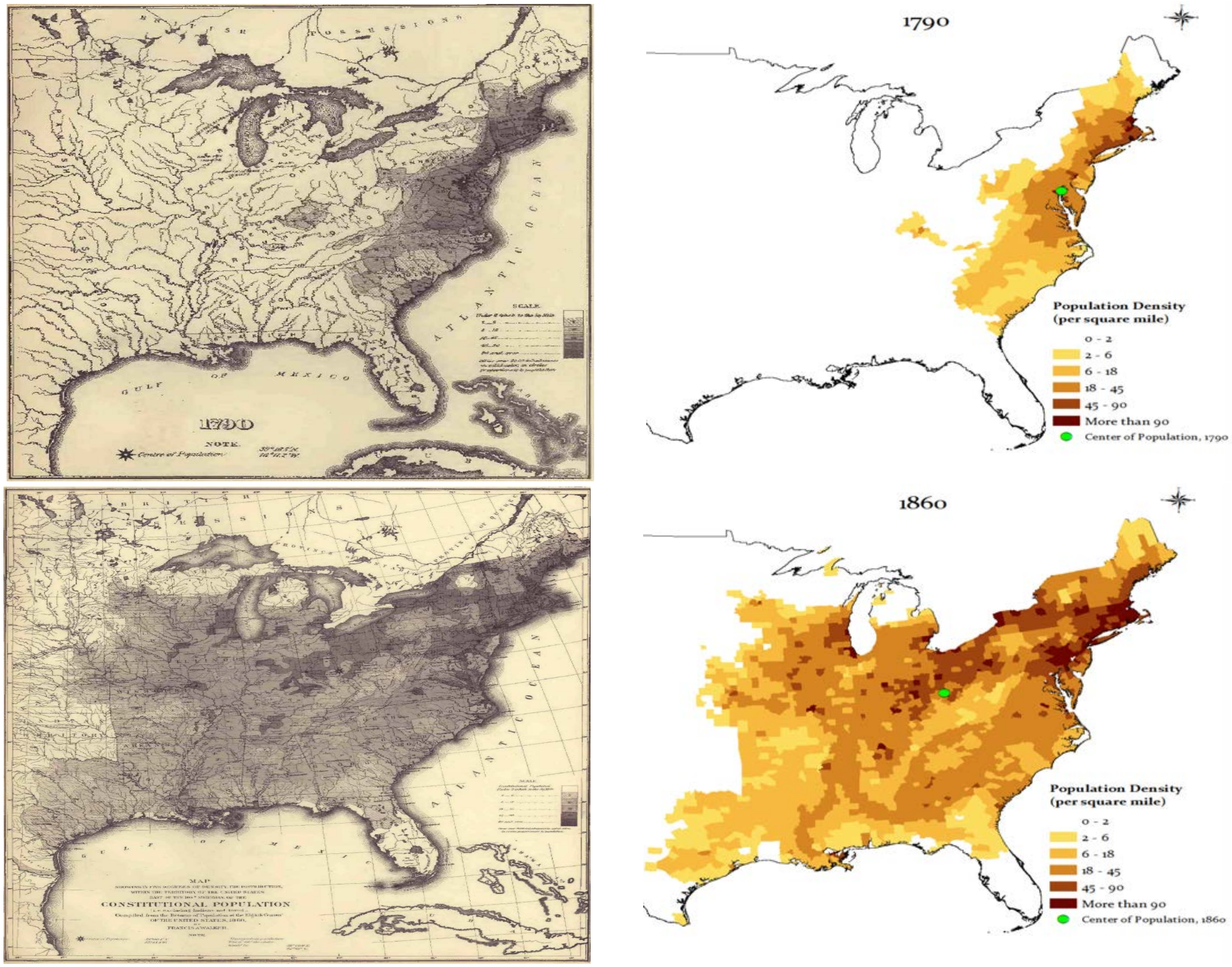
1. Calculate county level population density per square mile for each year in 1790–1890 using the 2010 county boundaries. First, we harmonize the county-level population data from each year to the 2010 county boundaries using the procedure discussed in Section 2. For intercensal years, we interpolate county-level population by assuming a constant annual population growth rate that matches the decadal growth rate (replacing initial zeros with 0.01 to avoid infinite growth rates). Then, using the 2010 county boundaries shapefile, we calculate the county-level population density as the ratio of population over county area in square miles.

2. Draw a contour line at population density equal to 2 people per square mile for each year. We use ArcGIS and the 2010 county boundaries. First, for each year, we convert the polygon containing the county level population density data into a raster file using ***PolygonToRaster*** tool and set population density for the given year as the ***"value field"*** for the conversion. Then, using the ***ContourList*** tool, select the raster file created in the preceding step as an input and set the ***"contour value"*** to "2" to create contour lines at population density equal to 2. The resulting lines delineate the counties that have a population density below 2 people per square mile from those counties that have a population density above 2.

3. Clean the contour lines to retain only the significant frontier lines. With the purpose of capturing historical notions of the frontier as "margins of civilization," we discard all contour line segments less than 500 km and also discard isolated pockets of relatively sparse populations within the main area of settled territory. These isolated pockets are the "inner islands" formed by counties with population density below 2 people per square mile surrounded by counties with population density above 2 people per square mile. A second set of frontier lines emerge in the West Coast in mid-19th century. This process of settlement was marked by the Gold Rush and different historical forces than the main east-to-west expansion, so for our baseline analysis we focus on the territory spanned by east-to-west expansion. We do this by keeping only those frontier lines that are east of the westernmost east-to-west frontier line in 1890. In the robustness analysis, we add the West Coast to our baseline sample.

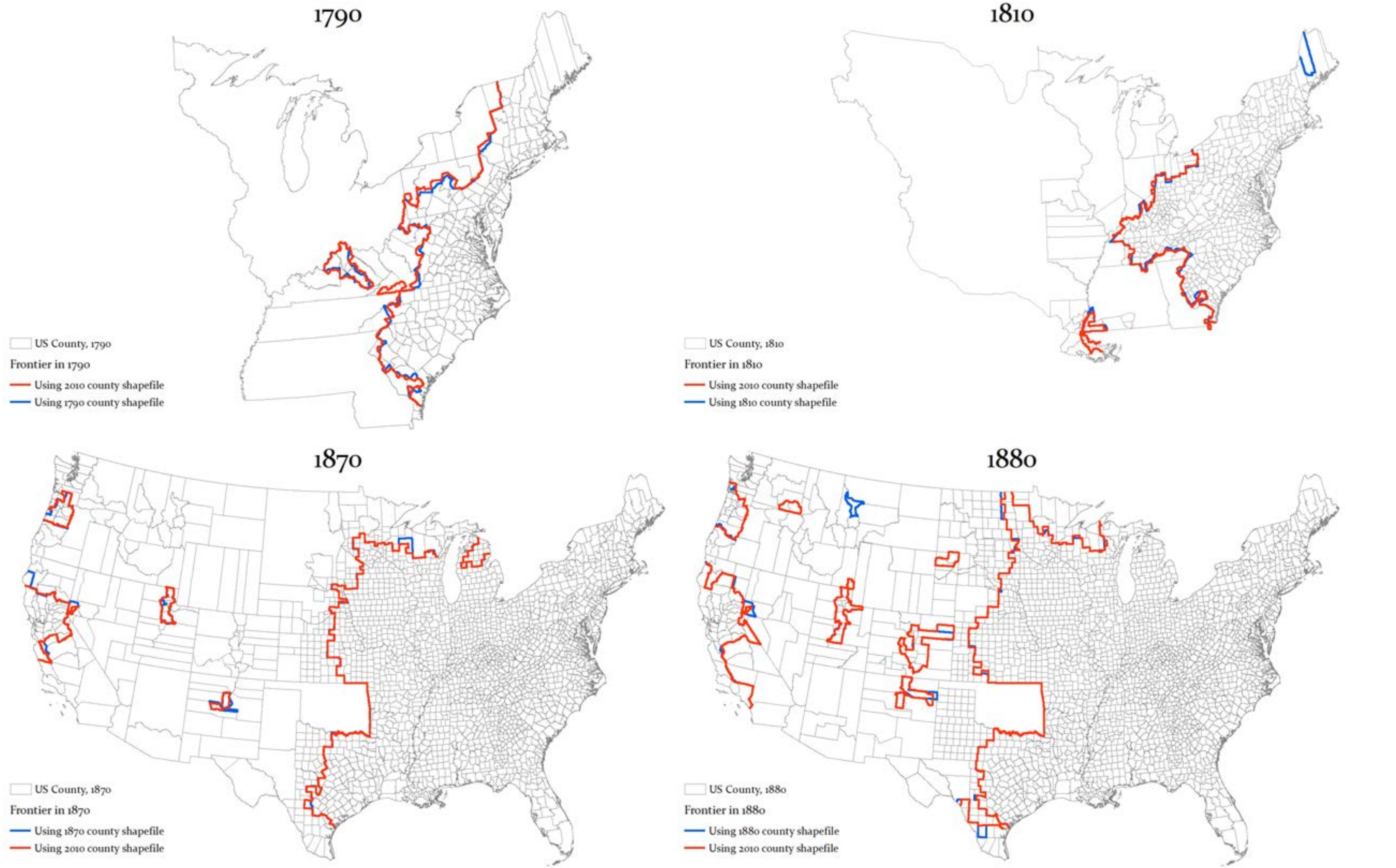
We select line segments based on length and location (e.g., X centroid of the line midpoint) in ArcGIS using the ***SelectLayerByAttribute*** tool, and apply ***CopyFeatures*** to keep only the selected lines. In the detailed robustness checks in Section 4.4, we also consider various alternatives to the frontier definition such as changing the line cutoffs, restricting to single westernmost frontier line, including the "inner island" lines, and considering the frontier lines that emerge from the West Coast.

Figure A.1: Population Density Maps from the 1890 Census Report and Our Maps (1790 and 1860)



Notes: This figure compares the maps of population density in 1790 and 1860 from the *Progress of the Nation* Census report (on the left) with the maps we constructed for 2010 county boundaries using the procedure described in Section 2 (on the right).

Figure A.2: Frontier Lines Using Contemporaneous vs 2010 County Boundaries for selected years



Notes: Based on county level Population Census data from 1790-1880 and NHGIS county shapefiles. The figures provide the county boundaries for selected years and the frontier lines for the corresponding years drawn using the contemporaneous county boundaries as well as the 2010 county boundary. The frontier lines delineate the counties that had population density of two persons or higher. The frontier lines in blue are drawn using the contemporaneous county boundaries whereas the frontier lines in red are drawn using the 2010 county boundaries (after the data harmonization discussed in Section 2.1).

B Further Robustness Checks

B.1 Robustness of the Historical Frontier Differential in Individualistic Names

Table B.1: Robustness Check on Historical Names Measures in Table 1

	(1)	(2)	(3)	(4)
	Raw Reported Name		Metaphone-Adjusted Name	
	Baseline OLS	Foreign-Born NN-Matching	Baseline OLS	Foreign-Born NN-Matching
frontier county	0.022	0.014	0.018	0.011
	(0.007)	(0.006)	(0.006)	(0.005)
Number of County-Years	6,907	6,905	6,907	6,905
Mean Dep. Var. in Non-Frontier Counties	0.631	0.631	0.602	0.602

Notes: Columns 1 and 3 report the baseline estimates from Table 1. Columns 2 and 4 report, for these same outcomes, nearest-neighbor matching estimates of the frontier differential in individualistic names. We match on the county with the most similar foreign-born population share in the given Census year. These estimates are based on the single nearest-neighbor. Standard errors in odd columns are clustered using the grid cell approach of Bester et al. (2011) as described in Section 4.1 and in even columns are bias-adjusted and robust following best practice in the matching literature.

B.2 Alternative Survey-Based Proxy for Contemporary Individualism

Beyond infrequent names, we draw upon a well-suited measure from the ANES data to provide further evidence of the link between TFE and high levels of individualism. Specifically, we use the 1990 ANES round in which respondents were asked whether (1) "it is more important to be a cooperative person who works well with others", or (2) "it is more important to be a self-reliant person able to take care of oneself." While this question was designed explicitly for studies of American individualism (see Markus, 2001), unfortunately, it was only asked in a single round.

Table B.2 below provides evidence that self-reliant preferences are stronger today in counties with longer exposure to the frontier historically. Around 55 percent of individuals respond in support of the cooperative answer. However, across different specifications, each decade of additional TFE is associated with around 2–6 percentage points lower support for cooperation over self-reliance. While the results with the full set of controls are noisy, we nevertheless view these findings as at least suggestive of long-standing claims about the rugged individualism pervasive on the frontier. In linking to results elsewhere in the paper, it is worth noting that individuals that identify as Republican in the ANES data are around 15–20 percent more likely to believe that it is better to be a self-reliant than a cooperative person.

Given the small number of counties, we retain this outcome in the appendix rather than in the main tables. Nevertheless, it is reassuring that the results align with our findings for other outcomes with more systematic coverage.

Table B.2: Total Frontier Experience and Contemporary Cooperation vs. Self-Reliance

	(1)	(2)	(3)	(4)
total frontier experience	-0.019	-0.025	-0.041	-0.026
	(0.009)	(0.009)	(0.014)	(0.012)
Oster δ for $\beta = 0$	-2.77	-2.61	-15.37	-249.36
Number of Individuals	567	567	567	567
Number of Counties	48	48	48	48
Mean of Dependent Variable	0.549	0.549	0.549	0.549
R^2	0.01	0.02	0.02	0.03
Individual Demographic Controls	✓	✓	✓	✓
Division Fixed Effects		✓	✓	✓
State Fixed Effects			✓	
Geographic/Agroclimatic Controls				✓

Notes: This table reports estimates for a dependent variable based on a proxy for individualism in the 1990 round of ANES, covering 567 individuals in 48 counties across 17 states in our sample. The measure asks individuals whether (1) "it is more important to be a cooperative person who works well with others", or (2) "it is more important to be a self-reliant person able to take care of oneself." The dependent variable equals one if they answer (1). We report the same set of specifications in columns 1–4 as in Table 2 to demonstrate the statistically and economically significant effect sizes despite the coverage limitations. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. The Oster (2019) tests are with reference to a baseline specification with no controls.

B.3 Robustness of the Long-Run Effects of TFE on Individualistic Names

We present below several tables with results discussed in the paper. Appendix Tables B.3 and B.4 demonstrate robustness of the infrequent names measure to alternative specifications of the reference group (national, division, state, county), reported name versus phonetic sound (metaphone), and top 10 versus top 100 in terms of defining infrequency. These tables also report results for the non-patronymic/-matronymic measure, including a version that is based solely on first-born children of each gender.

Appendix Table B.3 reports county-level results. Panel (a) reports estimates of the specification in column 2 of Table 2 for the different outcomes listed at the top of each column. Panels (b) and (c) report analogous estimates for columns 3 and 4 of Table 2 which include fixed effects for county pairs with, respectively, the most similar population density and foreign-born population shares in 1940.

Appendix Table B.4 reports individual-level results rather than county-level mean outcomes. Panels (a)–(c) are as in Table B.3 with added fixed effects for child age, birth order, and gender. Panel (d) augments the panel (a) specification with nearly 400,000 fixed effects for family surnames. Panel (e) adds those surname fixed effects to the panel (b) specification.

Finally, Appendix Table B.5 shows that the baseline results for individualistic names look similar in each decade before 1940 but after the closing of the frontier.

Table B.3: TFE and Alternative Measures of Individualistic Names, County-Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Raw Reported Name					Metaphone-Adjusted Name					Non-Patronymic/Matronymic	
	Top 10 National	Top 10 Division	Top 10 State	Top 10 County	Top 100 Division	Top 10 National	Top 10 Division	Top 10 State	Top 10 County	Top 100 Division	All	1st Born
	Panel (a): Baseline Specification											
total frontier experience	0.144	0.141	0.149	0.163	0.142	0.150	0.141	0.148	0.159	0.132	0.224	0.241
	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.022)	(0.021)	(0.024)	(0.024)	(0.022)	(0.029)	(0.030)
Oster δ for $\beta = 0$	-7.93	-13.80	80.78	12.99	-34.76	-8.57	-16.20	98.75	19.66	-30.35	-6.82	-5.63
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.60	0.60	0.56	0.55	0.70	0.55	0.58	0.49	0.49	0.69	0.37	0.31
	Panel (b): Nearest-Neighbor Matching on Population Density in 1940											
total frontier experience	0.102	0.096	0.097	0.134	0.097	0.102	0.088	0.093	0.119	0.089	0.181	0.183
	(0.022)	(0.022)	(0.024)	(0.024)	(0.023)	(0.022)	(0.022)	(0.025)	(0.025)	(0.023)	(0.027)	(0.028)
Oster δ for $\beta = 0$	3.22	2.63	2.58	5.33	1.97	3.08	2.32	2.93	4.41	2.20	11.63	8.98
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.85	0.85	0.83	0.82	0.87	0.84	0.85	0.81	0.80	0.87	0.80	0.79
	Panel (c): Nearest-Neighbor Matching on Foreign-Born Share in 1940											
total frontier experience	0.090	0.086	0.098	0.114	0.107	0.095	0.089	0.103	0.119	0.106	0.126	0.139
	(0.018)	(0.018)	(0.019)	(0.020)	(0.017)	(0.020)	(0.019)	(0.022)	(0.022)	(0.017)	(0.025)	(0.027)
Oster δ for $\beta = 0$	2.16	1.96	2.73	2.85	3.02	2.52	2.41	4.02	4.50	4.81	2.96	3.45
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.86	0.87	0.85	0.84	0.90	0.84	0.85	0.81	0.81	0.90	0.81	0.80

Notes: This table reports estimates in panel (a) based on the column 2 specification in Table 2, in panel (b) based on the column 3 specification in Table 2, and in panel (c) based on the column 4 specification in Table 2. The outcomes from Table 2 are in columns 2 and 7. The other columns are based on alternative specifications of the dependent variable as listed at the top of the table. All other specification details are as in Table 2. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. Panels (b) and (c) additionally two-way cluster on the nearest-neighbor pair.

