Homeward Bound: How Migrants Seek Out Familiar Climates*

Marguerite Obolensky, Marco Tabellini, and Charles A Taylor†

November 2023

Abstract

This paper introduces the concept of "climate matching" as a driver of migration. Using historical censuses and administrative data, we document several novel findings. First, we show that climate strongly predicts the spatial distribution of international and domestic migrants in the US, both historically (1850-1940) and more recently (2011-2019), whereby movers select destinations with climates similar to those in their origin. Second, we analyze historical flows of German, Norwegian, and domestic migrants in the US and find that climate sorting also holds *within* countries. Third, we exploit variation in the long-run *change* in average US climate from 1900 to 2019 and find evidence of climate matching over time, with migration increasing between locations whose climate converged. Our results hold across time, geography, and migrant groups, and are not driven by the persistence of ethnic networks or other confounders. We provide evidence for two complementary mechanisms: climate-specific human capital and climate as amenity. We then back out the marginal value of climate by: *i*) exploiting the 1862 Homestead Act, a public policy that changed relative land prices; and, *ii*) estimating the effect of climate mismatch on life expectancy. We conclude by projecting how climate change can reshape the geography of US internal migration throughout the 21st century.

*We thank David Albouy, Chris Boulos, Tamma Carleton, Leonardo D'Amico, Tatyana Deryugina, James Feigenbaum, Martin Fiszbein, Nicola Gennaioli, Edward Glaeser, Joe Henrich, Andrew Hultgren, Peter Huybers, Kelsey Jack, William Kerr, Eliana La Ferrara, Ross Mattheis, Eoin McGuirk, Tom Nicholas, Maria Petrova, Ambra Seck, Marlon Seror, Jesse Shapiro, Andrei Shleifer, Felipe Valencia, Nico Voigtlaender, Matt Weinzierl, Guglielmo Zappalà, Richard Zeckhauser, and participants at conferences and seminars for useful comments. Alessio Corò, Jessica Greer, Marta Leva, Emanuele Licari, Marco Panunzi, Monia Tomasella, Francesco Toti, Hao Yang, and Zhengyang Zhou provided excellent research assistance. All remaining errors are ours.

†Obolensky: School of International and Public Affairs, Columbia University (m.obolensky@columbia.edu); Tabellini: Harvard Business School, NBER, CEPR, and IZA (mtabellini@hbs.edu); Taylor: Harvard Kennedy School, Harvard University (ctaylor@hks.harvard.edu).

"What a glorious new Scandinavia might not Minnesota become! The climate...agrees with our people better than that of any other of the American States." – Fredrika Bremer, 1850

1 Introduction

Climate change is an increasingly important driver of global migration (Stern, 2007; Missirian and Schlenker, 2017), but little is known about how average climate influences migration patterns more generally. Do migrants select their destination based on climate? Do people from wet and mild places seek out similar climates? Answering these questions is crucial as governments try to adapt to climate change. Recent work suggests that migration may mitigate the direct effects of climate change (Bosetti et al., 2021; Desmet and Rossi-Hansberg, 2023), but that climate-induced migration may also increase conflict (McGuirk and Nunn, 2023). Although individuals displaced by weather shocks are often forced to leave their home region (Mahajan and Yang, 2020; Bertoli et al., 2022), their preferences may determine where they move—or, where they would like to move if they were unconstrained. Likewise, preferences over climate may influence the location decisions of people who migrate in anticipation of climate change, as well as economic migrants more generally. Thus, understanding whether climate similarity matters for migration has important welfare implications both for migrants themselves and for receiving countries.

To our knowledge, this is the first paper to present systematic evidence on the relationship between climate similarity and migration. Our analysis spans US international and internal migration for both the historical and the modern periods. The US provides an ideal setting to study the role of climate preferences on migration for several reasons. First, the US is unique in its inherent climate variability: the continental US spans climates ranging from arid to temperate to tropical. Second, few restrictions on migration from European countries existed during the Age of Mass Migration (1850-1920), when more than 30 million people moved from Europe to the US.[1] Likewise, there are few formal impediments to domestic migration in the US, both historically and in the present day.[2] The relative lack of internal and international restrictions on mobility and the wide range of climate zones imply that migrants have a large variety of climatic options to choose from. Finally, rich datasets allow us to precisely measure the relationship between climate similarity and migration.

[1] In 1917, US Congress mandated a literacy test for all immigrants (Goldin, 1994). This was followed by the 1921 and 1924 Immigration Acts, which drastically reduced European immigration until 1965. The quotas did not apply to migrants from Mexico and Canada, but stringent restrictions had been introduced against Chinese and Japanese immigrants in 1882 and 1907, respectively (Abramitzky and Boustan, 2017).

[2] Although the movement of white individuals within the US was largely unregulated, formal and informal restrictions severely limited movement of African Americans and Native Americans (Alston and Ferrie, 1999; Nichols, 2014).

Figure 1 summarizes our central result in a scatterplot of the average temperatures where immigrants settled in the US in 1880 against the average temperature in each migrant-sending country of origin. This figure indicates that immigrants coming from warmer (colder) countries settled in similarly warmer (colder) parts of the US. The migrant-mix in the US in 1880 included large groups with well-established networks (e.g., the Irish and the Germans) as well as pioneers from countries that would later see mass migration (e.g., Italy, Japan, or Poland). While similar patterns hold in other years (e.g., 1900, 1920, 1940), this earlier period reduces concerns that Figure 1 merely captures the persistence of ethnic enclaves—a possibility that we further investigate and rule out.[3]

The paper will show that the positive association between climate and migration is remarkably robust, holding across time periods, geographies, migrant groups, and definitions of climate similarity. In Section 2, we provide descriptive evidence from a variety of contexts using different approaches. After verifying that results in Figure 1 hold across time periods, including today, and do not merely reflect factors such as geographic distance, country of origin GDP, or past migration, we turn to a novel dataset. We use US mortality records for 1959-1961, which track the country of origin of each foreign-born individual who died in the US. Comparing the temperature in the capital city of the country of origin to the temperature in the US county of death of foreign-born individuals, we confirm the positive association between climate at origin and climate at destination in Figure 1.

Next, we move beyond cross-country evidence to explore whether the climate-migration relationship holds *within* countries. First, we focus on German immigrants at the turn of the 20$^{\text{th}}$ century—when this group accounted for 30% of the US foreign-born population—and develop a novel measure of climate similarity based on the spatial distribution of surnames in Germany and in the US. Second, we consider Norwegian immigrants between 1865 and 1880, when around 250,000 individuals (or, 15% of the 1865 Norwegian population) embarked for the US. Using automated algorithms developed in the literature (Abramitzky et al., 2021), we link the 1865 Norwegian and the 1880 US Censuses to match individuals across locations (and the corresponding climates) over time. Third, we turn to US internal migration from 1850 to 1940—a period characterized by the westward expansion of the United States (Bazzi et al., 2020). We rely on linked datasets assembled by Abramitzky et al. (2020) to follow migrants moving across US counties from one decade to the next. In all cases, the relationship depicted in Figure 1 across countries also holds within countries. That is, German, Norwegian, and US domestic migrants each tended to match on climate: individuals from colder (warmer) origins settled in relatively colder (warmer) destinations in the US.

