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

Samuel Bazzi†
Boston University and CEPR

Martin Fiszbein‡
Boston University and NBER

Mesay Gebresilasse§
Boston University

November 2017

Abstract

In 1893, Frederick Jackson Turner 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 county-level measure of total frontier experience (TFE). Historically, frontier locations had distinctive demographics and greater individualism. Many decades after the closing of the frontier, counties with greater TFE exhibit more pervasive individualism and opposition to redistribution. Suggestive evidence on the roots of rugged individualism points to selective migration, the adaptive advantage of self-reliance, and opportunities for upward mobility through effort.

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

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

*We thank Romain Wacziarg for a helpful discussion of the paper at the NBER Summer Institute. We also thank Jeremy Atack, Michael Clemens, William Collins, James Feigenbaum, Ray Fisman, Oded Galor, Camilo Garcia-Jimeno, Paola Giuliano, Bob Margo, Nathan Nunn, Ömer Özak, Daniele Paserman, John Wallis, and David Weil, as well as seminar participants at ASREC, Boston University, Brown, Georgetown, the NBER Political Economy Summer Institute, the NBER Development of the American Economy Summer Institute, UPenn, and Vanderbilt for helpful comments. Yeonha Jung, Max McDevitt, 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. 270 Bay State Rd., Boston, MA 02215. Email: mmelese@bu.edu.

1 Introduction

Rapid westward expansion marked the early history of the United States. Until the late 19th century, its territory contained vast tracts of open land. According to the influential historian Frederick Jackson Turner, the frontier that divided settled and yet unsettled locations strongly influenced American culture, fostering the development of unique cultural traits. Salient among these were individualism and opposition to government intervention (Turner, 1920). The concept of "rugged individualism," famously used by Republican President Herbert Hoover, captures the combination of these traits.[1]

This paper shows that the American frontier shaped a culture of rugged individualism that persisted throughout time. First, using Census data from the 18th and 19th century, we establish the distinctive demographics and higher levels of individualism that historically characterized frontier locations. Then, using modern survey and Census data, we show that locations exposed to the frontier for a longer period historically exhibit higher contemporary levels of individualism, lower desired and actual levels of redistribution, and stronger opposition to government regulation. Finally, using linked Census records, we document empirical patterns that point to the underlying mechanisms. Frontier individualism is partly but not entirely explained by selective migration. Frontier conditions favored individualism through differentially higher socioeconomic returns, and they created expectations of high income growth through effort that fueled opposition to government intervention.

To understand the contemporaneous and long-run effects of the frontier across U.S. locations, we revisit the classic Frontier Thesis through the lens of modern political economy and social psychology. In our simple conceptual framework, the significance of the frontier can be explained by three factors. First, frontier locations attracted individualists able to thrive in harsh conditions. Second, the frontier experience, characterized by isolation and low population density, further promoted the development of self-reliance. At the same time, favorable prospects for upward mobility through effort nurtured hostility to redistribution. Finally, frontier populations affected local culture at a critical juncture, thus leaving a lasting imprint.

We determine the position of the frontier and track its evolution over time using population data from the Census and applying Geographic Information System (GIS) techniques. Following Turner's classic essay and the *Progress of the Nation* report from the 1890 Census, we define the frontier line as the line at which population density dropped below two people per square mile. We identify the frontier at each point in time as comprised of counties with low population density in close proximity to the frontier line. We measure total frontier experience as the time spent on the frontier during the 1790–1890 period. This provides a precise and comprehensive measurement of the history of the American frontier, and to our knowledge, the first measure of historical intensity of frontier exposure.

Consistent with historical narratives, we find systematic evidence of the demographic and cultural distinctiveness of frontier locations. Frontier settlers were disproportionately male, prime-age, illiterate, and foreign-born. Both the sparse population and the isolation of frontier locations are drivers of these distinctive traits. Using semiparametric regressions, we identify sharp structural breaks in these demo-

[1] See Hoover (1929), which compiles the 1928 presidential campaign speeches. The Merriam-Webster Online Dictionary defines rugged individualism as "the practice or advocacy of individualism in social and economic relations emphasizing personal liberty and independence, self-reliance, resourcefulness, self-direction of the individual, and free competition in enterprise." According to Wikipedia, rugged individualism "refers to the idea that individuals should be able to help themselves out and that the government does should not involve itself in the economic lives of people or the nation in general."

graphic variables close to the population density cutoff defining the frontier line in historical accounts. Moreover, event study specifications show how these traits evolve as counties exit the frontier.

Frontier locations also had sharply higher levels of individualism, as reflected in the prevalence of infrequent children's names. This result holds for several alternative ways of measuring infrequency and restricting to children with native-born parents or grandparents. The informational content of given names has been emphasized in economics (e.g., Abramitzky, Boustan and Eriksson, 2016; Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004; Olivetti and Paserman, 2015) as well as psychology and sociology (e.g., Gerrit and Onland, 2011; Gureckis and Goldstone, 2009; Lieberson and Bell, 1992). We borrow our names-based measure of individualism from social psychologists, who note that individualistic types are prone to give their children infrequent names, reflecting a desire to stand out, as opposed to common names, reflecting a desire to fit in (Twenge, Abebe and Campbell, 2010). This measure is strongly correlated with other proxies for individualism in multiple contexts (Beck-Knudsen, 2017; Ogihara et al., 2015; Varnum and Kitayama, 2011). In our setting, names have the crucial advantage of allowing us to measure individualism historically. Furthermore, name choices are particularly useful for studying cultural persistence as they represent a primordial act of cultural transmission by parents, which has lasting effects on children's identity and behavior (Nelson and Simmons, 2007; Yadin, 2016).

We investigate the long-run effects of frontier exposure on culture using our new measure of total frontier experience (TFE). First, we show that TFE positively correlates with infrequent naming patterns several generations after the closing of the frontier. Second, we find a robust association of TFE with opposition to redistribution and public spending based on several popular surveys capturing different notions of government intervention. In terms of actual policy outcomes, TFE is associated with lower local property tax rates. These long-run results are robust to different spatial fixed effects as well as a host of geographic and agroclimatic controls, including, among others, area, latitude and longitude, rainfall and temperature, distance to waterways, and potential agricultural productivity.

These long-run differences in preferences have translated into stronger contemporary support for the Republican Party. Each decade of TFE is associated with 3.5 percent more votes for Republican candidates in presidential elections since 2000. This association ratchets up over the 2000s as each election exhibits a significantly larger effect of TFE, with a particularly large frontier legacy in the 2016 election. Frontier exposure explains a significant part of the increase in Republican vote shares in the American heartland from 2000 to 2016, a period of sharp political polarization during which the conservative political agenda moved further to the right. The legacy of TFE remains significant when compared to the effect of Chinese import competition, a leading proximate explanation for the recent regional shift towards the Republican Party (Autor, Dorn, Hanson and Majlesi, 2016).

We provide deeper insights into why Republican party support is stronger in areas with higher TFE by considering preferences over a set of contentious policy issues: the Affordable Care Act, increases in the minimum wage, the ban on assault rifles, and the regulation of CO_2 emissions. The political discussions around these policies provide telling examples of the trend toward polarization. Moreover, Republican Party positions can be linked to salient aspects of the frontier culture described in the historical literature including opposition to state intervention, strong belief in effort versus luck in reward, necessity of self-defense, and notions of "manifest destiny." We show that locations with higher TFE exhibit stronger opposition to each of these policies. These effects survive even after conditioning on

strength of identification with the Republican Party, individual education, and income.

We take several steps towards a causal interpretation of these long-run effects. First, we show that the results survive a progressively richer set of controls, accounting for potential confounders of frontier experience and contemporary culture. These include, among others, population density, diversity, and historical access to railroads. Second, we rule out additional concerns about omitted variables by using the Oster (2016) approach to show that selection-on-unobservables would have to be implausibly large to explain the observed empirical patterns. Third, the key results hold when expanding the analysis to include regions exposed to frontier conditions after the Census declared the frontier closed in 1890, or the West Coast, including California, which experienced its own frontier expansion in the second half of the 19th century. Finally, we introduce an instrumental variables (IV) strategy.

Our IV strategy exploits variation in immigrant inflows to the United States over time. For each location, we consider the intensity of inflows in the period starting just before the onset of local frontier settlement. The ups and downs of immigration affected the speed of westward expansion and hence the length of time it took for frontier locations to become established settlements. The exclusion restriction requires that expected local frontier conditions were not a major pull factor affecting contemporaneous aggregate flows to the U.S. While this seems plausible from the perspective of a small frontier county, we also find similar results when using an alternative instrument based on predicted outflows from Europe due to climatic shocks, isolating push factors abroad as in Nunn, Qian and Sequeira (2017).

In an attempt to understand the roots of frontier culture, we first examine the selective migration of individualistic types to the frontier. Using a linked sample of households from the 1870 and 1880 Census, we are able to track people across locations and decompose county-level differences in individualism into components coming from migrants versus long-time residents. We find that selective migration was significant, though frontier conditions may have also contributed to rugged individualism.

We identify two factors that may have complemented and reinforced selective migration. First, individualism came to thrive on the frontier due to its adaptive advantage in a setting of wilderness and isolation, where self-reliance was necessary for survival and success. We find that infrequent names are associated with greater socioeconomic status on the frontier than elsewhere and also with lower rates of out-migration from the frontier. Second, frontier conditions presented opportunities for upward mobility through effort, which would hone opposition to redistribution. We show that intergenerational persistence in socioeconomic status was indeed weaker in frontier locations, building upon prior evidence of widespread access to land and high rates of wealth accumulation on the frontier.[2]

This paper contributes to a growing economics literature on culture, focusing on individualism and preferences for redistribution. Individualism has not attracted much attention in economics, with the exception of work by Greif (1994) and recent contributions by Gorodnichenko and Roland (2011, 2012, 2015, 2016). In contrast, preferences for redistribution are the subject of a large literature (see Alesina and Giuliano, 2010, for a survey). In an early contribution on differences between the U.S. and Europe, Alesina, Glaeser and Sacerdote (2001) conjecture that "American anti-statism" may be partly traced to the frontier, which "strengthened individualistic feelings and beliefs in equality of opportunities rather than equality of outcomes." To our knowledge, the only empirical examination of the Frontier Thesis

[2]We also consider a competing, disease-based explanation for the origins of individualism rooted in biology and known as the parasite-stress theory (Fincher and Thornhill, 2012). However, using data on disease and illness in the 1880 Census, we find little evidence in support of this mechanism as a key factor explaining frontier culture.

in economics is García-Jimeno and Robinson (2011), which links variation in the quality of democratic institutions across countries in the Americas to variation in the historical importance of the frontier. In this paper, we identify the long-run effects of the American frontier on individualism and preferences for redistribution using a wealth of subnational data and a novel measure of historical frontier experience.

Our results on the long-run effects of frontier experience add new evidence to a growing literature on the deep roots and persistence of cultural traits (e.g., Alesina, Giuliano and Nunn, 2013; Fernández, 2010; Guiso, Sapienza and Zingales, 2016; Nunn and Wantchekon, 2011; Spolaore and Wacziarg, 2013; Voigtländer and Voth, 2012). We do not address the roots of cultural traits in the U.S. as a whole; nor do we provide an explanation for cross-country cultural variation. However, our findings offer a new perspective on the process by which American culture and politics became imbued with high levels of individualism and opposition to redistribution. They may also shed new light on a puzzle in American political economy, namely the relative stability of preferences for redistribution over the last 40 years despite significant increases in inequality (see Ashok, Kuziemko and Washington, 2015).[3]

We contribute to a large literature outside economics that elaborates on Turner's influential work. Many studies in history and sociology describe the demographic characteristics of the frontier.[4] We provide a comprehensive and systematic analysis of its distinctive features, measuring the local prevalence of individualism for the first time. A small literature in social psychology, beginning with Vandello and Cohen (1999), documents higher levels of individualism in the Western United States using contemporary state-level data. Some recent studies reframe Turner's argument that individualism has an adaptive value in frontier conditions, and some adopt infrequent names as a proxy (e.g., Varnum and Kitayama, 2011). We not only identify the connection between historical and modern individualism but also provide empirical evidence on underlying mechanisms rooted in economic theory. Moreover, our results go beyond broad geographic correlations to offer evidence on the link from local frontier exposure to preference formation and, ultimately, the contemporary political landscape.

The paper is organized as follows. Section 2 provides a general discussion of individualism and opposition to redistribution as well as economic theories about their origins and consequences. We also link these theories to the Frontier Thesis and offer a simple conceptual framework to understand its significance. Section 3 explains how we locate the frontier and measure historical frontier exposure. Section 4 documents the distinctive features of frontier populations. Section 5 provides estimates of the long-run effects of frontier experience on culture. Section 6 then offers evidence for why the frontier may have favored individualism and opposition to redistribution. Section 7 concludes with key lessons and caveats about extrapolating to other countries or even the U.S. as a whole.

[3] *New York Times* journalist David Brooks nicely captures this possible connection in his July 4th, 2017 article, "What's the Matter with Republicans?", questioning voters' proclivities for supporting policies that are seemingly against their economic self-interest. He conjectures, "My stab at an answer would begin in the 18th and 19th centuries. Many Trump supporters live in places that once were on the edge of the American frontier. Life on that frontier was fragile, perilous, lonely and remorseless. [. . .] discipline and self-reliance were essential. [. . .] In their view, government doesn't reinforce the vigorous virtues. On the contrary, it undermines them. [. . .] I'd say they believe that big government support would provide short-term assistance, but that it would be a long-term poison to the values that are at the core of prosperity."

[4] We discuss this rich literature throughout the paper in Sections 2, 4, and 6.

2 A Modern Reading of The Frontier Thesis

This section provides a conceptual background for our study that connects Turner's ideas about the American Frontier with contemporary political economy, cultural economics, and social psychology. We start by discussing some contributions in this literature that provide insight into outcomes of interest. Then, we restate the Frontier Thesis, spelling out the potential channels for initial influence and subsequent persistence.

2.1 Individualism and Preferences for Redistribution

A large literature in social and cultural psychology portrays individualism as the most important dimension of cross-country variation in culture (e.g., Triandis, 1995; Heine, 2010). Following Hofstede (1980, 1991) and Triandis (1988, 1995, 2001), we think of individualism (in contrast to collectivism) as comprising several related traits: a view of the self as independent rather than interdependent, the emphasis on self-reliance, the primacy of self-interest, and the regulation of behavior by personal attitudes rather than social norms.[5]

Empirical measures of individualism illustrate the concept more concretely. Some studies use Hofstede's survey-based index while others propose coarse proxies like divorce rates or the percentage of people living alone. Social psychologists meanwhile argue that individualism can be captured by the use of first- and second-person singular pronouns or by the share of infrequent children names. Kashima and Kashima (1998) show that in individualistic cultures, "I" and "you" are never dropped, as that would de-emphasize the individual. Twenge, Abebe and Campbell (2010) argue that infrequent (common) children names reflect parents' desires to stand out (fit in). In this respect, name choices echo the behavior characterized by Kim and Markus (1999) when contrasting preference for uniqueness in American culture with preference for conformity in East Asia: given the choice among a set of colored pens, Americans chose the minority color while East Asians chose the majority color.

In economics, a small set of contributions has focused on individualism, starting with the work of Greif (1994) on how individualistic and collectivistic cultures shaped different trade institutions in the Middle Ages. The recent contributions of Gorodnichenko and Roland (2011, 2015, 2016) show that individualistic countries have higher levels of income, productivity, and innovation, as well as more democratic institutions. Gorodnichenko and Roland (2016) explain some of these effects through an endogenous growth model in which individualism fosters innovation by creating incentives to stand out.

Preferences for redistribution are distinct but closely related to individualism. Paul Samuelson (1965) once noted that "to an economist the word 'individualism' is tied up with *laissez faire*." In fact, Alesina and Giuliano (2010) measure preferences for redistribution using the same question from the General Social Survey that Di Tella, Dubra and MacCulloch (2008) use to measure individualism. Using data from the European Social Survey, Quattrociocchi (2014) shows that immigrants who were born in countries with a more individualistic culture tend to have weaker preferences for redistribution in their country of residence. Intuitively, the defining characteristics of individualism—self-interest and inclination toward self-reliance—may be associated with opposition to redistribution and other forms of government

[5]Some aspects of this notion go back to the classic works of Durkheim (1893), Tönnies (1887), and Weber (1905).

intervention.[6] The connection is explicit in the American ideology of "rugged individualism," which promotes self-reliance and opposes state intervention through taxes or regulations.

Preferences for redistribution are the subject of a rich literature in economics. A few contributions (e.g., Alesina and Angeletos, 2005; Alesina and La Ferrara, 2005; Benabou and Ok, 2001; Benabou and Tirole, 2006; Piketty, 1995) offer particularly useful insights for our understanding of frontier culture as elaborated in the following section.[7] One point concerns the role of expectations about future income. If there is inertia in tax rates, then favorable prospects of upward mobility tend to generate opposition to redistribution. This literature also shows how the importance of effort (relative to luck) in the income-generation process may lead to lower desired tax rates because of concerns about efficiency and fairness. The greater the importance of effort, the larger the negative effects of taxes due to adverse incentives, and the larger the perception that they are unfair. Finally, these studies offer models with multiple equilibria that can shed light on the persistent nature of cultural traits.

2.2 The Frontier Thesis: A Restatement

According to the classic thesis advanced by F.J. Turner, the presence of a frontier separating established settlements from vast tracts of open land during a formative period shaped the distinctive aspects of American culture. Quoting from his 1893 essay published in Turner (1920), the frontier was "the meeting point between savagery and civilization," and "the existence of an area of free land, its continuous recession, and the advance of American settlement westward, explain American Development." According to his thesis, "these free lands promoted individualism, economic equality, freedom to rise, democracy." He also observed that, on the frontier, the "tax-gatherer is viewed as a representative of oppression," since the environment "produces antipathy to control".

The conceptual framework guiding our analysis combines some of Turner's ideas with insights from economics and social psychology. We think of the contemporaneous and long-run effects of the frontier as the result of three main forces. First, frontier locations attracted people with distinctive characteristics, both in terms of demographics and the prevalence of individualism. Second, the frontier experience, characterized by isolation and wilderness, fostered the development of self-reliance and related cultural traits. Finally, the distinctive features of frontier populations affected preferences and social norms at a critical juncture of institutional formation and thus left a persistent imprint on local culture. While these mechanisms may be relevant beyond the context of our study, it is of course possible that certain preconditions were specific to American history. We discuss the three mechanisms below and revisit this question of external validity in the conclusion.[8]

Selective Migration. Traditional narratives characterize the frontier by the prevalence of young single men, mostly of low socioeconomic status. These distinctive demographics reflected the type of people

[6] Gorodnichenko and Roland (2012) note that "The individualist view of government would tend to be wary of possible infringements of government on the individual's drive to self-achievement." In the sociology literature, Celinska (2007) notes that an aspect of "utilitarian individualism and the consequence of a strong belief in self-reliance" is the "opposition toward governmental efforts to equalize citizens' economic position, to limit private business, and to build strong social programs that provide assistance to the most disadvantaged."

[7] While these studies focus on preferences for redistribution, their insights also apply more broadly to preferences about the size and scope of government in terms of taxation and labor market regulations.

[8] Note that our findings should be interpreted with caution even when considering the U.S. as a whole. While Turner's thought largely addressed culture and politics at the national level, our analysis focuses on subnational variation.

willing and able to undertake migration and settle on the frontier. Harsh living conditions and high crime rates were particularly hostile to women and the elderly, which helps explain the skewed sex ratio and age distribution. In addition, the frontier attracted workers from the lower end of the urban skill distribution, as suggested by the theory of the "safety valve" (see Ferrie, 1997; Goodrich and Davison, 1935, 1936; Steckel, 1989; Stewart, 2006; Turner, 1920).

Frontier residents also tended to exhibit a high degree of individualism. Migrants generally have independent mindsets. This trait may be amplified among those moving to the frontier, giving up their social environment to settle in remote and isolated contexts (see Beck-Knudsen, 2017; Jokela, 2009; Kitayama et al., 2006, 2010). Moreover, as discussed next, the adaptive advantage of self-reliance in such conditions would further hasten the self-selection of individualist types to the frontier.

Effects of Frontier Conditions. While frontier locations attracted people with specific traits, the frontier's unique natural and social conditions, in turn, influenced the settlers' values, beliefs, and behavior. In Turner's words, "a modification of the original stock occurred." Remoteness and isolation implied a particular set of opportunities and challenges. The abundance of land and other natural resources offered ample profit opportunities, insofar as they were deftly exploited. On the other hand, as Overmeyer (1944) argues, "life was rough, crude, hard, and dangerous." Frontier settlers often faced harsh climatic conditions and multiple types of danger, such as plagues, droughts, blizzards, and crop failure, as well as attacks from wild animals, Native Americans, and other settlers. Violence was commonplace, and social infrastructure providing protection and care was limited or nonexistent.[9]

These opportunities and threats on the frontier may have favored individualism through an adaptive mechanism. In the frontier context, people had to rely on themselves for protection and prevention, and to improve their living conditions.[10] Moreover, the resourcefulness associated with individualism would prove useful in a context characterized by novel and uncertain conditions. Thus, individualistic traits had an adaptive value: beliefs and behavior based on independence and self-reliance made people better suited to cope with the frontier environment (Kitayama et al., 2010; Plaut et al., 2002). In turn, the adaptive advantage of individualism may have increased its prevalence in the population through differential reproductive success, learning, or both (see Galor and Özak, 2016).

Moreover, land abundance and remoteness also offered favorable prospects of upward mobility and a large perceived importance of effort in income generation. Based on the political economy theories mentioned in 2.1, and as conjectured by Alesina, Glaeser and Sacerdote (2001), these conditions would naturally foster opposition to government intervention. This resonates with historical narratives. Billington (1974), a noted Turnerian, argued 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."

[9] Rampant violence, noted in many historical narratives of the frontier, are a common characteristic of contexts with low population density and high population mobility, lack of well-defined property rights, and absence of clear mechanisms for law enforcement (Couttenier, Grosjean and Sangnier, 2016; Grosjean, 2014; Nisbett and Cohen, 1996; Restrepo, 2015).

[10] While returns to cooperation may have been high, maintaining reciprocity would have been difficult in frontier settings with such high population mobility, as noted in the literature on social capital (see Munshi, 2014). Some 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."

Cultural Persistence. Culture can be remarkably persistent but can also change rapidly (Gershman, 2017; Giuliano and Nunn, 2017). In our view, the persistence of frontier culture long after frontier conditions abated can be linked to the distinctive traits of early settlers at critical junctures of institutional development. A seminal theory in cultural geography due to Zelinsky (1973) captures this potential channel. The "doctrine of first effective settlement" argues that 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."

The economics literature on culture offers several mechanisms by which the distribution of cultural traits in the population at a point in time can influence its subsequent evolution (see Bisin and Verdier, 2010). First, it affects the likelihood that new generations adopt these traits through horizontal transmission. Second, it can influence vertical transmission, for example, by affecting beliefs about the behavior of other members of society and thus the expected rewards for different traits (see Guiso, Sapienza and Zingales, 2008). Various models of intergenerational transmission imply that initial conditions determine the long-run cultural equilibrium. The initial conditions in our setting were precisely those distinctive traits of frontier populations during the inception of settlement.