Table B.4: TFE and Alternative Measures of Individualistic Names, Individual-Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Raw Reported Name					Metaphone-Adjusted Name					Non-Patronymic/Matronymic	
	Top 10 National	Top 10 Division	Top 10 State	Top 10 County	Top 100 Division	Top 10 National	Top 10 Division	Top 10 State	Top 10 County	Top 100 Division	All	1st Born
	Panel (a): FE: Age, State, Birth Order, Gender											
total frontier experience	0.015	0.014	0.013	0.011	0.018	0.014	0.013	0.011	0.010	0.013	0.011	0.018
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Number of Individuals	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	9,986,891	3,134,130
Dep. Var. Mean	0.735	0.732	0.723	0.710	0.353	0.702	0.693	0.679	0.665	0.227	0.925	0.888
R^2	0.03	0.03	0.03	0.02	0.04	0.02	0.02	0.02	0.01	0.03	0.06	0.09
	Panel (b): FE: Panel (a) + Neighbor Population Density Pair											
total frontier experience	0.005	0.004	0.004	0.004	0.007	0.005	0.003	0.003	0.003	0.005	0.004	0.007
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Number of Individuals	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	9,986,891	3,134,130
Dep. Var. Mean	0.735	0.732	0.723	0.710	0.353	0.702	0.693	0.679	0.665	0.227	0.925	0.888
R^2	0.04	0.03	0.03	0.03	0.05	0.02	0.02	0.02	0.02	0.03	0.07	0.10
	Panel (c): FE: Panel (a) + Neighbor Foreign-Born Share Pair											
total frontier experience	0.005	0.005	0.005	0.004	0.009	0.005	0.005	0.004	0.004	0.007	0.003	0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of Individuals	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	9,986,891	3,134,130
Dep. Var. Mean	0.735	0.732	0.723	0.710	0.353	0.702	0.693	0.679	0.665	0.227	0.925	0.888
R^2	0.04	0.03	0.03	0.03	0.05	0.02	0.02	0.02	0.02	0.03	0.07	0.10
	Panel (d): FE: Panel (a) + Last Name											
total frontier experience	0.014	0.013	0.012	0.011	0.017	0.014	0.012	0.010	0.009	0.012	0.010	0.016
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Number of Individuals	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,803,398	2,891,127
Dep. Var. Mean	0.735	0.732	0.724	0.710	0.353	0.703	0.694	0.680	0.665	0.228	0.926	0.889
R^2	0.07	0.07	0.07	0.06	0.08	0.06	0.06	0.06	0.05	0.07	0.10	0.14
	Panel (e): FE: Panel (b) + Last Name											
total frontier experience	0.005	0.005	0.004	0.004	0.006	0.005	0.003	0.003	0.003	0.004	0.004	0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Number of Individuals	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,803,398	2,891,127
Dep. Var. Mean	0.735	0.732	0.724	0.710	0.353	0.703	0.694	0.680	0.665	0.228	0.926	0.889
R^2	0.08	0.08	0.07	0.07	0.09	0.07	0.06	0.06	0.06	0.07	0.11	0.15

Notes: This table reports analogous individual-child-level regressions of Table B.3. This allows for the inclusion child age, gender, and birth order fixed effects. Panels (d) and (e) additionally include fixed effects for family surname. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. The sample size in column 11 is smaller than in prior columns because it excludes all boys with no father and girls with no mother present in the household.

Table B.5: Persistence of the Effect of TFE on Individualistic Names, 1910–1940

	1910	1920	1930	1940
	(1)	(2)	(3)	(4)
	Panel (a): Infrequent Names			
total frontier experience	0.170	0.157	0.138	0.141
	(0.025)	(0.023)	(0.023)	(0.021)
Number of Counties	2,036	2,036	2,036	2,036
R^2	0.50	0.53	0.56	0.60
	Panel (b): Infrequent Names Metaphone-Adjusted			
total frontier experience	0.169	0.171	0.140	0.141
	(0.027)	(0.024)	(0.023)	(0.021)
Number of Counties	2,036	2,036	2,036	2,036
R^2	0.41	0.48	0.51	0.58

Notes: This table reports analogous estimates of Table 2 but for each year since 1910. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1.

B.4 Robustness: Additional Controls, Regional Heterogeneity, and Extended Frontier Era

Appendix Table B.6 reports the checks on omitted confounders discussed in Section 4.4. The full elaboration of coefficients on the control variables can be seen in Appendix Table J.3.

Table B.6: Additional Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Baseline controlling for. . .														
ruggedness		✓												✓
rainfall risk			✓											✓
distance to nearest portage site				✓										✓
distance to nearest mine					✓									✓
distance to nearest Indian battle						✓								✓
slave population share, 1860							✓							✓
sex ratio, 1890								✓						✓
immigrant share, 1890									✓					✓
Scottish and Irish immigrant share, 1890										✓				✓
birthplace diversity, 1890											✓			✓
years connected to railroad by 1890												✓		✓
manufacturing employment share, 1890													✓	✓
	Panel (a): Infrequent Name Share in 1940 (standardized)													
total frontier experience	0.141	0.143	0.139	0.143	0.141	0.145	0.106	0.142	0.134	0.107	0.127	0.094	0.151	0.089
	(0.021)	(0.020)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.019)	(0.020)	(0.022)	(0.020)	(0.019)
Oster δ for $\beta = 0$	-13.80	-11.03	-40.85	-8.28	-14.43	-6.38	1.23	-13.21	10.07	2.03	3.86	1.16	-6.57	1.61
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.60	0.62	0.60	0.60	0.61	0.60	0.62	0.61	0.61	0.67	0.61	0.66	0.66	0.76
	Panel (b): Infrequent Name Share in 1940, Metaphone-Adjusted (standardized)													
total frontier experience	0.141	0.143	0.139	0.143	0.141	0.144	0.107	0.141	0.135	0.106	0.128	0.094	0.150	0.091
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)	(0.020)	(0.020)	(0.021)	(0.022)	(0.021)	(0.019)
Oster δ for $\beta = 0$	-16.20	-12.62	-36.47	-8.50	-17.00	-8.30	1.49	-15.37	16.69	2.21	5.20	1.30	-7.25	1.93
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.58	0.60	0.58	0.58	0.58	0.58	0.59	0.58	0.58	0.65	0.59	0.63	0.63	0.73
	Panel (c): Mean Government Preferences Outcomes (CCES), 2006–16													
total frontier experience	0.014	0.013	0.014	0.014	0.014	0.013	0.014	0.014	0.011	0.011	0.010	0.012	0.011	0.009
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Oster δ for $\beta = 0$	6.49	6.23	7.07	7.00	6.34	5.46	7.00	6.64	3.73	3.36	3.33	4.20	3.71	2.26
Number of Counties	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759
Mean of Dependent Variable	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
R^2	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	Panel (d): County Property Tax Rate in 2010													
total frontier experience	-0.034	-0.034	-0.034	-0.033	-0.034	-0.033	-0.033	-0.034	-0.028	-0.026	-0.027	-0.023	-0.036	-0.023
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Oster δ for $\beta = 0$	-27.45	-7.11	-3.23	2.54	-30.95	9.19	8.59	-30.17	0.55	0.48	0.49	0.28	-2.08	0.42
Number of Counties	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029
Mean of Dependent Variable	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
R^2	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.83	0.85
	Panel (e): Republican Vote Share, Average 2000–16													
total frontier experience	2.055	2.050	2.115	2.095	2.055	2.172	1.399	2.060	1.715	1.717	1.689	1.640	2.137	0.931
	(0.349)	(0.349)	(0.338)	(0.344)	(0.350)	(0.351)	(0.361)	(0.347)	(0.328)	(0.340)	(0.327)	(0.361)	(0.350)	(0.316)
Oster δ for $\beta = 0$	-8.55	-8.93	-6.43	-6.75	-8.56	-5.10	2.34	-8.92	7.33	7.40	6.72	4.53	-7.81	1.35
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
Mean of Dependent Variable	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04
R^2	0.33	0.33	0.34	0.33	0.33	0.34	0.38	0.34	0.38	0.38	0.39	0.36	0.36	0.49
State Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table augments the baseline specification, reproduced in column 1, with additional controls. The variables are defined in Section 4.4 and at the end of Appendix K, but we note here that the measure in column 5 is based on the known mining sites pre-1890. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. The Oster (2019) tests are with reference to a baseline specification with no controls.

Appendix Table B.7 reports the regional heterogeneity and extended time-frame results described in Section 4.4. We begin by adding West Coast frontier counties to our sample. These 105 counties were settled starting in the mid-19th century and were located to the west of the major frontier line on the West Coast in 1890 (the year in which the Census declared the frontier closed). As shown in column 1, for all key outcomes, the estimated effects of TFE remain effectively unchanged.

Then, we split the sample by Census region and show that the effects of TFE hold separately in the Midwest (column 2), the South (column 3), and the West (column 4). The coefficient estimates are generally smaller and noisier in the West, which can be explained in part by the small sample size (152 counties). In subsequent columns 5–8, we extend the frontier time period through 1950, incorporating in our sample counties that experienced frontier conditions beyond 1890. Here, the effects of TFE are economically and statistically significant across all regions.

Table B.7: West Coast, Extended Time Frame, Regional Heterogeneity

Frontier Time Frame:	Baseline (1790–1890)				Extended (1790–1950)			
Regional Sample Restriction:	Baseline + West Coast (1)	Only Midwest (2)	Only South (3)	Only West (4)	Extended Sample (5)	Only Midwest (6)	Only South (7)	Only West (8)
	Panel (a): Infrequent Name Share in 1940 (standardized)							
total frontier experience	0.138	0.247	0.163	0.103	0.087	0.124	0.114	0.068
	(0.020)	(0.044)	(0.030)	(0.069)	(0.014)	(0.034)	(0.026)	(0.020)
Number of Counties	2,141	987	936	152	2,500	1,038	1,074	322
	Panel (b): Infrequent Name Share in 1940, Metaphone-Adjusted (standardized)							
total frontier experience	0.138	0.242	0.148	0.112	0.085	0.125	0.093	0.071
	(0.021)	(0.043)	(0.029)	(0.072)	(0.014)	(0.033)	(0.024)	(0.019)
Number of Counties	2,141	987	936	152	2,500	1,038	1,074	322
	Panel (c): Mean Government Preferences Outcomes (CCES), 2006–16							
total frontier experience	0.015	0.019	0.010	0.026	0.013	0.017	0.010	0.013
	(0.003)	(0.004)	(0.003)	(0.012)	(0.002)	(0.004)	(0.003)	(0.004)
Number of Individuals	140,715	49,218	52,285	32,319	158,403	49,479	55,462	46,569
Mean of Dependent Variable	0.40	0.40	0.43	0.36	0.41	0.40	0.43	0.38
	Panel (d): County Property Tax Rate in 2010							
total frontier experience	-0.031	-0.051	-0.027	-0.006	-0.025	-0.042	-0.031	-0.009
	(0.006)	(0.014)	(0.007)	(0.013)	(0.004)	(0.012)	(0.006)	(0.004)
Number of Counties	2,134	981	935	152	2,491	1,029	1,074	322
Mean of Dependent Variable	1.01	1.24	0.75	0.76	0.98	1.23	0.78	0.72
	Panel (e): Average Republican Vote Share over 2000-2016							
total frontier experience	2.070	1.882	2.458	1.459	1.302	1.515	1.429	1.197
	(0.332)	(0.414)	(0.396)	(0.890)	(0.256)	(0.350)	(0.422)	(0.274)
Number of Counties	2,141	987	936	152	2,500	1,038	1,074	322
Mean of Dependent Variable	59.43	59.15	61.78	48.81	60.49	59.43	63.18	56.10
State Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Focusing on five key outcomes across panels (a)–(e), this table extends our baseline sample of counties and examines region-by-region sample splits. Column 1 adds 105 counties along the secondary West Coast frontier (see Figure 3). Column 2 restricts to counties in the Midwest Census region, column 3 restricts to the South region, and column 4 restricts to the West, which includes the 105 counties added in column 1 plus 47 others in states in the West region but falling inside the 1890 main east-to-west frontier line. Column 5 expands the column 1 sample to include counties beyond the (main and secondary) 1890 frontier lines but inside the eventual frontier line realized by 1950. Columns 6–8 then proceed with the same region-by-region sample splits. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1.

C Further Results on Voting and Partisanship

Appendix Table C.1 reports estimates for the effects of TFE on partisan legislative speech associated with opposition to big government and redistribution. We compute measures of speech intensity based on the corpus of legislative speech and associated bigrams provided in Gentzkow et al. (2019). In particular we estimate the following specification:

$$\left(\frac{\text{bigram } b}{\text{all words}}\right)_{d(\ell)t} = \alpha + \beta \text{total frontier experience}_d + \mathbf{x}_d'\gamma + \mathbf{FE} + \varepsilon_{d(\ell)t}, \tag{C.1}$$

where the dependent variable captures the share of bigrams related to topic b in all words used by Republican legislator ℓ from congressional district d in congress year t.[1] We include fixed effects $\mathbf{FE}$ for the Census division in which d lies as well as the congress year. Standard errors are clustered at the congressional district level.

We consider three topics b that are particularly relevant to the other anti-statist outcomes we consider in the paper. These include, across panels (a) "Big Government" based on the "big govern" bigram, (b) "Taxation" based on the top four Republican-leaning bigrams on this topic with highest average partisanship across all sessions identified by Gentzkow et al. (2019) (tax increas, rais tax, tax relief, american taxpay), and (c) "Budget" based on the top four Republican-leaning bigrams on this topic with highest average partisanship across all sessions identified by Gentzkow et al. (2019) (govern spend, feder spend, intern revenu, treasuri depart). TFE is computed at the congressional district level rather than the county level as in our core specifications in the paper.

The estimates in Table C.1 suggest that TFE amplified the supply of political opposition to big government among Republican legislators, especially beginning in the mid-1990s. While partisanship around these themes and issues grew for politicians everywhere around this time, our estimates suggest that this growth may have been differential in regions with greater TFE. While some of the differences with earlier periods are noisy (e.g., in panels (b) and (c)), this supply-side pattern is consistent with the growing demand-side differential seen in the Republican presidential vote shares in Figure 5. These findings paint a consistent picture of TFE capturing latent cultural attitudes that can be activated around salient political themes.

As noted in Section 4.3, these time patterns could be due to changes in the type of elected representative or to a change in the type of speech used by representatives that would have been elected otherwise. What's important here is that the patterns line up with the strong voter demand for attention to such issues seen in Table 3.

[1] We multiply the dependent variable by 1,000 for presentational purposes.

Table C.1: TFE and Opposition to Big Government in Republican Legislator Speech

	1902-30 (1)	1932-60 (2)	1962-90 (3)	1992-2016 (4)
	Dep. Var.: Share of Legislator Speech with Bigrams Including [...]			
	Panel (a): Big Government			
total frontier experience	-0.000	-0.008	0.001	0.021
	(0.000)	(0.015)	(0.009)	(0.010)
Number of Counties	2,100	1,510	1,447	1,638
R^2	0.01	0.02	0.04	0.06
	Panel (b): Taxation			
total frontier experience	-0.009	-0.003	0.079	0.122
	(0.013)	(0.055)	(0.105)	(0.136)
Number of Counties	2,100	1,510	1,447	1,638
R^2	0.04	0.04	0.12	0.10
	Panel (c): Budget			
total frontier experience	-0.034	0.099	0.067	0.079
	(0.070)	(0.097)	(0.056)	(0.036)
Number of Counties	2,100	1,510	1,447	1,638
R^2	0.02	0.03	0.09	0.06

Notes: This table reports estimates of equation (C.1) across four different time periods: 1902–30 (column 1), 1932–60 (column 2), 1962–90 (column 3), and 1992–2016 (column 4). Other details on the specifications can be found in the discussion above.

Table C.2: Total Frontier Experience and Preferences Over Partisan Policy Issues

Dependent Variable:	Opposes Affordable Care Act (1)	Opposes Increasing Minimum Wage (2)	Opposes Banning Assault Rifles (3)	Opposes Regulation of CO_2 Emissions (4)
total frontier experience	0.022	0.023	0.015	0.016
	(0.004)	(0.008)	(0.004)	(0.004)
Oster δ for $\beta = 0$	7.33	11.07	9.44	10.77
Number of Individuals	29,446	5,134	29,404	29,215
Number of Counties	1,728	1,066	1,723	1,718
Mean of Dependent Variable	0.53	0.31	0.37	0.32
R^2	0.06	0.06	0.09	0.08
Survey Wave Fixed Effects	✓	✓	✓	✓
Individual Demographic Controls	✓	✓	✓	✓
State Fixed Effects	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓

Notes: This table reports estimates of equation (4) for four measures of support for conservative issues that are particularly relevant to the frontier setting in historical accounts. The dependent variables are all binary indicators based on questions in the CCES across different years. The measure in Column 1 equals one if the individual in 2014 believes that the Affordable Care Act (ACA) should be repealed, in Column 2 equals one if the individual in 2007 opposes an increase in the minimum wage, in Column 3 equals one if the individual in 2014 opposes a ban on assault rifles, and in Column 4 equals one if the individual in 2014 opposes regulation of pollution by the Environmental Protection Agency (EPA). The set of specifications are otherwise the same as in Table 3; see the notes therein for details. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1. The Oster (2019) tests are with reference to a baseline specification that only includes state fixed effects.

D Instrumental Variables Strategy

This section presents a suite of results corresponding to our discussion of the IV specification in Section 4.4. Table D.1 presents IV estimates for the same five primary outcomes as in Tables 4–B.6. In Panel (a), we find large, significant effects of TFE that are slightly larger but statistically indistinguishable from the OLS estimates. Panel (b) shows similar results when using predicted rather than actual migrant flows in the IV construction. Overall, the IV exercises help clarify the identifying variation in TFE. The first stage results in Appendix Table D.2 show that there are various geoclimatic predictors of TFE, but national immigration inflows, which are unrelated to local conditions of any given county, also account for a sizable amount of variation in TFE. Coupled with the robustness checks in the previous section, the similarity of the IV and OLS results in Table D.1 suggests that our findings are not driven by local conditions determining both TFE and outcomes of interest today. Below, we further develop the IV and its identifying assumptions.

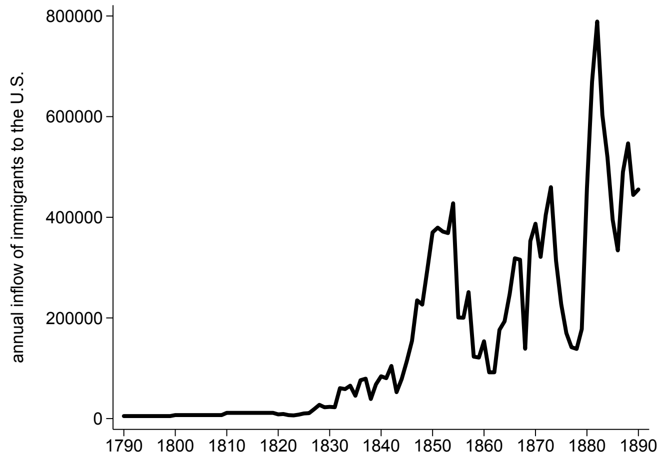
Table D.1: Instrumental Variables (IV) Estimates for Summary Outcomes

Dependent Variable:	Infrequent Name		Mean	County	Republican
	raw	metaphone	Gov. Prefs.	Property	Vote Share
	standardized		CCES	Tax Rate	Avg.
	1940	1940	2006–16	2010	2000–16
	(1)	(2)	(3)	(4)	(5)
	Panel (a): IV = Log Average Actual National Migration Inflows Over 30 Years				
total frontier experience	0.193	0.215	0.010	-0.045	3.407
	(0.039)	(0.060)	(0.004)	(0.014)	(0.585)
Number of Observations	2,036	2,036	112,759	2,029	2,036
Mean of Dependent Variable	0.00	0.00	0.41	1.02	60.04
First Stage F Statistic	193.64	193.64	40.34	194.13	193.64
	Panel (b): IV = Log Average Predicted National Migration Inflows Over 30 Years				
total frontier experience	0.223	0.232	0.007	-0.049	3.177
	(0.044)	(0.062)	(0.004)	(0.014)	(0.624)
Number of Observations	2,036	2,036	112,759	2,029	2,036
Mean of Dependent Variable	0.00	0.00	0.41	1.02	60.04
First Stage F Statistic	195.84	195.84	44.56	196.31	195.84
State Fixed Effects	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓

Notes: This table reports instrumental variables estimates of equation (4) based on the instruments described in Section 4.4. We again report results for the four summary outcomes examined in prior tables, and total frontier experience is measured in decades. Panel (a) reports the IV estimates for the baseline sample and specification using the log of the average national annual actual migration inflows over the 30 years since the frontier is within 110km from the county centroid. Panel (b) reports the estimates using the IV constructed based on annual migration inflows to the US predicted by weather shocks in Europe. The details on the construction of both instrumental variables are presented in the Appendix Section D. The first-stage F statistics are cluster-robust, and standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1.