[3] In 1880, approximately 85% of the US foreign-born population came from Europe (Abramitzky and Boustan, 2017). Results in Figure 1 are unchanged when restricting attention to European immigrants.

Motivated by the patterns described thus far, in Section 3 we estimate the effects of climate distance on migration at the origin-by-destination level with gravity equations that are common in the trade literature (Anderson and Van Wincoop, 2003). We focus on the two settings for which we can measure origin and destination climates at a highly granular level. First, we consider the linked dataset of Norwegian migrants mentioned above. We collapse the individual-level data to the Norwegian municipality by US county level, defining the climate distance between origin and destination as the absolute value of the difference in the average climate (temperature or precipitation) in each area. We show that municipality-to-county migration is higher for pairs with more similar climates. Specifically, we find that a decline in temperature and precipitation distance of 7.5°C and 40 mm per month—roughly equivalent to the mean sample distance, or the annual temperature difference between New York City and Oslo—increases migration between origin and destination by .86% and .24%, respectively.

When examining the relationship between climate and migration, a key concern is that climate is spatially correlated.[4] For this reason, any association between the two variables may simply reflect geographic, rather than climate, distance. Importantly, our Norway-US estimates are unlikely to suffer from this problem because the transatlantic component of migration breaks the potential spatial correlation of climate. Nevertheless, we show that our results hold when controlling for US county and Norwegian municipality fixed effects as well as for the difference in several county-municipality pair variables—including other geographic features (e.g., elevation and ruggedness) and measures of economic activity (e.g., employment shares across sectors)—and when using alternative linking methods to match immigrants across Norwegian and US Censuses. Results are also stable across Census-linking time periods (e.g., 1865-1870; 1865-1900; 1900-1910; and, 1900-1920).

Second, we consider US internal migration over the century spanning 1850 to 1940, collapsing and stacking the data at the county pair-decade level to estimate gravity regressions. Concerns about the spatial correlation of climate may be more relevant for US internal migration, since one might expect climate and geographic distances to be correlated. In our setting, however, it turns out that geographic distance explains at most 22% and 31% of the variation in, respectively, temperature and precipitation between county-pairs. To err on the side of caution, all our analyses control for county-pair geographic distance. In addition, we *i*) account for non-linearities in geographic distance, *ii*) consider the direction of move (e.g., east-west; north-south) separately, *iii*) control for distance in terms of latitude and longitude, *iv*) drop migration between neighboring counties and within states, *v*) exclude potentially anomalous states (such as California) and counties, *vi*) use travel cost measures other than physical distance that vary

[4] Climate is no exception to the first law of geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

over time due to rail expansion, and *vii*) adjust standard errors to account for spatial correlation in multiple ways.

Regardless of the specification and approach used, climate distance strongly predicts bilateral migration flows—consistent with our international migration results. Reducing temperature and precipitation distance between any county-pair by 5.5°C and 31 mm per month (approximately equal to the mean distances in the internal migration sample) increases migration by 1.4% and .31%, respectively. These magnitudes are roughly equivalent to the effect of increasing the geographic distance between counties by 1,000 km (e.g., New York City to Detroit) and 220 km (e.g., Baltimore to Richmond), respectively. Compared to the US-Norwegian context, coefficients are somewhat larger for domestic migration, but of overall similar magnitudes; moreover, in both cases the effect of temperature is stronger than precipitation.

Besides the exercises already described to account for the potentially confounding effect of climate distance, we verify that results hold when: *i*) considering each decade separately; *ii*) accounting for the county-pair difference in several economic, demographic, and geographic variables (including ruggedness, elevation, coastal access, and soil type); *iii*) controlling for other forces shaping population movement and economic activity during our sample period, such as exposure to the frontier (Turner, 2017; Bazzi et al., 2020), market access (Donaldson and Hornbeck, 2016), and connection to railroads (Atack and Margo, 2011); and, *iv*) defining climate distance in different ways. Because our results are obtained using linked samples, we address concerns about potential bias due to false positive matches and lack of representativeness (Bailey et al., 2020). We use the full count 1940 US Census, which asked individuals where they lived 5 years earlier, to construct a county-pair migration matrix that does not rely on linking algorithms. Reassuringly, we continue to find that climate similarity strongly predicts migration.

One may be worried that our estimates partly reflect the persistence of cultural, social, or economic factors correlated with both climate and migration—even after controlling for our large battery of county-pair covariates. To address this potential concern, we leverage shifts in average climate across counties throughout the 20$^{\text{th}}$ century, which are influenced by multi-decadal oceanic oscillations (e.g., North Atlantic Oscillation) as well as anthropogenic climate change. We find that *changes* in climate distance from 1900 to 2019 predict the *change* in migration patterns between county-pairs over the same period. In other words, if the temperature converged between two counties over the 20$^{\text{th}}$ century, migration between those two counties increased. Besides reducing concerns of identification, these results have important policy implications in light of the projected climate change, which we discuss in further detail below.

In Section 4, we turn to the mechanisms. We start by exploring the role of ethnic enclaves—

examining the possibility that climate similarity determined the location of pioneers, and that subsequent migrants from the same origin merely followed the footsteps of the first movers without taking climate into account. While this scenario would not alter the validity of our findings, it would change their interpretation. To account for such forces, we replicate the analysis controlling for migration flows in previous decades from a given origin to a given destination. We also limit the sample to newly-settled destinations, where the scope for ethnic networks to influence migrants' location decision is limited. Notably, results remain unchanged: we continue to detect a strong relationship between climate similarity and migration.

Then, we consider two broad classes of complementary mechanisms. First are those capturing climate-specific human capital and skills (Steckel, 1982; Bazzi et al., 2016), such as farming expertise and techniques for surviving extreme heat or cold events. Supporting this channel, results are larger in magnitude for farmers than for non-farmers and become smaller over time as the share of the US population working in agriculture declined. Because skill transferability matters more in more weather-exposed occupations, such as agriculture, these patterns suggest that climate-specific human capital is an important, though not the exclusive, driver of our results.

The second class of factors driving climate matching involve climate as an amenity. Stable and predictable climates favor the transmission of cultural norms across generations (Giuliano and Nunn, 2021), and migrants might value a familiar climate to the extent that it facilitates the persistence of cultural practices. Consistent with this channel, we find that climate distance influences migration also for non-farmers, who work in sectors that are less tied to climate and where climate-specific human capital is less relevant than for farmers. Moreover, using present day data from the US Internal Revenue Service (IRS), we document that climate similarity predicts US domestic migration even at a time when US workers are far more insulated from climate than in the past.