In addition, the initial distribution of traits can influence the long-run cultural equilibrium by shaping early institutions (see Bisin and Verdier, 2017; Tabellini, 2008a,b). Several theoretical models suggest that initial preferences for redistribution lead to different institutional outcomes, which in turn affect the evolution of preferences (see Alesina and Angeletos, 2005; Alesina, Cozzi and Mantovan, 2012; Benabou and Tirole, 2006; Bisin and Verdier, 2005). These models feature multiple equilibria in cultural traits, which entails path dependence and persistence. Consistent with these notions, Turner (1920) himself noted 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."

3 Mapping the History of the Frontier

This section presents our method for mapping the history of the frontier. After providing historical background, we explain how to use U.S. Census data and GIS techniques to determine the position of the frontier line at each point in time. Finally, we explain the definition of frontier counties and our measure of total frontier experience. As a baseline, we consider the period of frontier settlement through 1890, as defined by the U.S. Census Bureau, but, for robustness, we also consider an extension through 1950 in Section 5.5.

From colonial times until the late 19th century, America underwent rapid population growth and a massive westward expansion. Historical sources document this process, and the noteworthy 1890 Census report on the *Progress of the Nation* (Porter, Gannett and Hunt, 1890) provides a key source of inspiration for Turner's classic 1893 essay. The authors observe that the Thirteen Colonies, already settled communities by 1790, 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 describes in great detail the decade-by-decade evolution of population growth, the extent of the settled area, and the westward movement of the center of population throughout the 1790–

1890 period.[11] It also suggested the notion of a "frontier," which appeared, in various ways, in previous works (see references in Juricek, 1966; Mood, 1945).

The population of the Thirteen Colonies was concentrated in close proximity to the East coast. From 1790 to 1890, as the nation's total population increased from 3.9 million to 62.6 million, the extent of settled area went from under 240,000 square miles to almost 2,000,000. In the same period, the mean center of population shifted westward over 500 miles, from a point in Maryland, just east of Washington D.C., to Decatur, Indiana.

The process of frontier settlement advanced continuously throughout the 1800s. However, the speed of westward shifts in the center of population varied considerably from decade to decade. The magnitude of the shift decreased from 1790–1800 to 1800–1810, only to speed up again in the next decade. The maximum east-to-west shift was 127 km in 1850–1860, a decade of large immigrant inflows to the U.S. (2.8 million) and very fast overall population growth (a 35.5 percent increase). In the following decade, which witnessed the Civil War, the westward shift hit a 3-decade low of around 70 km, but it bounced back in the 1870s to around 90 km.

The Porter et al. (1890) Census report considered the process of westward expansion complete, and the frontier closed, by 1890. In a passage famously 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."[12] This eventual breaking down of the frontier line can be seen clearly in our new measure of frontier exposure that we describe next.

3.1 Locating the Frontier and Tracking its Movements

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

We take a different approach, adopting a definition of the frontier line that follows the Porter et al. (1890) report and Turner's classic essay. While acknowledging the fluidity of the concept, Turner embraced the Porter et al. definition of the frontier line as the the line dividing settlements with population density of two or more per square mile from those with less.[13] We therefore define frontier counties as those (i) in close proximity to the frontier line (100 kilometers in our baseline), and (ii) with population density below six people per square mile, a cutoff suggested by Porter et al..[14] While these cutoffs are

[11] The Census report included some remarkable maps of population density, to which our own maps bear a close resemblance (see Appendix Figure A.1). We include in the figure the mean center of population, i.e., 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.

[12] Turner would go on to discuss some of the forces driving variation in the shape and speed of westward expansion. We revisit these underlying forces in Section 5 when developing an empirical strategy for studying the long-run effects of the frontier.

[13] Turner (1920) 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. The term is an elastic one, and for our purposes does not need sharp definition. We shall consider the whole frontier belt including the Indian country and the outer margin of the "settled area' of the census reports."

[14] Traveling 100 kilometers in a covered wagon would have required about 2–3 days in the early 19th century.

necessarily arbitrary, we offer empirical support for these definitions in Section 4.2, and our primary results in Section 5 are qualitatively unchanged using different distance and density thresholds.

For each Census year from 1790 to 1890, we calculate county-level population density per square mile. For intercensal years, we interpolate county-level population density by assuming a constant annual population growth rate that matches the decadal growth rate (replacing initial zeros with arbitrarily small values to avoid infinite growth rates). We maintain consistent units of observation over time by harmonizing all data to the 2010 boundaries using an approach suggested in Hornbeck (2010) and detailed in Appendix C. While the harmonization and interpolation procedures entail measurement error, any resulting attenuation bias should work against finding systematic effects of the frontier and is moreover eliminated in the instrumental variable estimates.[15]

Using the annual county-level population densities between 1790–1890, we locate the frontier line for each year by drawing contour lines that divide counties with population densities above and below two people per square mile. Figure 1 plots the resulting lines for 1790, 1820, 1850, and 1890, and full details on the underlying GIS procedure can be found in Appendix A. In order to closely approximate historical notions of the frontier as "margins of civilization," we discard all line segments less than 500 km, as well as isolated pockets of relatively sparse populations within the main area of settled territory (i.e., to the east of the main frontier line).[16] Figure 2 shows the evolution of the main frontier line in red from 1790 to 1890. A second major frontier line emerges on the West Coast, starting in California, in the mid-19th century (see Figure 2). This process of settlement was marked by the Gold Rush and different historical forces than the east-to-west expansion. We leave the locations spanned by this secondary frontier out of the baseline analysis but consider them for robustness in Section 5.5.

3.2 Total Frontier Experience

The uneven speed of westward expansion gives rise to differences across counties in time exposed to the frontier that we exploit in our investigation of long-run persistence.

To measure the intensity of historical frontier experience for each location, we calculate the number of years spent within the frontier belt from 1790 to 1890. For each year between 1790–1890, we assign each county a dummy variable equal to one if it is on the frontier according to the abovementioned definitions of proximity to the frontier line and density. Then, the total frontier experience (TFE) for each county is the sum of indicators of frontier status from 1790 to 1890. We exclude from our sample counties to the east of the 1790 east-to-west main frontier line for which we do not observe *total* frontier experience given the available data. Our baseline also excludes counties west of the 1890 east-to-west main frontier line, which for the most part had not yet begun to be settled by the time the frontier closed.[17]

Figure 3 shows the spatial distribution of TFE, measured in years and using the 100 km frontier cutoff, for the counties included in our baseline sample. Total frontier experience ranges from 0 to 63 years with

[15] Moreover, Appendix Figure A.2 suggests that the use of harmonized 2010 county boundaries has little effect on the location of the frontier lines relative to an approach based on contemporaneous county boundaries historically.

[16] This 500 km cutoff discards many contour lines but retains some large unconnected lines off of the main east-to-west frontier line, e.g., the ones spanning Maine in 1820 and Michigan in 1850. Like other cutoffs we are forced to specify, this one is arbitrary but also robust to other rules, including having no cutoff at all.

[17] We follow the same logic in robustness checks adding the West Coast frontier, discarding locations east of the 1890 west-to-east frontier line (but west of the 1890 east-to-west frontier line), and similarly when we extend the time frame to 1950.

a mean of 18.2 years and a standard deviation of 11.2 years. TFE exhibits considerable variation both across and within states and bands of latitude and longitude more generally. Within Illinois, for instance, Cass County has TFE of 10 years, while Johnson County stayed on the frontier for 32 years. While some of the greatest TFE is found in the South, it is important to note that we find similar long-run effects across different Census regions as discussed in Section 5.5.

4 The Distinctive Features of the Frontier

This section systematically documents the unique demographic features of the frontier and its higher levels of individualism. Historians and sociologists have devoted considerable attention to analyzing the demographics of frontier locations.[18] However, these studies usually focus on a specific place at a particular time, making it difficult to establish empirical regularities. In contrast, we characterize the demographics of the frontier using data from all Census rounds from 1790 to 1890. Moreover, we provide the first empirical validation of differential individualism on the frontier. We do this adopting a names-based measure of individualism suggested by social psychologists and using historical data on names recently made available by `ancestry.com`.

We document the distinctive features of the frontier using three complementary strategies. Section 4.1 offers a simple cross-sectional comparison that establishes the basic differences between frontier and non-frontier counties (east of the frontier line), and it shows that both remoteness and sparsity of frontier counties matter in explaining these differences. Section 4.2 validates these two defining features of frontier counties by identifying nonlinear relationships between these population traits and both density and distance. Finally, Section 4.3 exploits time-series variation comparing counties before and after exiting the frontier.

We focus on a set of demographic characteristics associated with the frontier in historical accounts. These include sex ratios, age distributions, literacy rates, and foreign-born population shares, all of which we draw from historical Census data in Haines and ICPSR (2010). With the exception of immigrant shares, we calculate all variables over the white population as this helps with consistency across time periods and ensures that results are not driven by racial composition. For each variable, we use data for all available years from 1790 to 1890. Further details can be found in Appendix C.

We measure individualism based on children's names. As suggested by social psychologists, the share of children with infrequent names reflects the prevalence of individualism, correlating strongly with other proxies.[19] We use full-count, historical Census data from several decades beginning in 1850 to measure the share of children in a given county under 10 years of age with infrequent names. Appendix C 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).[20] Our findings are robust (i) to defining infrequent names

[18] See, for example, Bowen (1978), Coombs (1993), Demos (1968), Easterlin, Alter and Condran (1978), Eblen (1965), Modell (1971), Moller (1945), Schaefer (1985), and Smith (1928).

[19] For example, Varnum and Kitayama (2011) shows a strong cross-country correlation between infrequent names and Hofstede's widely used index of individualism. Beck-Knudsen (2017) shows correlations 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), which includes divorce rates and the percentage of people living alone.

[20] As an example of infrequent names on the frontier from historical fiction, consider the Luckett family at the center of the

in various ways, as those outside the top 10, 25, 50, or 100 in the county, state, or Census division,[21] and (ii) to restricting the sample to children born in the U.S., with U.S.-born parents, or with U.S.-born grandparents.[22]

Overall, the results below provide new, systematic empirical support for historical narratives about the rugged individualists settling the American frontier. The three approaches offer a stark and consistent picture of the frontier being a qualitatively different type of society. In Section 5, we show that these historical differences have long-run implications for contemporary culture. Then, in Section 6, we explore how rugged individualism came to thrive on the frontier.

4.1 Demographics and Individualism on the Frontier: Basic Patterns

We begin by documenting the basic differences in demographics and individualism on the frontier using the following specification:

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

where x_{cdt} is one of the population traits of interest in county c in Census division d at time t, frontier_{ct} is time-varying 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 x outcomes.[23]

Across columns, we find evidence consistent with historical narratives as frontier populations tend to have significantly more (1) males, (2) prime-age adults, (3) illiterates, and (4) foreign-born. Frontier counties have 0.14 additional males for every female relative to non-frontier counties where the average sex ratio is 1.09. Age distributions also differ in frontier counties, where the share of prime-age adults (15–49 years old) in the population is 2.3 percentage points (p.p.) higher than in non-frontier counties, for which that share is around 46 percent. Additionally, literacy rates are significantly lower on the frontier, which is consistent with the negative selection of westward migrants during this period, in line with the safety valve theory. Finally, frontier counties have nearly 6 p.p. higher foreign-born population shares than the average non-frontier county where only 7 percent of residents are immigrants.

Furthermore, columns 5 and 6 show that individualism is more pervasive in frontier counties as reflected in the share of children with infrequent names. In frontier counties, around 2 p.p. more girls (boys) have infrequent names relative to the average non-frontier county with around 66 (58) percent of children having infrequent names. These measures capture the share of children aged 0–10 with names that are outside the top 10 most popular names in that decade's birth cohort within the Census division. We restrict here to white children with native-born parents, but results are similar using the other ancestry restrictions and measures of popularity noted above.

We further clarify the frontier differential by distinguishing the two defining attributes of frontier

celebrated trilogy, *The Awakening Land*, by novelist Conrad Richter. Members of this white, native-born family on the Ohio frontier in 1795 had first names that included Chancey, Wyitt, and Worth for boys and Ascha, Sayward, Sulie for girls.

[21] There are nine Census divisions, the boundaries of which are seen in Figure 3.

[22] A potentially confounding naming practice lies in the passing on of parental first names to children. Using data discussed in Section 6.1, we find that while only around three (five) percent of girls (boys) have such matronymics (patronymics), this practice is less common on the frontier and significantly so for boys. Choosing novel names for one's children rather than passing on one's own arguably reflects a desire to instill independence. As such, this finding is consistent with our broader claim that the higher prevalence of infrequent names on the frontier reflects greater individualism.

[23] These outcomes are all measured at the 2010 county-level but are not all available in every Census round, which explains why the sample varies across columns.

locations: (i) proximity to the frontier line and (ii) low population density. Panel B of Table 1 reports estimates of the following specification:

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

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.

Overall, the results in Panel B suggest that both isolation and sparsity contribute to the distinctive demographics and higher rates of individualism in frontier counties. With the exception of illiteracy, proximity to the frontier line and low population density are positively (and mostly significantly) correlated with the defining features of frontier counties highlighted in Panel A. This suggests that the frontier differentials are not merely an artifact of their low population density. Even within the set of low density counties, remote areas in proximity to the frontier line exhibit more distinctive population traits associated with narratives about frontier culture. Nor are these differences in frontier counties due to arbitrary density or proximity cutoffs, as we show next.

4.2 The Frontier Is Qualitatively Different: Semiparametric Evidence

In this section, we identify a set of nonlinear relationships that validate the notion that populations residing in frontier areas were qualitatively different. The evidence suggests that the cultural distinctiveness of the frontier seen in Table 1 is not merely explained by a linear trend as we move towards lower population density and greater remoteness. Rather, there are sharp differences in demographics and individualism that arise only at sufficiently low levels of density and proximity to the frontier line.

Using semiparametric regressions, we substantiate the seemingly arbitrary population density cutoff defining the frontier in historical sources. We estimate the following cross-sectional specification:

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

where $g(\cdot)$ is a nonlinear function recovered using the partially linear (Robinson, 1988) estimator, but the specification is otherwise similar to equation (1). While we estimate $g(\cdot)$ across all counties in the sample, we restrict the graphs presented in this section to counties with less than 50 people/mi^2 in order to focus on changes close to the assumed frontier threshold.[24]

Figure 4 provides a stark illustration of the qualitative differences in demographics and individualism in low density areas. Each graph shows the local linear regression function and 95 percent confidence interval around $g(\cdot)$. In graph (a), the sex ratio displays levels around 1.6 in the most sparsely populated counties and declines sharply as population density rises to 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 graph (b), the prime-age adult share declines sharply as we move towards density levels of 2–3 people/mi^2 and eventually levels off at around 0.37. Graphs (c) and (d) show similar downward-sloping albeit less starkly nonlinear $g(\cdot)$ curves. Finally, graphs (e) and (f) show that in the most sparsely populated counties, more than 65 (80) percent of boys (girls) have infrequent names based on the baseline definition

[24] In 1840, the midpoint of our study period, in the sample there are 437 counties in the range of 0 to 2 people/mi^2, 303 counties from 2–6 people/mi^2, 201 from 6–10, 540 from 10–50 people/mi^2, and 41 from 50–100 people/mi^2.

used in Table 1. However, this ratio decreases sharply as we move towards counties with more than 2–3 people/mi^2, leveling off thereafter. Appendix Figure B.1 provides analogous semiparametric evidence for proximity to the frontier, though these nonlinearities are less stark than those for density.

Overall, the results in Figure 4 point to a structural break in demographics and individualism at levels of population density consistent with the seemingly arbitrary cutoffs in the historical literature. In fact, using the Chow (1960) test, we can easily 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 Eblen, 1965; Porter et al., 1890), or above and below any cutoff in the 2–6 range. We can also be agnostic about the relevant cutoff, using the Zivot and Andrews (2002) test to identify unknown structural break points in each decade. In 1850, for example, we find a break in the sex ratio at 2.7 people/mi^2, the prime-age adult share at 2.0, and infrequent names for boys (girls) at 3.2 (2.6).[25]

4.3 Frontier Transitions: Event-Study Evidence

Building on the prior cross-sectional results, we now exploit time-series variation as counties transition from frontier conditions to established settlements. We estimate a panel event-study analogue to equation (1) based on the following form:

$$x_{cdt} = \alpha + \sum_{j=-20}^{40} \gamma_j \mathbf{1}(\text{years since exiting frontier} = \text{j}) + \theta_d + \theta_t + \varepsilon_{cdt}, \tag{4}$$

where the γ_j coefficients identify the average x for counties that have exited or will exit the frontier (as defined in Section 3.1) 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 5 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 around 0.25 higher sex ratios in the two decades prior to exiting the frontier whereas those decades thereafter exhibit slightly lower ratios but effectively level off by the second decade. Panel (b) provides similar evidence of convergence towards a lower prime-age adult share as counties exit the frontier. We find a similar albeit less pronounced transition path for illiteracy rates in panel (c). Finally, panel (d) shows that the foreign-born population share exhibits a steady and roughly linear decline along the frontier transition path.

Alongside these demographic changes, panels (e) and (f) demonstrates the a declining prevalence of infrequent children names as counties approach the decade in which they exit the frontier. Thereafter, we see naming patterns stabilize around that less individualistic equilibrium in which popular names becomes more common at the local level. Together, the results in Figure 5 provide consistent evidence of changes in demographic composition and the pervasiveness of individualism as frontier conditions subside. However, as we show next, the length of exposure to frontier culture in these early stages of settlement has profound implications for the culture observed today in these counties.

[25] While the results for illiteracy and foreign-born share are less stark, we strongly reject different slopes above and below 6 people/mi^2 based on the Chow (1960) test.

5 Long-Run Effects of Frontier Experience on Culture

In this section, we examine the long-run effects of frontier experience on culture and discuss their implications for modern political economy debates. We start by presenting our empirical framework, discuss key data sources, and then move to our main cross-county results, which establish the lasting effects of historical frontier experience on contemporary cultural outcomes.

Our motivation stems from the theories of cultural persistence discussed in Section 2.2. While the high levels of individualism on the frontier historically could have dissipated, it is also possible that frontier experience shaped the long-run evolution of local culture. The duration of exposure to frontier conditions determined the scope for the mechanisms through which rugged individualism came to thrive on the frontier, which we analyze in Section 6. In turn, the composition of cultural traits produced at early stages of settlement established the initial conditions for subsequent cultural evolution. In the presence of multiple equilibria and path dependence, these early stages would represent a critical juncture, and frontier experience could have a lasting legacy.

5.1 Estimating Equation

We relate historical frontier exposure to modern measures of individualism and preferences for individuals in county c. In particular, our main county-level, cross-sectional estimating equation is given by:

$$y_c = \alpha + \beta \text{ total frontier experience}_c + \mathbf{x}_c'\boldsymbol{\gamma} + \theta_{FE(c)} + \varepsilon_c, \tag{5}$$

where y_c is some long-run outcome capturing cultural traits (e.g., individualism or preferences for redistribution). Total frontier experience (TFE) is the amount of time *in decades* a given county remained on the frontier according to our baseline definition in Section 3.2. Our baseline sample, seen in Figure 3, is restricted to those counties for which the 1790–1890 period contains the whole extent of frontier experience.[26] In Section 5.5, we show robustness to alternative measures of TFE, including an expanded sample with the secondary West Coast frontier. The vector $\mathbf{x}_c$ contains a set of county characteristics as well as individual characteristics where possible. In baseline specifications, $\mathbf{x}_c$ includes latitude, longitude, county area, average rainfall and temperature, elevation, predetermined agricultural yield, and distance to rivers, lakes, and the coast.[27] The $\theta_{FE(c)}$ term includes state or Census division fixed effects (see Figure 3 for these boundaries). The coefficient β therefore identifies a local effect of TFE after accounting for geographic or agroclimatic factors that may have given rise to particular cultural norms and also determined the speed of frontier transitions historically. Following the approach suggested by Bester, Conley and Hansen (2011), standard errors in all specifications are clustered on 60-square-mile grid squares that completely cover counties in our sample.[28] When considering several correlated

[26] Recall that we exclude counties along the East Coast as we are not able to appropriately measure their *total* frontier experience given the lack of Census data before 1790.

[27] See Appendix C for details on each of these measures. All results are robust to the exclusion of controls, which often reduce the estimated coefficient on TFE. We retain them in the baseline specification to be conservative.

[28] Inference remains unchanged when using the computationally more intensive Conley (1999) spatial HAC estimator with a bandwidth of 300 km. We retain the arbitrary grid-cell approach as it is considerably easier to implement and less prone to instabilities, which becomes important as we move to an instrumental variables specification.

outcomes, we also estimate mean effects based on the Kling, Liebman and Katz (2007) approach.[29]

The main threat to causal identification of β lies in omitted variables. There are a number of factors that may be correlated with TFE as well as contemporary culture. We address this concern in three ways. First, we progressively add a richer set of controls to $\mathbf{x}_c$ aimed at removing as much of the confounded variation as possible. Second, we use the Oster (2016) approach to show that selection-on-unobservables is unlikely to drive our results. Third, in Section 5.6, we pursue an instrumental variables (IV) strategy that exploits variation in the speed of the frontier's westward movement induced by changes in the intensity of national immigration flows over time.

In Appendix B.2, we revisit the Illinois counties of Cass and Johnson noted in Section 3.2 to illustrate the link from historical TFE to contemporary differences in cultural and political economy outcomes of interest. This case study also clarifies how our empirical strategy, robustness checks, and IV approach help to isolate variation in TFE that is not confounded with other drivers of long-run cultural differences.

5.2 Data on Contemporary Culture and Political Economy

We measure contemporary culture and policy outcomes using several data sources. We draw upon 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 on political preferences and social norms (see, e.g., Acharya, Blackwell and Sen, 2016; Alesina and Giuliano, 2010). The three surveys ask different questions about similar underlying preferences. Their geographic coverage differs, and in the cases of GSS and ANES, it is quite narrow. Nevertheless, for our key outcomes we are able to identify relatively precise effects across the three surveys. Appendix C describes details of each survey, discusses advantages and disadvantages, and also provides definitions and coverage maps for all outcomes discussed in this section.

These sample coverage limitations of survey data do not extend to the measure of infrequent names from the post-1890 Census rounds or to two salient policy outcomes. First, we measure the Republican vote share in recent presidential elections using data from Leip's Atlas. Second, we take estimates of property tax rates from the American Community Survey in 2010 as prepared by the National Association of Home Builders. Together, these survey data and local voting records allow us to paint a rich picture of the persistent culture of individualism and small-government norms in areas exposed to the frontier for a longer time.