Figure D.1 shows immigration inflows to the U.S. over the study period.

Figure D.1: Annual Migration Inflows



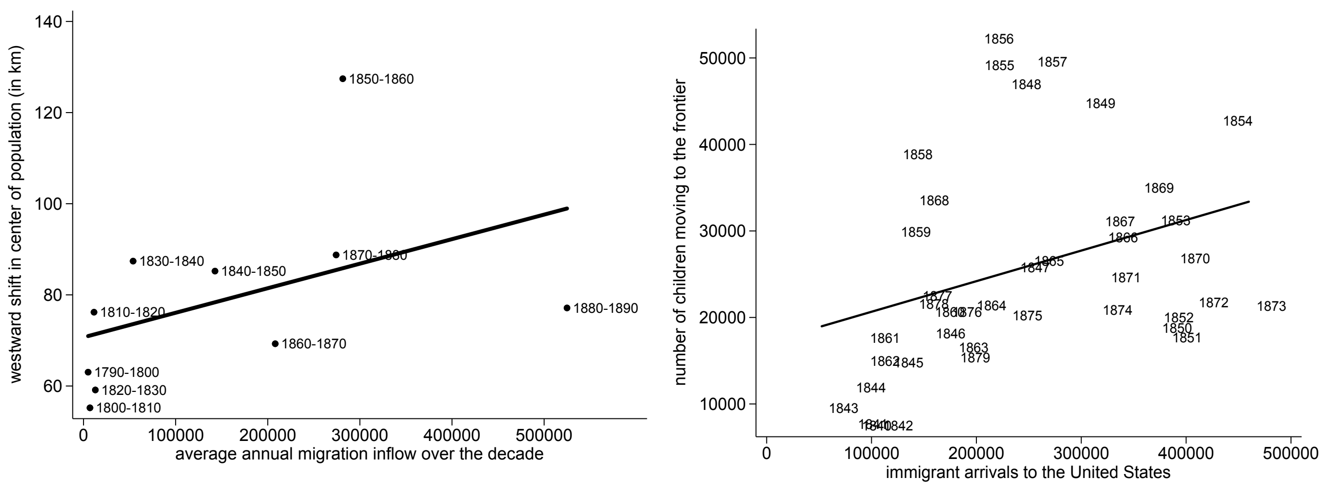
Notes: This figure plots the total number of migrants entering the United States, 1790-1890. The data for 1820–1890 is available from the Migration Policy Institute (2016), while the data for 1790-1819 is imputed from Tucker (1843).

Figure D.2(a) then shows the strong positive correlation between these inflows by decade and the speed of westward expansion, proxied by the east-to-west distance traveled by the country's population centroid (the green dot in Figure A.1(b) for 1860). Figure D.2(b) shows that the scale of native-born migration to the frontier is greater in years with more immigrants arriving to the United States. We identify migrants moving to the frontier using the same procedure based on differences in children's birthplaces as detailed in Section 5.

Figure D.2: Immigration and Westward Expansion

(a) Center-of-Population Shift

(b) Migration of Native-Born Children to the Frontier



Notes: Graph (a) plots the length of the decadal westward shift of the center of population (in km) against the average annual immigrant inflow during the decade. The center of population is the point at which weights of equal magnitude corresponding to the location of each person in an imaginary flat surface representing the U.S. would balance out. Graph (b) plots the relationship between the number of immigrant arrivals to the U.S. in a given year and the number of children brought to the frontier with their parents, a sample that we use throughout Section 5.

These scatterplots help visualize the process by which immigrants arriving in the U.S. pushed the edges of settlement westward, which in turn hastened the onward march of the frontier line. In periods with low immigrant inflows, this push slowed down, leading some counties to remain part of the frontier for longer than those that just happened to be getting closer to the frontier line at a time of rapid inflows into the U.S. Table D.2 demonstrates the strong first stage in our main IV regressions from Table D.1.

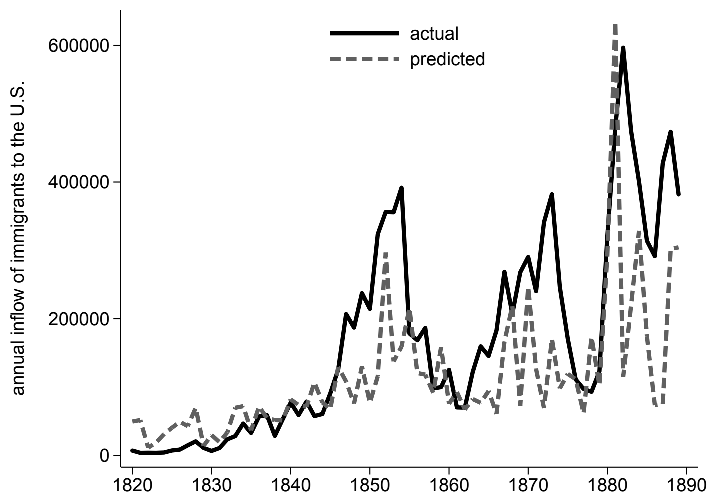
Table D.2: First Stage Results for the Instrumental Variables Estimates in Table D.1

Dependent Variable (in first stage):	total frontier experience	
	(1)	(2)
Log Average Actual National Migration Inflows	-1.016	
	(0.073)	
Log Average Predicted National Migration Inflows		-2.010
		(0.144)
Log county area	0.234	0.232
	(0.072)	(0.073)
Latitude	-0.091	-0.076
	(0.079)	(0.075)
Longitude	-0.153	-0.176
	(0.029)	(0.030)
Mean Annual Temperature	-0.144	-0.102
	(0.071)	(0.068)
Mean Annual Rainfall	-0.002	-0.002
	(0.001)	(0.001)
Mean of Median Altitude	-0.001	-0.001
	(0.000)	(0.000)
Distance to Coast	-0.001	-0.001
	(0.000)	(0.000)
Distance to Rivers	0.003	0.004
	(0.001)	(0.001)
Distance to Lakes	-0.000	-0.000
	(0.000)	(0.000)
Average Agricultural Suitability	2.686	0.876
	(0.788)	(0.752)
Number of Counties	2036	2036
First Stage F Statistic	193.64	195.84

Notes: This table reports the first stage results corresponding to the baseline IV regressions presented in Table D.1.

Section 4.4 shows that the main results are hold in an instrumental variable specification exploiting time series variation in national migration inflows. To address concerns regarding the excludability of the baseline instrument due to pull factors associated with immigrant inflows, we show in Panel B of Table D.1 that the IV results are qualitatively unchanged when using an instrument based on push factors unrelated to frontier conditions. For this version of the IV, we draw on the approach in Sequeira, Nunn and Qian (2020), using country-year level data on migrant inflows from 16 European countries to the US from 1820–1890 and constructing predicted migration outflows induced by weather shocks. First, using country-specific regressions, we predict the annual migrant outflows from each country to the US as a function of country-specific shocks to temperature and rainfall in the prior year (see Sequeira, Nunn and Qian, 2020, for details on these measures). Second, we aggregate across countries to obtain the total predicted migrant inflows to the US for each year. Analogous to our baseline instrument, we then construct the IV for each county in our sample by calculating the average annual predicted migrant inflow to the US over the 30 years starting from the first year in which the given county is just west of the frontier. Figure D.3 shows how the predicted inflows, which isolate push factors, compare to the actual inflows, which naturally include both push and pull. While the data on migrant inflows from Europe to the US is available only starting in 1820, we retain the full sample of counties in the IV regressions by imputing the inflows for 1790-1819 using linear extrapolation of the post-1819 predicted inflows.[1]

Figure D.3: Actual vs. Predicted Immigration Inflows from Europe to the United States



Notes: This figure compares the actual migration inflows from Europe from 1820–1890 to the predicted flows based on the total country-specific predicted outflows using the climatic shocks approach in Sequeira, Nunn and Qian (2020) as described above.

Probing Instrument Validity. While this IV strategy addresses some concerns about omitted variables, the exclusion restriction may not hold. In particular, the immigration flows underlying the IV affect both the scale and composition of migrants to the frontier.[2] During periods of greater immigration (lower predicted TFE) frontier settlers may include relatively more foreigners and non-individualistic native-born. For example, in periods with many immigrants arriving to the east coast, a large number of native-born Americans flowed westward, many of whom could have been non-individualistic types. By a similar logic, large immigration inflows out of Europe (induced by weather shocks) may lead to a greater stock of non-individualistic, foreign-born arriving on the frontier. These population flows would directly lower TFE and reduce the prevalence of individualism. While these types of selective migration could invalidate the IV, neither seems pervasive during the frontier era as we argue here.

We explore this potential bias in the latter half of the 1800s when it is possible to separately relate scale and composition of frontier migrant flows to the national immigration shocks underlying our IV. We identify migrants moving to the frontier using the same procedure based on differences in children's birthplaces as detailed in Section 5. We measure individualism based on the names of children born prior to moving.

In certain contexts, one might expect the prevalence of individualists to be inversely related to the scale of migration. However, Appendix Figure D.4(a) suggests that this is not the case when looking at native-born migration flows to the frontier. Each point on the graph reflects a given year's number of children under the age of 20 migrating to the frontier with their parents and the mean prevalence of

[1] Restricting the sample to counties just west of the frontier after 1820—for which the IV is solely based on predicted flows without extrapolation—delivers similar results, though the estimates are noisier due to the smaller sample size.

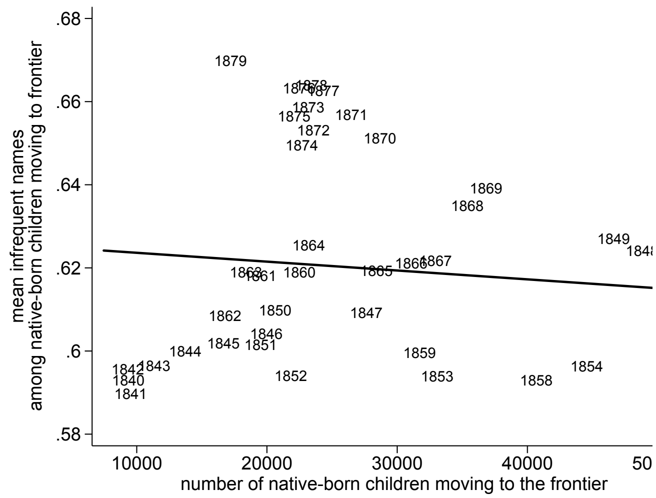
[2] In a standard Roy-Borjas model of migration, the size and composition of migration flows are in general jointly determined and not independent (see, e.g., Grogger and Hanson, 2011).

individualistic names among those children. The lack of correlation between scale and individualism suggests that there is not a mechanical relationship between the two in our setting.

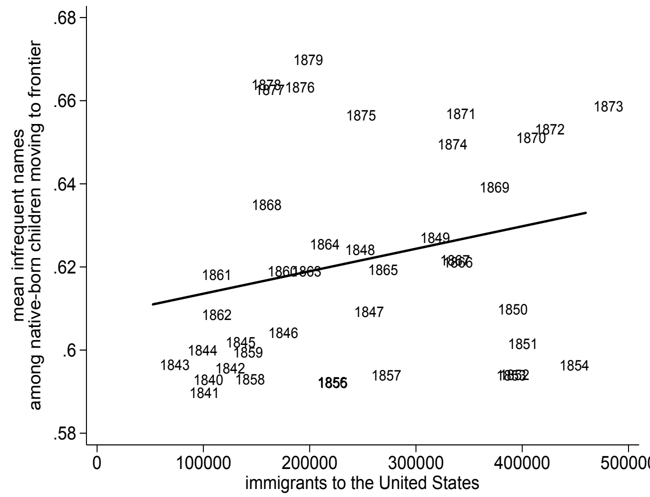
Appendix Figure D.4(b) shows that the prevalence of individualism among frontier migrants is not lower in years with more immigrants arriving to the United States, but rather, weakly higher. This goes against the intuition that selective migration of individualists would be weaker in periods with greater push factors in settled areas. One explanation might be that non-individualists have stronger social networks that allow them to deal with adverse labor market shocks and ultimately remain in settled areas. Without deep social networks, individualists might be more readily pushed to move to the frontier. In any case, the observed patterns tend to alleviate a salient concern about the exclusion restriction in the IV estimation.[3]

Figure D.4: U.S. Immigrant Arrivals and Native-Born Frontier Migration Scale and Individualism

(a) Frontier Migrant Scale and Individualism



(b) U.S. Immigration and Frontier Migrant Individualism



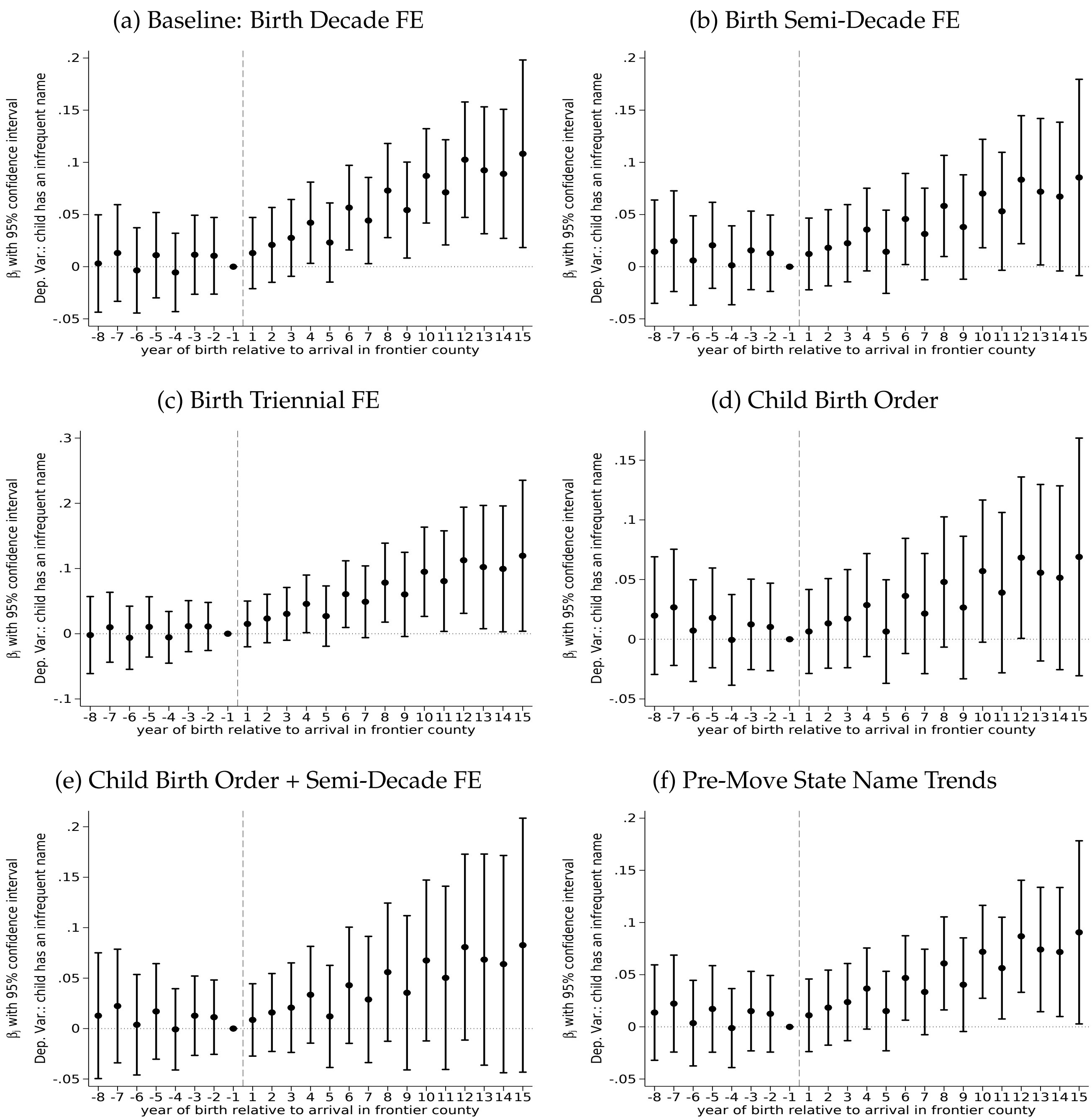
Notes: Graph (a) plots the relationship between the number of children brought to the frontier with their parents and the prevalence of individualistic names among them. Each point is a given year of migration computed based on the procedure used in Section 5 based on the 1850–80 Censuses. Graph (b) plots the relationship between the number of immigrant arrivals to the U.S. in a given year and the prevalence of individualistic children's names among frontier migrants.

[3] Note that each of the graphs in Figure D.4 looks similar when allowing for (cumulative) lags in the number of immigrants to the U.S. and/or when using the alternative, predicted immigrant flows based on weather shocks in Europe.

E Further Robustness Checks on the Exposure Effects in Section 5.2

E.1 Additional Results: Adulthood Exposure Event Study

Figure E.1: Identifying Exposure Effects: Adulthood Exposure (I)



Notes: These graphs report estimates of β_j +/- 95% confidence intervals in equation (6) for $j = -8, \ldots, 15$ (with other j suppressed for presentational purposes). Each β_j can be interpreted as the differential likelihood of an infrequent name being given to a child born j years before/after their parents moved to the frontier, relative to the child born one year prior to moving. The sample includes 57,097 children born to 16,901 families headed by white, native-born parents that moved with at least one child to a frontier county as we observe them in the Census in 1850, 1860, 1870 or 1880. All estimates control for household fixed effects and child gender. Graph (a) additionally includes child birth decade FE, (b) includes 5-yearly birth cohort, (c) includes 3-yearly birth cohort FE, (d) controls for child birth order, (e) controls for child birth order, and (f) controls for the mean gender-specific infrequent name share in each child birth year in the state from which each family migrated from before arriving on the frontier. Standard errors are clustered by contemporaneous county.