In Section 5, we discuss the implications of our findings. First, we estimate the marginal value of climate for migrants. In a first exercise, we leverage temporal and spatial variation across US counties generated by the 1862 Homestead Act, which transferred 10% of US land from the federal government to 1.6 million farmers (Edwards et al., 2017). We find that the Homestead Act, by subsidizing land, reduced the role of climate distance in governing migration decisions. Contrasting counterfactual migration flows predicted by our estimates, we calculate that the Homestead Act increased the average temperature mismatch of migrants by .1°C. Using price differentials between Homestead land and other federal lands available for purchase, we estimate the value of an additional 1°C in climate similarity to be about $4,500 in current dollars.

In another exercise, we use the mortality records encompassing the universe of foreign-born

individuals who died in the US between 1959 and 1961 to test the relationship between climate mismatch and life expectancy. We estimate individual-level regressions that relate age at death to the temperature distance in the country of birth vs the US county of death, controlling for county of death and country of origin fixed effects as well as for individual demographic characteristics. Our analysis indicates that 1°C of climate mismatch is associated with approximately one month reduction in lifespan. Expressing these results in terms of the Environmental Protection Agency (EPA)'s value of a statistical life of $9.6 million (in 2020 dollars) implies that one additional Celsius degree similarity is worth around $15,800 per person in today's dollars.

Second, we illustrate how our estimates can be used to predict the future of US domestic migration patterns induced by global warming. Since temperature will not rise uniformly across space, climate change will alter the geography of county-pair climate distances and, in turn, county-pair connectivity and migration flows. To formalize this intuition, we rely on localized climate change projections for the 21st century to calculate the percent change in average temperature distance between any county-pair. Combining this with the elasticity of migration with respect to climate distance obtained in Section 3, we compute a measure of climate-connectivity that is similar in spirit to the notion of (rail-induced) "market access" proposed in Donaldson and Hornbeck (2016). This exercise indicates that, because of global warming, the US South will become more connected to the US Northeast, with potentially important implications for future US domestic migration patterns.

Taken together, our evidence indicates that climate similarity has significant economic value to people. This has important policy implications. For instance, refugee resettlement programs (within and between countries) designed without taking climate into account may generate large welfare losses, both for refugees and for residents of receiving countries. In this respect, our paper is also relevant to the growing literature on the global impacts of climate change. Recent papers seek to understand the effects of climate shocks on the global economy through general equilibrium models that assume that the migration costs underlying individuals' location and occupational choices are exogenous to climate change (Cruz and Rossi-Hansberg, 2021; Desmet and Rossi-Hansberg, 2023; Bilal and Rossi-Hansberg, 2023). Our findings indicate that migrants place a large weight on climate similarity between origin and destination. The value that migrants place on climate distance is a travel cost that depends on the spatial distribution of climate change. Thus, our results can be used to enrich these models, improving the precision of welfare estimates and the assessment of the macroeconomic effects of climate change.

Our paper also speaks to recent work that has used hedonic approaches to cross-sectionally estimate the value of climate as an amenity, finding that, on average, people prefer mild temperatures and seek to avoid excess heat and cold (Albouy et al., 2016; Sinha et al., 2021). These papers assume climate preferences to be homogeneous across individuals, and abstract

from the notion that migrants might value climate similarity. A notable exception is Albouy et al. (2021), who compare the location choice of domestic and international migrants across US cities between 2000 and 2014, finding a positive correlation between selected features in the country of origin and city of destination—from safety to coastal proximity to number of winter days with mild temperatures. We complement this work in several ways. First, we provide systematic evidence on the relationship between origin and destination climate using a broad and generalizable measure like mean annual temperature—which is easily measured, understood, and translatable into climate change models. Second, we show that this relationship is robust to controlling flexibly for the spatial correlation of climate and a wide array of non-geographic characteristics, and that it holds across settings, time periods, and geographic areas. Third, we explore the channels behind the relationship between climate similarity and migration. Fourth, we use two alternate approaches to estimate the marginal value of climate similarity, and we illustrate how our results can be applied in the context of US domestic migration in response to future climate change.

Finally, our findings are consistent with long-discussed anecdotes about migration patterns, like why Scandinavians settled in cold upper Midwestern states like Minnesota and North Dakota. As remarked in 1925 by President Calvin Coolidge: "It is frequently noted regarding immigration that the newcomers from Europe commonly sought climatic conditions here like those in which they had been raised. So the Scandinavians are found chiefly in the northern parts of this country" (Coolidge, 1926, p. 255). Similar arguments have been made about early English settlement patterns in the colonial US (Fischer, 1989). Our results also resonate with the discussion in Steckel (1982), who conjectured that historical patterns of east-west (as opposed to north-south) US internal migration were driven by climate matching. To the best of our knowledge, we provide the first piece of systematic evidence in support of these ideas. By shedding light on a new, important driver of migration—namely, climate similarity—at a critical point in American history, we also contribute to the large literature on US internal (Bazzi et al., 2020, 2023a; Zimran, 2023) and international (Eriksson, 2020; Abramitzky and Boustan, 2022; Collins and Zimran, 2023) migration between 1850 and 1940.

2 Motivating Evidence: Climate and Migration

Cross-country evidence from US Censuses. We first provide additional evidence in support of our headline Figure 1 to showcase the robustness of the positive relationship between climate at origin and climate at destination among migrants. In Panels B to D of Figure A1, we show that results hold for subsequent decades (1900, 1920, and 1940). To account for the possibility that confounders are driving the raw correlation shown in the figures above, Figure

A2 controls for: GDP per capita in the country of origin (Panel B); average geographic distance between the capital city of the country of origin and the US counties where individuals from each immigrant group lived in 1880 (Panel C); and, the number of immigrants from each origin country living in each US county in 1870 to proxy for the potential role of ethnic enclaves (Panel D).[5] In all cases, results are virtually unchanged and closely resemble those from the baseline specification. Finally, we replicate Figure A2 for the more recent period. In Figure A3, we document that the positive association between origin-destination temperatures holds as of 2015, when the composition of US immigrants differed starkly from that prevailing in the early 20th century.[6]

Evidence from US death certificates. Next, we use publicly-available US mortality data, derived from individual-level death certificates, for the three years from 1959 to 1961 (National Center for Health Statistics, 2021). US mortality records typically only report whether an individual was born in the US, a US territory, or the "rest of the world." However, for the years 1959-1961, they also include an individual's country (or region) of birth for all locations outside the continental US. Mortality data complement the full count US Census datasets by tracking where in the US the person died—a proxy for where she ended up settling, as opposed to the location reported in the Census, which instead tells us where the person was living at a given point in time. Figure 2 compares the temperature in the capital city of the country of birth to the average temperature in the US county of death of foreign-born individuals from each birth country. Consistent with Figure 1, we find a strong and positive relationship between temperature at origin and temperature at destination.