5.3 Total Frontier Experience and Persistent Individualism

We begin by documenting a long-run link between total frontier experience (TFE) and contemporary individualism. Nearly five decades after the closing of the frontier, infrequent children names are more pervasive in counties exposed to the frontier for a longer historical period. In Table 2, we report the effect of TFE (β from equation 5) on the share of white boys (Panel A) and girls (Panel B) age 0–10 given infrequent names in the 1930s. The data come from the full count 1940 Census and capture naming

[29] This effectively takes a weighted average of the estimates of β for each of K related outcome variables (placed on the same scale) with the weights equal to inverse sample standard deviation of that variable for a suitable control group. The choice of that control group affects the mean effect size but not significance. Results are very similar across three alternative approaches, including the few counties with zero frontier experience, those with less than a decade, and those with below the median.

choices multiple generations after counties exited the frontier. Our baseline measure of infrequent names considers those outside the top 10 within the county's Census division, and we report robustness to a battery of alternative measures below. On average, 72 percent of boys and 79 percent of girls have infrequent names with standard deviations of 0.07 and 0.04, respectively. In this and subsequent tables, we normalize non-binary dependent variables so that standard deviation effect sizes can be read directly from the coefficients.

The most demanding OLS specification in column 4 of Table 2 suggests that each additional decade of TFE is associated with 0.11 (0.16) standard deviations higher share of infrequent names for boys and girls, respectively. Comparing counties across the interquartile range of TFE (11 vs. 24 years) implies 1.4 (1.1) percent more boys (girls) with infrequent names. We build up to this results by progressively expanding the set of control variables, starting in column 1 with no controls. Columns 2 and 3 add Census division and state fixed effects, respectively, to rule out broad regional differences in TFE and culture. Our main specification in column 4 includes the full set of geographic and agroclimatic controls noted in Section 5.1. Comparing across columns 1–4, the coefficient remains relatively stable despite large changes in the R^2. This pattern is consistent with limited selection-on-unobservables according to the test parameter δ reported at the bottom of the table; Oster (2016) suggests that $|\delta| > 1.3$ implies limited scope for unobservables to explain observational results. In Section 5.5, we show that these findings hold up to a further set of (non-predetermined) controls for population density, diversity, and year of connection to the railroad.

Furthermore, these results are not sensitive to the particular measure of infrequent names or the national background of the parents assigning names. We document this robustness in Appendix B.3. After replicating the baseline result in column 1, Appendix Table B.2 restricts columns 2–4, respectively, to children with native-born fathers, native-born parents, native-born grandparents. Together, these help address concerns about immigrants having infrequent name preferences and being more likely to settle in frontier areas historically.[30] Column 5 defines infrequent names based on the top 10 names nationally while columns 6 and 7 do so at the state and county level, respectively. Columns 8 and 9 define infrequent names as those outside the top 25 and top 100 names, respectively. Finally, column 10 restricts to non-biblical names to account for the fact that religiosity may be confounded with TFE and naming choices. All measures reveal a similar effect of TFE.

Together with the findings in Section 4, these results suggest that infrequent name choices were not only more common in frontier areas historically but also are 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 (see Appendix Table B.4). This points to the persistence of the early frontier culture of individualism long after frontier conditions abated.[31]

In Appendix B.3, we further validate the link between TFE and individualism today using a well-suited measure from the ANES data in 1990. In particular, greater TFE is associated with respondents identifying more strongly with self-reliance as opposed to cooperative behaviors. We turn now to iden-

[30] The robustness to these alternative measures of ancestry is consistent with the rapid speed of assimilation to American name choices reported in Abramitzky, Boustan and Eriksson (2016). They show that the immigrant–native gap in Americanized name choice is halved within 20 years after parents arrive in the U.S.

[31] Ideally, we could carry these results through to the contemporary period, but, 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 as required for our empirical strategy.

tify the closely related link between frontier experience and opposition to government intervention.

5.4 Total Frontier Experience and Opposition to Redistribution and Regulation

This section uses a wide array of survey data to identify a long-run effect of total frontier experience (TFE) on contemporary preferences for small government. First, greater TFE is associated with opposition to redistribution, preferences for limited government intervention, and low levels of local taxation. Second, these differences in preferences translate into stronger Republican Party support today. Finally, we identify a link between TFE and opposition to government regulations surrounding issues that were salient in frontier culture historically. We view all of these outcomes as closely connected measures capturing the same underlying opposition to government intervention. In all cases, we report estimates of equation (5) controlling for individual demographics (age, age squared, gender, and race dummies), survey wave fixed effects, and the full set of geographic and agroclimatic characteristics in column 4 of Table 2. We continue reporting supportive Oster (2016) tests for selection-on-unobservables while other robustness checks and instrumental variables results are discussed in Sections 5.5 and 5.6.

Redistribution and Limited Government. Our first set of results in Table 3 shows that greater TFE is associated with stronger opposition to income redistribution today. In column 1, we draw upon ANES data from 1992 and 1996, which asks respondents whether they 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. Columns 2 provides complementary evidence, showing that each decade of TFE is associated with 0.7 p.p. higher support for cutting state spending on welfare as reported in the CCES. Although seemingly small relative to the mean of 40 percent, this effect size is roughly equivalent to the gap in opposition to such spending between people five years apart (with older respondents more in favor of welfare spending cuts, reflecting well-known cohort differences in conservatism). Following Alesina and La Ferrara (2005), column 3 uses a measure from the GSS indicating the intensity of preferences 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, an effect size akin to a ten year age gap among respondents.

Turning to broader measures of opposition to big government, columns 4 and 5 show that residents of areas with greater TFE exhibit 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 based on the principal components of a set of questions from the GSS on whether the government spends too much on an array of public goods and social transfers. In both cases, we find that individuals are significantly more opposed to high levels of government spending in areas with greater TFE. The Kling, Liebman and Katz (2007) mean effects analysis yields similar insights as the individual outcomes, e.g., combining the CCES measures in columns 2 and 4 into a single index yields a statistically significant effect of around 0.02.

Finally, column 6 of Table 3 shows that these reported preferences line up with actual policy dif-

ferences across counties. In particular, each decade of TFE is associated with around 3.4 percent lower reported property tax rates, which range from 0.1 to 2.9 percent across counties in our study. Given that much of the variation in tax rates lies across rather than within states, this is not a small effect.[32] In fact, it is roughly akin to the within-state difference between counties that are 10 percent more versus less aligned with the Republican party, another policy outcome we consider next.

Party Identification. We show in Table 4 that the persistent effects of TFE have strong implications for the growing strength of the Republican party, the contemporary bearer of conservative politics. Republican party platforms have been increasingly associated with broad opposition to government intervention and aversion to redistribution. For simplicity, we consider the average vote share across the five elections since 2000, but results are similar when pooling and including year fixed effects. Column 1 shows that each decade of TFE is associated with around a 2 p.p. greater Republican vote share relative to the mean of 60 percent. This effect size is plausible and in line with individual-level regressions using degree of stated support for the Republican party in CCES.[33] 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.

This average effect across the 2000s masks an interesting ratcheting up over time as seen in columns 2–6. An additional decade of TFE is associated with a significantly higher Republican vote share in each subsequent election, based on cross-equation tests of the relative effect sizes (i.e., coefficients divided by mean outcomes). Moreover, the effects increased over time, with the 2016 election exhibiting a uniquely large frontier legacy.

Putting these results together, the estimates in columns 7 and 8 provide marked evidence of the relatively larger shift towards the Republican party in areas with greater TFE. The average heartland county in our long-run analysis exhibits a 9 p.p. shift towards Republican candidates from 2000 to 2016. Each decade of TFE is associated with an additional 1.6 p.p. relative to that mean. Alternatively, comparing a county at the 25th percentile of TFE (11 years) to a county at the 75th percentile of TFE (24 years), implies an additional 2.2 p.p. Republican party shift. As a benchmark, Autor, Dorn, Hanson and Majlesi (2016) find that an interquartile shift in exposure to import competition from China implies a 1.7 p.p. Republican shift relative to a mean of -0.6 p.p. over the same period in their full-country sample of commuting zones.[34] A similarly large shift can be seen in column 8, which shows the frontier effect on the differential between 2012 and 2016. Overall, these findings offer suggestive evidence of a potential link between frontier culture and the growing strength of the Republican Party in swathes of

[32] More than half of the variation in tax rates lies across rather than within states. Including division instead of state fixed effects, we find an even larger effect size around 5.5 percent.

[33] 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. Using this individual-level measure of Republican party support is helpful as it allows for us to control for potential confounders of political preferences such as age and race that may also be correlated with TFE. The estimates imply that an additional decade of TFE is associated with around 4.5 percent greater intensity of strong Republican support. As a benchmark, consider that with each additional year of age, individuals are around 2 percent more likely to report strong Republican support. Further note that the rich set of county-level controls helps rule out concerns that TFE is simply capturing broad regional variation in political preferences. In Section 5.5, we discuss robustness to controlling for population density.

[34] In a more direct comparison, we use the original data from Autor, Dorn and Hanson (2013) and map the China shock to our sample of counties. Estimating a single equation with both measures, we find that the TFE effect is around one-quarter as large as the effect of the more proximate China shock.

the American heartland.[35]

In sum, the results in Table 4 suggest relatively more conservative political preferences in areas of the United States today that were part of the frontier for a longer period of time in the 19th century. While people identify with and vote Republican for many reasons, one consistent theme across Republican party platforms is the view that government should not be too heavily relied upon and hence government should be small. Of course, this notion is sometimes vague and selectively applied in political discourse. However, it does bear an interesting similarity to the individualistic norms described in historical accounts of the frontier. It is in this respect that we view these voting outcomes as reflecting preferences shaped by frontier culture.

Using the CCES, we provide further insight into why historical frontier experience is associated with increasing Republican Party support today. In Table 5, we relate TFE to measures of opposition to (1) the Affordable Care Act (ACA or "Obamacare"), (2) increases in the minimum wage, (3) the ban on assault rifles, and (4) Environmental Protection Agency (EPA) regulations on pollution. These policy issues have been sharply contentious, with the main political parties adopting increasingly polarized positions. Moreover, they can be connected to norms and beliefs pervasive on the frontier in terms of the link between effort and reward (ACA and minimum wage), the salience of manifest destiny (EPA regulations), and the right to bear arms (ban on assault rifles). The results in Table 5 show that places with greater TFE display significantly stronger opposition to each of these government regulations. Combining all estimates into a single index implies a mean effect size of around 0.04 that is significant at the 1 percent level. Moreover, as discussed in Appendix B.4, these CCES results are also robust to controlling for individual-level education, family income, and reported strength of party identification. Together, these findings shed new light on the local growth of Republican Party support.

Summary. Overall, the findings in Tables 3–5 paint a rich picture of the cultural and political legacy of historical frontier exposure. While we are unable to observe these modern outcomes throughout the 20th century, it is plausible that the early settlers left a lasting imprint on frontier locations and that the degree of that imprint increased with duration of exposure. As a summary takeaway, the Kling, Liebman and Katz (2007) mean effect on individualism (infrequent name share), conservative political preferences (Republican vote share) and policy (property tax rate) suggests that each decade of TFE is associated with roughly 0.15 standard deviations more frontier culture today.

5.5 Further Robustness Checks

In Table 6, we conduct three main robustness checks for the main outcomes of infrequent children's names, property tax rates, and the Republican vote share. First, we expand our analysis to include the secondary frontier on the West Coast, incorporating a number of Western counties that experienced frontier settlement in the late 19th century (Panel A). Second, we control for a quadratic in population density in 1890, birthplace diversity in 1870, and the year in which the county was first connected to a

[35]The augmented effect of frontier experience on the share of Republican votes in Donald Trump's election may be surprising, as some of his platform suggests stronger government intervention. However, a segment of the libertarian right makes a connection between Trump and frontier culture, in some cases directly likening some of his stances and peculiarities to frontier behavior and values (see, e.g., Mendenhall, 2017).

rail line (Panel B).[36] We restrict to covariates measured during the frontier era or near its end, but results also survive when controlling for contemporary density (see Appendix Table B.7). Although all such controls may be outcomes of frontier experience and hence "bad controls," robustness to their inclusion alleviates specific concerns about omitted variable bias (e.g., urban–rural status, early exposure to public investment). In Panel C, we measure TFE through 1950 instead of through 1890, the year in which the Census Bureau declared the frontier closed. In all cases, the key takeaways about the persistence of frontier culture remain.

Furthermore, Appendix B provides additional robustness checks. Appendix B.6 shows that the effects of TFE are similar across major Census regions, including the West. Appendix B.7 shows robustness to a battery of alternative ways of measuring TFE. Appendix B.8 reports the same set of checks in Table 6 for the other survey-based outcomes explored in Section 5.4.

5.6 Instrumental Variable Strategy

Despite the battery of robustness checks, omitted variables may undermine a causal interpretation of our core findings. Several factors may be correlated with TFE as well as contemporary culture. For example, areas with higher TFE may have had unfavorable natural endowments, high levels of conflict with Native Americans, or less exposure to beneficial public investments. Additionally, regulations facilitating access to land may have reduced TFE and caused favorable views of government intervention. This section introduces an instrumental variables (IV) strategy to address these types of concerns.

We exploit the timing of immigration inflows to isolate plausibly exogenous variation in TFE. The basic idea is that, while the process of westward movement may be thought of as a wave with time-varying speed, the time that a given location spends on the frontier depends on how fast the wave is going when passing through. To construct the instrument, we first identify the relevant time frame for each county. The beginning of that period is the first year in which the location is just west of the frontier, i.e., 110 km west of the frontier line. At this time, the county is still largely unsettled, and its local conditions do not affect the contemporaneous process of westward movement. We then consider the average annual immigrant inflow in the next 30 years, noting that nearly 85 percent of counties exit the frontier within 30 years.[37] These immigrants would contribute directly or indirectly to westward expansion by going west themselves or by exerting population pressure on the Eastern seaboard. Both forces would increase the speed of the frontier wave as it approaches and ultimately moves through the given county. Appendix B.9 elaborates on this logic, documenting the connection between the intensity of migrant inflows and the speed of westward expansion.

Table 7 presents IV estimates for the same summary set of measures capturing individualism and small government preferences as in Table 6. In Panel A, the IV estimates for these core outcomes reveal a large and statistically significant effect of TFE. The estimates are generally indistinguishable from the OLS estimates. Panel B shows similar results when controlling for population density, diversity, and

[36] It is important to note that less than 5 (25) percent of counties were connected to a rail line prior to entering (exiting) the frontier. Hence, for most counties, the railroad did not contribute to variation in TFE.

[37] Results are largely unchanged regardless of the window, ranging from 20 to 60 years, the latter spanning the range of TFE in our sample. We augment the widely-used Migration Policy Institute (2016) data on annual migration inflows (collected by the Office of Immigration Statistics) with data from Tucker (1843) for the pre-1820 period (see Appendix C).

year of first connection to the railroad as in Table 6. The instrument is strong across all specifications.[38]

Our approach can be thought of as a Bartik (1991)-inspired strategy, since we interact variation in local initial conditions—the year at which the relevant time frame starts for each county—with a national shock exogenous to local conditions. The exclusion restriction implies that (i) the position of the frontier line at a point in time is unaffected by the conditions of locations that are over 100 km farther west, and (ii) the intensity of immigrant flows to the U.S. is unaffected by the prevailing culture and conditions in any given frontier county. For instance, the instrument might not be valid if, at a given point in time, individualistic people in Europe knew that frontier locations were particularly fertile, and this increased the total flow of migrants. To address this concern, we aim to eliminate the potential pull factors and to isolate push factors unrelated to conditions on the frontier. Following Nunn, Qian and Sequeira (2017), we predict migrant outflows from Europe based on climate shocks. In Appendix Table B.11, we use these predictions to construct the instrument and find estimates that are very similar to the baseline IV results.

6 Mechanisms: The Roots of Frontier Culture

This section explores mechanisms through which the frontier shaped a persistent individualism and opposition to government intervention. Section 6.1 shows that there was significant selective migration of individualistic types to the frontier, though other factors complementing selection were arguably important as well.[39] We then document empirical patterns consistent with frontier conditions shaping people's values and behavior. Section 6.2 shows that individualism was differentially rewarded on the frontier, which may have fostered the prevalence of this trait over time. Section 6.3 shows that frontier conditions implied favorable prospects of upward mobility and a large perceived importance of effort in income generation, which would hone opposition to redistribution. In Appendix B.10, we consider another potential mechanism, the prevalence of infectious diseases, and show that the evidence does not support its relevance in explaining differential individualism on the frontier.

While the results presented here are meant to be suggestive rather than conclusive, they enrich our understanding of the long-run relationship between frontier experience and culture established in the previous section. We do not try to disentangle the quantitative importance of the different mechanisms, which we view as complementary and mutually reinforcing. For instance, a greater adaptive advantage of individualism on the frontier would induce more selective migration of individualists. And conversely, selective migration of individualists to the frontier would likely increase the advantage of this trait given that collectivistic norms would be of limited value in a society of individualists.[40]

The persistence of these mechanisms is due in part to their salience in local cultural development at an early stage of settlement. The fact that the frontier arrived in each county at this critical juncture has important implications not only for the cross-sectional distribution of individualism but also perhaps

[38] This can be seen in the large Kleibergen and Paap (2006) first-stage Wald statistics and graphically in Appendix Figure B.5. Moreover, note that the alternative measures of infrequent names are also robust to the IV approach (see Appendix Table B.3).

[39] While emphasizing the implications of selective migration of individualists for frontier culture, selection on other attributes might have been important as well. For example, selective migration of men, reflected in the high sex ratios seen in Section 4, might have also contributed to frontier culture as men are less inclined toward cooperation and interdependence (Cross and Madson, 1997; Gabriel and Gardner, 1999) and more opposed to redistribution (Ashok, Kuziemko and Washington, 2015).

[40] Moreover, given the selective migration of individualists to the frontier, the higher their differential advantage, the more favorable their wealth accumulation prospects. If this process leads to the establishment of local institutions characterized by low levels of redistribution, it would tend to further reinforce selective migration of individualists.

for its prevalence at the national level. Consider selective migration. At any given time, reshuffling individualists across locations cannot change the national prevalence of individualism. However, over the course of westward expansion, each location received an influx of individualistic types—in varying degrees, depending on TFE—at a time in which they could exert a strong influence on the formation of norms and institutions. This may well have increased individualism in all locations, and thus also in the U.S. as a whole.

6.1 Selective Migration

This section investigates the role of selective migration in explaining the differences in individualism between frontier and non-frontier locations. Using linked Census records, we show that selection was significant but may not fully explain the pervasiveness of individualism on the frontier.

Our basic strategy is to distinguish the relative contributions of early versus later frontier settlers to the overall differential in individualistic naming patterns. The key intuition is that because the latter have lived in the given location for a longer period of time, local conditions have a greater scope for affecting their preferences by the time we observe them. To estimate selection patterns, we need to track households across time. For this purpose, we use full count data from the 1870 and 1880 Censuses available from `ancestry.com` (including location, names, and demographics), and link individuals across rounds using an algorithm developed by Feigenbaum (2016).[41]

Table 8 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:

$$\text{child has infrequent name}_{ic,1880} = \alpha + \beta\, \text{frontier}_{c,1880} + \mathbf{x}_{ic}'\boldsymbol{\zeta} + \varepsilon_{ic,1880}, \tag{6}$$

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×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 Table 8 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.

[41] The base sample in 1880 is restricted to male household heads, native-born, aged 30–50, white, and who have at least one (biological) child aged 0–10. 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, Boustan and Eriksson, 2012; 26 percent in Collins and Wanamaker, 2017; and 22 percent in Long and Ferrie, 2013). Although the matching on names leaves scope for systematic sample selection, our core results in Tables 8 and 9 look very similar when reweighting the sample using the inverse probability of being linked across Census rounds (following Bailey, Cole and Massey, 2017). We estimate the underlying probabilities based on the same characteristics used for linking as well as an interaction of infrequent name status and frontier location in 1880. These latter controls help to re-balance the linked sample to account for differential missing-ness along our key variables of interest.

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.

In column 2, we capture 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. In Column 3, we decompose this 7.5 p.p. differential into differences coming from early versus later frontier settlers. We find that 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. Together, these results suggest that greater time on the frontier is associated with more individualistic naming patterns.

Overall, the findings in Table 8 provide suggestive evidence that selection was significant but may not fully explain the frontier differential in individualism. It is of course still possible that selective migration before 1870, which cannot be observed in this data, helps explain some of the differential.[42] For example, pre-1870 frontier migrants may be more individualistic than post-1870 frontier migrants. However, for this to be the case, the degree of differential selection would have to be nearly three times as large, which seems unlikely given that both groups of individuals migrated when the county was characterized by similar frontier conditions. This suggests that frontier conditions might reinforce and amplify the already individualistic tendencies of migrants. We now explore one potential explanation for why: individualism has differential returns on the frontier.

6.2 The Adaptive Advantage of Individualism

The opportunities and threats faced by frontier settlers may have favored individualism through an adaptive mechanism. Because people on the frontier primarily had to rely on themselves for protection and material progress, the independent, self-reliant types would arguably have fared better (Kitayama et al., 2010).[43] Moreover, frontier settlers faced novel agroclimatic conditions, and there was little local knowledge about how best to approach the harsh and unfamiliar setting (see Baltensperger, 1979; Libecap and Hansen, 2002; Shannon, 1977). Thus, adherence to old traditions and norms was less suited to the environment than non-conformism and innovation, two traits associated with individualism.[44]

This section presents evidence consistent with an adaptive advantage of individualism in frontier conditions. Using data from the linked Census sample, we show that households exhibiting greater individualism were more successful economically and more likely to stay in frontier locations.

First, we estimate the relationship between father i's economic status in county c in 1880, $y_{ic,1880}$, and

[42] Abramitzky, Boustan and Eriksson (2014) make a similar point about differential selection among early versus later immigrants to the U.S.

[43] This view can be framed within a notion of culture as decision-making rules-of-thumb used in uncertain environments, as proposed by evolutionary anthropologists (Boyd and Richerson, 1985, 2005). In their models, a process of natural selection governed by the payoffs from different rules-of-thumb determines which rule prevails.

[44] The connection between innovation and individualistic culture is discussed at length in Gorodnichenko and Roland (2012). In characterizing the traits of frontier populations, Turner (1920) himself mentions individualism along with the "coarseness and strength combined with acuteness and inquisitiveness," the "practical, inventive turn of mind, quick to find expedients," and the "masterful grasp of material things."

infrequent names according to the following difference-in-difference type specification:

$$\begin{aligned} y_{ic,1880} &= \alpha + \beta \text{ own infrequent name}_i + \eta(\text{own infrequent name}_i \times \text{frontier}_{c,1880}) \\ &\quad + \delta \text{ children infrequent name}_i + \zeta(\text{children infrequent name}_i \times \text{frontier}_{c,1880}) + \theta_c + \varepsilon_{ic}, \end{aligned} \tag{7}$$

where β captures the hedonic returns to the father's own infrequent name outside the frontier and η the differential effect on the frontier. At the same time, δ captures the association of infrequent name choices for children born during the 1870s and the father's economic well-being outside the frontier, and ζ the frontier differential. We restrict attention to white, native-born fathers that did not move between counties from 1870 to 1880 and had at least one child in 1880. Again, we define infrequent names as those outside the top 10 nationally, but other definitions yield similar results.