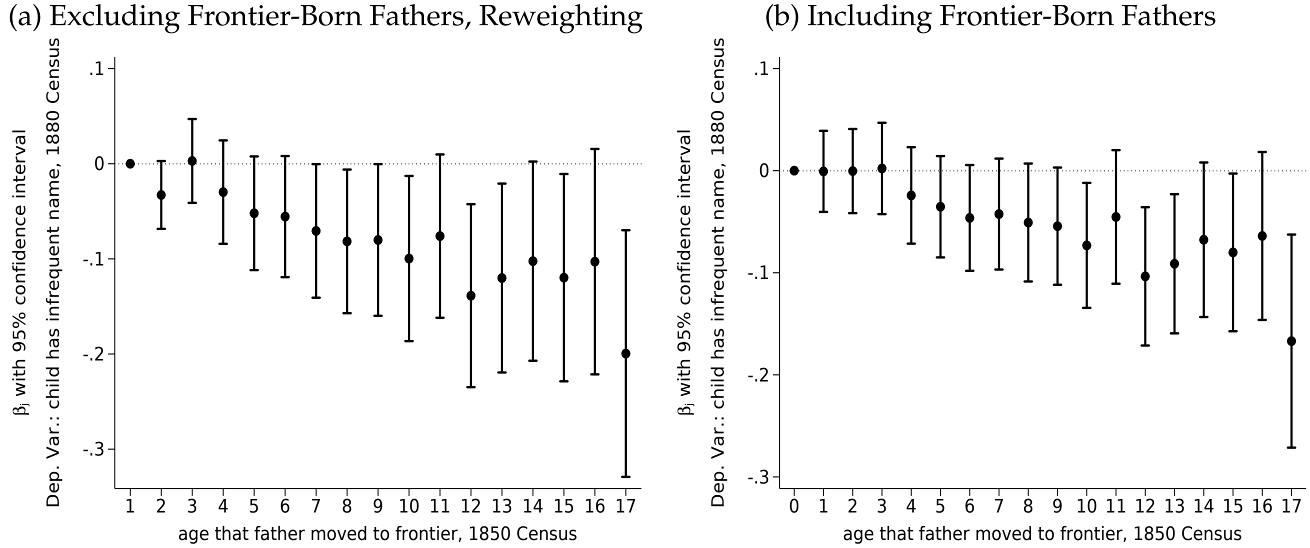
E.2 Additional Results: Childhood Exposure Age-at-Move

As discussed in Section 5.2, our results are also robust to accounting for measurement error in the linking procedure. First, in Figure E.2(a) and panel (a) of Table E.1 below, we reweight the baseline sample by the odds of a successful link estimated as a flexible function of the father's age interacted with whether or not the father himself has infrequent name. Following Bailey et al. (forthcoming), we use these propensity scores to construct inverse probability weights. The results are very similar and in some specifications more precisely estimated than the baseline in Figure 7 and Table 8. This helps to rule out a sample selection bias wherein fathers with more individualistic names, and hence greater inherited individualism, might be more likely to be linked across Censuses. Second, we can further restrict our baseline sample to the children of fathers with unique matches between the 1850 and 1880 Census. This substantially cuts the sample by more than half, which leads to sizable reductions in statistical power. The resulting estimates of continuous age-at-move effects range from -0.010 (0.007) to -0.007 (0.007) across the analogous specifications 1 and 5 in Table 8. In other words, while discarding the considerable information in non-unique matches we use in the baseline (which includes a dummy indicator for such matches), we find estimates that are quantitatively similar but noisy. We cannot reject that the coefficients equal zero, but we also cannot reject that they equal the baseline estimates.

Figure E.2: Identifying Exposure Effects: Adulthood Exposure (I)

(a) Excluding Frontier-Born Fathers, Reweighting

(b) Including Frontier-Born Fathers



Notes: Graph (a) reports a reweighted estimate of Figure 7. We reweight each child observation by the estimated odds (inverse probability weights) that the father was successfully linked across Census rounds. These weights are estimated as a function of the interaction of father's age in 1850 and whether or not the father has infrequent name. In graph (b), the estimates are with respect to children born to fathers who were themselves born on the frontier. The sample in graph (a) (graph (b)) consists of 81,823 (146,085) children age 0–20 in the 1880 Census with fathers hailing from 17,778 (28,776) families observed in the 1850 Census and where at least two brothers (one brother) were born before the family moved to the frontier.

Moreover, the age-at-move estimates are robust to including children whose fathers were born on the frontier. We generalize equation (8) to include all brothers who were born after their parents moved to the frontier, normalizing their ages-at-move to $j = 0$. Doing so in Figure E.2(b) and panel (b) of Table E.1 suggests similar patterns, despite the substantial increase in the sample size.[1] We omit the children of fathers born on the frontier from the baseline in Figure 7 since they may only appear in the sample as a result of selective fertility among parents that found high returns to individualism after arrival to the

[1]This procedure adds a large number of 1850 households for whom a first son was born prior to moving while a second (and higher-order) son was born after moving to the frontier.

frontier. This need not introduce a source of bias per se as these fathers would also have experienced more years of frontier conditions than their older siblings. However, they do introduce a source of sample selectivity just as they did in the event study. That the results look similar with and without these additional fathers suggests that this type of sample selectivity is not a first-order concern.

Table E.1: Identifying Exposure Effects: Childhood Exposure (II)

	(1)	(2)	(3)	(4)	(5)
	(a) Excluding Frontier-Born Fathers and Reweighting by Link Probability				
age-at-move to frontier	-0.009	-0.008	-0.007	-0.006	-0.006
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	81,823	81,823	81,823	81,823	81,823
Number of Families	17,778	17,778	17,778	17,778	17,778
Mean of Dependent Variable	0.68	0.68	0.68	0.68	0.68
R^2	0.27	0.27	0.27	0.27	0.27
	Panel (b): Including Frontier-Born Fathers				
age-at-move to frontier	-0.007	-0.006	-0.005	-0.004	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	146,085	146,085	146,085	146,085	146,085
Number of Families	28,776	28,776	28,776	28,776	28,776
Mean of Dependent Variable	0.69	0.69	0.69	0.69	0.69
R^2	0.25	0.25	0.25	0.25	0.25
Extended Family (1850 Household) FE	✓	✓	✓	✓	✓
State of Residence FE in 1880	–	✓	–	–	–
Child Birth Cohort FE	–	–	decade	5-yearly	3-yearly

Notes: This table reports reweighted estimate of panel (a) in Table 8. We reweight each child observation by the estimated odds (inverse probability weights) that the father was successfully linked across Census rounds. These weights are estimated as a function of the interaction of father's age in 1850 and whether or not the father has infrequent name. Standard errors are two-way clustered by 1850 family and 1880 county.

F Robust Inference

Appendix Table F.1 demonstrates the robustness of our baseline approach to inference in county-level regressions. We consider seven alternative approaches to accounting for spatial autocorrelation in the error terms. Column 1 is our baseline approach of clustering at the level of arbitrary 60 square-mile grid cells following the procedure in Bester et al. (2011). Columns 2–7 adopt the Conley (1999) GMM-based spatial heteroskedasticity-autocorrelation (HAC) consistent procedure that allows for arbitrary correlation in unobservables across all counties within 100, 200, 300, 500, and 1000 km respectively. Column 7 uses two-way clustering on both the arbitrary grid cells in column 1 and the year in which each county entered the frontier. This procedure, based on Cameron et al. (2011), allows for possible correlated unobservable across counties subject to shocks at the same time historically that would have led their county to become relevant to frontier settlers. Finally, column 8 clusters at the state level, using a wild bootstrap procedure to account for the relative small number of states (30 in our main sample). Across all columns, we find no meaningful departures from the precisely estimated effects in our baseline approach. We therefore retain this computationally simple baseline throughout the paper.

Table F.1: Alternative Approaches to Inference

	Base	Spatial HAC: [...] km Bandwidth					Two-Way	State
		100	200	300	500	1000	Base + Yr. Entry	Wild Clust.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel (a): Infrequent Name Share in 1940							
total frontier experience	0.141	0.141	0.141	0.141	0.141	0.141	0.141	0.141
	(0.021)	(0.027)	(0.027)	(0.027)	(0.032)	(0.040)	(0.018)	[0.004]
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
	Panel (b): Infrequent Name Share in 1940, Metaphone-Adjusted							
total frontier experience	0.141	0.141	0.141	0.141	0.141	0.141	0.141	0.141
	(0.021)	(0.028)	(0.029)	(0.030)	(0.033)	(0.041)	(0.019)	[0.005]
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
	Panel (c): County Property Tax Rate in 2010							
total frontier experience	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034
	(0.007)	(0.009)	(0.009)	(0.009)	(0.012)	(0.009)	(0.010)	[0.006]
Number of Counties	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029
	Panel (d): Average Republican Vote Share over 2000–2016							
total frontier experience	2.055	2.055	2.055	2.055	2.055	2.055	2.055	2.055
	(0.349)	(0.438)	(0.469)	(0.477)	(0.404)	(0.249)	(0.376)	[0.003]
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036

Notes: This table reports alternative approaches to inference for the four main county-level outcomes used in the paper. The baseline estimates from the paper are reproduced in column 1 for reference. The bracketed numbers in the final column correspond to p-values from a wild bootstrap procedure clustering at the state level.

G Further Background Characterizing Frontier Life

G.1 Demographics and Individualism Along the Transition out of the Frontier

Figure G.1 reports estimates time-varying estimates of equation (1):

$$x_{cdt} = \alpha + \sum_{j=-20}^{40} \gamma_j \mathbf{1}(\text{years since exiting frontier = j}) + \theta_d + \theta_t + \varepsilon_{cdt}, \tag{G.1}$$

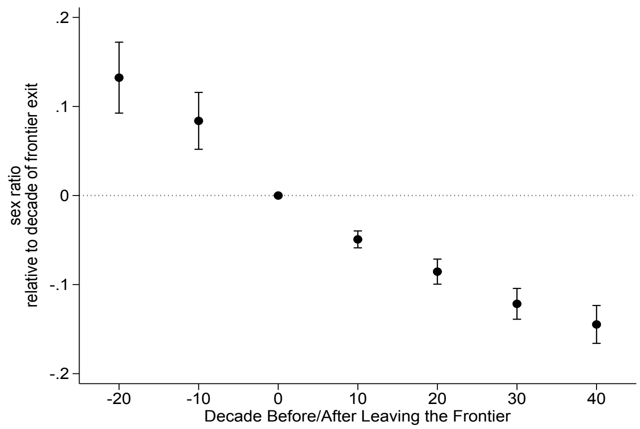
where the γ_j coefficients identify the average x for counties that have exited or will exit the frontier j years prior or in the future, respectively. We plot 95 percent confidence intervals for the γ terms, each of which are estimated with reference to the decade in which the county transitioned out of the frontier.

The estimates in Figure G.1 provide additional insight into the process of demographic and cultural change along the frontier. Panel (a) reveals an abrupt shift in the sex ratio as counties exit the frontier. On average, counties have 0.25 higher sex ratios in the two decades prior to exiting the frontier whereas those decades thereafter exhibit lower ratios that stabilize by the second decade. Panel (b) provides similar evidence of convergence towards a lower prime-age adult share as counties exit the frontier. Panel (c) shows that the foreign-born population share exhibits a steady and roughly linear decline along the frontier transition path. Results are noisier for illiteracy rates in panel (d).

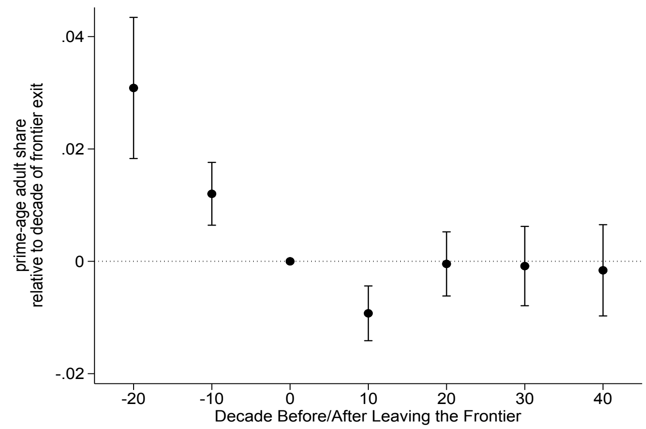
Panels (e) and (f) demonstrate the declining prevalence of individualistic children's names as counties approach the decade in which they exit the frontier. Thereafter, we see naming patterns stabilize around a less individualistic equilibrium in which popular names becomes more common at the local level. Note, however, that these estimates are with reference to the year of exiting the frontier and do not address the long-run differences in individualistic names as a function of years of exposure to frontier conditions. This is precisely what the analysis in Section 4 provides.

Figure G.1: Demographics and Individualism Along the Transition out of the Frontier
Estimates with Respect to Decade of Exiting the Frontier

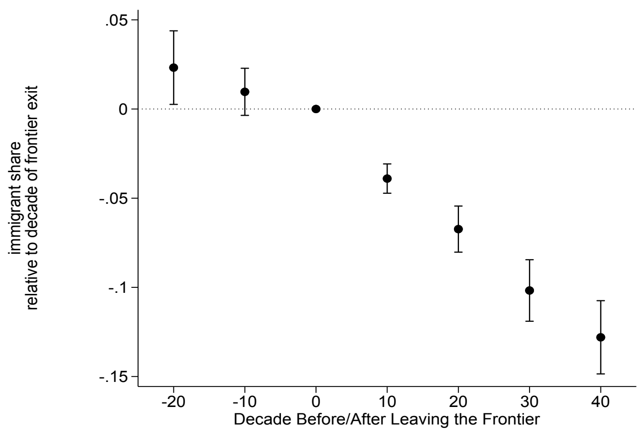
(a) Sex Ratio



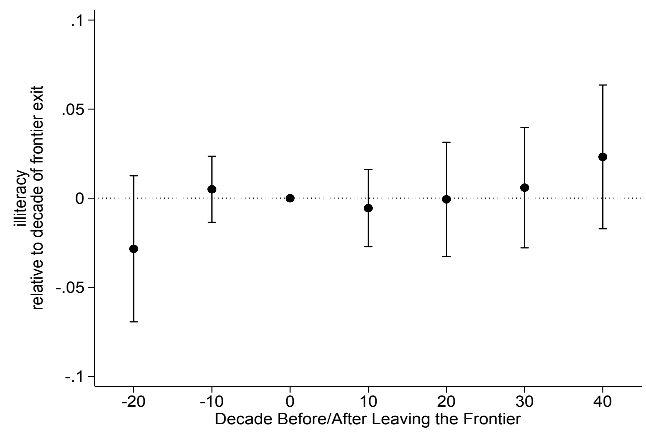
(b) Prime-Age Adult Share



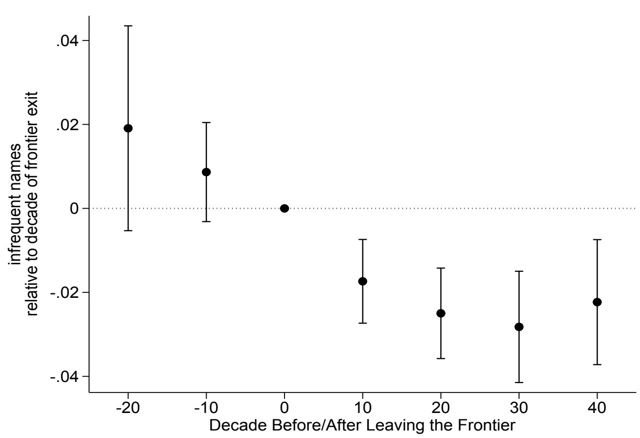
(c) Foreign-Born Share



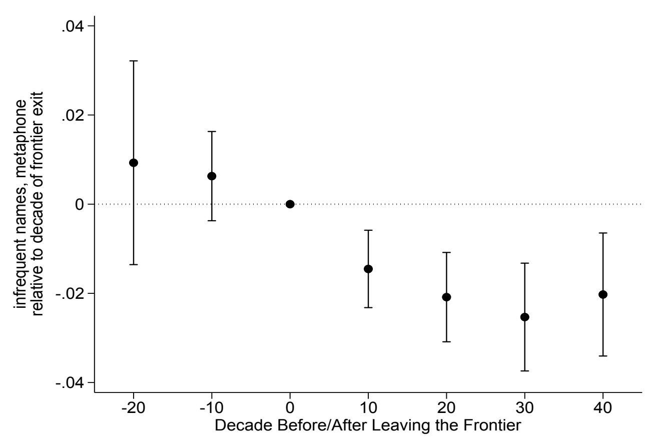
(d) Illiteracy



(e) Infrequent Names, Raw



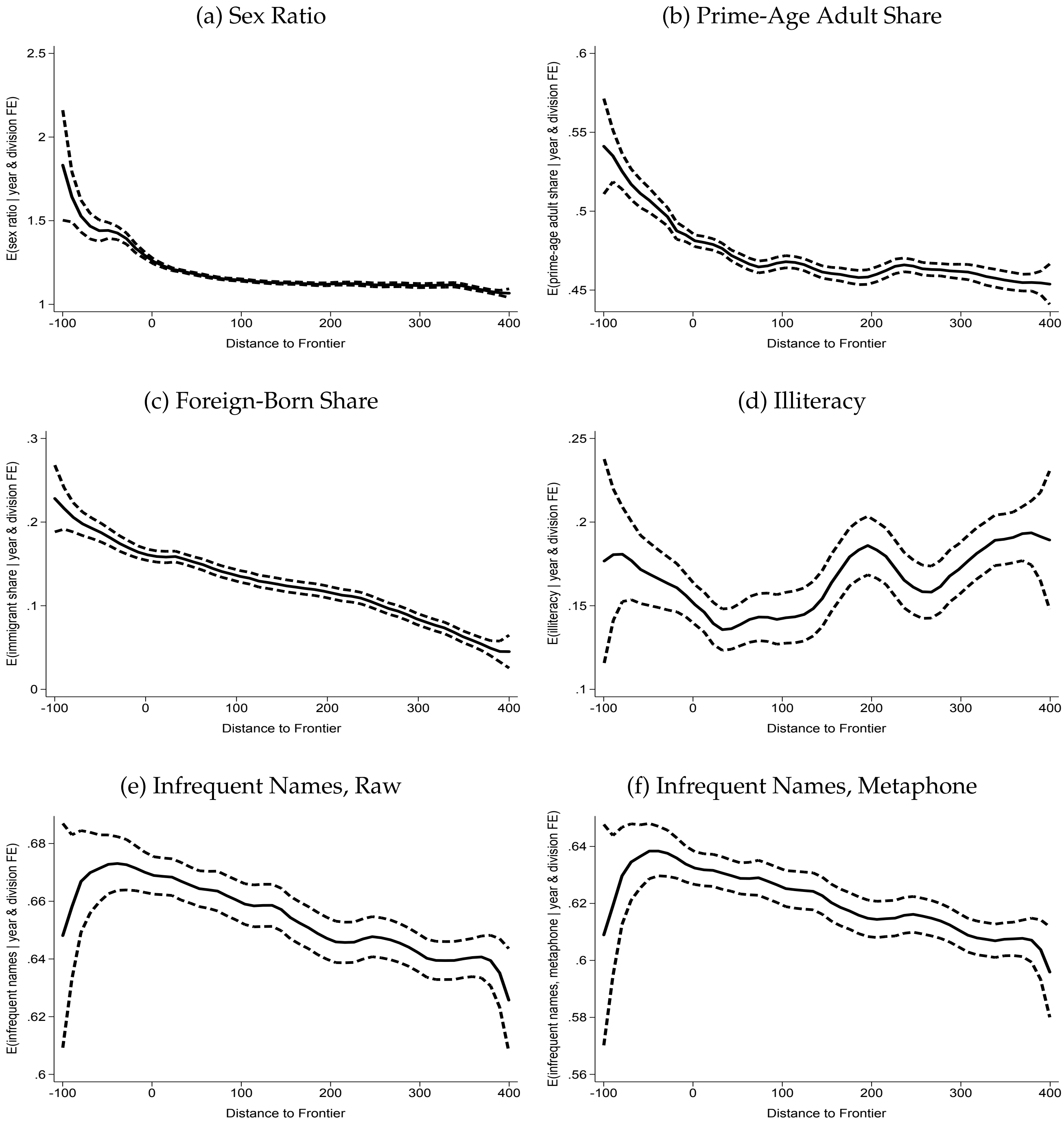
(f) Infrequent Names, Metaphone



Notes: This figure plots coefficients from the event study regressions in equation (G.1) for each of the outcomes in Table 1. The decade-specific point estimates and 95 percent confidence intervals are each with reference to the county-specific decade of exiting the frontier. All regressions include division and year fixed effects. Standard errors are clustered using the grid cell approach of Bester, Conley and Hansen (2011) as described in Section 4.1.