Surname-level analysis for German immigrants. We complement the cross-country evidence described thus far by exploring whether climate and migration are correlated also *within* countries. We begin by focusing on German immigration to the US in the late 19th century. Between 1850 and 1880, approximately 2,450,000 Germans moved to the US, and, by 1880, they represented the largest immigrant group in the country, accounting for 30% and 4% of the foreign-born and the total population, respectively. Since full count Censuses do not exist for Germany for this historical period, we assemble a novel dataset to assign to German immigrants in the US the climate in their place of origin. We use German WWI casualty data from Verein für Computergenealogie e.V. (2014), which records each individual's name, surname, and place of birth. This allows us to derive the geographic distribution of surnames in Germany at the

[5] GDP data come from Bolt et al. (2018). To maximize the sample size, we use GDP measured in 1920, which is available for 50 countries. Results are unchanged when using GDP for other years.

[6] In 1920, 85.2% of the US foreign-born population came from Europe, while only 3.8% and 1.6% came from Latin America and Asia, respectively. In 2015, Asian and Latin American immigrants accounted for 29.7% and 51.8% of the US foreign-born population, respectively. Only 11.5% immigrants living in the US were born in Europe, instead. Panels B to D of Figure A3 verify that results hold when controlling for other variables—most notably, the historical number of immigrants from each origin in each US county.

turn of the 20th century. We collapse the data to obtain the average temperature associated with each surname. Then, we leverage the distribution of German immigrant men in the 1880 full count US Census to compute the average temperature (across US counties) associated with each German surname (see Appendix B.5 for more details). In Figure 3, we show a binned scatterplot estimated with a surname-level regression that correlates the average temperature in Germany with the average temperature in the US. We find that *within* countries there is a positive and statistically significant relationship between the climate at origin and destination.[7] That is, Germans with names highly prevalent in colder parts of German tended to live in colder parts of the US, and vice versa.

Climate matching at the individual level for Norwegian immigrants. Finally, we consider the experience of individual migrants. We use a linked dataset of Norwegian immigrants, described in more details in Section 3.2 below, that reports their municipality of origin in 1865 and their US county of residence in 1880. In Panel A of Figure 4, we present a binned scatterplot produced by regressing the average temperature of the US county where the immigrant was living in 1880 against the average temperature prevailing in the Norwegian municipality of origin in 1865. The positive and statistically significant coefficient indicates that Norwegian immigrants from colder municipalities within Norway tended to locate in colder counties within the US, as compared to Norwegians coming from milder parts of the country.[8]

Climate matching at the individual level for US domestic migrants. In Panel B of Figure 4, we turn to US domestic migration using a linked sample of individuals who moved across counties in each decade between 1850-1860 and 1930-1940 (see Section 3.3 and Appendix B for more details). We estimate a regression analogous to the one presented in Panel A for Norwegian immigrants, correlating average temperature in the county of destination with average temperature in the county of origin.[9] Confirming our previous results, US domestic migrants originating from colder counties systematically moved to colder destinations. To assuage potential concerns about spatial correlation of climate in the context of internal migration, Panel B of Figure A4 verifies that results are unchanged when controlling for county-pair geographic distance.

[7] While not reported for brevity, results are robust to including a large set of control variables defined as surname-level averages (e.g., distance from New York City; individual and county-level characteristics of German immigrants in the US).

[8] We cluster standard errors at the US state by Norwegian province level. Notably, results hold when controlling for the distance between destination counties and New York City and between origin municipalities and the closest major Norwegian port (Figure A4, Panel A), or any other proxy for the physical distance between origin and destination.

[9] Standard errors are clustered at the state of origin by state of destination by decade level.

3 Main Results

The evidence in Figures 1 to 4 points to the existence of a positive relationship between climate at origin and climate at destination that holds across datasets, at the individual and at the aggregate level, and between and within countries. Motivated by these patterns, we formally estimate the link between climate similarity and migration. We focus on the two contexts that allow for the measurement of climate at origin and destination at a highly granular level: Norwegian immigration to the United States during the Age of Mass Migration and US domestic migration from 1850 to 1940. In Section 3.1, we describe the main variables used in the analyses.[10] In Sections 3.2 and 3.3, we build on the trade literature and estimate gravity equations (Anderson and Van Wincoop, 2003) that model migration flows as a decreasing function of the climate distance between origins and destinations. Finally, in Section 3.4, we leverage long-difference variation in the distribution of climate across US counties from 1900 to 2019 to study if changes in climate distance over time predict changes in migration.

3.1 Variable Definition and Data Sources

Climate. We use temperature and precipitation yearly averages to proxy for a location's climate, and define climate distance as the absolute value of the difference in temperature or precipitation between origin o and destination d.[11] We obtain measures of historical climate for US counties by taking the 1895-1920 average of yearly climate from the NOAA Monthly US Climate Divisional Database (Vose et al., 2014). Because data prior to 1895 are not available at a sufficiently fine resolution for our purposes, we are unable to cover the entire period for which we measure US historical (international and domestic) migration. However, the turn of the 20$^{\text{th}}$ century was not characterized by large shifts in climate, and the climate between 1850 and 1900 was rather similar to that prevailing in the subsequent two decades. To measure climate in Norwegian municipalities, we use TerraClimate monthly gridded data (Abatzoglou et al., 2018). Because there are no historical data that are sufficiently granular to define climate at the municipality level, we take the average of yearly temperature and precipitation between the first available year, 1958, and 1980 to proxy for the climate prevailing in Norway at the turn of the 20$^{\text{th}}$ century. We deem this approximation reasonable, given that the effects of anthropogenic climate change only became noticeable in Norway and elsewhere after 1980 (Burke and Emerick, 2016; Kirchner, 2021).

Figures A5 and A6 plot the distribution of climate in the US and in Norway, respectively. In

[10] We refer the interested reader to Appendix B for more details.

[11] Concretely, we define temperature distance as: $\text{Temperature distance}_{od} = |\text{Temperature}_o - \text{Temperature}_d|$. We use the same approach for precipitation.

both countries, climate gradients follow well-known patterns: it gets colder as one goes north and drier as one goes west (US) and east (Norway). However, geographic features such as mountains and water bodies lead to significant within-country variation in climate. The North Sea and the Norwegian Sea tend to warm neighboring land, whereas low average temperatures are recorded in the Rocky Mountains in the US. The east-west precipitation gradient is interrupted in the US Southeast by southerly winds blowing from the Gulf of Mexico and bringing moisture. These disruptions in otherwise continuous climate patterns provide useful variation for our analysis. Figure A7 shows the geographic distribution of climate distances from two example locations—one in the US South and the other one in the Northeast. Figure A8 displays the climate distance between Oslo (Norway) and each US county. In both cases, it is clear that climate distance and geographic distance are not perfectly correlated—something we discuss in more detail when presenting the results below.