We measure economic status y_{ic} using data on occupation from the 1880 Census recorded in the linked sample. We consider the Duncan (1961) socioeconomic index (*sei*) and the occupational score provided by the NAPP (*occscore*). Both measures range from 0 to 100 and capture the income returns associated with occupations in the 1950 Census, and *sei* additionally captures education and occupational prestige. These measures are widely used in the economic history literature and capture broad differences in economic status across individuals (see Olivetti and Paserman, 2015, for a discussion). Finally, we cluster standard errors at the county level, and the county fixed effects, θ_c, account for all differences in outcomes common across individuals within the same county.

The estimates in Panel A of Table 9 provide evidence of differential returns to individualism on the frontier. Across all specifications, fathers with infrequent names outside the frontier exhibit socioeconomic status that is nearly 0.05 standard deviations lower than fathers with more common names. This is roughly the typical difference between a farmer and a blacksmith or a blacksmith and a carpenter.[45] This apparent economic penalty might be due to various types of discrimination or other mechanisms favoring conformity. However, this penalty is more than offset on the frontier where infrequent names exhibit a differentially positive association with economic status. We find a similar differential for infrequent names of children, which exhibit a positive correlation with father's status outside the frontier and an even stronger positive correlation on the frontier. These results, which hold for both *sei* and *occscore*, suggest that individualists are relatively better off on the frontier.

In Appendix Table B.13, we show that these differential hedonic returns arise not only for levels but also for changes in socioeconomic status. The NAPP linked sample for 1870–1880 allows us to investigate changes in occupational standing for 1 percent of the entire population (i.e., a small subset of all individuals in the prior analyses). The results show that fathers with infrequent names exhibit significantly faster growth in *sei* and *occscore* on the frontier but not outside the frontier.

In Panel B of Table 9, we provide a second piece of evidence consistent with an individualist advantage on the frontier. We estimate the following equation relating infrequent names to migration choices

[45] These results are robust to including a farmer dummy regressor. A list of the top 10 occupations with employment shares in frontier versus non-frontier counties can be found in Appendix Table B.1.

for household h living in frontier county c in 1870:

$$\begin{aligned} \text{outmigrate}_{hc} \quad = \quad & \alpha + \beta_f \text{ father has infrequent name}_h + \beta_m \text{ mother has infrequent name}_h \\ & + \eta \text{ any children with infrequent name}_h + \theta_c + \varepsilon_{hc}, \end{aligned} \tag{8}$$

where outmigrate_{hc} is a binary outcome indicating whether the household moved from a frontier county in 1870 to a non-frontier county by 1880. The key explanatory variables are defined as above, with the mother's infrequent name status defined similarly. The results suggest that, within a given frontier county, households in which fathers have infrequent names are around 4 percentage points less likely to leave the frontier by 1880. This is a sizable magnitude given that 40 percent of linked households in our sample left the frontier during this period.[46] We observe little relationship to mother's names, but households with children with infrequent names are also significantly less likely to leave the frontier.[47]

Overall, the findings in Table 9 suggest that inherited and revealed individualism are associated with a higher likelihood of socioeconomic success on the frontier. This may explain the self-selection of individualists to the frontier as well as their apparent increase in individualism after arrival.

The adaptive value of individualism may have favored its prevalence in the formation and evolution of local culture through different mechanisms. While differential fertility and mortality may have mattered, selective immigration and emigration likely played a larger role in driving compositional differences across counties. In addition, the differential returns to individualism may have boosted its prevalence among existing settlers through intergenerational cultural transmission.

6.3 Effort as the Road to Riches

This final section provides suggestive evidence that the opportunities and challenges on the frontier contributed to a culture of opposition to government intervention in society. The frontier's favorable prospects of upward mobility and a large perceived importance of effort in income generation may have fostered opposition to redistribution in particular, as suggested by the literature discussed in Section 2.2). This connection between the American frontier and theories of preferences for redistribution, hinted at by Alesina, Glaeser and Sacerdote (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 reading 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."

[46] In Appendix Table B.14, we show that this migration accounts for the lion share of departures from frontier counties during this period. The other direction of migration—from frontier counties to other frontier counties—is not associated with infrequent names.

[47] We acknowledge, though, that this latter result is less straightforward to interpret as we do not observe the timing of migration within the decade.

Numerous historical studies present stylized facts consistent with the frontier presenting both prospects for upward mobility and a large perceived importance of effort.[48]. As summarized by Stewart (2006), the frontier was "a place of economic opportunity," where settlers had low levels of initial wealth (consistent with higher rates of illiteracy documented in Section 4), but land-holding was widespread and rates of wealth accumulation were high, especially for early settlers. Indeed, as shown in Appendix Figure B.2, historical census data on land-holdings is consistent with the idea that frontier locations offered a more level playing field. There, we show that land inequality (captured by the Gini coefficient) was significantly lower on the frontier, but this difference dissipates over time as counties exited the frontier and the usual forces giving rise to inequality took hold.

In a more novel set of results, we show in Table 10 that intergenerational mobility was relatively higher on the frontier. Adopting a standard specification, we relate father's socioeconomic status (*sei* or *occscore*) in 1870 to the status of their children in 1880 who have become heads of their own households. Outside the frontier, the intergenerational persistence elasticity is nearly 0.38 for *sei* (0.45 for *occscore*), which is broadly similar to estimates from prior work in the United States (e.g., Feigenbaum, 2016, forthcoming; Long and Ferrie, 2013). However, persistence is significantly lower on the frontier, falling by as much as one-half to two-thirds.[49]

The greater intergenerational mobility on the frontier is consistent with two salient perceptions of the frontier economy: (i) a relatively high importance of effort as opposed to luck (of being born into a given class), and (ii) equality of opportunity offering a level playing field. While we acknowledge other possible interpretations, these results suggest a relatively limited role for inherited social class as a key determinant of income and wealth generation. Together with the selection and cultivation of individualism, these conditions plausibly contributed to the origins and persistence of frontier culture.

7 Discussion

This paper provides new evidence on the historical and long-run effects of the American frontier on culture at the subnational level. We show that frontier locations exhibited strikingly different demographics and a higher prevalence of individualism as reflected in name choices for children. We then identify a long-run effect whereby the initial culture of rugged individualism survived long after frontier conditions subsided. Today, counties that remained on the frontier for a longer period historically exhibit stronger opposition to government intervention in the form of redistribution, taxation and various regulations. Guided by historical record and insights from social psychology and political economy, we offer empirical evidence on the origins of frontier culture, identifying the importance of selective migration, the adaptive advantage of self-reliance, and expectations of high income growth through effort.

Our findings shed new light on the historical roots and persistence of rugged individualism in the United States both in terms of results and methods. We provide some of the first systematic evidence on a prominent theme in American history and lend credence to Turner's famous thesis. Our method for locating and tracking the frontier historically should prove useful in other attempts to understand the legacy of the frontier.

[48] See Curti (1959), Galenson and Pope (1989), Gregson (1996), Kearl, Pope and Wimmer (1980), and Schaefer (1987).
[49] We find similar results when estimating in levels rather than logs, which retains zero value status for children and/or fathers.

Our findings have suggestive implications about the sharp contrast between the U.S. and Europe in terms of preferences for redistribution and redistributive policies, a recurring topic in the literature (e.g., Alesina, Glaeser and Sacerdote, 2001; Alesina and Glaeser, 2004; Alesina, Cozzi and Mantovan, 2012; Benabou and Tirole, 2006). According to Turner, initially "the Atlantic coast ... was the frontier of Europe," but "the advance of the frontier has meant a steady movement away from the influence of Europe," and ultimately "moving westward, the frontier became more and more American." Intuitively, 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 results also offer new perspective on contemporary political debates. The deep roots of opposition to redistribution in the United States may explain why their levels have remained uniquely high even in the face of sharply rising inequality. Our findings suggest that expressions of stark opposition to government intervention amidst growing political polarization reflect not only a reaction to current events but also a rekindling of long-standing elements of American culture.

Nevertheless, we acknowledge that our study is designed to identify the effects of frontier experience at the local level. Hence, the implications based on extrapolation to higher levels of social and political organization, no matter how suggestive, remain speculative. Understanding the broader economic implications of frontier culture as well as its coevolution with institutions remain important topics for future research.

The relevance of our findings in other settings is also an important topic for future work. The national institutions of the United States, which ensured relatively high levels of geographic mobility, access to land, and security of property rights, may have been preconditions for the operation of the mechanisms we emphasize. In their analysis of how historical frontiers affected later democratic quality across countries in the Americas, García-Jimeno and Robinson (2011) show that the positive effects of frontiers depend on the quality of initial institutions. Thus, our findings may be less relevant in the context of other countries such as Argentina or Russia that also underwent massive territorial expansion in their early history but were ruled by elites that built very different institutions.

While some unique features of American history may be at the roots of rugged individualism, its persistence contains broader insights about the dynamics of cultural formation. In our view, this persistence points to the relevance of critical junctures. The frontier only represented a small fraction of the country's territory and population at any given time; most places were on the frontier for a few decades or less, and that was more than a century ago. However, its influence was ubiquitous and did not fade away. Over the course of America's westward expansion, the frontier covered most locations for at least a few years in their early history. Since this early stage of settlement was formative for local culture, the frontier was bound to leave a long-lasting imprint.

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.

_ , _ , and _ , "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration," *Journal of Political Economy*, 2014, *122* (3), 467–506.

_ , _ , and _ , "Cultural Assimilation during the Age of Mass Migration," *NBER Working Paper 22381*, 2016, (22381).

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

Alesina, A. and E. L. Glaeser, *Fighting Poverty in the US and Europe: A World of Difference*, Oxford, U.K.: Oxford University Press, 2004.

_ 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*, Vol. 1, North Holland,

_ , 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.

_ , G. Cozzi, and N. Mantovan, "The Evolution of Ideology, Fairness and Redistribution," *The Economic Journal*, 2012, *122* (565), 1244–1261.

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

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

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.

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.

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," 2016. Mimeo, MIT.

Bailey, M., C. Cole, and C. Massey, "Representativeness and False Links in the 1850-1930 IPUMS Linked Representative Historical Samples," Technical Report, Working Paper 2017.

Baltensperger, B. H, "Agricultural Adjustments to Great Plains Drought: The Republican River Valley, 1870-1900," in B. W. Blouet and F. C. Luebke, eds., *The Great Plains: Environment and Culture*, Lincoln: University of Nebraska Press, 1979, pp. 43–59.

Bartik, T. J., *Who Benefits from State and Local Economic Development Policies?* 1991.

Beck-Knudsen, A. S., "Historical Individualism: Selective Migration and Cultural Persistence," 2017.

Benabou, R. and E. A. Ok, "Social Mobility and the Demand for Redistribution: The Poum Hypothesis," *The Quarterly Journal of Economics*, 2001, *116* (2), 447–487.

_ and J. Tirole, "Belief in a Just World and Redistributive Politics," *The Quarterly Journal of Economics*, 2006, *121* (2), 699–746.

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

Bisin, A. and T. Verdier, "Work Ethic and Redistribution: A Cultural Transmission Model of the Welfare State," *Unpublished Manuscript, New York University*, 2005.

_ **and** _ , "The Economics of Cultural Transmission and Socialization," in J. Benhabib, A. Bisin, and M. O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, North-Holland, 2010, pp. 339–416.

_ **and** _ , "On the Joint Evolution of Culture and Institutions," Technical Report, National Bureau of Economic Research 2017.

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

Bowen, W. A., *The Willamette Valley: Migration and Settlement on the Oregon Frontier*, Seattle: University of Washington Press, 1978.

Boyd, R. and P. J. Richerson, *Culture and the Evolutionary Process*, Chicago: University of Chicago Press, 1985.

_ **and P. J Richerson**, *The Origin and Evolution of Cultures*, New York: Oxford University Press, 2005.

Celinska, K., "Individualism and Collectivism in America: The Case of Gun Ownership and Attitudes Toward Gun Control," *Sociological Perspectives*, 2007, *50* (2), 229–247.

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

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.

Coombs, J., "Frontier Patterns of Marriage, Family, and Ethnicity: Central Wisconsin in the 1880s," *Journal of Family History*, 1993, *18* (3), 265–282.

Couttenier, M., P. Grosjean, and M. Sangnier, "The Wild West is Wild: The Homicide Resource Curse," *Journal of the European Economic Association*, 2016.

Cross, S. E. and L. Madson, "Models of the Self: Self-Construals and Gender," *Psychological Bulletin*, 1997, *122* (1), 5–37.

Curti, M. E., *The Making of an American Community: A Case Study of Democracy in a Frontier County*, Stanford, CA: Stanford University Press, 1959.

Demos, J., "Families in Colonial Bristol, Rhode Island: An Exercise in Historical Demography," *The William and Mary Quarterly: A Magazine of Early American History*, 1968, *25* (1), 40–57.

Duncan, O. D., "A Socioeconomic Index for All Occupations," *Class: Critical Concepts*, 1961, *1*, 388–426.

Durkheim, E., *De la division du travail social: étude sur l'organisation des sociétés supérieures*, Alcan, 1893.

Easterlin, R. A., G. Alter, and G. A. Condran, "Farms and Farm Families in Old and New Areas: The Northern States in 1860," *Quantitative Studies in History*, 1978, pp. 22–84.

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

Feigenbaum, J. J., "Automated Census Record Linking: A Machine Learning Approach," *Unpublished Manuscript*, 2016.

_ , "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940," *Economic Journal*, forthcoming.

Fernández, R., "Does Culture Matter?," in J. Benhabib, A. Bisin, and M. O. Jackson, eds., *J. Benhabib, A. Bisin, and M. O. Jackson, eds.*, Vol. 1, North-Holland, 2010, pp. 481–510.

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.

Fryer, R. G. and S. D. Levitt, "The Causes and Consequences of Distinctively Black Names," *The Quarterly Journal of Economics*, 2004, *119* (3), 767–805.

Gabriel, S. and W. L. Gardner, "Are There" His" and" Hers" Types of Interdependence? The Implications of Gender Differences in Collective versus Relational Interdependence for Affect, Behavior, and Cognition.," *Journal of Personality and Social Psychology*, 1999, *77* (3), 642–655.

Galenson, D. W. and C. L. Pope, "Economic and Geographic Mobility on the Farming Frontier: Evidence from Appanoose County, Iowa, 1850–1870," *The Journal of Economic History*, 1989, *49* (3), 635–655.

Galor, O. and Ö. Özak, "The Agricultural Origins of Time Preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

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.

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

Gershman, B., "Long-Run Development and the New Cultural Economics," in M. Cervelatti and U. Sunde, eds., *Demographic Change and Long-Run Development*, Cambridge, MA: MIT Press, 2017, pp. 221–261.

Giuliano, P. and N. Nunn, "Understanding Cultural Persistence and Change," Technical Report, Harvard University, mimeo 2017.

Goodrich, C. and S. Davison, "The Wage-Earner in the Westward Movement I," *Political Science Quarterly*, 1935, *50* (2), 161–185.

_ **and** _ , "The Wage-Earner in the Westward Movement II," *Political Science Quarterly*, 1936, *51* (1), 61–116.

Gorodnichenko, Y. and G. Roland, "Which Dimensions of Culture Matter for Long-Run Growth?," *American Economic Review: Papers & Proceedings*, 2011, *101* (3), 492–498.

_ **and** _ , "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 Democratization," *NBER Working Paper 21117*, 2015.

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

Gregson, M. E., "Wealth Accumulation and Distribution in the Midwest in the Late Nineteenth Century," *Explorations in Economic History*, 1996, *33* (4), 524–538.

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.

Guiso, L., P. Sapienza, and L. Zingales, "Social Capital as Good Culture," *Journal of the European Economic Association*, 2008, *6* (2-3), 295–320.

Guiso, L, P. Sapienza, and L. Zingales, "Long-term Persistence," *Journal of the European Economic Association*, 2016, *14* (6), 1401–1436.

Gureckis, T. M. and R. L. Goldstone, "How You Named Your Child: Understanding the Relationship Between Individual Decision Making and Collective Outcomes," *Topics in Cognitive Science*, 2009, *1* (4), 651–674.

Hahn, M. W. and R. A. Bentley, "Drift as a Mechanism for Cultural Change: An Example from Baby Names," *Proceedings of the Royal Society of London B: Biological Sciences*, 2003, *270* (Suppl 1), S120–S123.

Haines, M. R. and ICPSR, *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*, Inter-University Consortium for Political and Social Research, 2010.

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., *Culture's Consequences: International Differences in Work-Related Values*, London: Sage Publications, 1980.

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

Hoover, H., *The New Day: Campaign Speeches of Herbert Hoover, 1928*, Stanford, CA: Stanford University Press, 1929.

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.

Jokela, M., "Personality Predicts Migration Within and Between US states," *Journal of Research in Personality*, 2009, *43* (1), 79–83.

Juricek, J. T., "American Usage of the Word "Frontier" from Colonial Times to Frederick Jackson Turner," *Proceedings of the American Philosophical Society*, 1966, *110* (1), 10–34.

Kashima, E. S. and Y. Kashima, "Culture and Language: The Case of Cultural Dimensions and Personal Pronoun Use," *Journal of Cross-Cultural Psychology*, 1998, *29* (3), 461–486.

Kearl, J. R., C. L. Pope, and L. T. Wimmer, "Household Wealth in a Settlement Economy: Utah, 1850–1870," *The Journal of Economic History*, 1980, *40* (3), 477–496.

Kim, H. and H. R. Markus, "Deviance or Uniqueness, Harmony or Conformity? A Cultural Analysis," *Journal of Personality and Social Psychology*, 1999, *77* (4), 785–800.

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.

Kleibergen, F. and R. Paap, "Generalized Reduced Rank Tests using the Singular Value Decomposition," *Journal of Econometrics*, 2006, *133* (1), 97–126.

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

Libecap, G. D. and Z. K. Hansen, ""Rain Follows the Plow" and Dryfarming Doctrine: The Climate Information Problem and Homestead Failure in the Upper Great Plains, 1890–1925," *The Journal of Economic History*, 2002, *62* (1), 86–120.

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.

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

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.

Mendenhall, A., "Donald Trump The Cowboy," *The Daily Caller*, January 2017.

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

Minnesota Population Center, "National Historical Geographic Information System: Version 2.0.," *Minneapolis, MN: University of Minnesota*, 2011.

Modell, J., "Family and Fertility on the Indiana Frontier, 1820," *American Quarterly*, 1971, *23* (5), 615–634.

Moller, H., "Sex Composition and Correlated Culture Patterns of Colonial America," *The William and Mary Quarterly: A Magazine of Early American History, and Culture*, 1945, *2* (2).

Mood, F., "The Concept of the Frontier, 1871-1898: Comments on a Select List of Source Documents," *Agricultural History*, 1945, *19* (1), 24–30.

Munshi, K., "Community networks and the process of development," *The Journal of Economic Perspectives*, 2014, *28* (4), 49–76.

Nelson, L. D. and J. P. Simmons, "Moniker maladies: When names sabotage success," *Psychological Science*, 2007, *18* (12), 1106–1112.

Nisbett, R. E. and D. Cohen, "Culture of Honor: The Psychology of Violence in the South," 1996.

Nunn, N. and L. Wantchekon, "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 2011, *101* (7), 3221–3252.

_ , N. Qian, and S. Sequeira, "Migrants and the Making of America," *Unpublished Manuscript*, 2017.

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*.

Olivetti, C. and M. D. Paserman, "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850–1940," *American Economic Review*, 2015, *105* (8), 2695–2724.

Oster, E., "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*, 2016, *0* (0), 1–18.

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.

Paul Samuelson, "Modern Economic Realities and Individualism," in G. H. Mills, ed., *Innocence and Power: Individualism in Twentieth-century America*, Austin: University of Texas Press, 1965, pp. 51–71.

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

Plaut, V. C., H. R. Markus, and M. E. Lachman, "Place Matters: Consensual Features and Regional Variation in American Well-Being and Self," *Journal of Personality and Social Psychology*, 2002, *83* (1), 160–184.

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.

Quattrociocchi, Jeff, "Culture and Redistribution," *PhD Dissertation, York University*, 2014.

Restrepo, P., "The Mounties and the Origins of Peace in the Canadian Prairies," 2015.

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

Schaefer, D. F., "A Statistical Profile of Frontier and New South Migration: 1850-1860," *Agricultural History*, 1985, *59* (4), 563–578.

_ , "A Model of Migration and Wealth Accumulation: Farmers at the Antebellum Southern Frontier," *Explorations in Economic History*, 1987, *24* (2), 130–157.

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

Smith, G.H., "The Populating of Wisconsin," *Geographical Review*, 1928, *18* (3), 402–421.

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

Spolaore, E. and R. Wacziarg, "How Deep are the Roots of Economic Development?," *Journal of Economic Literature*, 2013, *51* (2), 325–369.

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.

Tabellini, G., "Institutions and Culture," *Journal of the European Economic Association*, 2008, *6* (2-3), 255–294.

_ , "The Scope of Cooperation: Values and Incentives," *The Quarterly Journal of Economics*, 2008, *123* (3), 905–950.

_ , "Culture and Institutions: Economic Development in the Regions of Europe," *Journal of the European Economic Association*, 2010, *8* (4), 677–716.

Tella, R. Di, J. Dubra, and R. MacCulloch, "A Resource Belief-Curse? Oil and Individualism," *NBER Working Paper 14556*, 2008.

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.

Tönnies, F., *Gemeinschaft und Gesellschaft: Abhandlung des Communismus und des Socialismus als empirischer Culturformen*, Fues, 1887.

Triandis, H., "Collectivism v. Individualism: A Reconceptualisation of a Basic Concept in Cross-Cultural Social Psychology," in G. K. Verma and C. Bagley, eds., *Cross-Cultural Studies of Personality, Attitudes and Cognition*, London: Palgrave Macmillan UK, 1988, pp. 60–95.

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

_ , "Individualism and Collectivism: Past, Present, and Future," 2001, pp. 35–50.

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.

Turner, F. J., "The Frontier in American History (New York, 1920)," *The Significance of Sections in American History*, 1920.

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.

Voigtländer, N. and H.J. Voth, "Persecution Perpetuated: the Medieval Origins of Anti-Semitic Violence in Nazi Germany," *The Quarterly Journal of Economics*, 2012, *127* (3), 1339–1392.

Weber, Max, *Die protestantische Ethik und der Geist des Kapitalismus* 1905.