G.2 Demographics and Individualism by Distance to the Frontier

Figure G.2: Distribution of Demographics and Individualism by Distance to the Frontier



Notes: This figure reports the distance-based analogues to the density plots in Figure 4. Distance to the frontier, measured in kilometers, is the distance from the county's centroid to the nearest frontier line. The distance is negative if the county centroid is to the west of the nearest main frontier line. Figures (a)-(f) provide the semiparametric estimates of the corresponding dependent variables, with 95 percent confidence intervals, as a function of distance to the frontier estimated using county-level pooled data and applying a nonlinear function recovered using the partially linear Robinson (1988) estimator. The specification includes Census division and year fixed effects and are based on an Epanechnikov kernel and rule-of-thumb bandwidth.

G.3 Taxes and Public Goods on the Frontier

We present evidence here that demonstrates two important facts about the prevalence of taxation and public goods on the frontier historically.

First, *local* taxation and public spending were significantly lower in frontier counties historically. We demonstrate this for tax revenue per capita in column 1 of Appendix Table G.1 using county- and city-level tax revenue per capita, which is available for 1870 and 1880 as reported by Manson et al. (2019). Similarly, column 2 demonstrates lower government spending per capita in frontier than in non-frontier counties in 1870. Moreover, graphs (a) and (b) in Figure G.3 demonstrate the same sort of structural break in these outcomes at levels of population density around the cutoffs defining frontier locations historically, i.e., 2–6. Together, these results are consistent with smaller local government and weaker taxation on the frontier historically, consistent with the lack of social infrastructure described in most accounts of frontier life.

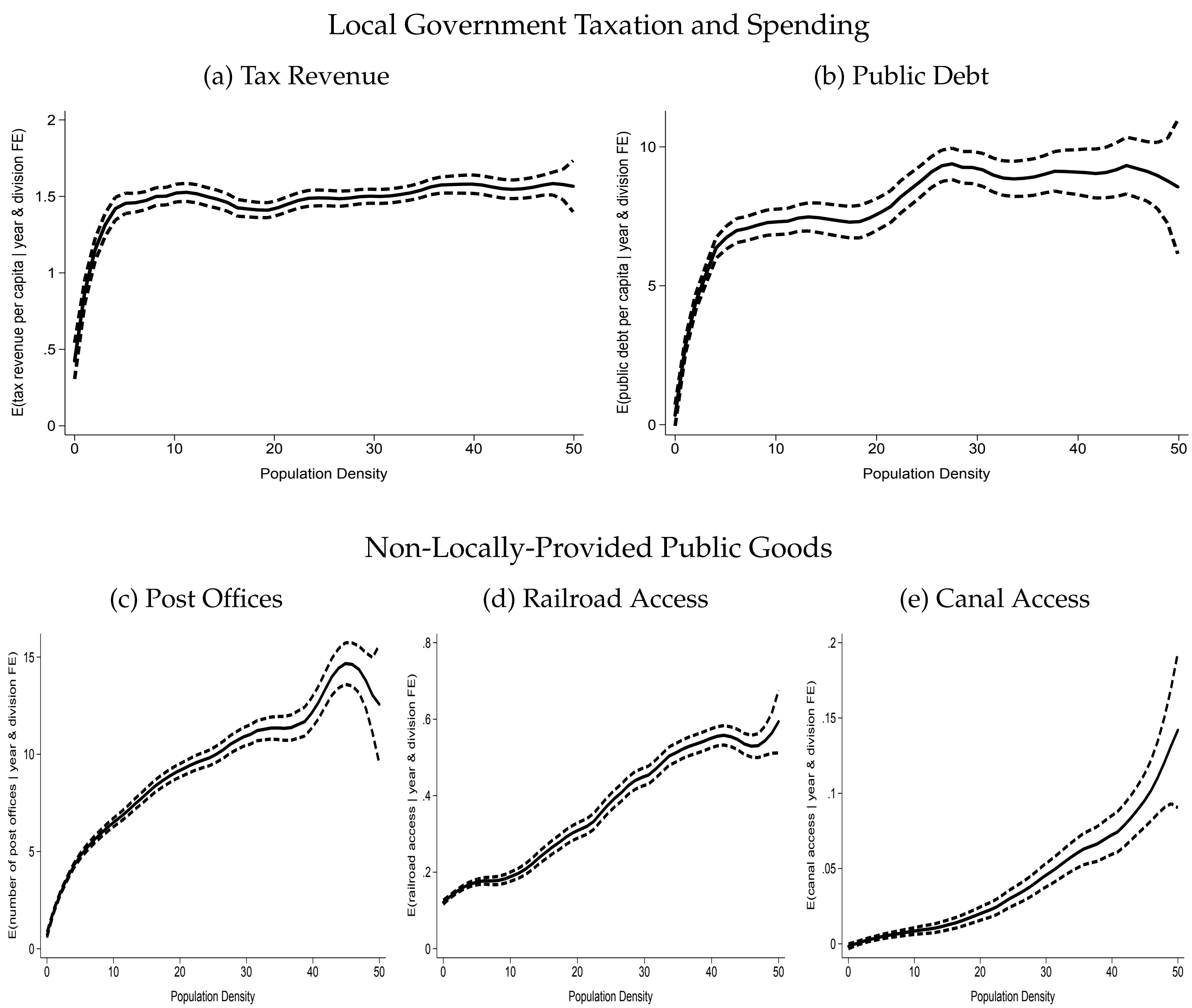
Second, *non-local* public goods were also less pervasive on the frontier, but this infrastructure did not exhibit the same sort of structural break at very low levels of population density seen for local public goods. Columns 3–5 of Appendix Table G.1 show, respectively, that frontier counties had a smaller number of post offices and more limited railroad and canal access historically. Appendix Figure G.3 shows that these non-locally-provided public goods were more pervasive in counties with greater population density. However, unlike local government efforts to raise taxes and make public investments, neither post offices nor railroads and canals exhibit structural changes at low levels of population density characterizing frontier conditions. All three outcomes vary smoothly if not roughly linearly from counties with population density ranging from just above zero to around 40. Importantly, this suggests that non-local government investments were not necessarily leading the westward expansion of the frontier as seen through the types of individuals underlying our results in Sections 3 and 5.

Table G.1: Taxes and Public Goods on the Frontier

	Local Government		Non-Local Government		
Dependent Variable:	Tax Revenue Per Capita	Public Debt Per Capita	Number of Post Offices	Railroad Access	Canal Access
	(1)	(2)	(3)	(4)	(5)
frontier county	-0.565	-4.914	-4.708	-0.114	-0.010
	(0.089)	(0.438)	(0.360)	(0.013)	(0.006)
Mean Dep. Var. in Non-Frontier Counties	1.49	8.18	10.02	0.43	0.05
Number of County-Years	3,756	1,789	9,801	13,458	14,043
R^2	0.42	0.16	0.44	0.54	0.15

Notes: This table reports estimates of the historical frontier differential in public goods using the same specification as in Table 1. Column 1 is the inverse hyperbolic sine of the sum of `taxcountyppop` and `taxcityetcppop` reported by Manson et al. (2019) for 1870 and 1880. Column 2 is the inverse hyperbolic sine of the sum of `pubdebtcounty` and `pubdebtcityetc`. Column 3 is the number of post offices in the given county–decade as reported in Acemoglu et al. (2016). Columns 4 and 5 are indicators equal to one if the given county–decade has railroad and canal access, respectively. See the notes below that table for details on the specification.

Figure G.3: Taxes and Public Goods by Population Density, 1790–1890



Notes: This figures reports semiparametric estimates of the relationship between population density and public goods measures from Table G.1 using the same specification as in Figure 4. See the notes below that table for details on the specification.

G.4 Low Inequality and Effort as the Road to Riches on the Frontier

This section argues that the opportunities and challenges on the frontier contributed to a culture of opposition to government intervention. The frontier's favorable prospects of upward mobility and a large perceived importance of effort in income generation may have fostered opposition to taxes, as suggested by the political economy theories of preferences for redistribution. This connection between the American frontier and theories of preferences for redistribution, hinted at by Alesina et al. (2001), echoes Billington (1974), who argued that the frontiersman "wanted not government interference with his freedom as he followed the road to riches."

In his analysis of the Turner thesis, Billington (1974) emphasizes the implications of the frontier's land abundance and "widespread property holdings." In these conditions, "a man's capacities, not his ancestry, determined his eventual place in the hierarchy, to a greater degree than in older societies." The frontiersman believed that "his own abilities would assure him a prosperous future as he exploited the natural resources about him." Access to land offered profit of opportunities, even for settlers with low initial wealth. Class distinctions were also weakened by the ubiquity of threats characterizing frontier life. As Overmeyer (1944) argues, since everyone "had to face the same hardships and dangers," the frontier was a "great leveling institution."

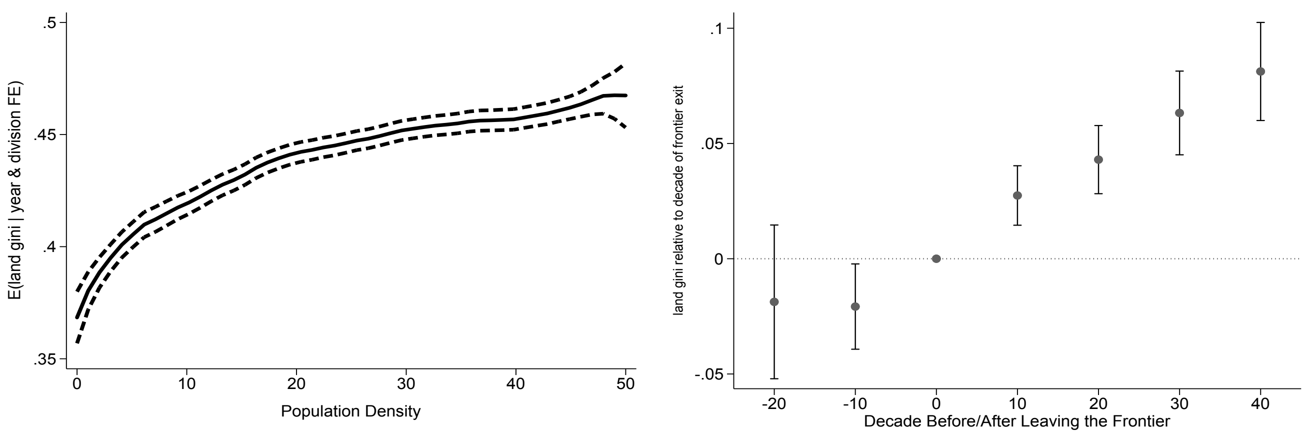
Numerous historical studies present stylized facts consistent with favorable prospects of upward mobility on the frontier and also with a large perceived importance of effort. As summarized by Stewart (2006), the frontier was "a place of economic opportunity," where settlers had low levels of initial wealth, but land-holding was widespread and rates of wealth accumulation were high, especially for early settlers and those that were able to endure.

Indeed, as shown in Appendix Figure G.4, historical Census data on landholdings is consistent with the idea that frontier locations offered a more level playing field. Land inequality, captured by the Gini coefficient, was significantly lower on the frontier, with sharp differences resembling what we documented for key demographics and individualism in Section 3. Moreover, this difference dissipated over time as counties exited the frontier and the usual forces giving rise to inequality took hold.

In sum, the stylized facts summarized above suggest a relatively limited role for inherited social class as a key determinant of income and wealth generation in the frontier economy. This implied a level playing field offering equality of opportunity, and a relatively high importance of effort as opposed to luck (of being born into a given class). Together with the selection and cultivation of individualism, these conditions plausibly contributed to the origins and persistence of frontier culture.

Figure G.4: Inequality is Lower on the Frontier

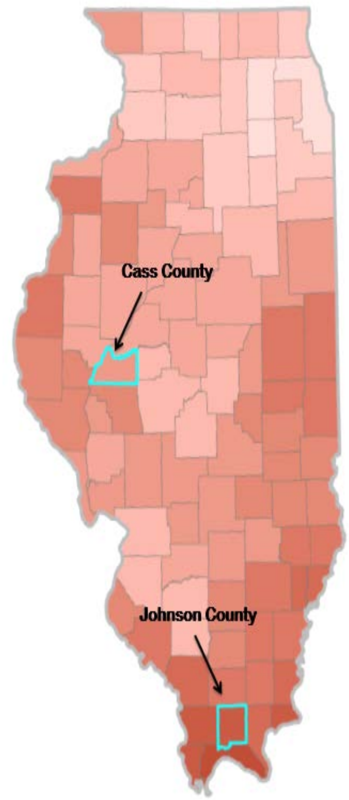
(a) Semiparametric (b) Event Study



Notes: Based on county level data from Manson et al. (2019) from 1790-1890. Land inequality is measured using the county level gini coefficient based on the number of farms in seven bins of farm size. The semiparametric specification in (a) is the same as in Figure 4, and the event study specification in (b) is the same as in Figure G.1. See the notes therein for details.

H Case Study Illustrating Long-Run Effects

To fix ideas, consider the two counties of Cass and Johnson mentioned in Section 2.2 and seen in the TFE map on the right, which is a snapshot of Illinois from Figure 3. Both are roughly equidistant from the Mississippi River and the important historical city of St. Louis. Today, the two rural counties look very similar in terms of population density: Cass has 36.3 people/mi^2, and Johnson has 36.6 people/mi^2. These two counties had very similar population density in 1890 as well. However, they differ significantly in their total frontier experience historically. Cass was on the frontier for 10 years, and Johnson for 32 years. This difference may be explained by any number of factors shaping the westward movement of the frontier through this area of the midwest in the early 1800s as seen in Figure 2. One potentially important contributor lies in our instrumental variable. Johnson entered the frontier in 1803 whereas Cass entered in 1818. While only 15 years apart, this implied a considerable difference in exposure to subsequent immigration-induced pressure on the westward expansion of the frontier over the next few decades as evidenced in Figures D.1, D.2, and especially D.3.



These historical differences in TFE translate into substantial long-run differences in the prevalence of rugged individualism in local culture. In Cass, 75 (64) percent of girls (boys) have infrequent names in 1940, Republican presidential candidates captured 55 percent of the vote in the average election since 2000, and local property tax rates are around 1.9 percent in 2010. Meanwhile, in Johnson, 78 (71) percent of girls (boys) have infrequent names in 1940, 68 percent average Republican vote shares since 2000, and 1.3 percent local property tax rates in 2010. This is striking insomuch as the two counties have such similar contemporary population density.

I Other Evidence on Selective Migration

This section provides additional evidence on selective migration and returns to individualism using a linked sample of men from 1870 to 1880. These results are discussed at greater length in an earlier working paper, but we retain them here given that they complement the more comprehensive results in Section 5 of the paper.

Compared to the results in the paper, those presented below are based on a narrower, linked sample of households from 1870 to 1880. We create this sample using complete-count restricted-access data from IPUMS (Ruggles et al., 2019) for 1870 and the US 100% sample (Ruggles et al., 2019) from the North Atlantic Population Project (NAPP) from Minnesota Population Center (2019) for 1880. We focus on the latest consecutive rounds available within the frontier period to ensure a large sample. We link individuals across rounds using an algorithm developed by Feigenbaum (2016). The base sample in 1880 is restricted to male household heads, native-born, aged 30–50, white, and who have at least one child aged 0–10.[1]

Appendix Table I.1 reports estimates of the frontier differential in infrequent naming patterns based on versions of the following equation for different sub-populations of movers and stayers:

$$\text{child has infrequent name}_{ic,1880} = \alpha + \beta \, \text{frontier}_{c,1880} + \mathbf{x}_{ic}'\boldsymbol{\zeta} + \varepsilon_{ic,1880}, \tag{I.1}$$

where the binary dependent variable equals one if child i residing in county c in 1880 has a name that falls outside the top 10 nationally in that decade, and the frontier indicator equals one if county c lies on the frontier according to our baseline definition. We restrict attention to white children aged 0–10 with native-born parents and cluster standard errors at the county level. The $\mathbf{x}_{ic}$ vector includes age$\times$gender and birth order fixed effects as well as indicators for whether the parents have infrequent names, but results are identical without these controls.

Column 1 of Appendix Table I.1 identifies the significance of selective migration. Children in households that migrated to the frontier between 1870 and 1880 are 4.2 p.p. more likely to have infrequent names than those remaining in non-frontier areas during that period, 71 percent of whom have infrequent names. While we do not observe whether these children were born before or after arriving on the frontier, this differential points to the self-selection of individualist types.

Column 2 captures the overall frontier differential in individualism. Children in frontier counties in 1880 are 7.5 p.p. more likely to have an infrequent name relative to children in non-frontier locations. Next, we show that the longer-term frontier residents (stayers) exhibit stronger individualism than recent arrivals from other counties. Column 3 decomposes the 7.5 p.p. differential into differences coming from early versus later frontier settlers. Early settlers in frontier counties are nearly three times more likely to give their children infrequent names than those that arrived more recently during the 1870s. Column 4 corroborates this differential, restricting the sample to those living in frontier counties in 1880. These results suggest that greater time on the frontier is associated with more individualistic naming patterns, a result that is rigorously borne out in Section 5.2 of the paper.

[1] The target year is 1870. The set of potential matches for these men are first identified based on first and last name, birth state and birth year. A random training sample is then drawn from among the potential matches and manually trained. The importance of each match feature is quantified using a probit model, and used to estimate a probability score for each link. A true match is defined as one with a sufficiently high score both in absolute and relative terms. The match rate was 25 percent, which is comparable with the rates achieved by recent studies linking records with broadly comparable data albeit different target populations (e.g., 29 percent in Abramitzky et al., 2012; 26 percent in Collins and Wanamaker, 2017; and 22 percent in Long and Ferrie, 2013). Although matching on names leaves scope for sample selection, the results look similar when reweighting using the odds of being linked across Census rounds (following Bailey et al., forthcoming). We estimate these probabilities using the same characteristics used for linking as well as an interaction of infrequent name status and frontier location in 1880. These interactions re-balance the linked sample to account for differential missing-ness along our key variables of interest.