Migration. To study the role of climate similarity in the context of historical Norwegian migration to the US, in Appendix B.2, we use automated algorithms developed in the literature (Abramitzky et al., 2020, 2021) to link the 1865 Norwegian and the 1880 US full count Censuses, made available by IPUMS (The Digital Archive, 2008; Ruggles et al., 2021). We also present results obtained when considering alternative linking periods (1865-1870; 1865-1900; 1900-1910; and, 1900-1920). For domestic migration, we rely on linked samples from 1850 to 1940 assembled by Abramitzky et al. (2020) and available through the Census Linking Project.[12] Either in the 1865-1880 Norwegian or in the US domestic linked samples, we do not know the exact year of migration. We thus impose the restriction that individuals must be at least 15 years old in the previous decade.[13] We refer the interested reader to Appendix B.2 and B.3 for more details on the construction and the characteristics of the linked samples. We complement the domestic migration analysis using the 1940 full count US Census, which asked individuals their county of residence 5 years before, to derive a county-to-county migration matrix that does not rely on linked data. When presenting additional results throughout the paper, we also derive migration flows using different datasets, which are introduced when they become relevant and are described in more detail in Appendix B.

[12] Because women often changed surnames upon marriage, standard linking algorithms focus on men (however, see Olivetti and Paserman, 2015 and Althoff et al., 2022 for exceptions). Following Abramitzky et al. (2021), we restrict our analysis to men.

[13] This implies that, for the Norwegian sample, we restrict attention to men who were at least 10 years old in 1865. Since year of immigration becomes available in the US Census from 1900 onwards, when focusing on Norwegian linked samples that are derived using US Censuses for 1900 or later, we restrict attention to men who were 15 years old in the immigration year. See Appendix B.2 for more details. Results are unchanged when removing any age restriction or using different age ranges.

3.2 Norwegian Immigration

Historical context. The first record of Norwegian immigrants to the US dates back to 1825, when 30 families moved to Pennsylvania. More migrants followed in the 1850s, although their numbers remained relatively limited. From the onset, Norwegian pioneers sent letters to their families and friends back home praising the quality of the US soil and emphasizing the similarity between the destination and the origin climates. Describing the conditions prevailing in Vernon, Wisconsin, in 1842, Norwegian immigrant Ole Knudsen Trovatten noted that there was "[...] no difference between the climate here and in Norway. To be sure, some days in the summer are warmer here—but the warmth is not excessive" (Blegen, 1955, p. 433). Along similar lines, Paul Hjelm-Hansen, who settled in Alexandria, Minnesota, wrote to his relatives that "Winter sets in in November with about the same degree of cold as in western Norway, and it continues, as a rule, into March" (Blegen, 1955, p. 441).

Norwegian mass migration did not take off until the 1860s, when severe frosts led to pervasive crop failures and pushed thousands of individuals out of the country. Between 1865 and 1880 alone, 250,000 individuals (15% of Norway's population in 1865) migrated to the United States (Semmingsen, 1960). In 1880, Norwegian immigrants accounted for about .4% and 2.7% of the total and the foreign-born US population, respectively. The size of the Norwegian population in the US continued to increase until the early 1910s, and another large wave of migration was recorded between 1903 and 1910, when around 200,000 individuals left Norway for America (Semmingsen, 1978). After 1910, the number of Norwegians living in the US began to decline, partly as a result of the Immigration Acts of the 1920s. Overall, it is estimated that, between 1825 and 1920, more than 1 million individuals moved to the United States (Eriksson, 2020).

At the onset of Norwegian mass migration, Norway was a predominantly rural country: in 1865, only 9.6% of the male population (10+) resided in urban areas, and about 50% of men (15-64) in the labor force worked in agriculture (Table A1, columns 1-3).[14] Norwegians who can be linked to either the 1880 US or the 1900 Norwegian Census are similar to those in the full count, along these and other characteristics (Table A1, columns 4-6). In Table A2, we compare the baseline characteristics of individuals in the linked sample who stayed in Norway to those of individuals who migrated to the US. Consistent with the evidence in Abramitzky et al. (2012), some differences emerge. Movers were more likely to live in urban areas, to be in the labor force, and to work in manufacturing (rather than in agriculture or farming).

Figure A9 contrasts the geographic distribution of Norwegian men in the full count (Panel A) and that of stayers (Panel B) and migrants (Panel C) in the linked sample. It shows that

[14] Numbers are very similar when considering both men and women and dropping age restrictions, which are introduced to more meaningfully compare the characteristics of individuals in the linked sample with those of people in the full count. See also Appendix B.2 for more details on the construction and the characteristics of the linked sample.

migrants tended to live in the southeast of Norway, whereas stayers were more evenly distributed across all Norwegian municipalities. Given east-west precipitation gradients in Norway, this might explain why migrants experienced lower precipitation levels than stayers (97 vs 99 mm per month on average). Because temperature does not follow clear gradients in Norway, the average temperature of stayers and migrants was instead very similar (Table A2).

The residential choices and the occupations of Norwegian men in the US mirrored those prevailing in the country of origin. In 1880, about 17% of Norwegian men in the US lived in urban areas, and approximately 63% of men in the labor force worked in agriculture (Table A3). In 1880, Norwegian enclaves were concentrated in Minnesota, Wisconsin, and Iowa, although some immigrant communities could also be found in parts of California and Washington state (Figure A10). Over time, immigrant settlements expanded, and by 1920, Norwegians lived in most states and counties, except for the South and the East (Eriksson, 2020). Notably, as for the Norwegian Census, the characteristics of Norwegian immigrants in the full count US Census and in the linked samples are very similar to each other in 1880 along both socio-economic and environmental dimensions (Table A3).[15]

Baseline estimates. Using the sample of individuals linked across the 1865 Norwegian and the 1880 US Censuses, we collapse the data at the Norwegian municipality of origin (o) by US county of destination (d). We restrict attention to US counties that, as of 1880, had at least one foreign-born individual, to proxy for the set of destinations available to Norwegian immigrants.[16] Since the migration matrix is sparse, as commonly done in the literature (Silva and Tenreyro, 2006), we use Poisson Pseudo Maximum Likelihood (PPML), rather than OLS. We estimate:

$$M_{od} = \exp[\alpha_p + \gamma_s + X_o + X_d + \beta_1 \text{Dist}_{od}^{\text{Temp}} + \beta_2 \text{Dist}_{od}^{\text{Precip}}]\epsilon_{od} \tag{1}$$

where M_{od} is the number of immigrants from Norwegian municipality o to US county d between 1865 and 1880; $\text{Dist}_{od}^{\text{Temp}}$ and $\text{Dist}_{od}^{\text{Precip}}$ are, respectively, absolute temperature and precipitation differences between o and d; X_o and X_d are vectors of Norwegian municipality and US county controls described below; and, α_p and γ_s are Norwegian province and US state fixed effects. We cluster standard errors at the Norwegian province by US state level.

We present results in Table 1, where we also report standardized beta coefficients in square brackets. In column 1, we estimate a parsimonious specification that only includes temperature and precipitation distance. Coefficients are negative and statistically significant, indicating

[15] This holds also for the other years considered in our analysis below. Statistics for these additional Census decades are not reported for brevity.