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

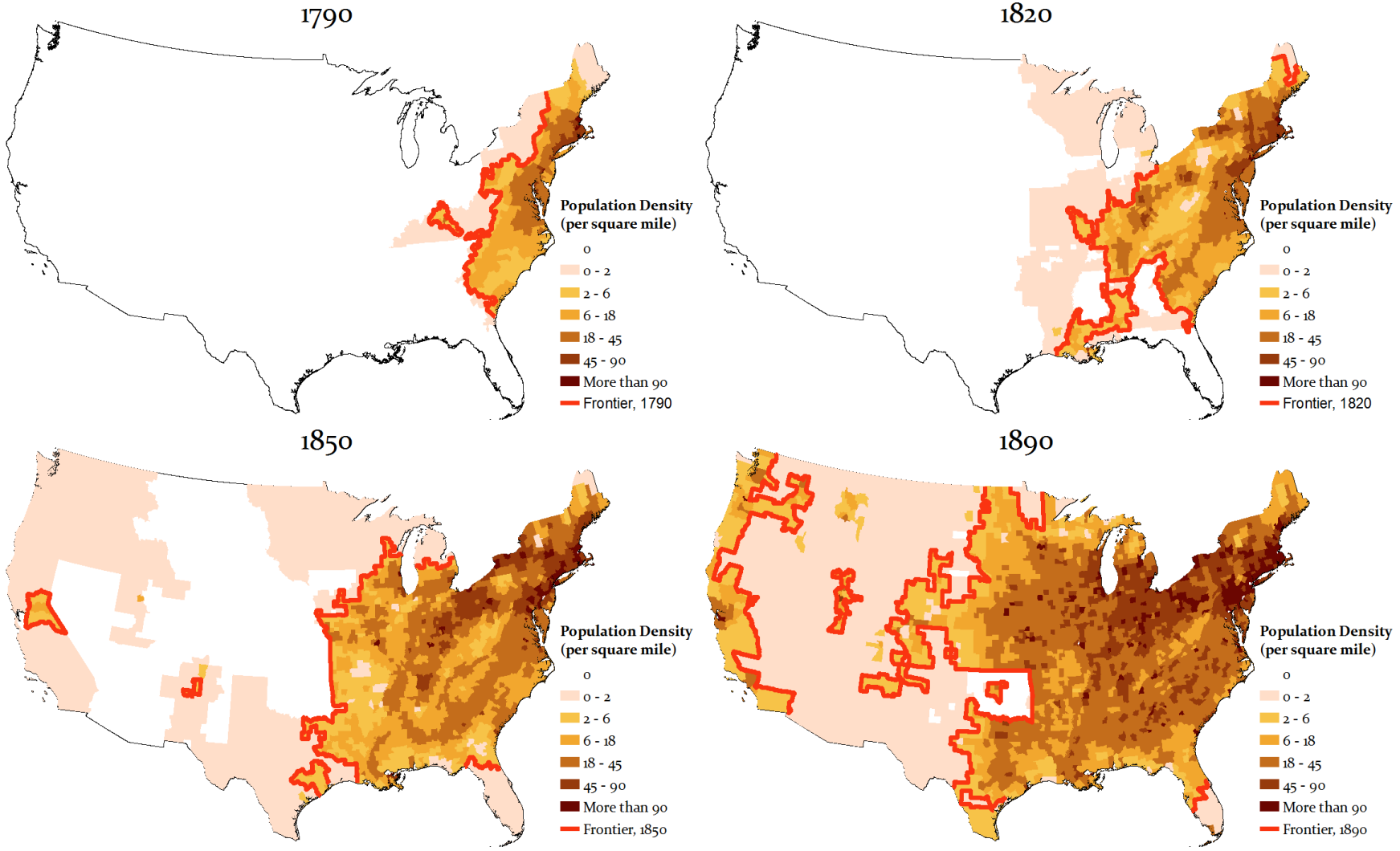
Yadin, Z. S., "Analyzing the Patient's First Name in the Search for Identity," *Contemporary Psychoanalysis*, 2016, *52* (4), 547–577.

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.

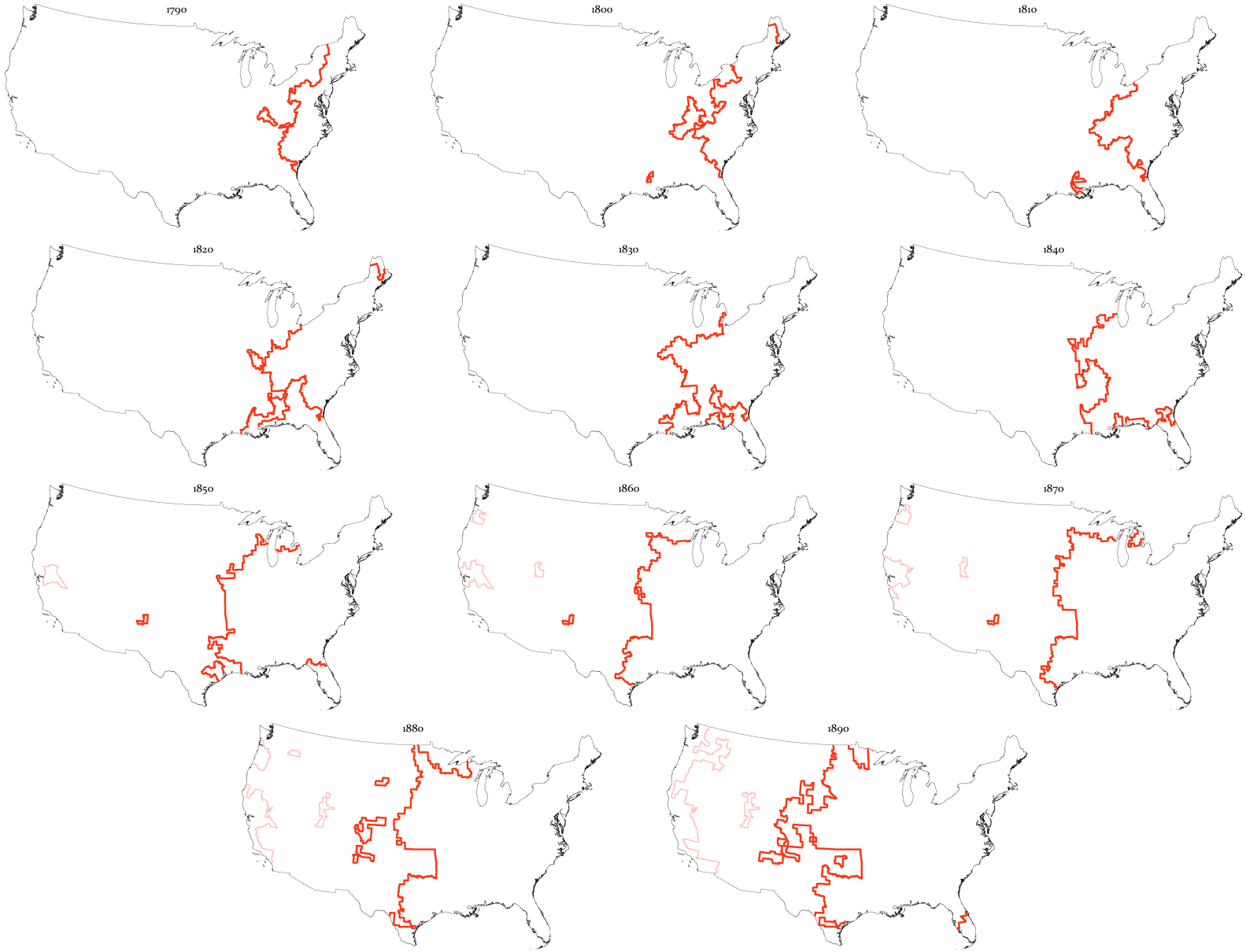
Figures

Figure 1: Population Density and the Frontier for Selected Years



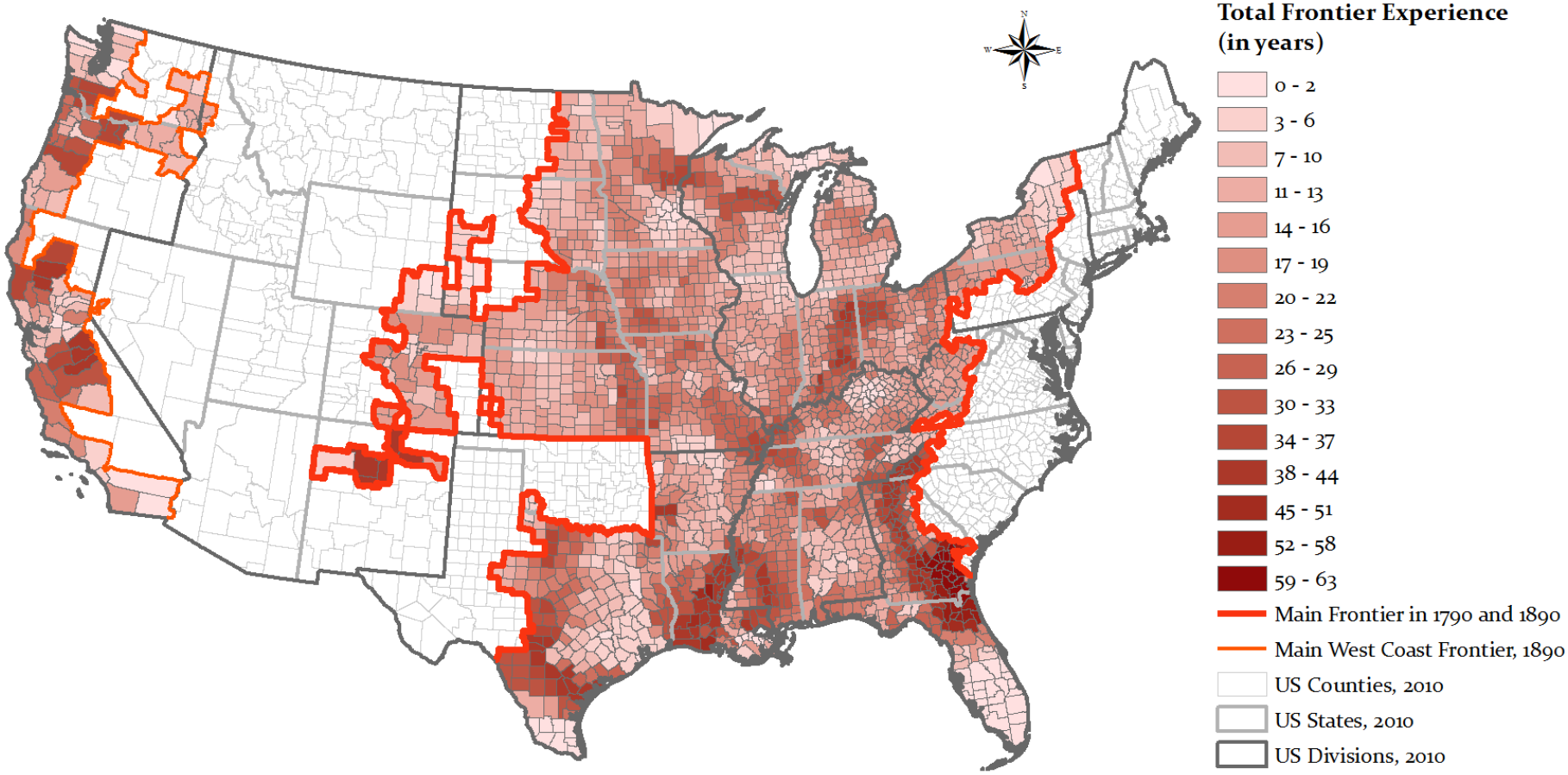
Notes: Based on county-level data from National Historical Geographic Information System: Version 11.0. Population is allocated across years and counties based on the procedure described in Section 3.1, which builds upon Hornbeck (2010). The red frontier line is based on the algorithm described in Section 3.1 and Appendix A.

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



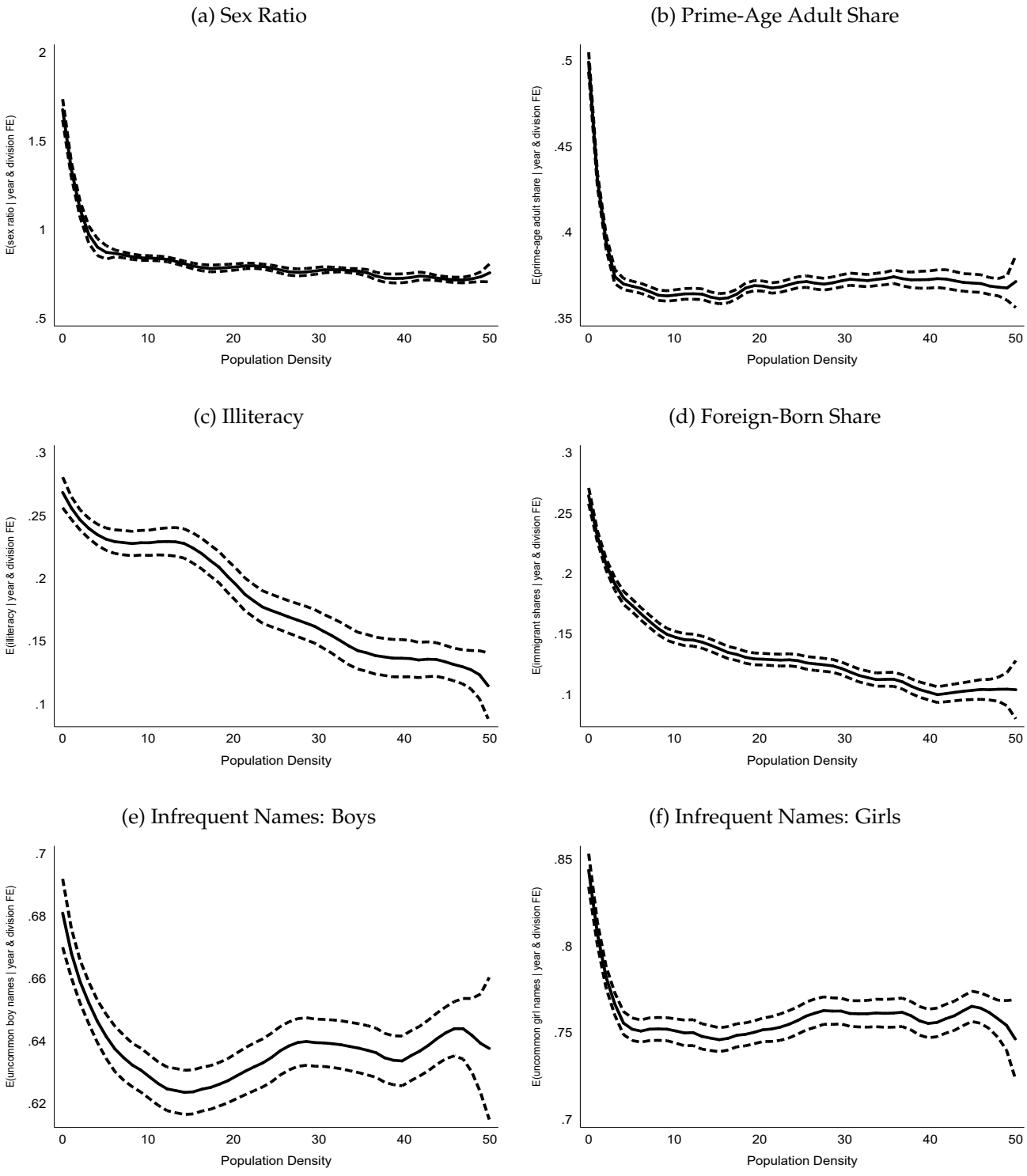
Notes: Based on county-level data from National Historical Geographic Information System: Version 11.0. The frontier lines demarcate the contour of counties with population density below and above 2 people per square mile. The dark red lines correspond to the main frontier lines emerging form east-to-west 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" lines in the interior. Full details on the frontier line algorithm can be found in Appendix A.

Figure 3: Total Frontier Experience, 1790 to 1890



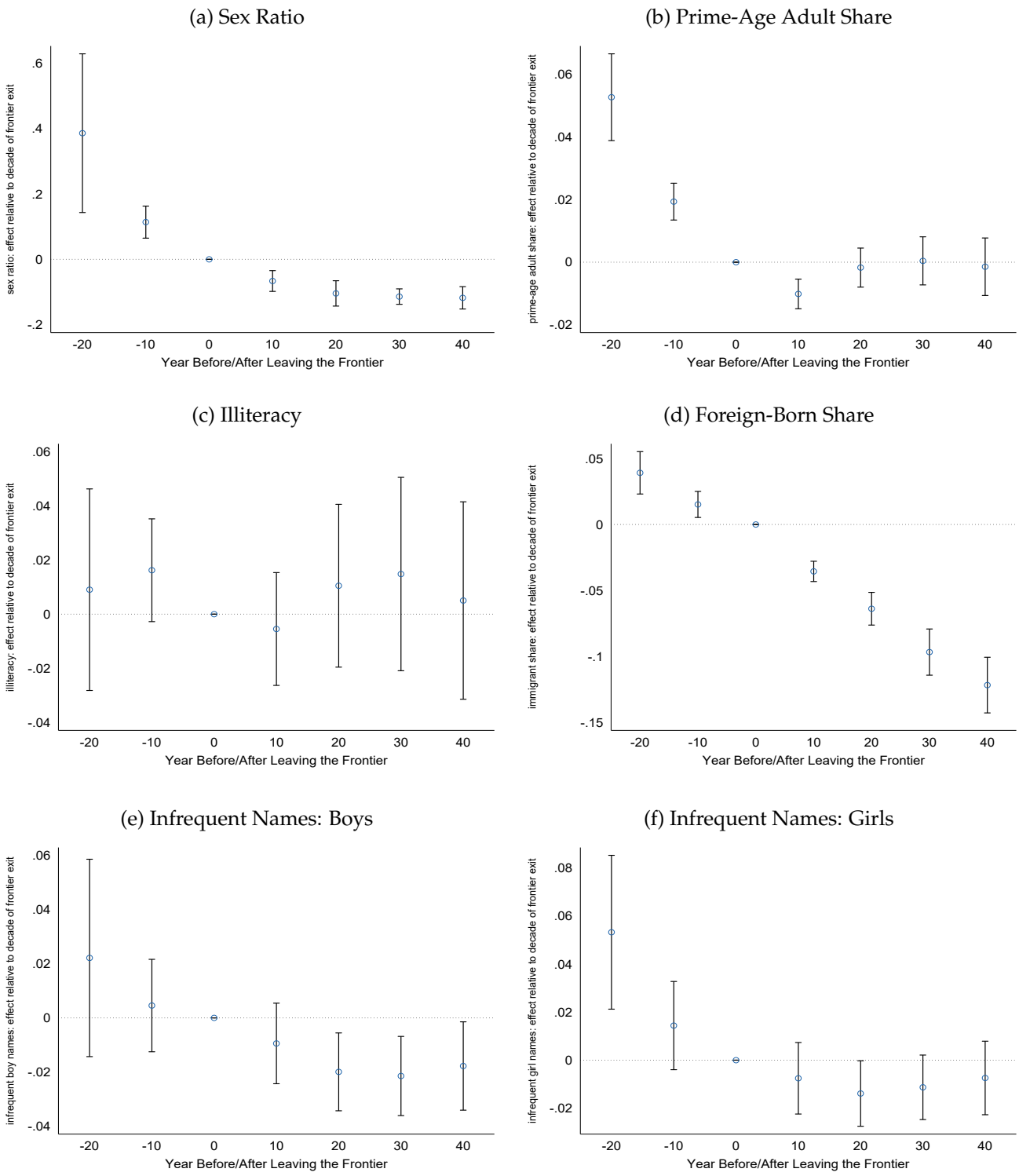
Notes: Based on county-level data from National Historical Geographic Information System: Version 11.0. Total frontier experience is the total number of years the county was within 100 km of the frontier line and its population density was below 6 people per square mile, between 1790–1890.

Figure 4: Distribution of Demographics and Individualism by Population Density



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) *Illiteracy* is the illiteracy rate for whites aged 20 or older. (d) *Foreign-Born Share* is the ratio of foreign-born persons over total population. (e) and (f) *Infrequent Names* are the share of boys and girls, respectively, 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.

Figure 5: Demographics and Individualism Along the Transition out of the Frontier
Event Study Estimates with Respect to Year of Exiting the Frontier



Notes: This figure plots coefficients from the event study regressions of the form in equation (4) for each of the outcomes in the semiparametric regressions presented in Figure 4. 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 5.1.