Table I.1: Frontier Individualism and Selective Migration

	Dependent Variable: Child Has Infrequent Name in 1880			
	(1)	(2)	(3)	(4)
omitted reference group:	non-frontier resident, 1870–80	non-frontier resident, 1880	non-frontier resident, 1880	frontier immigrant, 1870–80
frontier county resident in 1880, immigrant	0.042		0.055	
	(0.012)		(0.011)	
frontier county resident in 1880		0.075		
		(0.018)		
frontier county resident in 1880, non-immigrant			0.186	0.118
			(0.035)	(0.026)
Number of Individuals	1,223,600	1,239,513	1,239,513	12,630
Mean Infrequent Name Share	0.707	0.708	0.708	0.767
R^2	0.02	0.02	0.02	0.05
Gender×Age Fixed Effects	✓	✓	✓	✓
Birth Order Fixed Effects	✓	✓	✓	✓

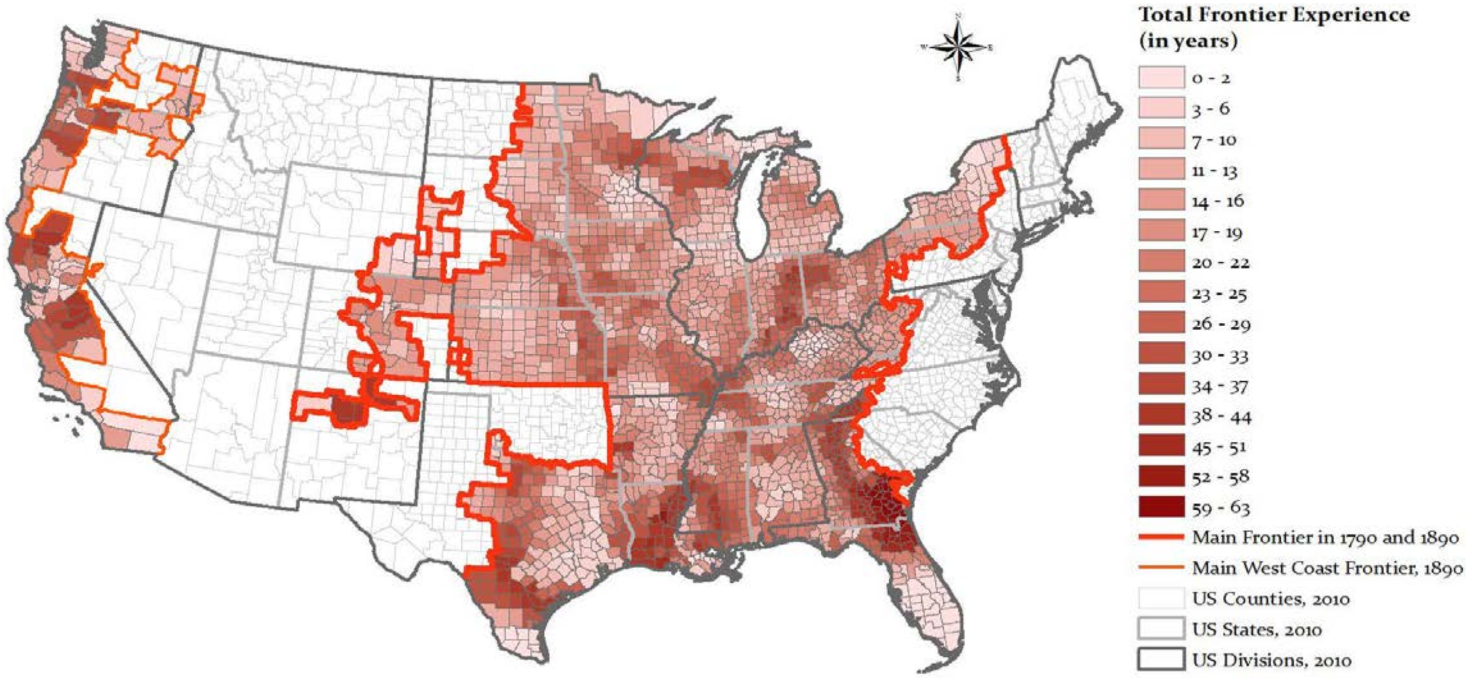
Notes: This table reports estimates of equation (I.1) based on the linked historical Census data from 1870 to 1880 for households with white, native-born fathers age 30–50 and children aged 0–10 in 1880. The dependent variable is an indicator equal to one if the child is given a name that falls outside the top 10 most popular names nationally in the 1870s. The top of each column reports the omitted reference group and the mean infrequent name share among them. We define immigrant status here based on whether the father switched counties between 1870 and 1880. Frontier counties are as defined in 1870 and 1880 based on the main east-to-west frontier line. Column 1 reports the selective migration differential between migrants from non-frontier to frontier counties and those that remained in non-frontier counties in both 1870 and 1880. Column 2 reports the overall differential in infrequent names between frontier and non-frontier counties in 1880, i.e., inclusive of stayers in frontier counties. Column 3 breaks down the overall differential into the component due to migrants between 1870 and 1880 and those that resided in the frontier county prior to 1870 (either by birth or earlier migration). Column 4 then restricts to frontier county residents, identifying the differential between recent immigrants and longer-term residents. In addition to gender×age and birth order fixed effects, all regressions control for indicators for whether the mother and father have infrequent names. Standard errors are clustered by county in 1870.

J Additional Empirics

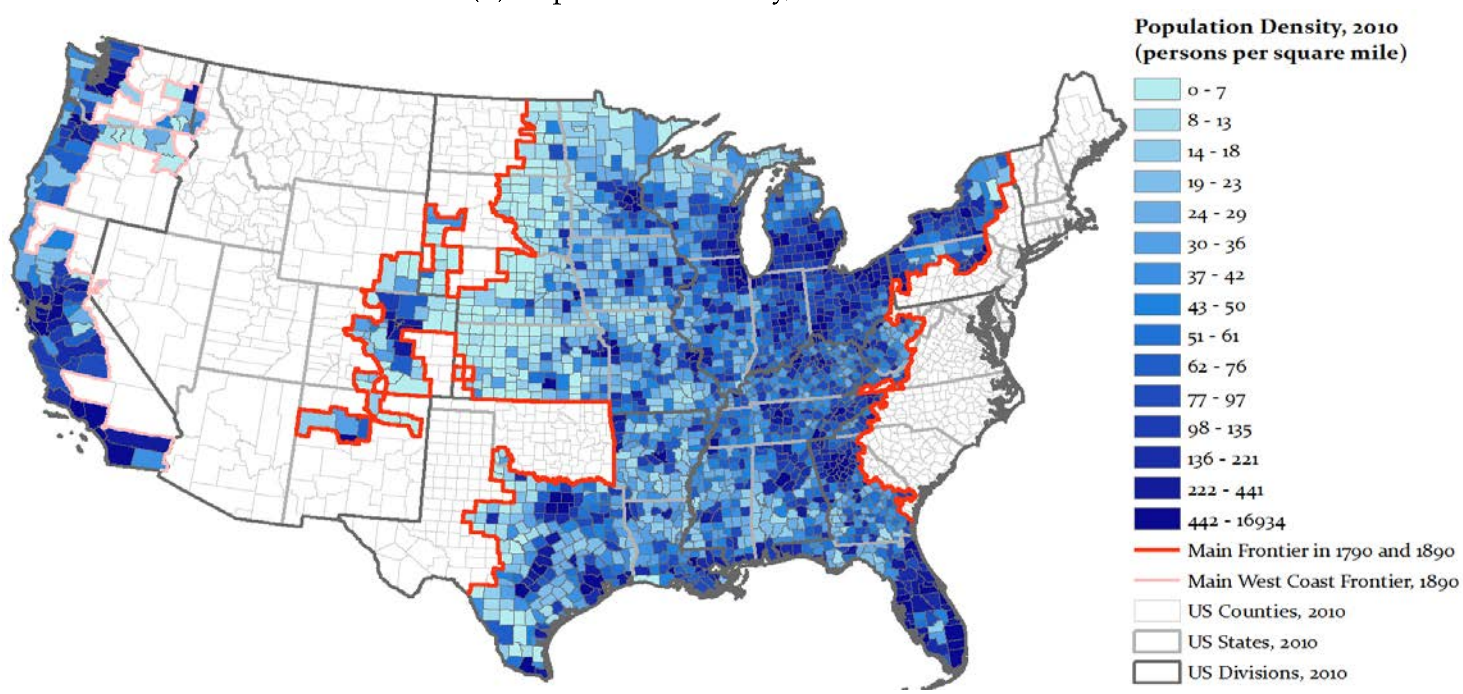
J.1 Comparing Total Frontier Experience and Current Population Density

Figure J.1: TFE is Distinct from Current Population Density

(a) Total Frontier Experience, 1790 to 1890



(b) Population Density, 2010



Notes: Panel (a) reproduces Figure 3, and (b) presents a similarly scaled map of population density in 2010 for the same counties.

J.2 Alternative Measures of Total Frontier Experience

Appendix Table J.1 reports baseline results excluding the 40 counties with zero TFE in the data. As noted in the paper, the findings are not sensitive to excluding this extensive margin of variation in TFE.

Our baseline measure of TFE closely followed definitions in the historical literature as discussed in Section 2. In Table J.2, we demonstrate the robustness of our results to three relevant margins of adjustment to our measure of TFE. In each case, we redefine what it means for county c to be on the frontier at time t. First, we reduce the catchment area from 100 km to 50 km in proximity to the frontier line. Second, we adjust the density restriction to include counties with > 2 people/mi^2 but still less than 6, counties with ≤ 18 people/mi^2, and then remove the population density restriction altogether. Finally, we consider defining the frontier line as including only the main, westernmost extent of all contour lines identified by the GIS algorithm. The overall message is that our particular choice of the frontier definition based on the historical record is not driving the main findings.

Table J.1: Robustness to Excluding Counties with Zero TFE

Dependent Variable:	Infrequent Name raw	standardized metaphone	Mean Gov. Prefs. CCES	County Property Tax Rate	Republican Vote Share Avg.
	(1)	(2)	(3)	(4)	(5)
	Panel (a): Baseline				
total frontier experience	0.141	0.141	0.014	-0.034	2.055
	(0.021)	(0.021)	(0.002)	(0.007)	(0.349)
Number of Observations	2,036	2,036	112,759	2,029	2,036
Mean of Dependent Variable	0.00	0.00	0.41	1.02	60.04
	Panel (b): Excluding Counties with Zero TFE				
total frontier experience	0.137	0.135	0.013	-0.031	1.664
	(0.020)	(0.021)	(0.002)	(0.007)	(0.340)
Number of Observations	1,996	1,996	105,623	1,990	1,996
Mean of Dependent Variable	0.00	0.00	0.42	1.02	60.30

Notes: This table reports estimates in panel (b) of our baseline specification (reproduced in panel (a)) excluding up to 40 counties with zero TFE. The specification is otherwise

Table J.2: Robustness to Alternative Measures of TFE for Summary Outcomes

Dependent Variable:	Infrequent Name raw standardized 1940 (1)	Infrequent Name metaphone standardized 1940 (2)	Mean Gov. Prefs. CCES 2006–16 (3)	County Property Tax Rate 2010 (4)	Republican Vote Share Avg. 2000–16 (5)
TFE: 100 km, $\leq 6/\text{mi}^2$, no inner or outer islands	0.141	0.141	0.014	-0.034	2.055
	(0.021)	(0.021)	(0.002)	(0.007)	(0.349)
TFE: 50 km, $\leq 6/\text{mi}^2$, no inner or outer islands	0.144	0.147	0.017	-0.035	2.051
	(0.021)	(0.022)	(0.003)	(0.007)	(0.358)
TFE: 100 km, $\leq 18/\text{mi}^2$, no inner island lines	0.106	0.105	0.010	-0.027	1.575
	(0.021)	(0.022)	(0.002)	(0.007)	(0.339)
TFE: 50 km, $\leq 18/\text{mi}^2$, no inner island lines	0.093	0.100	0.011	-0.025	1.458
	(0.020)	(0.020)	(0.003)	(0.006)	(0.351)
TFE: 100 km, 2-6/mi^2, no inner island lines	0.099	0.109	0.025	-0.014	1.877
	(0.035)	(0.035)	(0.006)	(0.008)	(0.485)
TFE: 50 km, 2-6/mi^2, no inner island lines	0.080	0.086	0.026	-0.012	1.771
	(0.039)	(0.039)	(0.006)	(0.009)	(0.530)
TFE: 100 km, no density restriction, no inner island lines	0.034	0.026	0.006	-0.011	1.001
	(0.021)	(0.021)	(0.002)	(0.007)	(0.335)
TFE: 50 km, no density restriction, no inner island lines	0.057	0.062	0.006	-0.018	1.078
	(0.019)	(0.019)	(0.003)	(0.006)	(0.339)
TFE: 100 km, $\leq 6/\text{mi}^2$, including inner island lines	0.162	0.162	0.014	-0.032	2.048
	(0.019)	(0.020)	(0.002)	(0.006)	(0.320)
TFE: 50 km, $\leq 6/\text{mi}^2$, including inner island lines	0.172	0.178	0.017	-0.035	2.098
	(0.021)	(0.023)	(0.003)	(0.007)	(0.335)
TFE: 100 km, $\leq 6/\text{mi}^2$, main single contour line	0.112	0.106	0.011	-0.037	1.872
	(0.027)	(0.028)	(0.004)	(0.008)	(0.436)
TFE: 50 km, $\leq 6/\text{mi}^2$, main single contour line	0.106	0.106	0.018	-0.043	1.787
	(0.029)	(0.029)	(0.004)	(0.008)	(0.460)
TFE: 50 km, $\leq 6/\text{mi}^2$, no inner or outer island lines	0.139	0.136	0.012	-0.034	2.133
	(0.021)	(0.022)	(0.002)	(0.007)	(0.357)
TFE: 50 km, $\leq 6/\text{mi}^2$, no inner or outer island lines	0.139	0.136	0.016	-0.035	2.116
	(0.022)	(0.023)	(0.003)	(0.007)	(0.373)
State Fixed Effects	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓

Notes: This table reports estimates of equation (4) for three measures of infrequent names for white children, age 0–10 in the 1940 Census. Each cell is a different regression based on the given dependent variable in the column and the measure of total frontier experience in the given row. The frontier lines considered in the baseline are countour lines longer than 500km after removing all "inner island lines" that are east of the main frontier line. The alternative measures of frontier experience considered above vary (i) the catchment area from 100 to 50 km from the contour lines, (ii) the density restriction from ≤ 6 people/mi^2 to $2\geq$people/mi$^2\leq 6$ to no restriction, (iii) including inner island lines, and (iv) including only the longest single contour line. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1.

J.3 Full Elaboration of Additional Controls in Table B.6

Table J.3: Robustness to Additional Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Panel (a): Infrequent Name Share in 1940 (standardized)														
total frontier experience	0.141	0.143	0.139	0.143	0.141	0.145	0.106	0.142	0.134	0.107	0.127	0.094	0.151	0.089
	(0.021)	(0.020)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.019)	(0.020)	(0.022)	(0.020)	(0.019)
ruggedness		4.965												3.411
		(0.873)												(0.602)
rainfall risk			2.515											-0.529
			(1.800)											(1.428)
distance to nearest portage site				0.047										0.038
				(0.045)										(0.032)
distance to nearest mineral discovery site					0.093									0.098
					(0.026)									(0.022)
distance to nearest Indian battle						-0.043								-0.060
						(0.037)								(0.025)
slave population share, 1860							-0.013							-0.008
							(0.002)							(0.001)
sex ratio, 1890								-0.391						-0.258
								(0.130)						(0.080)
immigrant share, 1890									-0.008					0.033
									(0.005)					(0.009)
scottish/irish immigrant share, 1890										-0.274				-0.192
										(0.021)				(0.020)
birthplace diversity, 1890											-1.256			-1.774
											(0.272)			(0.611)
years connected to railroad by 1890												-0.021		-0.012
												(0.002)		(0.002)
manufacturing employment share, 1890													-0.079	-0.056
													(0.006)	(0.005)
Oster δ for $\beta = 0$	-13.80	-11.03	-40.85	-8.28	-14.43	-6.38	1.23	-13.21	10.07	2.03	3.86	1.16	-6.57	1.61
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.60	0.62	0.60	0.60	0.61	0.60	0.62	0.61	0.61	0.67	0.61	0.66	0.66	0.76
Panel (b): Infrequent Name Share in 1940, Metaphone-Adjusted (standardized)														
total frontier experience	0.141	0.143	0.139	0.143	0.141	0.144	0.107	0.141	0.135	0.106	0.128	0.094	0.150	0.091
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)	(0.020)	(0.020)	(0.021)	(0.022)	(0.021)	(0.019)
ruggedness		5.282												3.865
		(0.910)												(0.642)
rainfall risk			2.150											-0.820
			(1.914)											(1.530)
distance to nearest portage site				0.065										0.058
				(0.046)										(0.036)
distance to nearest mineral discovery site					0.105									0.109
					(0.028)									(0.024)
distance to nearest Indian battle						-0.035								-0.052
						(0.038)								(0.028)
slave population share, 1860							-0.012							-0.008
							(0.002)							(0.001)
sex ratio, 1890								-0.394						-0.264
								(0.141)						(0.096)
immigrant share, 1890									-0.007					0.025
									(0.005)					(0.010)
scottish/irish immigrant share, 1890										-0.281				-0.213
										(0.021)				(0.021)
birthplace diversity, 1890											-1.109			-1.016
											(0.278)			(0.723)
years connected to railroad by 1890												-0.021		-0.012
												(0.002)		(0.002)
manufacturing employment share, 1890													-0.078	-0.055
													(0.007)	(0.006)
Oster δ for $\beta = 0$	-16.20	-12.62	-36.47	-8.50	-17.00	-8.30	1.49	-15.37	16.69	2.21	5.20	1.30	-7.25	1.93
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.58	0.60	0.58	0.58	0.58	0.58	0.59	0.58	0.58	0.65	0.59	0.63	0.63	0.73
Panel (c): Mean Government Preferences Outcomes (CCES), 2006–16														
total frontier experience	0.014	0.013	0.014	0.014	0.014	0.013	0.014	0.014	0.011	0.011	0.010	0.012	0.011	0.009
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
ruggedness		-0.342												-0.306
		(0.111)												(0.129)
rainfall risk			-0.371											-0.203
			(0.296)											(0.323)
distance to nearest portage site				0.000										-0.000
				(0.000)										(0.000)
distance to nearest mineral discovery site					0.009									0.007
					(0.004)									(0.004)
distance to nearest Indian battle						0.005								0.002
						(0.005)								(0.005)
slave population share, 1860							0.000							0.000
							(0.000)							(0.000)
sex ratio, 1890								0.008						-0.003
								(0.018)						(0.019)
immigrant share, 1890									-0.002					0.001
									(0.000)					(0.002)
scottish/irish immigrant share, 1890										-0.013				-0.005
										(0.002)				(0.003)
birthplace diversity, 1890											-0.150			-0.094
											(0.025)			(0.118)
years connected to railroad by 1890												-0.001		-0.001
												(0.000)		(0.000)
manufacturing employment share, 1890													-0.004	-0.002
													(0.001)	(0.001)
Oster δ for $\beta = 0$	6.49	6.23	7.07	7.00	6.34	5.46	7.00	6.64	3.73	3.36	3.33	4.20	3.71	2.26
Number of Counties	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759
Mean of Dependent Variable	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
R^2	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Panel (d): County Property Tax Rate in 2010														
total frontier experience	-0.034	-0.034	-0.034	-0.033	-0.034	-0.033	-0.033	-0.034	-0.028	-0.026	-0.027	-0.023	-0.036	-0.023
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
ruggedness		-0.959												-0.724
		(0.219)												(0.208)
rainfall risk			0.959											0.736
			(0.581)											(0.516)
distance to nearest portage site				0.026										0.006
				(0.019)										(0.016)
distance to nearest mineral discovery site					0.026									0.024
					(0.009)									(0.008)
distance to nearest Indian battle						-0.004								0.002
						(0.011)								(0.009)
slave population share, 1860							0.000							-0.001
							(0.000)							(0.001)
sex ratio, 1890								-0.010						-0.024
								(0.027)						(0.024)
immigrant share, 1890									0.007					-0.003
									(0.001)					(0.003)
scottish/irish immigrant share, 1890										0.059				0.029
										(0.008)				(0.006)
birthplace diversity, 1890											0.636			0.541
											(0.084)			(0.244)
years connected to railroad by 1890												0.005		0.003
												(0.000)		(0.000)
manufacturing employment share, 1890													0.016	0.008
													(0.002)	(0.002)
Oster δ for $\beta = 0$	-27.45	-7.11	-3.23	2.54	-30.95	9.19	8.59	-30.17	0.55	0.48	0.49	0.28	-2.08	0.42
Number of Counties	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029	2,029
Mean of Dependent Variable	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
R^2	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.83	0.85
Panel (e): Republican Vote Share, Average 2000–16														
total frontier experience	2.055	2.050	2.115	2.095	2.055	2.172	1.399	2.060	1.715	1.717	1.689	1.640	2.137	0.931
	(0.349)	(0.349)	(0.338)	(0.344)	(0.350)	(0.351)	(0.361)	(0.347)	(0.328)	(0.340)	(0.327)	(0.361)	(0.350)	(0.316)
ruggedness		-11.710												-34.668
		(11.247)												(9.508)
rainfall risk			-77.882											-108.600
			(42.010)											(36.779)
distance to nearest portage site				0.930										0.966
				(0.777)										(0.623)
distance to nearest mineral discovery site					0.191									0.137
					(0.406)									(0.344)
distance to nearest Indian battle						-1.393								-0.794
						(0.603)								(0.471)
slave population share, 1860							-0.237							-0.270
							(0.035)							(0.034)
sex ratio, 1890								-5.000						-1.948
								(1.856)						(0.940)
immigrant share, 1890									-0.417					-0.217
									(0.051)					(0.198)
scottish/irish immigrant share, 1890										-2.738				-1.427
										(0.263)				(0.243)
birthplace diversity, 1890											-32.582			-4.970
											(3.563)			(13.286)
years connected to railroad by 1890												-0.186		-0.089
												(0.026)		(0.025)
manufacturing employment share, 1890													-0.689	-0.422
													(0.090)	(0.075)
Oster δ for $\beta = 0$	-8.55	-8.93	-6.43	-6.75	-8.56	-5.10	2.34	-8.92	7.33	7.40	6.72	4.53	-7.81	1.35
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
Mean of Dependent Variable	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04
R^2	0.33	0.33	0.34	0.33	0.33	0.34	0.38	0.34	0.38	0.38	0.39	0.36	0.36	0.49
State Fixed Effects	✓	✓	Yes	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	Yes	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table reproduces the estimates from Table B.6, showing the coefficient estimates for the additional variables listed at the top of that table. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 4.1.