[16] As we show below, results are unchanged when dropping this restriction or when further narrowing the set of potential US destinations.

that higher temperature and precipitation distances are associated with lower migration. The transatlantic component of migration generates a "geographic break" that reduces concerns that our estimates may be confounded by the spatial correlation of climate. In other words, all migrants must cross an ocean greater in distance than what they would travel either in Norway or within the US. The large travel distance also eliminates the possibility that we are capturing the effect of migrants moving to a neighboring town that happens to have the same climate.[17] To more explicitly account for the effect of physical distance, in column 2, we include the distance between the US destination county and New York City, as well and the Norwegian municipality and the closest transatlantic port.[18] Coefficients are unchanged.

In column 3, we add US state and Norwegian province fixed effects. The point estimates, especially for temperature distance, become smaller in absolute value, but they remain negative and highly statistically significant. The drop in the magnitude of coefficients should not be surprising. When adding province and state fixed effects, we are effectively comparing migration flows: *i*) between municipalities within the same Norwegian province and any US county; and *ii*) between any Norwegian municipality and counties within the same US state.[19] It is possible that, even within a US state or a Norwegian province, climate similarity might be correlated with other origin and destination characteristics, which may, in turn, influence bilateral migration flows. For example, migrants from municipalities with more limited economic opportunities may select US counties with a stronger economy. Likewise, migrants from more rural and agricultural municipalities might sort into destinations with similar economic and residential structures.

If such forces are also correlated with climate distance, our estimates may be biased. For this reason, in column 4, we add a large vector of US county and Norwegian municipality controls.[20] We include variables that can be measured consistently in the Norwegian and in the US Censuses: population density, the urban population share, sex ratios, the share of men in the labor force, and the employment share in agriculture and manufacturing. For the US specifically, we also add: the immigrant, the Norwegian, and the Black population share; total frontier experience from Bazzi et al. (2020); and the measure of market access from Donaldson and Hornbeck (2016).[21] Coefficients remain negative and statistically significant.

[17] Recall that in Figure A8, which displays the climate distance between Oslo (Norway) and each US county, we see significant spatial variation.

[18] We consider the two major ports of this period—Kristiana (Oslo) and Kristiansand—but results are unchanged when including other ports (e.g., Bergen).

[19] In our sample, there are only 20 provinces and 435 (historical) Norwegian municipalities, with an average of 21 municipalities within each province.

[20] Because the economic, demographic, and urban structure may be endogenous to climate, these variables may be "bad controls" (Angrist and Pischke, 2009).

[21] All Norwegian and US controls are measured, respectively, in 1865 and 1880. See Appendix B for more details. Results are robust to including more controls or measuring US variables in 1870 or (when available) 1860.

According to the estimates in column 4, which we take as our preferred specification, reducing temperature and precipitation distance by 7.5°C and 40 mm per month in absolute levels (approximately equal to the mean sample distance, or the annual temperature difference between New York City and Oslo) increases migration by .80% and .24%, respectively. The effect of temperature distance is quantitatively relevant and comparable to that of increasing the physical distance between New York City and a US county by about 1,600 km (e.g., from New York City to Minneapolis). One interesting feature of our results is that the effects of climate distance on migration are larger for temperature than for precipitation. Indeed, the standardized beta coefficient on temperature distance (which is very similar to that on physical distance) is 2.2 times larger (in absolute value) than that on precipitation distance. As we show below, this is a pattern that we systematically observe across settings and time periods.

Climate similarity vs other forces. The stability of coefficients to the inclusion of Norwegian municipality and US county controls reduces concerns that our results might reflect forces other than climate similarity. To more cleanly isolate the role of climate, in Table A4 we estimate more stringent specifications. In column 2, we replace the Norwegian municipality (X_o) and US county (X_d) controls with their difference (X_{od}).[22] In column 3, we simultaneously include X_o, X_d, and X_{od}. In both cases, coefficients remain very similar to those from our preferred specification (reported in column 1 to ease comparison).

In column 4, we replace the vector of controls X_o with municipality fixed effects, thereby absorbing any origin characteristic that might trigger migration to destinations that are systematically a better or worse climate match. Coefficients remain negative and precisely estimated—if anything, the effect becomes stronger. In column 5, we replicate our baseline specification by replacing X_d with US county fixed effects, which account for county-specific pull forces that might attract Norwegian immigrants from origins with a more (or less) similar climate. Coefficients drop, but remain statistically significant and, especially for temperature, quantitatively large. In column 6, we estimate a very demanding specification that simultaneously includes both Norwegian municipality and US county fixed effects. In column 7, we further add the vector of origin-destination difference controls, X_{od}. Even though the coefficient on precipitation distance drops to zero and becomes statistically insignificant, that on temperature distance remains precisely estimated and quantitatively relevant.

Findings in Tables 1 and A4 support the notion that climate distance is an important driver of migration. However, one may wonder whether our results also pick up the effect of other geographic features correlated with climate. For instance, Albouy et al. (2021) find that immigrants moving to the US in the present day tend to select cities that are similar to their

[22] We collect in X_{od} only those variables that can be measured for both Norway and the US. We continue to include in the regression the set of controls that can be measured only at the county-level.

countries of origin in terms of distance from coast, elevation, temperate winters, and number of sunny days. To address this possibility, in Table A5 we augment the preferred specification (reported in column 1) by controlling for the origin-destination difference in elevation (column 2) and ruggedness (column 3). We also control for the coastal status of US counties (column 4). In all cases, the coefficients on climate distance are in line with those from the preferred specification. Results also hold when including all additional geographic features simultaneously (column 5).

Sample restrictions and alternative linking methods. Our results point toward the importance of climate similarity in migration decisions. This relationship is unlikely to be confounded by other factors and, as documented in Section 4.1, holds across samples and time periods. We conclude this section by presenting two additional sets of robustness checks. First, in Table A6 we replicate the analysis conducted in Table 1 by considering different sets of origins and destinations. In the baseline specification (restated in column 1 of Table A6 to ease comparisons), we consider all US counties with at least one immigrant in 1880 as a proxy for destinations available to Norwegian migrants. In column 2, we consider all US counties, irrespective of the presence of foreign (or native) born individuals. Columns 3 and 4 restrict attention to counties with at least one Norwegian immigrant in the 1880 full count US Census and in the 1880 linked sample, respectively. Columns 5 and 6 (resp., column 7) drop US counties and Norwegian municipalities (resp., county-municipality pairs) in the top 1% of the linked sample population (resp., number of migrants). Finally, column 8 excludes all US counties within Minnesota—the most common destination of Norwegian immigrants during the 1865-1880 period. In all cases, results remain in line with those from the baseline specification.

Second, in Table A7 we address the potential concern that our estimates may be biased due to false positive matches obtained in our linking procedure (Bailey et al., 2020). Column 1 replicates the baseline specification (Table 1, column 4), where the linked sample used to derive the number of immigrants is restricted to men 10+ in 1865. In columns 2, 3, and 4, we consider men: 15-40 in 1865 (in Norway), 18-65 in 1880 (in the US), and without age restrictions. Columns 5 and 6 derive the linked sample using, respectively, the Jaro-Winkler (JW) and the conservative version of the New York State Identification and Intelligence System (NYSIIS) algorithms from Abramitzky et al. (2012).[23] In all cases, results are unchanged.