Tables

Table 1: Demographics and Individualism on the Frontier

Dependent Variable:	Male/Female Ratio	Prime-Age Adult Share	Illiterate Share	Foreign-Born Share	Share of Infrequent Child Girl Names	Share of Infrequent Child Boy Names
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Baseline Frontier Definition: Low Density and Proximity to Frontier Line					
frontier county	0.140***	0.023***	0.020*	0.056***	0.020***	0.020***
	(0.018)	(0.004)	(0.012)	(0.008)	(0.007)	(0.007)
Mean Dep. Var. in Non-Frontier Counties	1.09	0.46	0.17	0.07	0.66	0.58
Number of County-Years	9,641	5,508	2,779	10,840	6,884	6,885
R^2	0.06	0.19	0.17	0.34	0.49	0.13
	Panel B: Distinguishing Low Density and Proximity to Frontier Line					
near frontier line	0.086***	0.018***	-0.022*	0.053***	0.015**	0.012*
	(0.012)	(0.003)	(0.012)	(0.009)	(0.006)	(0.007)
low population density	0.095***	0.006*	0.041***	0.031***	0.006	0.011
	(0.011)	(0.003)	(0.010)	(0.008)	(0.006)	(0.007)
Mean Dep. Var. in Non-Frontier Counties	1.09	0.46	0.17	0.07	0.66	0.58
Number of County-Years	9,641	5,508	2,779	10,840	6,884	6,885
R^2	0.07	0.19	0.18	0.36	0.49	0.13

Notes: This table reports OLS estimates of equations (1) and (2) in Panels A and B, respectively. The dependent variables and sample are the same as in Figures 5 and 4. 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 boys and girls, respectively, 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. *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. All regressions include year and Census division FE. Standard errors are clustered using the grid cell approach of Bester, Conley and Hansen (2011) as described in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 2: Total Frontier Experience and 20th Century Individualism

	Dep. Var.: Infrequent Names Among White Children Aged 0-10 with Native-Born Parents, 1940 Census			
	(1)	(2)	(3)	(4)
	Panel A: Boys with Infrequent Names (normalized share)			
total frontier experience	0.220***	0.134***	0.112***	0.112***
	(0.034)	(0.025)	(0.026)	(0.022)
Oster δ for $\beta = 0$		3.36	1.75	1.70
Number of Counties	2,036	2,036	2,036	2,036
R^2	0.06	0.48	0.54	0.61
	Panel B: Girls with Infrequent Names (normalized share)			
total frontier experience	0.202***	0.157***	0.161***	0.161***
	(0.033)	(0.028)	(0.030)	(0.024)
Oster δ for $\beta = 0$		5.12	3.35	3.42
Number of Counties	2,036	2,036	2,036	2,036
R^2	0.05	0.28	0.33	0.42
Division Fixed Effects	No	Yes	Yes	Yes
State Fixed Effects	No	No	Yes	Yes
Geographic/Agroclimatic Controls	No	No	No	Yes

Notes: This table reports estimates of equation (5) for our leading proxy of individualism in the 20th century, the share of boys and girls age 0–10 with infrequent names in the 1940 Census. The dependent variable is normalized 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. The baseline measure of infrequent names is given by the share of children with native-born parents in county c with a name that falls outside the top 10 names for children with native-born parents born in the same Census division within the given decade. Other measures of infrequent names are considered in Appendix Table B.2. Frontier experience is expressed in decades. Column 1 is the simple bivariate regression. Columns 2 and 3 add Census division and state fixed effects, respectively. Column 4 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. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

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
Scale:	binary	binary	normalized	binary	normalized	[0, 100]
Data Source:	ANES	CCES	GSS	CCES	GSS	ACS
	(1)	(2)	(3)	(4)	(5)	(6)
total frontier experience	0.010***	0.007**	-0.022*	0.014***	0.028**	-0.034***
	(0.004)	(0.003)	(0.012)	(0.002)	(0.011)	(0.007)
Oster δ for $\beta = 0$	5.59	6.86	5.79	2.40	2.28	1.67
Mean of Dependent Variable	0.09	0.40	0.00	0.43	-0.00	1.02
Number of Individuals	2,322	53,472	9,085	169,630	5,739	–
Number of Counties	95	1,863	255	2,001	253	2,029
R^2	0.04	0.04	0.06	0.04	0.07	0.82
Survey Wave Fixed Effects	Yes	Yes	Yes	Yes	Yes	–
Individual Demographic Controls	Yes	Yes	Yes	Yes	Yes	–
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports estimates of equation (5) for several measures capturing preferences for redistribution and state spending as well as actual property tax rates. Total frontier experience is expressed in decades. Full details on the outcomes can be found in Appendix C. 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. 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 normalized 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 normalized 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 from American Community Survey data from 2010. Combining estimates from different columns and related outcomes across subsequent tables yields mean effects estimates based on the Kling, Liebman and Katz (2007) approach that retain statistical and economic significance of the frontier effect. All columns are based on the specification in column 4 of Table 2 with additional individual-level controls for age, age squared, gender, and race in columns 1–5. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 4: Total Frontier Experience and Republican Vote Share

Dependent Variable:	Republican Vote Share in Recent Presidential Election							
	2000–16 (Avg.)	2000	2004	2008	2012	2016	Δ '16–'00	Δ '16–'12
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
total frontier experience	2.055***	1.215***	1.580***	1.979***	2.329***	3.171***	1.956***	0.842***
	(0.349)	(0.312)	(0.327)	(0.364)	(0.390)	(0.416)	(0.265)	(0.134)
Oster δ for $\beta = 0$	13.01	10.47	6.44	6.12	18.68	-24.08	-7.38	-3.65
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
Mean of Dependent Variable	60.0	56.6	60.3	57.4	60.6	65.4	8.9	4.9
R^2	0.33	0.32	0.33	0.38	0.37	0.32	0.35	0.33
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports estimates of equation (5) for measures of the county-level Republican vote share in the last five presidential elections with data from the Leip Atlas. Total frontier experience is expressed in decades. Column 1 averages across all five elections. Columns 2–6 report year-specific effects. The sample and measure of frontier experience are as described in the notes to Table 2, and all estimates are based on the specification in column 4 from that table. Cross-equation tests reveal that the effect sizes are statistically different in each subsequent year and each year is statistically different from 2016. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 5: Total Frontier Experience and Preferences Over Contentious Policy Issues

Dependent Variable:	Opposes Affordable Care Act	Opposes Increasing Minimum Wage	Opposes Banning Assault Rifles	Opposes Regulation of CO_2 Emissions
	(1)	(2)	(3)	(4)
total frontier experience	0.022***	0.023***	0.015***	0.016***
	(0.004)	(0.008)	(0.004)	(0.004)
Oster δ for $\beta = 0$	2.96	3.05	2.46	2.22
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	Yes	Yes	Yes	Yes
Individual Demographic Controls	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes

Notes: This table reports estimates of equation (5) for four measures of support for conservative issues that are particularly relevant to the frontier setting in historical accounts. Total frontier experience is measured in decades. 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 2; 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 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 6: Robustness Checks for Summary Outcomes

Dependent Variable:	Infrequent Name Share normalized		County Property Tax Rate	Republican Vote Share
	Boys 1940	Girls 1940	2010	Avg. 2000–16
	(1)	(2)	(3)	(4)
	Panel A: Adding West Coast to the Baseline Sample			
total frontier experience	0.111***	0.156***	-0.031***	2.070***
	(0.021)	(0.023)	(0.006)	(0.332)
Number of Counties	2,141	2,141	2,134	2,141
Mean of Dependent Variable	0.00	0.00	1.01	59.43
	Panel B: Adding Controls for Population Density Birthplace Diversity, and Initial Railroad Access			
total frontier experience	0.068***	0.125***	-0.019***	1.431***
	(0.021)	(0.024)	(0.006)	(0.352)
Number of Counties	1,839	1,839	1,837	1,839
Mean of Dependent Variable	0.00	0.00	1.03	59.57
	Panel C: Extending Historical Frontier Period to 1950			
total frontier experience	0.074***	0.092***	-0.025***	1.302***
	(0.013)	(0.016)	(0.004)	(0.256)
Number of Counties	2,500	2,500	2,491	2,500
Mean of Dependent Variable	0.00	0.00	0.98	60.49
State Fixed Effects	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes

Notes: This table augments the baseline specifications for three summary outcomes examined in prior tables in three ways. Again, total frontier experience is measured in decades. Panel A expands the sample to include counties in the West Coast frontier sample seen in Figure 3. Panel B includes a set of additional controls: quadratic population density in 1890, birthplace diversity (fractionalization) in 1870, and the first year the county was connected to the railroad. Results are similar when adding "bad controls" for current density and diversity. The sample size is slightly reduced here relative to the baseline sample due to coverage gaps in the 1870 Census. Panel C extends the historical frontier window to 1950, thereby including additional counties beyond the 1890 east-to-west frontier line seen in Figure 3 with population density greater than 2 in 1950. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 7: Instrumental Variables Estimates for Summary Outcomes

Dependent Variable:	Infrequent Name Share normalized		County Property Tax Rate	Republican Vote Share Avg.
	Boys 1940	Girls 1940	2010	2000–16
	(1)	(2)	(3)	(4)
	Panel A: Baseline Sample and Specification			
total frontier experience	0.121**	0.113**	-0.035***	1.492**
	(0.051)	(0.057)	(0.014)	(0.715)
Number of Counties	2,036	2,036	2,029	2,036
Mean of Dependent Variable	0.00	0.00	1.02	60.04
KP Wald Statistic	216.3	216.3	216.1	216.3
	Panel B: Adding Controls for Population Density Birthplace Diversity, and Initial Railroad Access			
total frontier experience	0.154***	0.191***	-0.045***	1.797***
	(0.043)	(0.051)	(0.012)	(0.655)
Number of Counties	1,839	1,839	1,837	1,839
Mean of Dependent Variable	0.00	0.00	1.03	59.57
KP Wald Statistic	245.0	245.0	244.8	245.0
State Fixed Effects	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes

Notes: This table reports instrumental variables estimates of equation (5) based on the Bartik (1991)-style instrument described in Section 5.1. We again report results for three 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. Panel B reports the IV estimates for the specification with the additional controls as used in the prior OLS estimates in Panel B of Table 6. The first-stage KP Wald Statistic is due to Kleibergen and Paap (2006). Given the large Wald stats, similar inference arises using weak-instrument robust estimators. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 8: 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
mean infrequent name, omitted group:	0.707	0.708	0.708	0.767
frontier county resident in 1880, arrived between 1870 and 1880	0.042***		0.055***	
	(0.012)		(0.011)	
frontier county resident in 1880		0.075***		
		(0.018)		
frontier county resident in 1880, arrived before 1870			0.186***	0.118***
			(0.035)	(0.026)
Number of Individuals	1,223,600	1,239,513	1,239,513	12,630
R^2	0.02	0.02	0.02	0.05
Gender×Age Fixed Effects	Yes	Yes	Yes	Yes
Birth Order Fixed Effects	Yes	Yes	Yes	Yes

Notes: This table reports estimates of equation (6) 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. This linked sample is detailed in footnote 41 in the paper. 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.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 9: Individualism, Socioeconomic Success, and Endurance on the Frontier

	(1)	(2)	(3)	(4)
	Panel A: Dep. Var.: Father's Economic Status in 1880 (normalized)			
	sei			*occscore*
at least one child 0-10 has infrequent name	0.077***		0.080***	0.065***
	(0.007)		(0.007)	(0.007)
at least one child 0-10 has infrequent name × frontier county	0.077*		0.073*	0.066
	(0.042)		(0.042)	(0.042)
father has infrequent name		-0.047***	-0.049***	-0.044***
		(0.005)	(0.005)	(0.005)
father has infrequent name × frontier county		0.072**	0.069**	0.075**
		(0.033)	(0.033)	(0.033)
Number of Individuals	264,038	264,038	264,038	264,038
R^2	0.09	0.09	0.09	0.10
County Fixed Effects	Yes	Yes	Yes	Yes
	Panel B: Dep. Var.: Migrated from Frontier County in 1870 to Non-Frontier County in 1880			
father has infrequent name	-0.042***			-0.041***
	(0.009)			(0.009)
mother has infrequent name		-0.010		-0.009
		(0.007)		(0.007)
at least one child 0-10 has infrequent name			-0.026**	-0.023**
			(0.011)	(0.011)
Number of Individuals	27,066	27,066	27,066	27,066
Mean of Dependent Variable	0.410	0.410	0.410	0.410
R^2	0.08	0.07	0.07	0.08
Origin County Fixed Effects	Yes	Yes	Yes	Yes

Notes: This table reports estimates of equation (7) in Panel A and equation (8) in Panel B based on the same linked sample of households from the 1870 and 1880 Census described in the notes to the previous table and at length in Section 6.1. Infrequent name measures are as defined elsewhere and based on the top 10 nationally for all family members. The frontier dummy in both panels is as defined earlier. In Panel A, the dependent variable in columns 1–3 is the normalized Duncan (1961) socioeconomic index (*sei*) and in column 4 is the normalized occupational score (*occscore*), both as observed in 1880 and as provided by the North Atlantic Population Project (NAPP). The sample in Panel A includes all white native-born male household heads (fathers) aged 30–50 that did not migrate across counties between 1870 and 1880. Standard errors are clustered at the county level. The sample in Panel B is restricted to all white native-born households aged 30–50 residing in frontier counties in 1870, and the dependent variable equals one if the household moved to a non-frontier county by 1880. Standard errors are clustered at the origin county. All regressions include dummies for the number of children born in the 1870s.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 10: Differential Intergenerational Mobility on the Frontier

	Dependent Variable: Child's			
	log *sei*, 1880		log *occscore*, 1880	
	(1)	(2)	(3)	(4)
frontier1880	0.322	0.247	0.547*	0.606*
	(0.369)	(0.391)	(0.312)	(0.322)
father's log seius, 1870	0.384***	0.375***		
	(0.042)	(0.041)		
frontier × father's log seius	-0.243**	-0.219*		
	(0.118)	(0.125)		
father's log occscore, 1870			0.450***	0.443***
			(0.036)	(0.036)
frontier × father's log occscore			-0.228**	-0.249**
			(0.115)	(0.120)
Number of Individuals	777	777	805	805
Mean of Dependent Variable	2.846	2.846	2.892	2.892
R^2	0.25	0.26	0.35	0.36
County Fixed Effects	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	No	Yes

Notes: This table reports estimates of intergenerational mobility based on the NAPP Linked Sample covering 1 percent of the population in the 1870 and 1880 Population Censuses. The sample is comprised of individuals who were children in an 1870 household and household heads in an 1880 household. The sample is further restricted to the white population that did not switch counties between 1870 and 1880. The frontier, *sei*, and *occscore* measures are as defined in prior tables. The outcome is the child's log *sei* or *occscore* in 1880, and the father's log values in 1870 are the regressors. All regressions control for nativity status, though most of the sample is native-born. Columns 2 and 4 additionally control for a quadratic in age as well as fixed effects for the number of children in the new household. Standard errors are clustered at the county level.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

For Online Publication

Appendix

Contents

A Retracing the Frontier

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

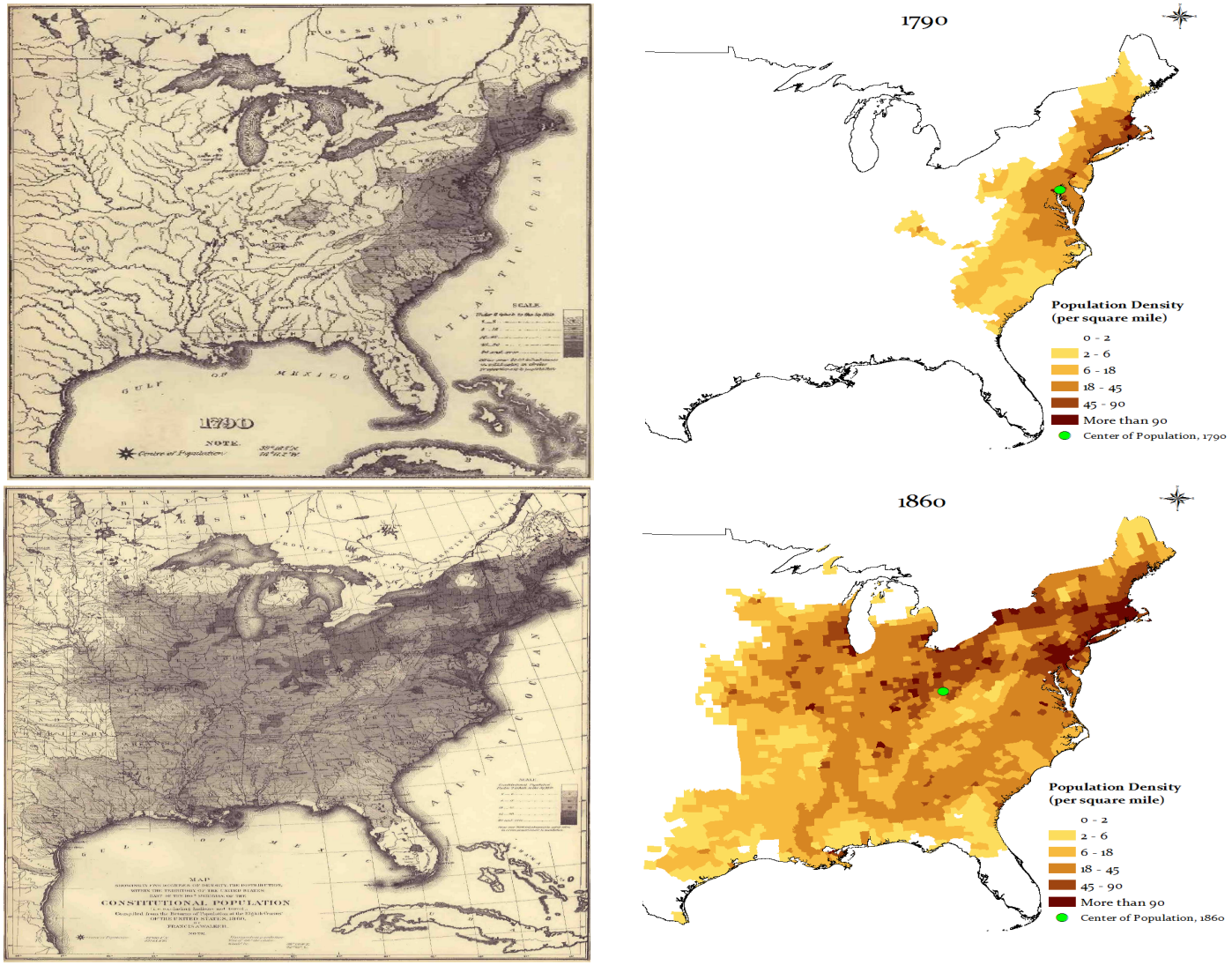
1. Calculate county level population density per square mile for each year in 1790–1890 using the 2010 county boundaries. First, we harmonize the county-level population data from each year to the 2010 county boundaries using the procedure discussed in Section 3. For intercensal years, we interpolate 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, as well as 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.

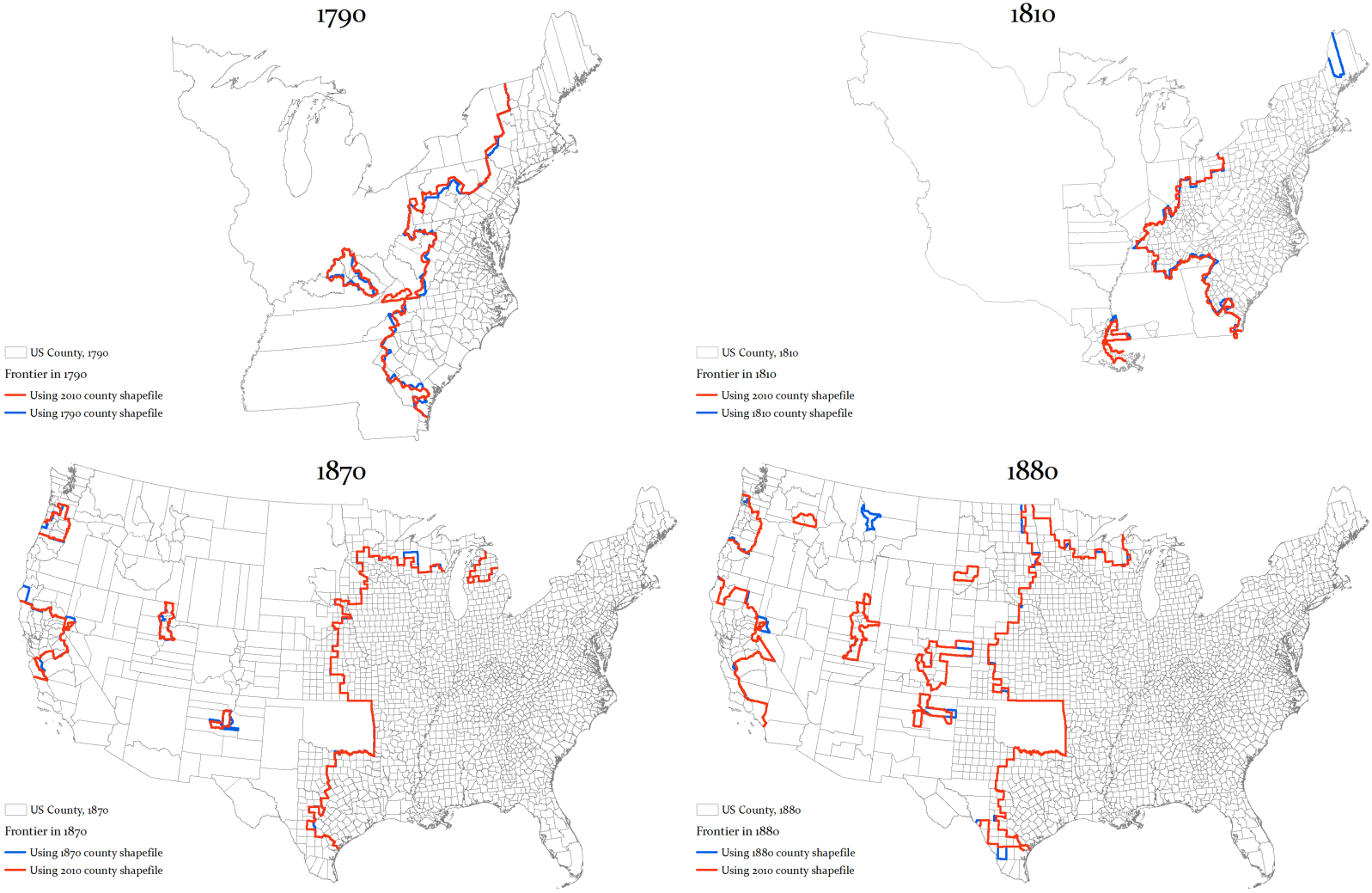
We select line segments based on length and location (e.g., X centroid of the line midpoint) in ArcGIS using the ***SelectLayerByAttribute*** tool, and apply ***CopyFeatures*** to keep only the selected lines. In the detailed robustness checks in Section 5.5, we also consider various alternatives to 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 our estimates on the right of population density in 1790 and 1860 at the (2010) county level using the harmonization procedure described in Section 3 to the historical estimates in on the left based on the the noteworthy *Progress of the Nation* Census report on which Turner based his thesis.

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

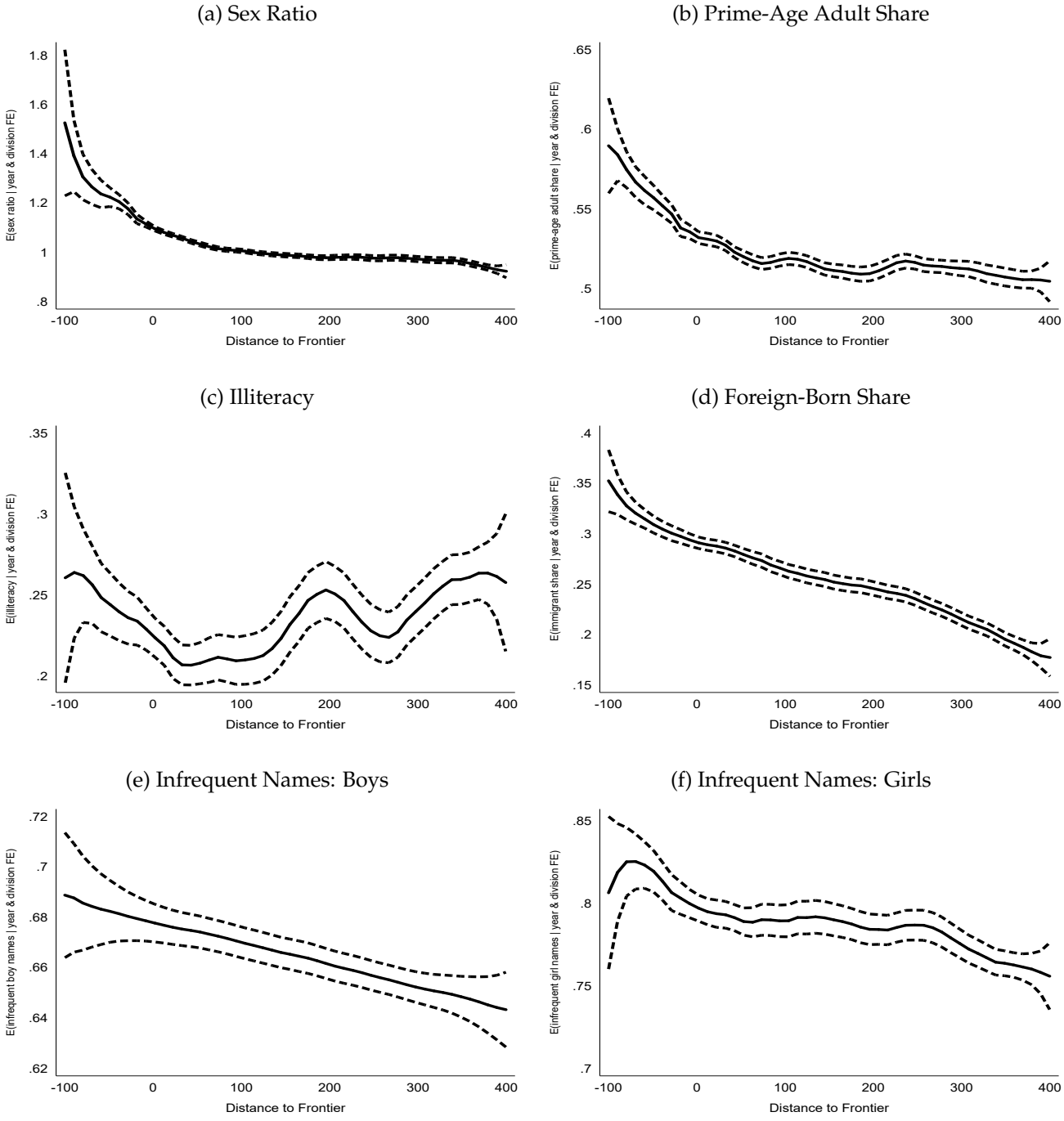


Notes: Based on county level Population Census data from 1790-1880 and NHGIS county shapefiles. The figures provide the county boundaries for selected years and the frontier lines for the corresponding years drawn using the contemporaneous county boundaries as well as the 2010 county boundary. The frontier 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 where as the frontier lines in red are drawn using the 2010 county boundaries (after the data harmonization discussed in Section 3.1).