J.4 The Parasite-Stress Theory of Values

The parasite-stress theory of values due to Thornhill and Fincher (2014) argues that the prevalence of infectious diseases leads to higher levels of in-group assortative sociality, which they associate with collectivism, as an adaptive response that minimizes contagion. In the context of our study, this theory might suggest that frontier individualism resulted from the low prevalence of infectious diseases on the frontier. However, this potential mechanism does not arise in historical narratives. Nor do we find evidence of differential disease prevalence or morbidity on the frontier. As seen in Table J.4 below, the prevalence of pathogens—associated with tuberculosis, malaria, and typhoid, among other diseases considered in Gorodnichenko and Roland (2016)—does not exhibit any differential intensity on the frontier. We can measure the incidence of these specific infectious diseases as well as a broad array of other illnesses for the first time in the 1880 Population Census. Adopting specifications similar to Table 1, we find little evidence that individuals living on the frontier had differential (infectious) disease or illness. If the parasite-stress mechanism were salient, we would find that frontier locations exhibit significantly less prevalence of infectious diseases. While the relatively precise zeros in the table may be specific to 1880, this provides suggestive evidence that the parasite-stress channel is not a first-order factor in explaining the differential individualism on the frontier.

Table J.4: No Differential Infectious Diseases or Sickness on the Frontier

Dependent Variable:	Share of Pop. with Infectious Disease		Share of Pop. with Any Illness	
	(1)	(2)	(3)	(4)
on the frontier	0.0001		0.0009	
	(0.0003)		(0.0013)	
near frontier line		-0.0001		0.0011
		(0.0001)		(0.0007)
low population density		0.0000		0.0001
		(0.0002)		(0.0010)
Mean Dependent Variable	0.001	0.001	0.009	0.009
Number of Counties	1,780	1,780	1,780	1,780
R^2	0.05	0.05	0.08	0.08

Notes: This table reports estimates of the relationship between frontier definitions and the share of the county with any of the infectious diseases considered in Gorodnichenko and Roland (2016) (columns 1–2) and any illness (column 3–4). The infectious diseases of interest include tuberculosis, malaria, and typhus. The specification is otherwise similar to that in Table 1, with Census division FE and standard errors clustered using the grid-cell approach of Bester, Conley and Hansen (2011).

K Data Sources and Construction

Harmonization to 2010 Boundaries

We harmonize all historical Census data to the 2010 boundaries using an approach suggested in Hornbeck (2010). First, we intersect the county shapefiles from each of the decadal census years with the 2010 county shapefile and calculate the area of each intersection. When the 2010 county falls in one or more counties of the earlier shapefile, each piece of the 2010 county is assigned a value equal to the share of the area of the piece in the earlier county multiplied by the total value of the data for the earlier county. Then, the data for each county in 2010 is the sum of all the pieces falling within its area. This harmonization procedure would be exact if all the data from the various years are evenly distributed across county areas.

Demographic Variables and Individualism

Population density. Population/area. Digitized U.S. Census data on population for every decade in 1790–2010, from Manson et al. (2019). The data on area is calculated using the 2010 county shapefiles from NHGIS (Manson et al., 2019) using GIS software. The county-level population data along with other pre-2010 data are harmonized to the 2010 county boundaries and the data for intercensal years is imputed using the procedure detailed in Section A. The population density figures include slaves in the antebellum period but exclude most Native Americans throughout the frontier era as they were not counted by the Census.

Sex Ratio. Whites males/white females. The data is available for every decade in 1790-1860, 1880, and 1890. Data source: Manson et al. (2019).

Prime Age Adult Share. Whites aged 15–49/all whites. The data used is consistently available for every decade in 1830-1860. Data source: Manson et al. (2019).

Illiteracy. Illiterate whites aged above 20/whites aged over 20. The variable is available consistently for 1840 and 1850. Data source: Manson et al. (2019).

Immigrant Share. Foreign born/population. The variable used is available for every decade in 1820-1890 (excluding 1840). Data source: Manson et al. (2019).

Out of State Born Share. Out-of-state born/population. The variable is consistently available for every decade in 1850-1880. Data source: Manson et al. (2019).

Land inequality. Gini index using distribution of farm sizes, based on county level data on the number of farms of sizes 0–10, 10–19, 20–49, 50–99, 100–499, 500–1000, and above 1000 acres. Available for every decade in 1860-1890. Data source: Manson et al. (2019).

Infrequent Children Names. White Children Aged 0–10 with Non-Top-10 First Names in Division/White Children Aged 0–10 With Native-Born Parents. In addition, for the same sample, we construct additional variables by calculating the popularity of names at the county, state, and national level instead of the Census division. We use the following procedure to generate the name shares: start by restricting the sample as desired (e.g. white children aged 0-10 with native parents), then calculate the number of children in the county for each given name, then using that value identify the top 10 given names within the given administrative unit, and then accordingly count the number of children in that county with the identified top 10 names in their corresponding census division. The variable restricting to white children aged 0–10 is available for every decade in 1850–1940 (excluding 1890), with further native-parent restriction for 1850 and 1880-1940 (excluding 1890). In 1940, only 5% of individuals are asked about parental birthplace. Hence, for 95% of kids, we can only identify native-born parents if those parents co-reside with the children. For the remaining 5% of kids, we can identify native-born parents even if one of the parents does not reside with the children. Data source: US 100% samples (Ruggles et al., 2019) of the North Atlantic Population Project (NAPP) from Minnesota Population Center (2019) for 1850 and 1880, and complete-count restricted-access data from IPUMS (Ruggles et al., 2019) for other years.

To give some examples of names, in 1850 the top 10 boy names nationally in descending order of popularity were John, William, James, George, Charles, Henry, Thomas, Joseph, Samuel and David. Meanwhile, a random sample of less common names (outside the top 25) includes ones like Alfred, Nathan, Patrick, Reuben, Herbert, Matthew, Thaddeus and Luke. For girls, the top 10 include Mary, Sarah, Elizabeth, Martha, Margaret, Nancy, Ann, Susan, Jane, and Catherine while less common names (outside the top 25) include ones like Rachel, Susannah, Nina, Olive, Charlotte, Lucinda, and Roxanna. By 1880, the rankings shifted only slightly for boys with Samuel falling outside the top 10 and Harry entering. For girls, the changes were a bit more dramatic with the new top 10 list being Mary, Sarah, Emma, Ida, Minnie, Anna, Annie, Martha, Cora, and Alice.

Non-Patronymic/-Matronymic Names. Using the same sample definitions as with infrequent names, we compute the share of kids with names that are distinct from their parents' names of the given gender. This measure can only be defined for children with co-resident parents of the given gender. Data source: US 100% samples (Ruggles et al., 2019) of the North Atlantic Population Project (NAPP) from Minnesota Population Center (2019) for 1850 and 1880, and complete-count restricted-access data from IPUMS (Ruggles et al., 2019) for other years.

Economic Status. We measure economic status using the occupational score (*occscore*) provided by IPUMS (Ruggles et al., 2019). Both measures range from 0 to 100, and capture the income returns associated with specific occupations in the 1950 Census while the *sei* measure additionally captures notions of prestige as well as educational attainment. Data source: Ruggles et al. (2019).

Public Goods, Local Taxation and Spending

Local Tax Revenue Per Capita. County taxes and other local taxes (town, city, etc) divided by county population. Data Source: Manson et al. (2019).

Local Public Debt Per Capita. County-level public debt and other local debt (town, city, etc) divided by county population. Data Source: Manson et al. (2019).

Infrastructure. Rail Access and Canal Access are indicator variables for whether the county is intersected by a railroad line or a canal, respectively. Data source: Atack et al. (2010).

Access to Irrigation Dam is an indicator on whether the county has an irrigation dam within its boundaries. We also generate similar variables separately by ownership of the dams (federal, state, local and private). Data source: Washington (2018).

Survey-Based Cultural Outcomes

Some of our key measures of contemporary preferences for government policy are based on data from multiple rounds of three widely used, nationally representative surveys: the Cooperative Congressional Election Study (CCES), the General Social Survey (GSS), and the American National Election Study (ANES). These surveys are staples in the social science literature on political preferences and social norms. For instance, Acharya, Blackwell and Sen (2016) uses CCES and ANES in a related methodological setting, and Alesina and Giuliano (2011) conducts a thorough investigation of the determinants of preferences for redistribution using the GSS. The CCES is a web-based survey conducted every two years, the ANES is an in-person survey conducted annually since 1948, and the GSS is an in-person survey conducted annually since 1972. All three are repeated cross-sections.

One advantage of working with three surveys is that we can cross-validate the findings across surveys that ask different questions about similar underlying preferences. For example, the CCES asks respondents if and how respondents would like state-level welfare spending to change whereas the ANES asks respondents if and how federal spending on the poor should change. The CCES also includes a set of questions on policy issues such as gun ownership that are particularly relevant to some of the mechanisms driving the persistence of frontier culture. For all measures, we link county-level identifiers in the underlying data to the 2010 county boundaries.

Despite their rich level of detail, these surveys have one important limitation for our purposes, namely the limited geographic scope. The three surveys are nationally representative, but their coverage differs. While the CCES has broad spatial coverage, the GSS and ANES do not (see Appendix Figures K.1). Despite its broader coverage, the CCES has the potential disadvantage that it captures an internet-savvy sample that may not be reflective of the underlying population in the way that an in-person survey generally would. This is particularly disadvantageous given our focus on county-level variation in TFE across a swathe of the United States outside of major coastal population centers.

Prefers Cutting Public Spending On Poor. The Prefers Cutting Public Spending On Poor is an indicator variable based on the following survey question: "*Should federal spending be increased, decreased, or kept about the same on poor people?*" The variable takes a value of 1 if the respondent answered "*decreased*" and 0 otherwise, and it is available for 1992 and 1996. Data source: The American National Election Studies Cumulative Data (2012). The ANES is a large, nationally-representative survey of the American electorate in the United States taken during the presidential and midterm election years. See Appendix Figure K.1(a) for the map of the maximum survey coverage in the final sample of ANES data merged with the frontier related data.

Prefers Cut Public Spending on Welfare. This is an indicator variable based on the following survey question: "*State legislatures must make choices when making spending decisions on important state programs. Would you like your legislature to increase or decrease spending on Welfare? 1. Greatly Increase 2. Slightly Increase 3. Maintain 4. Slightly Decrease 5. Greatly Decrease.*" Prefers Cut Public Spending on Welfare takes a value of 1 if the respondent answered "*Slightly Decrease*" or "*Greatly Decrease*" and 0 otherwise. The data is available in the 2014 and 2016 waves. Data source: Cooperative Congressional Election Study (Ansolabehere and Schaffner, 2017) Common Content surveys. The CCES was formed in 2006, through the cooperation of several academic institutions, to study how congressional elections, representation and voters' behavior and experiences vary with political geography and social context using very large scale national surveys. The 2014 and 2016 CCES surveys were conducted over the Internet by YouGov using a matched random sample methodology. The Common Content portion of the survey, which contains our variables of interest, surveyed 56,200 adults in 2014 and 64,600 adults in 2016. See Appendix Figure K.1(b) for the map of the maximum survey coverage in the final sample of GSS data merged with frontier-related data.

Believes Government Should Redistribute. Based on the following survey question: "*Some people think that the government in Washington ought to reduce the income differences between the rich and the poor, perhaps by raising the taxes of wealthy families or by giving income assistance to the poor. Others think that the government should not concern itself with reducing this income difference between the rich and the poor. Here is a card with a scale from 1 to 7.*" We have recoded the variable so that it is increasing in preference for redistribution, where a score of 1 means that the government should not concern itself with reducing income differences and a score of 7 means the government ought to reduce the income differences between rich and poor. The Believes Government Should Redistribute is a normalized version of the above variable, and it is available in our sample for 1993 and all even years between 1994-2016. Data source: The General Social Survey (Smith, Marsden, Hout and Kim, 2015). The GSS is a repeated cross-sectional survey of a nationally representative sample of non-institutionalized adults who speak either English or Spanish. The surveys has been conducted since 1972, almost every year between 1972-1993 and biennial since 1994. While the sample size for the annual surveys was 1500, since 1994 the GSS administers the surveys to two samples in even-numbered years, each with a target sample size of 1500. The surveys provide detailed questionnaires on issues such as national spending priorities, intergroup relations, and confidence in institutions. See Appendix Figure K.1(c) for the map of the maximum survey coverage in the final sample of CCES merged with frontier related data.

Prefers Reducing Debt by Cutting Spending. The variable is based on the CCES survey question: *"The federal budget deficit is approximately [$ year specific amount] this year. If the Congress were to balance the budget it would have to consider cutting defense spending, cutting domestic spending (such as Medicare and Social Security), or raising taxes to cover the deficit. Please rank the options below from what would you most prefer that Congress do to what you would least prefer they do: Cut Defense Spending; Cut Domestic Spending; Raise Taxes."*. While this question varies slightly from year to year, the underlying theme is the same. The Prefers Reducing Debt by Cutting Spending variable takes a value of 1 if the respondent chose "Cut Domestic Spending" as a first priority. The data is available for 2006-2014 (excluding 2013). Data source: Ansolabehere and Schaffner (2017).

Index of Preferences for Spending Cuts. The index is the principal component of nine dummy variables that take the value of 1 if the respondents answers "too much" to the following questions: *"We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (READ ITEM A) . . . are we spending too much, too little, or about the right amount on (ITEM)?"*. The items considered are improving and protecting the environment, improving healthcare, solving big city problems, halting increasing crimes, dealing with drug addictions, improving the education system, improving conditions for blacks, military spending, foreign aid, welfare, and roads. The variable is available in our sample for 1993 and all even years between 1994-2016. Data source: Smith et al. (2015).

Opposes Affordable Care Act. Based on the CCES survey question: *"The Affordable Health Care Act was passed into law in 2010. It does the following: Requires all Americans to obtain health insurance, Prevents insurance companies from denying coverage for pre--existing condition, Allows people to keep current health insurance and care provider, and Sets up national health insurance option for those without coverage, but allows states the option to implement their own insurance system. Would you have voted for the Affordable Care Act if you were in Congress in 2010?"* The Prefers Repealing Affordable Care Act variable takes a value of 1 if the respondent answers "Yes" and 0 if the answer is "No". The data is available for 2014. Data source: Ansolabehere and Schaffner (2017).

Opposes Increasing Minimum Wage. Based on the survey question:*"As you may know, the federal minimum wage is currently $5.15 an hour. Do you favor or oppose raising the minimum wage to $7.25 an hour over the next two years, or not?"*. The variable Opposes Increasing Minimum Wage takes a value of 1 if the respondent choses "oppose" and 0 otherwise. Available in 2007. Data source: Ansolabehere and Schaffner (2017).

Opposes Banning Assault Rifles. Based on the CCES survey question: *"On the issue of gun regulation, are you for or against for each of the following proposal? proposal: banning assault rifles"*. Opposes Banning Assault Rifles takes value 1 if the respondent is against banning assault rifles and 0 otherwise. Available for 2014. Data source: Ansolabehere and Schaffner (2017).

Opposes EPA Regulations of CO_2 Emissions. Based on the CCES survey question *"Do you support or oppose each of the following proposals? proposal: Environmental Protection Agency regulating Carbon Dioxide emissions."* The Opposes EPA Regulations of CO_2 Emissions takes one if the respondent supports the proposal and 0 the respondent opposes. Available for 2014. Data source: Ansolabehere and Schaffner (2017).

Cooperation vs. Self-Reliance. Based on the survey question: *"I am going to ask you to choose which of two statements I read comes closer to your own opinion. You might agree to some extent with both, but we want to know which one is closer to your views: ONE, it is more important to be a cooperative person who works well with* *others; or TWO, it is more important to be a self-reliant person able to take care of oneself"*. The Cooperation vs. Self-Reliance variable takes a value of 1 if the respondent chooses "cooperative" and 0 otherwise. Available in 1990. Data source: The American National Election Studies.

Identifies As A Strong Republican. An indicator variable that takes 1 if the respondent identifies as a "Strong Republican." Available for 2007, 2012, 2014 and 2016. Data source: (Ansolabehere and Schaffner, 2017).