[23] Relative to the baseline linking method, the JW algorithm further restricts the sample to matches with names that are similar enough according to the JW similarity measure (we set the same threshold as in Abramitzky et al., 2012, but results are unchanged when using alternatives). The conservative version of the NYSIIS algorithm selects only individuals unique within a 5-years window around their birth date in terms of the matching variables.

3.3 US Internal Migration

Historical context. Between 1860 and 1940, the geographic distribution of the US population changed dramatically. This is shown in Figure A11, which plots population density across counties from 1860 (Panel A) to 1940 (Panel D). In 1860, of the 31.4 million people enumerated by the US Census, 36% and 20% lived in the Northeast and the South, respectively.[24] By 1900, fueled by mass European immigration, the US population had increased to 76.3 million, and its center of gravity had shifted from the East to the Midwest. The US westward expansion, promoted by the diffusion of the railroad networks especially between 1860 and 1900 (Fogel, 1964; Donaldson and Hornbeck, 2016), continued well into the 20th century. The first four decades of the 20th century were also marked by the Great Migration of 5 million whites and 1.5 million African Americans from the US South to the rest of the country (Gregory, 2006; Bazzi et al., 2023a). Internal migration patterns changed in the 1930s, due to the Great Depression (Rosenbloom, 2002; Fishback et al., 2006) and environmental shocks like the Dust Bowl (Hornbeck, 2012, 2023).

Demographic and geographic changes were accompanied by massive economic and social transformations. In Table A8, we report the characteristics of the male population in 1860 (Panel A), 1900 (Panel B), and 1940 (Panel C), for the full count (columns 1-3) and the linked sample (columns 4-6).[25] In 1860, 16.7% of men 15 or older lived in urban areas, and 52% (resp., 12.6%) of men (15-64) in the labor force were employed in agriculture (resp., manufacturing). Over time, cities grew and manufacturing employment rose, due to the rural-urban gradient of migration (Zimran, 2023) and to the process of structural transformation, which was stronger in initially rural counties (Eckert and Peters, 2022; Eckert et al., 2023). In 1940, at the end of our sample period, the majority of the US population lived in urban areas and manufacturing had become as important as agriculture.

As pointed out in the literature (Bailey et al., 2020; Abramitzky et al., 2021), the linked sample is not fully representative of the (male) US population. This is true in our context as well. Compared to the full count, men in the linked sample are more likely to be white, native born, and literate (Tables A8 and A9). In the first part of our sample period, they are also more likely to be farmers, although this trend reverses after 1900. However, the differences are rather small, suggesting that, overall, the linked sample and the full count are comparable to each other.[26]

[24] Note that the 1860 US Census did not include enslaved Black Americans and Native Americans.

[25] As for Norwegian immigrants, we restrict attention to the male population to facilitate the comparison between the full count and the linked sample (see also Appendix B.3). In both the full count and the linked sample, we also impose the restriction that men are 15 or older to select those who were likelier to make the decision of migrating. All results are unchanged when using different age or sample restrictions.

[26] See Appendix B.3 for more details.

In Table A10, we focus on the linked sample and compare the characteristics of stayers (columns 1-3) and migrants (columns 4-6), over the entire period (Panel A) and at different points in time (Panels B and C). Between 1860 and 1900, migrants were more likely to be foreign-born (20.6% vs 11.1%) and live in urban areas (25.3% vs 22.1%), and less likely to be farmers (49.6% vs 58.4%), relative to stayers. However, consistent with the evidence documented in Zimran (2023), the urban-rural gradient of domestic migration changed during the 20$^{\text{th}}$ century: between 1910 and 1940, 50% and 52.8% of migrants and stayers lived in urban areas, respectively, and about 31% of men in both groups were farmers.[27]

Baseline estimates. Using the linked sample described above, we stack the data at the county-pair by decade level, from 1850-1860 to 1930-1940, to derive the number of migrants between any origin and destination in each decade. Then, using PPML as in equation (1), we estimate:

$$M_{odt} = \exp[\alpha_{ot} + \gamma_{dt} + \beta_1 \text{Dist}_{od}^{\text{Physical}} + \beta_2 \text{Dist}_{od}^{\text{Temp}} + \beta_3 \text{Dist}_{od}^{\text{Precip}}]\epsilon_{odt} \tag{2}$$

where migration flows from county o to county d are measured between decade $t-10$ and t, climate distances are defined as before, and $Dist_{od}^{Physical}$ is the physical distance between counties o and d. An important advantage of the US domestic migration setting is that we can rely on panel data (at the county-pair level) for several decades. This allows us to include in our regressions a stringent set of fixed effects. In our preferred specification, we interact decade dummies with both county of origin and county of destination fixed effects, α_{ot} and γ_{dt}. We cluster standard errors at the state of origin by state of destination by decade level.[28]

We report results in Table 2. In column 1, we only include temperature and precipitation distance. As in Table 1, there is a strong, negative, and precisely estimated relationship between climate distance and migration. In column 2, we control for physical distance to reduce concerns that the spatial correlation of climate may bias our results. The point estimates on temperature and precipitation distance decline from -.375 to -.302 and from -.039 to -.014, respectively, but remain highly statistically significant. In column 3, we add county of origin, county of destination, and decade fixed effects. In column 4, we further interact decade dummies with state of origin and state of destination dummies.

In column 5, we present our preferred specification, which includes county of origin by decade as well as county of destination by decade fixed effects. This set of fixed effects absorbs any temporal and location-specific push and pull forces that might be correlated with county-pair climate distances. For instance, suppose that an economic downturn affecting the dairy

[27] All characteristics are measured in the baseline decade.

[28] Results are robust to using alternative clustering structures and to adjusting standard errors with the procedure in Conley (1999).

industry in western New York state (causing out-migration) coincided temporally with an iron ore mining boom in Minnesota (causing in-migration)—and that the particular regions of New York and Minnesota had similar climates (by chance). County of origin by decade and county of destination by decade fixed effects isolate the variation in migration patterns between counties that remained *after* partialling out the average effect of such economic shocks.

In all cases, coefficients remain statistically significant and quantitatively relevant. According to the point estimate in column 5, lowering temperature and precipitation distances by 5.5°C and 31 mm per month (or, about the average distances in our sample) increases migration by 1.4% and .31%, respectively. This is similar to the effect of increasing the physical distance between counties by 1,000 km (e.g., New York City to Detroit) and 220 km (e.g., Baltimore to Richmond), respectively. Compared to the US-Norwegian context, coefficients are now somewhat larger but of overall similar magnitudes. Moreover, as before, we find that the climate effect is stronger for temperature than for precipitation. This result resonates with the climate change economics literature, which has consistently found temperature to be an important factor in influencing health and economic outcomes (Carleton and Hsiang, 2016; Heal and Park, 2016; Deryugina and Hsiang, 2017). The precipitation link is less well established. Furthermore, in the agricultural context, temperature is a better predictor of crop yields than precipitation (Lobell and Burke, 2008; Schlenker and Roberts, 2009). Another reason for the more limited effect of precipitation might be that rainfall is spatially heterogeneous, less precisely measured than temperature, and subject to bias when spatially aggregated (Fezzi and Bateman, 2015). In addition, unlike temperature, the impacts of precipitation anomalies are mediated through soil moisture, and can be managed through cropping practices and irrigation (Proctor et al., 2022; Taylor, 2022).