B Additional Results and Robustness Checks

B.1 Further Background Characterizing Frontier Life

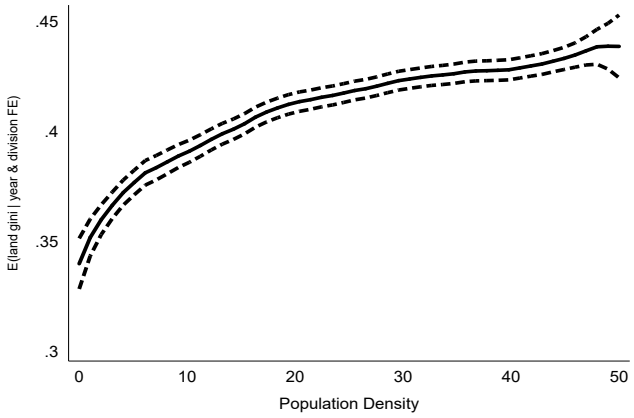
Figure B.1: Distribution of Demographics and Individualism by Distance to the Frontier



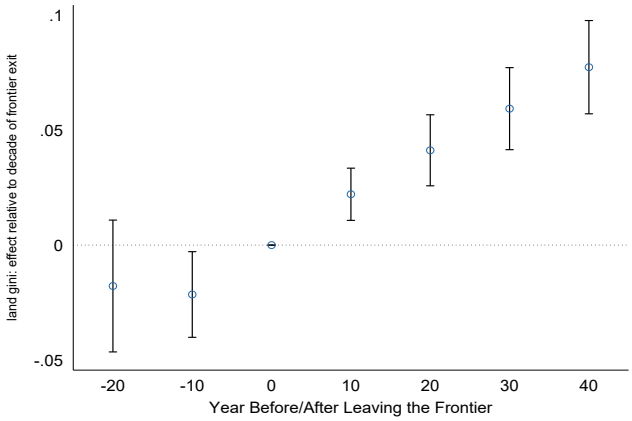
Notes: Distance to the frontier, measured in kilometers, is the distance from county's centroid to the nearest frontier line. The distance is negative if the county centroid is to the west of the nearest main frontier line. Figure (a)-(f) provide the semiparametric estimates of the corresponding dependent 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.

Figure B.2: Inequality is Lower on the Frontier

(a) Semiparametric



(b) Event Study



Notes: Based on county level data from National Historical Geographic Information System: Version 11.0 Database 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 5. See the notes therein for details.

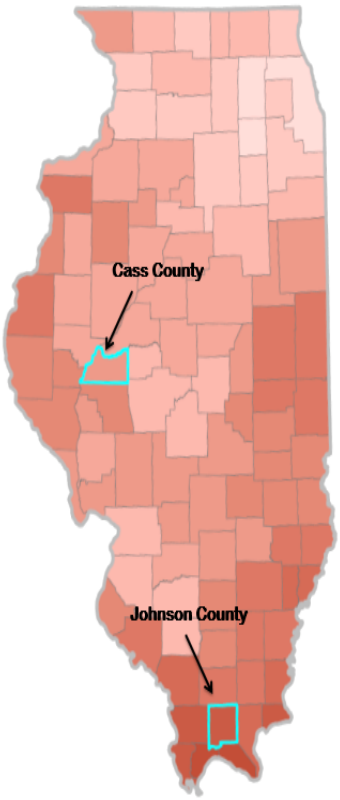
Table B.1: Occupational Composition in Frontier and Non-Frontier Counties

	Employment Share
Frontier Counties	
Farmers (owners and tenants)	.606
Laborers (n.e.c.)	.125
Farm laborers, wage workers	.047
Managers, officials, and proprietors (n.e.c.)	.040
Carpenters	.023
Truck and tractor drivers	.014
Blacksmiths	.013
Operative and kindred workers (n.e.c.)	.012
Other non-occupational response	.012
Lawyers and judges	.009
Non-Frontier Counties	
Farmers (owners and tenants)	.534
Managers, officials, and proprietors (n.e.c.)	.063
Laborers (n.e.c.)	.061
Operative and kindred workers (n.e.c.)	.0465
Carpenters	.037
Farm laborers, wage workers	.028
Salesmen and sales clerks (n.e.c.)	.015
Blacksmiths	.014
Other non-occupational response	.012
Physicians and surgeons	.010

Notes: This table reports the top 10 occupational shares in frontier and non-frontier counties in 1870 using the 1870–1880 linked sample that we use in our main analysis in Section 6.

B.2 Case Study Illustrating Long-Run Effects

To fix ideas, consider the two counties of Cass and Johnson mentioned in Section 3.2 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. Cass has 36.3 people/mi^2, median income is around US$ 41,544, and 86 percent of the population is white. Johnson also has 36.6 people/mi^2, median income around US$ 41,619, and 89 percent white population. 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 B.3, B.4, and especially B.6 below.



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, median income, and racial composition.

B.3 Further Robustness Checks on Individualism

Tables B.2 and B.3 show the robustness of the OLS and IV results, respectively, for infrequent names to alternative measures of infrequency and restrictions on (grand)parental ancestry. Table B.4 shows that the baseline OLS results for infrequent names look similar in each decade before 1940 but after the official closing of the frontier. Table B.5 validates the long-run relationship of TFE with individualism using an alternative survey-based proxy from 1990 ANES data.

Power Law Property of Names. One important dimension of robustness that we corroborate in Tables B.2 and B.3 is that the results are not sensitive to the cutoff for defining infrequent names (10, 25, 100, ...). This is likely due to the fact that naming frequencies in the United States follow a power law (see Hahn and Bentley, 2003; Gureckis and Goldstone, 2009), and hence the share of people with each of the top 10 names can be characterized by the same shape parameter as the share with the top 25 names, and so on. Adopting this parametrization, the literature on names has documented an increase in the national power law exponent over time, which suggests a growing trend towards less concentration among popular names. Our results here identify mid-20th century differences in name (in)frequency across counties.

Auxiliary Measures of 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.5 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.[1] While the results with the full set of controls are noisy, we nevertheless view these findings as at least suggestive of longstanding 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.

[1] In a related result using the CCES, we find that residents in counties with greater TFE are significantly less likely to have ever belonged to a union. While this result may be explained in part by differences in sectoral composition, it is also consistent with weaker collectivist tendencies.

Table B.2: Robustness to Other Measures of Infrequent Names (OLS Estimates)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Sample: White Children Aged 0-10, 1940 Census									
Further Sample Restriction	None	Native Father	Native Parents	Native Grand-parents	None	None	None	None	None	None
Infrequent Measure	Top 10 Division				Top 10 National	Top 10 State	Top 10 County	Top 25 Division	Top 100 Division	Top 10 National Non-Biblical
	Panel A: Share of Boys with Infrequent Names									
total frontier experience	0.109***	0.113***	0.112***	0.089***	0.123***	0.121***	0.098***	0.098***	0.115***	0.100***
	(0.022)	(0.022)	(0.022)	(0.024)	(0.020)	(0.022)	(0.034)	(0.023)	(0.022)	(0.017)
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.63	0.62	0.61	0.25	0.63	0.61	0.20	0.64	0.67	0.73
	Panel B: Share of Girls with Infrequent Names									
total frontier experience	0.167***	0.162***	0.161***	0.085***	0.166***	0.175***	0.047	0.148***	0.157***	0.155***
	(0.025)	(0.024)	(0.024)	(0.023)	(0.023)	(0.023)	(0.034)	(0.020)	(0.020)	(0.021)
Number of Counties	2,036	2,036	2,036	2,035	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.43	0.43	0.42	0.11	0.48	0.41	0.22	0.58	0.64	0.53

Notes: This table reports analogous estimates of Table 2 for alternative measures of the prevalence of individualistic naming patterns. The dependent variable is normalized across all columns. Column 1 show the results with no further sample restriction. Column 2 restricts the measure to children with native-born fathers, column 3 restricts to those with native-born mothers and fathers (the baseline in Table 2), and column 4 restricts to those with native-born grandparents. Column 5 changes the definition of infrequency of names to be based on the top 10 nationally, column 6 changes to the top 10 at the state level, and column 7 to the top 10 at the given county level. Column 8 increases the uncommon threshold to the top 25, and column 9 increases that to the top 100. Column 10 restricts to top names that are do not have biblical roots. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table B.3: Robustness to Other Measures of Infrequent Names (IV Estimates)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Sample: White Children Aged 0-10, 1940 Census									
Further Sample Restriction	None	Native Father	Native Parents	Native Grand-parents	None	None	None	None	None	None
Infrequent Measure	Top 10 Division				Top 10 National	Top 10 State	Top 10 County	Top 25 Division	Top 100 Division	Top 10 National Non-Biblical
	Panel A: Share of Boys with Infrequent Names									
total frontier experience	0.124**	0.125**	0.121**	0.111**	0.094*	0.124**	0.137**	0.125**	0.140**	0.070*
	(0.052)	(0.052)	(0.051)	(0.053)	(0.049)	(0.054)	(0.054)	(0.056)	(0.055)	(0.042)
Number of Counties	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036	2,036
KP Wald Statistic	216.3	216.3	216.3	216.3	216.3	216.3	216.3	216.3	216.3	216.3
	Panel B: Share of Girls with Infrequent Names									
total frontier experience	0.145**	0.118**	0.113**	0.027	0.099*	0.108**	0.085	0.117**	0.118**	0.124**
	(0.060)	(0.058)	(0.057)	(0.045)	(0.054)	(0.051)	(0.054)	(0.046)	(0.050)	(0.052)
Number of Counties	2,036	2,036	2,036	2,035	2,036	2,036	2,036	2,036	2,036	2,036
KP Wald Statistic	216.3	216.3	216.3	216.2	216.3	216.3	216.3	216.3	216.3	216.3

Notes: This table reports analogous instrumental variables estimates of the OLS specifications in Table B.2.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table B.4: Persistence of Individualism

	Sample: White Children Aged 0-10 with Native-Born Parents							
	1910		1920		1930		1940	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Share of Boys with Infrequent Names							
total frontier experience	0.155***	0.092	0.150***	0.118**	0.119***	0.079	0.112***	0.121**
	(0.029)	(0.076)	(0.026)	(0.058)	(0.025)	(0.054)	(0.022)	(0.051)
Number of Counties	2,029	2,029	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.28	0.02	0.39	0.03	0.49	0.02	0.61	0.03
KP Wald Statistic		222.5		216.3		216.3		216.3
	Panel B: Share of Girls with Infrequent Names							
total frontier experience	0.114***	0.054	0.152***	0.078	0.146***	0.052	0.161***	0.113**
	(0.022)	(0.046)	(0.021)	(0.054)	(0.024)	(0.053)	(0.024)	(0.057)
Number of Counties	2,029	2,029	2,036	2,036	2,036	2,036	2,036	2,036
R^2	0.49	0.02	0.55	0.03	0.51	0.02	0.42	0.03
KP Wald Statistic		222.5		216.3		216.3		216.3

Notes: This table reports analogous OLS and IV 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 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table B.5: Total Frontier Experience and Contemporary Cooperation vs. Self-Reliance

	OLS	OLS	OLS	OLS	IV
	(1)	(2)	(3)	(4)	(5)
total frontier experience	-0.019*	-0.025**	-0.041***	-0.026**	-0.039
	(0.009)	(0.009)	(0.014)	(0.012)	(0.030)
Oster δ for $\beta = 0$	-2.77	-2.61	-15.37	-249.36	
Number of Individuals	567	567	567	567	567
Number of Counties	48	48	48	48	48
Mean of Dependent Variable	0.549	0.549	0.549	0.549	0.549
KP Wald Stat					12.7
KP Underidentification Test p-value					0.005
R^2	0.01	0.02	0.02	0.03	0.01
Individual Demographic Controls	Yes	Yes	Yes	Yes	Yes
Division Fixed Effects	No	Yes	Yes	Yes	Yes
State Fixed Effects	No	No	Yes	No	No
Geographic/Agroclimatic Controls	No	No	No	Yes	Yes

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 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

B.4 Ruling out Individual-Level Confounders of Policy Preferences

Many of the policies in Tables 3 and 5 elicit strong partisanship within the U.S. as Republicans and Democrats hew closely to the party line. However, as seen in Tables B.6, greater TFE is associated with stronger opposition to government intervention even after controlling for the strength of Republican party support reported in the CCES. Moreover, these results survive further controls for individual education and family income. Again, although these covariates are "bad controls," their inclusion helps rule out the concern that all of the observed effects are driven by prolonged frontier experience simply leading to tribal party- and class-based identity unrelated to the deep roots of frontier culture.

Table B.6: Robustness to Controls for Income, Education, and Partisan Identification

Dependent Variable:	Prefers Cutting Public Spending on Welfare		Prefers Balancing Budget By Cutting Spending		Prefers Repealing Affordable Care Act		Opposes Increasing Minimum Wage		Opposes Banning Assault Rifles		Opposes EPA Regulation of CO_2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
total frontier experience	0.007**	0.006*	0.014***	0.009***	0.022***	0.019***	0.023***	0.015**	0.015***	0.015***	0.016***	0.016***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.008)	(0.007)	(0.004)	(0.004)	(0.004)	(0.003)
identifies as very strong Republican		0.299***		0.379***		0.415***		0.457***		0.284***		0.338***
		(0.008)		(0.006)		(0.009)		(0.017)		(0.008)		(0.009)
family income > USD 50,000		0.099***		0.048***		-0.019***				-0.004		0.019***
		(0.006)		(0.005)		(0.007)				(0.005)		(0.006)
education > high school		0.007		-0.007		-0.080***		0.076***		0.015**		0.018***
		(0.006)		(0.005)		(0.006)		(0.014)		(0.007)		(0.006)
Oster δ for $\beta = 0$	6.86	4.63	2.40	1.76	2.96	2.26	3.05	1.85	2.46	2.19	2.22	2.16
Number of Counties	53,472	47,851	169,630	80,155	29,446	26,131	5,134	4,618	29,404	26,093	29,215	25,938
Mean of Dependent Variable	0.40	0.39	0.43	0.38	0.53	0.53	0.31	0.30	0.37	0.36	0.32	0.31
R^2	0.04	0.10	0.04	0.11	0.06	0.14	0.06	0.21	0.09	0.13	0.08	0.14

Notes: This table subjects the results in Table 5 to additional, non-predetermined controls for education, family income, and Republican Party identification as described in Appendix C. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

B.5 Robustness Check Controlling for Contemporary Population Density

Whereas Tables 6 and 7 controlled for historical population density at the end of the frontier era, this table controls for contemporaneous population density. While this is of course a "bad control" specification, it nevertheless provides some reassurance that the long-run effects are not purely explained by persistent population density differentials in areas with greater TFE.

Table B.7: Robustness Checks for Summary Outcomes Controlling for Contemporary Population Density

Dependent Variable:	Infrequent Name Share normalized		County Property Tax Rate	Republican Vote Share Avg.
	Boys 1940	Girls 1940	2010	2000–16
	(1)	(2)	(3)	(4)
	Panel A (OLS): Adding Controls for Quadratic in Population Density in 1940/2010			
total frontier experience	0.092***	0.142***	-0.013**	1.427***
	(0.021)	(0.024)	(0.006)	(0.350)
Number of Counties	2,036	2,036	2,029	2,036
Mean of Dependent Variable	-0.00	0.00	1.02	60.04
	Panel B (IV): Adding Controls for Quadratic in Population Density in 1940/2010			
total frontier experience	0.125**	0.118**	-0.024**	1.164*
	(0.049)	(0.055)	(0.011)	(0.705)
Number of Counties	2,036	2,036	2,029	2,036
Mean of Dependent Variable	-0.00	0.00	1.02	60.04
KP Wald Statistic	228.91	228.91	218.56	218.82
State Fixed Effects	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes

Notes: This table re-estimates the baseline OLS and IV specifications for our main outcomes conditioning on a quadratic in contemporary population density where contemporary means 1940 for columns 1–2 and 2010 for columns 3–4. The specifications are otherwise identical to those in the main regression tables in the paper. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

B.6 Robustness Checks for Summary Outcomes by Census Regions

In Table B.8, we show the robustness of the summary outcomes by Census regions. Census regions are the groupings of states into four areas—Northeast, Midwest, South, and West—where each region is divided into two or more census divisions. We show results for the Midwest, South, and West regions for which we have enough observations in our baseline sample. In Panel A, we use the baseline TFE measure and baseline sample including the West Coast. In Panel B, we extend the historical frontier window to 1950. In both panels, the outcomes for each region are consistent with the baseline results. Furthermore, the results are similar across regions suggesting that the main findings are not driven by any particular region.

Table B.8: Robustness Checks for Summary Outcomes by Census Regions

Dependent Variable:	Infrequent Name Share normalized		County Property Tax Rate	Republican Vote Share Avg.
	Boys 1940	Girls 1940	2010	2000–16
	(1)	(2)	(3)	(4)
Panel A: Using Baseline Total Frontier Experience				
	I. Mid-West			
total frontier experience	0.236***	0.220***	-0.051***	1.882***
	(0.044)	(0.049)	(0.014)	(0.414)
Number of Counties	987	987	981	987
Mean of Dependent Variable	0.00	0.00	1.24	59.15
	II. South			
total frontier experience	0.111***	0.162***	-0.027***	2.458***
	(0.030)	(0.029)	(0.007)	(0.396)
Number of Counties	936	936	935	936
Mean of Dependent Variable	0.00	-0.00	0.75	61.78
	III. West			
total frontier experience	0.113*	0.108	-0.006	1.459
	(0.059)	(0.080)	(0.013)	(0.890)
Number of Counties	152	152	152	152
Mean of Dependent Variable	0.00	-0.00	0.76	48.81
Panel B: Extending the Historical Frontier Window to 1950				
	I. Mid-West			
total frontier experience	0.125***	0.108***	-0.042***	1.515***
	(0.035)	(0.038)	(0.012)	(0.350)
Number of Counties	1,038	1,038	1,029	1,038
Mean of Dependent Variable	-0.00	-0.00	1.23	59.43
	II. South			
total frontier experience	0.076***	0.096***	-0.031***	1.429***
	(0.025)	(0.027)	(0.006)	(0.422)
Number of Counties	1,074	1,074	1,074	1,074
Mean of Dependent Variable	0.00	-0.00	0.78	63.18
	III. West			
total frontier experience	0.056***	0.083***	-0.009**	1.197***
	(0.018)	(0.019)	(0.004)	(0.274)
Number of Counties	322	322	322	322
Mean of Dependent Variable	-0.00	0.00	0.72	56.10
State Fixed Effects	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes

Notes: This table reports estimates for the four summary outcomes examined in Table 6 by census regions. The results for the North East region are not provided above since the baseline sample does not contain sufficient counties in the North East. Panel A uses the baseline TFE measure and specification. Panel B extends the historical frontier window to 1950, thereby including additional counties beyond the 1890 main frontier lines seen in Figure 3 that have population density greater than 2. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

B.7 Alternative Measures of Total Frontier Experience

Our baseline measure of TFE closely followed definitions in the historical literature as discussed in Section 3. In Table B.9, we demonstrate the robustness of our results to three relevant margins of adjustment to our measure of TFE. In each case, we redefinie 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 B.9: Robustness to Alternative Measures of TFE for Summary Outcomes

Dependent Variable:	Infrequent Name Share normalized		County Property Tax Rate	Republican Vote Share
	Boys 1940	Girls 1940	2010	Avg. 2000–16
	(1)	(2)	(3)	(4)
TFE: 100 km, ≤ 6/mi^2, no inner island lines (baseline)	0.112***	0.161***	-0.034***	2.055***
	(0.022)	(0.024)	(0.007)	(0.349)
TFE: 50 km, ≤ 6/mi^2, no inner island lines	0.117***	0.173***	-0.035***	2.051***
	(0.022)	(0.026)	(0.007)	(0.358)
TFE: 100 km, ≤ 18/mi^2, no inner island lines	0.096***	0.105***	-0.027***	1.575***
	(0.022)	(0.025)	(0.007)	(0.339)
TFE: 50 km, ≤ 18/mi^2, no inner island lines	0.085***	0.105***	-0.025***	1.458***
	(0.021)	(0.023)	(0.006)	(0.351)
TFE: 100 km, 2-6/mi^2, no inner island lines	0.081**	0.110**	-0.014*	1.877***
	(0.033)	(0.044)	(0.008)	(0.485)
TFE: 50 km, 2-6/mi^2, no inner island lines	0.063*	0.105**	-0.012	1.771***
	(0.038)	(0.049)	(0.009)	(0.530)
TFE: 100 km, no density restriction, no inner island lines	0.033	0.030	-0.011	1.001***
	(0.021)	(0.025)	(0.007)	(0.335)
TFE: 50 km, no density restriction, no inner island lines	0.054***	0.068***	-0.018***	1.078***
	(0.020)	(0.023)	(0.006)	(0.339)
TFE: 100 km, ≤ 6/mi^2, including inner island lines	0.132***	0.188***	-0.032***	2.048***
	(0.020)	(0.023)	(0.006)	(0.320)
TFE: 50 km, ≤ 6/mi^2, including inner island lines	0.143***	0.205***	-0.035***	2.098***
	(0.022)	(0.025)	(0.007)	(0.335)
TFE: 100 km, ≤ 6/mi^2, main single contour line	0.087***	0.117***	-0.037***	1.872***
	(0.026)	(0.032)	(0.008)	(0.436)
TFE: 50 km, ≤ 6/mi^2, main single contour line	0.082***	0.113***	-0.043***	1.787***
	(0.028)	(0.034)	(0.008)	(0.460)
State Fixed Effects	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes

Notes: This table reports estimates of equation (5) 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 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

B.8 Robustness Checks for Additional Survey Outcomes

This table reports the full set of baseline robustness checks for the additional survey-based outcomes not reported in Table 6.