Other Long-run Outcomes

County Property Tax Rate. The average effective property tax rates per $100 of value, calculated at the county level as the ratio of the average real estate tax over the average house value. Data source: The data is obtained from the National Association of Home Builders, which calculated the average effective property tax rates based on the 2010–2014 American Community Survey (ACS) data from the Census Bureau.

Republican Vote Share in Presidential Elections. Votes for a GOP candidate/total votes, at the county level. Data source: ICPSR (8611) Presidential and Congressional Races, 1840–1972 for 1900–1972 elections, General Election Data for the United States 1950–1990 (ICPSR 13) for 1976–1996 elections, and Dave Leip's Atlas of U.S. Presidential Elections for 2000–2016 elections (2017).

Geographic and Agroclimatic Controls

Land productivity measures. Average of attainable yields for alfalfa, barley, buckwheat, cane sugar, carrot, cabbage, cotton, ax, maize, oats, onion, pasture grasses, pasture legumes, potato, pulses, rice, rye, sorghum, sweet potato, tobacco, tomato, and wheat. We normalize each product's values dividing it by the maximum value for that product in the sample. Measures of attainable yields were constructed by the FAO's Global Agro-Ecological Zones project v3.0 (IIASA/FAO, 2012) using climatic data, including precipitation, temperature, wind speed, sunshine hours and relative humidity (based on which they determine thermal and moisture regimes), together with crop-specific measures of cycle length (i.e. days from sowing to harvest), thermal suitability, water requirements, and growth and development parameters (harvest index, maximum leaf area index, maximum rate of photosynthesis, etc). Combining these data, the GAEZ model determines the maximum attainable yield (measured in tons per hectare per year) for each crop in each grid cell of 0.083×0.083 degrees. We use FAO's measures of agroclimatic yields (based solely on climate, not on soil conditions) for intermediate levels of inputs/technology and rain-fed conditions.

Area. The log of surface area in square miles, calculated using the 2010 county shapefiles from NHGIS using GIS software. Data source: Manson et al. (2019).

Temperature. County-level mean annual temperature measured in Celsius degrees. Data source: IIASA/FAO (2012).

Rainfall. County-level average annual precipitation measured in mm. Data source: IIASA/FAO (2012).

Elevation. County-level average terrain elevation in km. Data source: IIASA/FAO (2012).

Latitude. Absolute latitudinal distance from the equator in decimal degrees, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS. Data source: Manson et al. (2019).

Longitude. Absolute longitudinal distance from the Greenwich Meridian in decimal degrees, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS. Data source: Manson et al. (2019).

Distance to the coastline, rivers, and lakes. Minimum distance to a point in the coastline, rivers, and lakes in km, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS. Data source: Manson et al. (2019).

Additional Variables

Annual Migration Inflow. Total number of migrants entering the United States every year. The data for 1820–1890 is available from the Migration Policy Institute (2016), which tabulates data from the Office of Immigration Statistics, while the data for 1790–1819 is imputed from Tucker (1843). To construct the instrumental variable based on annual migration inflows predicted by weather shocks in Europe, we use the annual migration inflows to the U.S. from Belgium, Denmark, England, France, Germany, Greece, Ireland, Italy, Norway, Poland, Portugal, Russia, Scotland, Spain, Sweden, Wales from 1820–1890. Data source: Willcox (1929).

Years Connected to Railroad by 1890. The number of years since the county is first intersected by railroad to 1890. Data source: Atack et al. (2010).

Birthplace diversity, 1890. We take $1 - \sum_o (birthplace_{oc}/population_c)^2$, which is simply 1 minus the Herfindahl concentration index for origin o birthplace diversity in county c in 1890. Birthplaces include US or a given country or country grouping abroad. Data source: Manson et al. (2019)

Ruggedness. County-level average Terrain Ruggedness Index computed using 30-arc grid data on terrain variability. Data source: Nunn and Puga (2012).

Distance to nearest portage site. Minimum distance from county centroid to the nearest portage site, which is defined as the location where a river basin intersects the fall line. Data source: Bleakley and Lin (2012).

Manufacturing Employment Share. County-level percent of employment in manufacturing industries in 1890. Data source: Manson et al. (2019).

Distance to Nearest Mine. Minimum distance from county centroid to a site where there was a mineral discovery before 1890. The data is from the Mineral Resources Data System (MRDS) edited by the US Geological Survey. Data source: McFaul et al. (2000).

Distance to nearest Indian battle sites. Minimum distance from county centroid to major Indian battle sites. The battles sites are digitized using a map from McFaul et al. (2000).

Immigrant share, 1890. County-level percent of foreign born population in 1890. Data source: Manson et al. (2019).

Scottish and Irish immigrant share, 1890. County-level percent of population born in Scotland or Ireland in 1890. Data source: Manson et al. (2019)

Slave population share, 1860. County-level percent of slave population in 1860. Data source: Manson et al. (2019).

Rainfall Risk. Following Ager and Ciccone (2017), county level rainfall risk is constructed as the variance of the annual average log monthly rainfall from 1895-2000. Data source: Oregon State University (2018).

Linked Sample 1850–1880

This section provides a step-by-step description of how we construct the 1850 to 1880 linked sample and how we select the final sample used for the analyses in Section 5.2. The data comes from the 1850 and 1880 US 100% samples (Ruggles et al., 2019) of the North Atlantic Population Project (NAPP) from

Minnesota Population Center (2019). The linking procedure closely follows the approach detailed in Feigenbaum (2016).

Constructing the linked sample

1. Select the sample to be linked from the 1850 complete-count Census data: First, using the complete-count 1850 Census data from NAPP, we select all white males aged 0-20 in 1850. As is common in the literature, we restrict the linking sample to only males because last names are key in the linking process and the task of linking women across censuses is much more difficult given the frequency with which women change their last names in marriage. We keep the information on each individual's first name, last name, age and birth country. Call this dataset X1.

2. Identify all targets that could be potential matches from the 1880 census: The linking procedure requires that for each individual in 1850, the target person in 1880 has the same gender, birth country and within +/- 3 years of the predicted age from the 1850 census. In addition, to limit the set of possible links, we impose an additional restriction that the potential link has a first name Jaro-Winkler distance of less than .2, and a last name Jaro-Winkler distance of less than .2. Call this dataset X2.

3. Create all possible matching pairs between X1 and X2. For each male aged 0-20 in 1850, we create a dataset that contains all the possible matching pairs with the candidates we identified in Step 2. Call this dataset X1X2. Hence, X1X2 contains all the possible match pairs for the 5,329,750 white men we extracted from the 1850 census.

4. Create a training dataset based on manual linking of a random subset of all potential pairs: For a random sample of 10,000 individuals in X1, we extract all the possible matches from the X1X2 dataset. The paired data contains information on first name, last name and age of the individuals both in 1850 and 1880. After discarding matches with very small link probabilities based on first and last name similarity, we are left with a random sample of potential matching pairs for 8,557 white boys from 1850 matched with all potential targets in 1880.

Then, experienced research assistants examined the randomly sampled dataset and decided which pairs formed true matches by comparing the similarities in first name, last name and age for each possible link. If there are no good matches for a given unique record in X1, then all the links were marked as non-match. The manual linking procedure generates a training dataset which contains a random subset of X1X2 in which we have determined that a record is either a match or not a match. The research assistants were able to locate 42.7% (n=3,654) of our randomly selected 8,557 white boys from 1850 in the 1880 census.

5: Construct the training algorithm: The algorithmic approach aims to approximate the best efforts of the researcher assistants in the manual linking procedure to assess matches. In particular, using the sample of manually linked matches from Step 4, we train an algorithm that predicts the likelihood of a match between all potential match pairs from 1850 and 1880 in the X1X2 sample.

The algorithm considers several variables that are constructed to capture similarities based on first name, last name and age similarity which aim to describe the features of the potential matches. The algorithm, which is a probit model, provides us with weights that we can put on the aforementioned variables when deciding on which links to consider as matches.

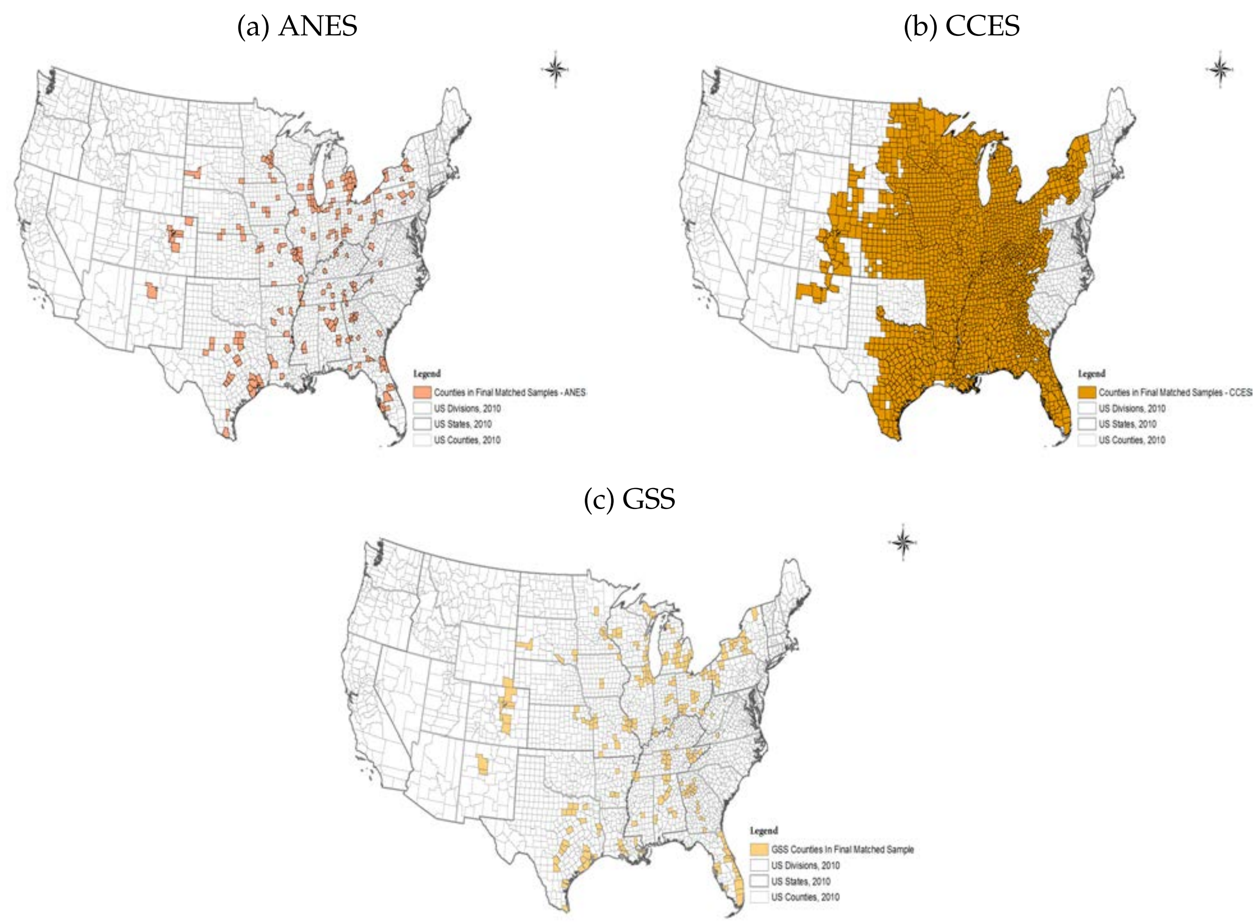
6: Apply the algorithm generated using the training dataset to the full X1X2 data: Using the same set of variables and the algorithm produced in Step 5, we predict the probability that each pair of matches in X1X2 is a true match. We then select which pair of matches are true matches based on an optimal cutoff that minimizes the likelihood of making false negative and false positive errors. See Feigenbaum (2016) for the details.

The matching process allows for multiple individuals in 1850 to be matched with a single individual in 1880. When we allow for multiple individuals from 1850 to match to a single individual in 1880, the algorithm has a match rate of 47.9% (n=2,552,950). When we restrict that only one individual in 1850 can match to an individual in 1880, the algorithm has a match rate of 33.04% (n=1,761,408). These compare favorably to other recent efforts in the literature (see footnote 1 in Appendix I).

Selection of Sample Used in Section 5.2.

7: Identify the individuals in the linked sample who are fathers in 1880: We take our sample of linked white men that we traced from 1850 to 1880, and we keep those that are household heads with children in 1880. We then bring all the relevant data for our linked men both from the 1850 and 1880 census and match it to their household members in 1880. Then, we keep the data on all their children, both boys and girls, who are under the age of 20 when we observe them in 1880. Hence, the final sample used in the analysis contains the 1880 record of the children who are descendants of the men we were able to link between 1850 and 1880.

Figure K.1: Geographic Coverage For Main Survey Data Sources



Notes: Figures (a), (b), and (c) provide the geographical distribution of the maximum number of counties available in our baseline sample matched with the ANES, CCES, and GSS data, respectively. Additional counties are included when incorporating the West Coast sample or extending the historical frontier window to 1950 (see Section 4.4). Due to varying data availability across rounds, not all the counties in the above map are included in every regression using the corresponding survey data.

References

Abramitzky, R., L. P. Boustan, and K. Eriksson, "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration," *American Economic Review*, 2012, *102* (5), 1832–1856.

Acemoglu, D., J. Moscona, and J. A. Robinson, "State capacity and American technology: evidence from the nineteenth century," *American Economic Review: Papers & Proceedings*, 2016, *106* (5), 61–67.

Acharya, A., M. Blackwell, and M. Sen, "The Political Legacy of American Slavery," *The Journal of Politics*, 2016, *78* (3), 621–641.

Ager, P. and A. Ciccone, "Agricultural risk and the spread of religious communities," *Journal of the European Economic Association*, 2017, *16* (4), 1021–1068.

Alesina, A. and P. Giuliano, "Preferences for Redistribution," in J. Benhabib, A. Bisin, and M. O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, North Holland, 2011, pp. 93–131.

_, E. L. Glaeser, and B. Sacerdote, "Why Doesn't the United States Have a European-Style Welfare State?," *Brookings Papers on Economic Activity*, 2001, *2001* (2), 187–277.

Ansolabehere, S. and B. F. Schaffner, "General Social Surveys, 1972-2014," *CCES Common Content, 2006-2016*, 2017.

Atack, J., F. Bateman, M. Haines, and R. A. Margo, "Did Railroads Induce or Follow Economic Growth?," *Social Science History*, 2010, *34* (2), 171–197.

Bailey, M., C. Cole, and C. Massey, "Simple Strategies for Improving Inference with Linked Data: A Case Study of the 1850-1930 IPUMS Linked Representative Historical Samples," *Historical Methods*, forthcoming.

Bester, C. A., T. G. Conley, and C. B. Hansen, "Inference with dependent data using cluster covariance estimators," *Journal of Econometrics*, 2011, *165* (2), 137–151.

Billington, R. A., *America's Frontier Heritage*, University of New Mexico Press, 1974.

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

Cameron, A. C., J. B. Gelbach, and D. L. Miller, "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*, 2011, *29* (2), 238–249.

Collins, W. J. and M. H. Wanamaker, "Up from Slavery? African American Intergenerational Economic Mobility Since 1880," *NBER Working Paper 23395*, 2017.

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

Feigenbaum, J. J., "Automated Census Record Linking: A Machine Learning Approach," *Unpublished Manuscript*, 2016. https://jamesfeigenbaum.github.io/research/census-link-ml/.

Gentzkow, M., J. M. Shapiro, and M. Taddy, "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 2019, *87* (4), 1307–1340.

Gorodnichenko, Y. and G. Roland, "Culture, Institutions and the Wealth of Nations," *Review of Economics and Statistics*, 2016, *99* (3), 402–416.

Grogger, J. and G. H. Hanson, "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 2011, *95* (1), 42–57.

Hornbeck, R., "Barbed wire: Property rights and agricultural development," *The Quarterly Journal of Economics*, 2010, *125* (2), 767–810.

IIASA/FAO, "Global Agro-ecological Zones (GAEZ v3.0)," *Unpublished Manuscript*, 2012.

Long, J. and J. Ferrie, "Intergenerational Occupational Mobility in Britain and the US since 1850," *American Economic Review*, 2013, *103* (4), 1109–1137.

Manson, S., J. Schroeder, D. Van Riper, and S. Ruggles, "IPUMS National Historical Geographic Information System: Version 14.0 [Database].," Minneapolis, MN: IPUMS. 2019.

Markus, G. B., “American Individualism Reconsidered,” in J. H. Kuklinski, ed., *Citizens and Politics: Perspectives from Political Psychology*, New York: Cambridge University Press, 2001, pp. 401–32.

McFaul, EJ, GT Mason, WB Ferguson, and BR Lipin, “US Geological Survey mineral databases; MRDS and MAS/MILS,” Technical Report, US Geological Survey, 2000.

Migration Policy Institute, “Tabulations of U.S. Department of Homeland Security, Office of Immigration Statistics, Yearbook of Immigration Statistics,” 2016.

Minnesota Population Center, “Integrated Public Use Microdata Series, International: Version 7.2 [dataset],” Minneapolis: IPUMS, 2019 2019.

Nunn, N. and D. Puga, “Ruggedness: The blessing of bad geography in Africa,” *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

of Engineers: FEMA’s National Dam Safety Program Washington, DC: U.S. Army Corps, “The National Inventory of Dams,” 2018.

Oregon State University, “PRISM Climate Group,” 2018.

Oster, E., “Unobservable selection and coefficient stability: Theory and evidence,” *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Overmeyer, P. H., “Westward Expansion Before the Homestead Act,” in H. F. Williamson, ed., *The Growth of the American Economy: An Introduction to the Economic History of the United States*, New York: Prentice-Hall, Inc, 1944, pp. 82–112.

Robinson, P. M., “Root-N-Consistent Semiparametric Regression,” *Econometrica*, 1988, *56* (4), 931–954.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek, “Integrated Public Use Microdata Series: Version 9.0 [dataset],” Minneapolis, MN: IPUMS, 2019. 2019.

Sequeira, S., N. Nunn, and N. Qian, “Immigrants and the Making of America,” *The Review of Economic Studies*, 2020, *87* (1), 382–419.

Smith, T. W., P. Marsden, M. Hout, and J. Kim, “General Social Surveys, 1972-2014,” 2015.

Stewart, J. I., “Migration to the Agricultural Frontier and Wealth Accumulation, 1860–1870,” *Explorations in Economic History*, 2006, *43* (4), 547–577.

The American National Election Studies, *THE ANES GUIDE TO PUBLIC OPINION AND ELECTORAL BEHAVIOR*, University of Michigan, Center for Political Studies: Ann Arbor, MI.

Thornhill, R. and C. L. Fincher, *The Parasite-Stress Theory of Values and Sociality: Infectious Disease, History and Human Values Worldwide*, New York: Springer, 2014.

Tucker, G., “Progress of Population and Wealth in the United States, in Fifty Years, as Exhibited by the Decennial Census Taken in that Period,” *The Merchants’ Magazine and Commercial Review*, 1843, *9*, 136–44.

Willcox, W. F., “International Migrations, Volume I: Statistics,” *National Bureau of Economic Research*, 1929.