Table B.10: Robustness Checks for Additional Outcomes

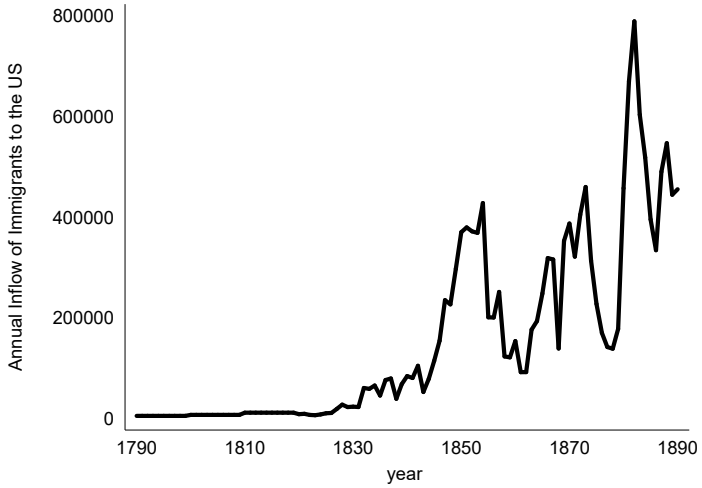
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Data Source:	ANES	CCES	GSS	CCES	GSS	CCES			
Dependent Variable:	Prefers Cut on Poor	Public Spending on Welfare	Government Should Redistribute	Prefer Cut Debt by Spending Cuts	Index of Preferences for Cut Spending	Opposes Affordable Care Act	Opposes Increasing Min. Wage	Opposes Banning Assault Rifle	Opposes EPA Regulation of CO_2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Adding West Coast to the Baseline Sample								
total frontier experience	0.008*	0.009***	-0.007	0.016***	0.033***	0.023***	0.023***	0.021***	0.017***
	(0.005)	(0.003)	(0.014)	(0.003)	(0.011)	(0.004)	(0.007)	(0.005)	(0.004)
Number of Individuals	2,810	66,254	11,271	210,948	7,109	36,768	6,553	36,711	36,479
Number of Counties	108	1,963	290	2,105	288	1,828	1,157	1,823	1,818
Mean of Dependent Variable	0.10	0.39	-0.00	0.42	0.00	0.51	0.30	0.36	0.31
	Panel B: Adding Controls for Population Density, Birthplace Diversity, and Initial Railroad Access								
total frontier experience	0.014***	0.003	-0.017	0.008***	0.017	0.016***	0.023***	0.008**	0.011***
	(0.004)	(0.003)	(0.014)	(0.002)	(0.012)	(0.004)	(0.008)	(0.004)	(0.003)
Number of Individuals	2,188	51,171	8,466	162,114	5,382	28,165	4,905	28,128	27,941
Number of Counties	87	1,711	242	1,820	240	1,591	1,004	1,589	1,582
Mean of Dependent Variable	0.09	0.40	0.00	0.43	-0.00	0.53	0.31	0.37	0.31
	Panel C: Extending Historical Frontier Period to 1950								
total frontier experience	0.007	0.011***	0.006	0.014***	0.035***	0.021***	0.016***	0.020***	0.016***
	(0.005)	(0.003)	(0.011)	(0.002)	(0.009)	(0.003)	(0.005)	(0.003)	(0.003)
Number of Individuals	3,035	74,260	12,566	237,598	7,916	41,211	7,252	41,151	40,895
Number of Counties	113	2,241	319	2,448	317	2,076	1,294	2,072	2,067
Mean of Dependent Variable	0.10	0.39	0.00	0.43	0.00	0.52	0.31	0.37	0.32
Survey Wave Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State/Division Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table augments the baseline specifications for additional main outcomes examined in prior tables in three ways. Panel A expands the sample to include counties in the West Coast frontier sample seen in Figure 3. Panel B includes a set of additional controls: quadratic population density in 1890, birthplace diversity (fractionalization) in 1870, and the first year the county was connected to the railroad. Results are similar when adding "bad controls" for current density and diversity. The sample size is slightly reduced here relative to the baseline sample due to coverage gaps in the 1870 Census. Panel C extends the historical frontier window to 1950, thereby including additional counties beyond the 1890 east-to-west frontier line seen in Figure 3 with population density greater than 2 in 1950. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.

B.9 Instrumental Variables Strategy: Further Background and Additional Results

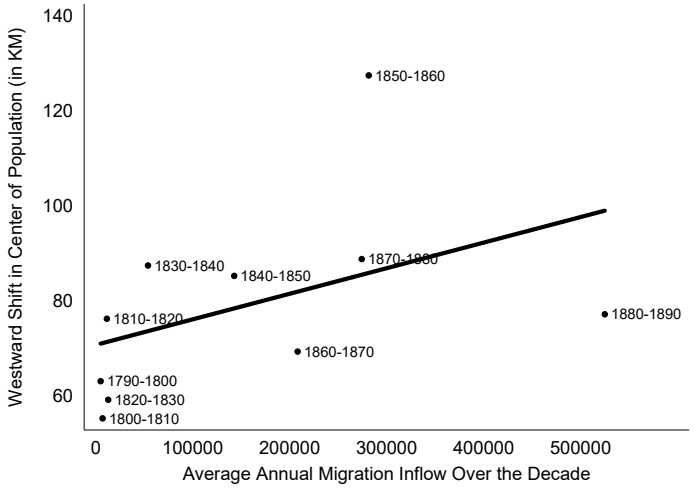
Figure B.3 shows the ebb and flow of immigration to the U.S. over the study period. Figure B.4 then shows the strong positive correlation between these immigrant inflows by decade and the speed of westward expansion, proxied by the east-to-west distance traveled by the country's population centroid (the green dot in Figure A.1(b) for 1860). This simple scatterplot helps visualize the process by which immigrants arriving in the U.S. (largely on the Eastern seaboard) pushed the edges of settlement farther westward, which in turn hastened the forward 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 happen to be getting closer to the frontier line at a time of rapid inflows into the U.S. Figure B.5 shows partial regression line that demonstrates the strong first stage in our main IV regressions from Table 7.

Figure B.3: Annual Migration Inflows



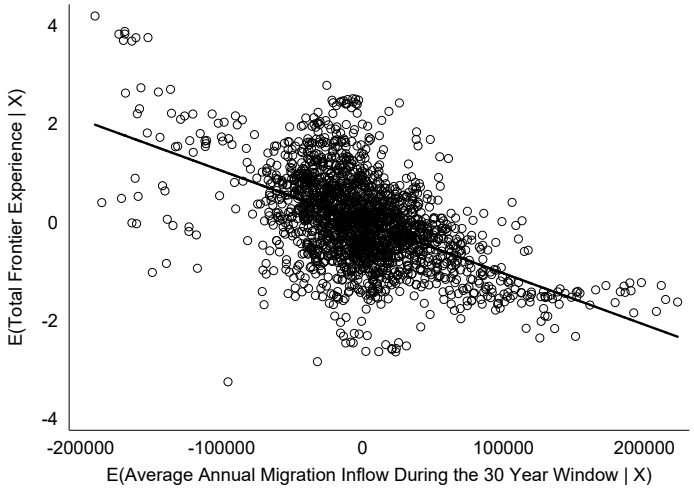
Notes: This figure plots the total number of migrants entering the United States, 1790-1890. The data for 1820–1890 is available from the Migration Policy Institute (2016), while the data for 1790-1819 is imputed from Tucker (1843).

Figure B.4: Immigration and Westward Expansion



Notes: This figure 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. This measure was reported historically by the U.S. Census Bureau (see footnote 11 in the paper).

Figure B.5: Immigration and Westward Expansion

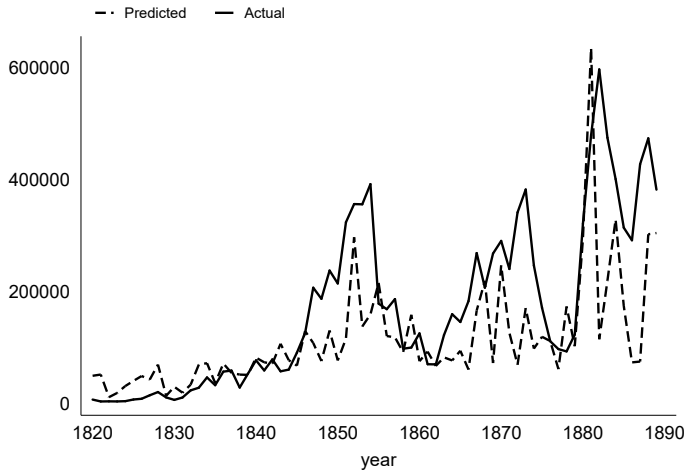


Notes: This figure plots the first stage partial regression coefficient corresponding to the baseline IV regression in Table 7.

Section 5.6 showed that the main results are robust to an instrumental variable estimation exploiting the plausibly exogenous variation in the timing of national migration inflows. To address concerns regarding the excludability of the baseline instrument due to pull factors associated with immigrant inflows, in Table B.11, we show that the main results are also robust to an instrument based on push factors unrelated to frontier conditions. Drawing upon the approach in Nunn, Qian and Sequeira (2017), we use country-year level data on migrant inflows from 16 European countries to the US from 1820–1890 to construct an instrument based on 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 Nunn, Qian and Sequeira, 2017, 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 B.6 shows how the predicted inflows, which isolate push factors, compare to the actual inflows, which naturally include both push and pull. We retain the full sample of counties in the regressions by using the imputed migration inflows before 1820 to augment the predicted inflows instrument thereafter.

Table B.11 below demonstrates that this alternative IV delivers results that are economically and statistically indistinguishable from the baseline IV results in Table 7. Restricting the sample to counties just west of the frontier after 1820—for which the IV is solely based on predicted flows—delivers similar signs and magnitudes, but the estimates are noisier due to the smaller sample size. Another alternative would be to simply use the post-1819 predicted flows in computing the average annual flows over the 30 years for counties beginning their frontier window in the given year, 1790–1819. This would mean having one predicted year for counties beginning in 1790, two predicted years for counties beginning in 1791 and so on. This approach delivers nearly identical results to the ones presented in Table B.11.

Figure B.6: 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 Nunn, Qian and Sequeira (2017) as described above.

Table B.11: Alternative Instrumental Variables Estimates for Summary Outcomes

Dependent Variable:	Infrequent Name Share normalized		County Property	Republican Vote Share
	Boys 1940	Girls 1940	Tax Rate 2010	Avg. 2000–16
	(1)	(2)	(3)	(4)
	Panel A: Baseline Sample and Specification			
total frontier experience	0.166***	0.154**	-0.038***	2.233***
	(0.055)	(0.060)	(0.013)	(0.725)
Number of Counties	2,036	2,036	2,029	2,036
Mean of Dependent Variable	-0.00	0.00	1.02	60.04
KP Wald Statistic	166.2	166.2	166.4	166.2
	Panel B: Adding Controls for Population Density Birthplace Diversity, and Initial Railroad Access			
total frontier experience	0.184***	0.223***	-0.046***	2.285***
	(0.044)	(0.050)	(0.011)	(0.654)
Number of Counties	1,839	1,839	1,837	1,839
Mean of Dependent Variable	-0.00	0.00	1.03	59.57
KP Wald Statistic	195.8	195.8	195.9	195.8
State Fixed Effects	Yes	Yes	Yes	Yes
Geographic/Agroclimatic Controls	Yes	Yes	Yes	Yes

Notes: This table re-estimates the baseline IV specifications 7 using the alternative generated instrument described above. The specifications are otherwise identical. Note that although the generated instrument adds some sampling variation to the regressions, it does not affect the consistency of the estimates. Standard errors are clustered based on the grid-cell approach of Bester, Conley and Hansen (2011) as detailed in Section 5.1.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

B.10 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 B.12 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 B.12: 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 County-Years	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).
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

B.11 Further Results on Individualism, Success, and Endurance on the Frontier

Table B.13: Individualism and Changes in Socioeconomic Status

Dependent Variable:	$sei_{1880} - sei_{1870}$				$occscore_{1880} - occscore_{1870}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
has infrequent name	0.374	0.338	0.205	-0.240	0.731	0.739	0.626	0.573
	(1.029)	(0.989)	(0.982)	(0.878)	(0.558)	(0.524)	(0.517)	(0.488)
has infrequent name × frontier county	21.463**	20.116**	20.944***	18.681***	9.012**	8.281**	8.990**	8.537**
	(9.689)	(8.088)	(7.616)	(6.466)	(4.517)	(4.139)	(4.212)	(4.050)
Number of Individuals	4,268	4,268	4,268	4,268	4,268	4,268	4,268	4,268
Mean of Dependent Variable	-0.245	-0.245	-0.245	-0.245	-0.548	-0.548	-0.548	-0.548
R^2	0.23	0.25	0.27	0.40	0.23	0.27	0.29	0.39
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Age Squared Control	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Number of Children Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Baseline Farmer Dummy	No	No	Yes	Yes	No	No	Yes	Yes
Baseline *sei* or *occscore*	No	No	No	Yes	No	No	No	Yes

Notes: This table reports estimates analogous to those in Panel A of Table 9 but based on the NAPP Linked Sample based on a 1 percent of the population in the 1870 and 1880 Population Censuses. The sample is restricted relative to the one in Table 9 as we are interested in the change in socioeconomic status (*sei*) and occupational standing (*occscore*) between 1870 and 1880, and this information is only available in both years for this smaller linked sample. The estimates are based on white, male household heads that reside in the same county in 1870 and 1880. The dependent variable is the change in *sei* or *occscore*; results are similar taking logs, but the levels allow us to retain individuals that switch from zero valued status to positive or vice versa. The frontier dummy equals one if the given county lies in the frontier in 1880. The infrequent name measure is based on the top 10 names nationally. The controls listed at the bottom of the table include a dummy for farmer occupations in 1870 in columns 3/4 and 7/8, and the baseline dependent variable in columns 4 and 8. Standard errors are clustered at the county level.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table B.14: Individualism and Endurance on the Frontier

	(1)	(2)	(3)	(4)
	Dependent Variable:			
	Panel A: Emigrated from Frontier County in 1870			
father has infrequent name	-0.051***			-0.050***
	(0.009)			(0.009)
mother has infrequent name		-0.005		-0.005
		(0.007)		(0.007)
at least one child 0-10 has infrequent name			-0.024**	-0.020*
			(0.010)	(0.010)
Number of Individuals	27,066	27,066	27,066	27,066
Mean of Dependent Variable	0.583	0.583	0.583	0.583
R^2	0.07	0.07	0.07	0.07
	Panel B: Migrated from Frontier County in 1870 to Non-Frontier County in 1880			
father has infrequent name	-0.042***			-0.041***
	(0.009)			(0.009)
mother has infrequent name		-0.010		-0.009
		(0.007)		(0.007)
at least one child 0-10 has infrequent name			-0.026**	-0.023**
			(0.011)	(0.011)
Number of Individuals	27,066	27,066	27,066	27,066
Mean of Dependent Variable	0.410	0.410	0.410	0.410
R^2	0.08	0.07	0.07	0.08
	Panel C: Onward Migrated from Frontier County in 1870 to Frontier County in 1880			
father has infrequent name	-0.006			-0.006
	(0.006)			(0.006)
mother has infrequent name		0.006		0.006
		(0.005)		(0.005)
at least one child 0-10 has infrequent name			0.001	0.001
			(0.008)	(0.008)
Number of Individuals	27,066	27,066	27,066	27,066
Mean of Dependent Variable	0.167	0.167	0.167	0.167
R^2	0.11	0.11	0.11	0.11

Notes: This table reproduces the estimates from Panel B of Table 9 alongside other outcomes in Panels A and C that clarify that the "return migration" effect comprises the full effect on outmigration destinations discussed with respect to that finding. Standard errors are clustered at the origin county level.
Significance levels: $*$: 10% $**$: 5% $***$: 1%.

C 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 Minnesota Population Center (2016). The data on area is calculated using the 2010 county shapefiles from NHGIS (Minnesota Population Center, 2011) 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.

Sex Ratio. Whites males/white females. The data is available for every decade in 1790-1860 and 1890. Data source: (Minnesota Population Center, 2011).

Prime Age Adult Share. Whites aged 15–49/all whites. The data used is consistently available for every decade in 1830-1860. Data source: (Minnesota Population Center, 2011).

Illiteracy. Illiterate whites aged above 20/whites aged over 20. The variable is available consistently for 1830-1860. Data source: (Minnesota Population Center, 2011).

Foreign Born Share. Foreign born/population. The variable used is available for every decade in 1820-1890 (excluding 1840). Data source: (Minnesota Population Center, 2011).

Out of State Born Share. Out-of-state born/population. The variable is consistently available for every decade in 1850-1880. Data source: (Minnesota Population Center, 2011).

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. (Minnesota Population Center, 2011).

Infrequent Children Names. White Children Aged 0–10 with Non-Top 10 First Names in Division/White Children Aged 0-10. We also construct similar variables further restricting to children aged 0–10 with native parents, and native grandparents. In addition, for the same sample, we construct additional variables by calculating the popularity of names at the 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 census division (or nationally), and then accordingly count the number of children in that county with the identified top 10 names in their corresponding census division. The variables 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), and with grandparent restriction for 1880–1940 (excluding 1890). To give some examples, 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. Data source: The NAPP full count census data for 1850 and the *Ancestry* data collected by NBER for 1860–1940.

Economic Status. We measure economics status using either the socioeconomic index (*sei*) or the occupational score (*occscore*) measures provided by the North Atlantic Population Project: Complete Count Microdata. 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: (Minnesota Population Center, 2011).

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 (2010) 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 C.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 C.1(a) for the map of the maximum survey coverage in the final sample of ANES data merged with the frontier related data.

Prefers State Decrease Welfare Spending. 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 C.1(b) for the map of the maximum survey coverage in the final sample of CCES 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 C.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, Marsden, Hout and Kim, 2015).

Prefers Repealing 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. For simplicity, we only consider the five presidential elections since 2000. Data source: Dave Leip's Atlas of U.S. Presidential 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 (Minnesota Population Center, 2011) using GIS software.

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 (Minnesota Population Center, 2011).

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 (Minnesota Population Center, 2011).

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: (Minnesota Population Center, 2011)

Additional Variables

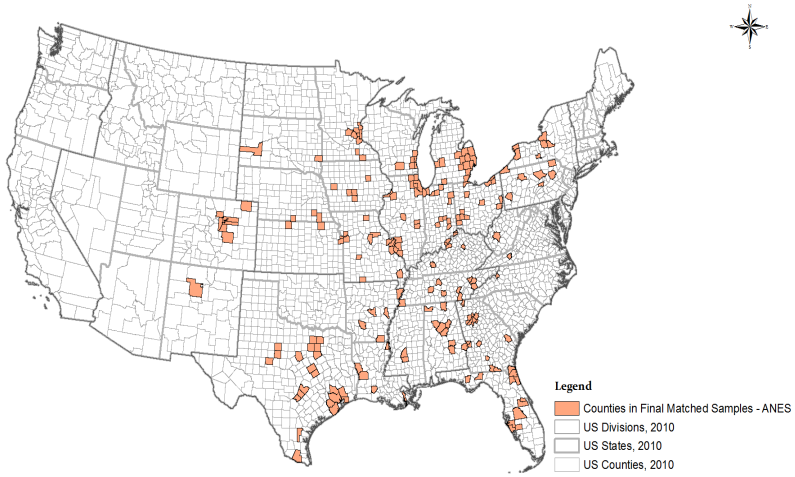
Annual Migration Inflow. Total number of migrants entering the United States every year. The data for 1820–1890 is available from the Migration Policy Institute (2016), which tabulates data from the Office of Immigration Statistics, while the data for 1790–1819 is imputed from Tucker (1843). To construct 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. The data on European migrant inflows comes from Willcox (1929).

First Year Connected to Rail. The first year when at least one railroad intersects a county. Data source: Atack, Bateman, Haines and Margo (2010).

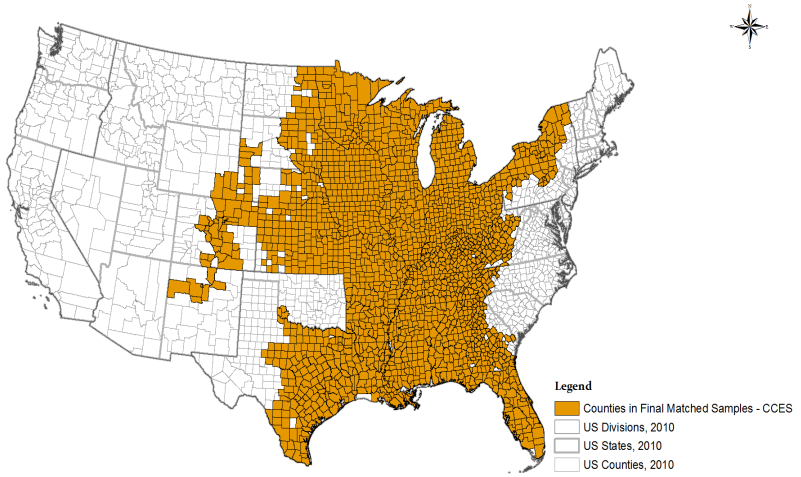
Birthplace Fractionalization. We take $1 - \sum_o (birthplace_{oc}/population_c)^2$, which is simply 1 minus the Herfindahl concentration index for origin o birthplace diversity in county c. Birthplaces are defined as either a given state within the U.S. or a given country or country grouping abroad. Data used is for 1870. Data source: The 1870 data uses the NAPP full count census data.

Figure C.1: Data Availability For Main Survey Data Sources

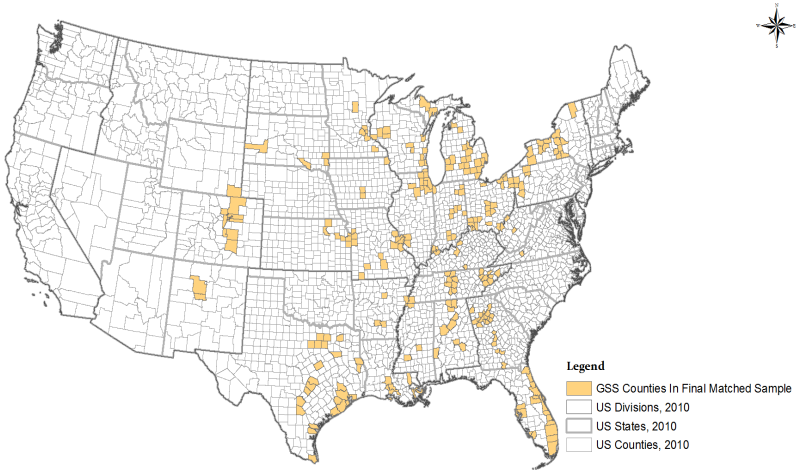
(a) ANES



(b) CCES



(c) GSS



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. Coverage expands to additional counties when incorporating the West Coast sample or extending the historical frontier window to 1950 (see Section 5.5). Note that not all the counties in the above map are included in every baseline regression using the corresponding survey data.