Spatial correlation of climate. Because climate tends to be spatially correlated, one may be worried that our estimates pick up the influence of physical—rather than climate—distance. This issue is potentially more prescient in the US domestic migration context, where we lack the Atlantic Ocean "break" that characterizes the international setting described earlier. However, while there exist general climate gradients within the US, geographic distance turns out to be a relatively weak predictor of climate. Figure A6 shows that in the US it gets colder as one goes north, and wetter as one goes west, but that geographic features such as mountains and water bodies lead to significant within-country variation in climate.[29] Indeed, a model aiming to predict the temperature distance (resp., precipitation distance) between county-pairs with a flexible function of geographic distance explains at most 22% (resp., 31%) of the variation.[30]

The primary way we address remaining concerns is through controlling for the geographic dis-

[29] To illustrate this point, Figure A7 plots the distribution of climate distances for two example locations within the US.
[30] See Appendix B.1 (Tables B1 and B2) for more details.

tance between county of origin and county of destination—accounting for the fact that most people make short-distance moves (e.g., to a neighboring county) to places that mechanically have similar climates. Since controlling linearly for physical distance may not adequately address all concerns related to the spatial correlation of climate, we perform a series of additional exercises.[31] First, in Table A11 we control non-linearly for geographic distance, allowing climate to vary across space in a non-linear way. Second, in Table A12 we consider each direction of move (e.g., east-west; north-south) separately. Third, in Table A13 we separately control for latitude and longitude distance (and thus for the horizontal or vertical spatial correlation in precipitation and temperature, respectively). Fourth, in Table A14 we drop: neighboring or next-to-neighboring counties, counties that belong to adjacent states, counties within 100 km of geographic distance, and California.[32] Fifth, in Table A15 we allow physical distance to have a time-varying impact on migration by interacting it with decade dummies; and, replace geographic distance with a newly developed measure that takes into account time or cost of travel, which varied over time due to the expansion of the railroads network.[33]

In all cases, the point estimates remain close to those from our baseline. This reaffirms the significance of climate in explaining migration patterns and suggests that results are robust to concerns related to the spatial correlation of climate.

Additional geographic features. So far, we focused on temperature and precipitation distance. However, it is possible that other geographic features correlated with climate might drive our results. As we did above with Norwegian immigration, in Table A16 we augment our baseline specification by controlling for the county-pair distance in the following geographic attributes: ruggedness (column 2), elevation (column 3), and coastal access (column 4). Reassuringly, results are virtually unchanged, also when we include all three variables simultaneously (column 5).

Another feature that might be correlated with climate and that might influence migration, especially for farmers, is soil type. For this reason, in Table A17, we augment the baseline specification by controlling for the bilateral distance of key soil characteristics. Using gridded soil data, we derive the county-area average of four soil measures: bulk density (column 2), organic matter concentration (column 3), soil pH (columns 4), and water availability (column 5).[34] In column 6, we include all variables together. In all cases, coefficients on precipitation

[31] In all the proceeding tables, we report our baseline results in column 1 to ease comparisons.

[32] California is a peculiar state; it attracted large flows of migrants during our sample period, but many of the early settlements were shaped by Pacific ports of entry or mining opportunities (i.e., the Gold Rush).

[33] See Appendix B.6 for a description of the railroad and transportation costs data.

[34] Bulk density is the weight of dry soil per unit volume, with higher values indicating compacted soil. Organic matter concentration is the percentage of decomposed plant and animal material in the soil—features that can influence soil fertility and structure. Soil pH is a measure of soil acidity that affects nutrient availability and the types of crops that can be grown. Water availability refers to the water holding capacity of the soil, which depends on soil texture and structure.

and temperature distance remain unchanged, suggesting that our results are not capturing the potential influence of soil characteristics. Interestingly, Table A17 also indicates that coefficients on soil distance are negative and statistically significant, though smaller in magnitude than those on climate distance. To our knowledge, this is the first piece of systematic evidence in support of the idea, often discussed in historical and anecdotal accounts, that farmer migrants sought out destinations with similar soils.[35]

County-pair economic and demographic differences. One may wonder whether county-pair differences in economic or demographic factors correlated with climate distance might influence the relationship between climate and migration. For instance, even after controlling for the set of fixed effects included in our preferred specification, it is possible that counties with higher climate similarity also have more similar economic structures, which make it easier for migrants to relocate. On the other hand, given the rural-urban migration gradient prevailing in the first part of our sample period (Zimran, 2023), if county-pairs with more similar climates happened to have (by chance) larger differences in the rural population share, then our estimates might not be entirely attributable to climate distance.

To address these and similar concerns, in Figure A12 we replicate our preferred specification (Table 2, column 5), displayed in the first dot from the left, by including the county-pair difference in several characteristics, measured at the beginning of each decade. The second, third, and fourth dots from the left control for the county-pair difference in labor force participation, manufacturing employment share, and agricultural employment share, respectively.[36] The subsequent dots include the difference in, respectively: the Black, the urban, and the immigrant population share; sex ratios; and population density. Finally, we consider three forces that have been shown to shape population movement and economic activity during our sample period: exposure to the frontier (Turner, 2017; Bazzi et al., 2020); market access (Donaldson and Hornbeck, 2016); and connection to railroads (Atack and Margo, 2011).[37] In the very last dot, we include all the above variables simultaneously. In all cases, the coefficients on climate distances remain negative, precisely estimated, and close to those from the baseline specification.

Census linking concerns. Results presented thus far were obtained using linked samples, which offer several advantages but also have two potential shortcomings (Bailey et al., 2020).

[35] Steckel (1982) conjectured that the east-west migration gradient prevailing in the US during the 19th century was partly explained by farmers seeking similar soil types. Another anecdotal example is that of the "Cajun Prairie" region of southwest Louisiana, which was first intensively farmed by Midwestern migrants in the 1880s who were attracted by the similar prairie soils (see https://www.loc.gov/item/sn88064676/).

[36] In all cases, the variables are defined for men 15–64 in the baseline decade. Results are unchanged when using different age thresholds, extending the sample to women, or when considering a larger set of economic outcomes.

[37] We measure frontier exposure as the total number of years a county was on the frontier, according to Bazzi et al. (2020). Because this is a time-invariant control, we interact it with decade fixed effects to allow for differential trends over time. Both market access and dummies for being connected to railroads in a given decade are taken from Donaldson and Hornbeck (2016). The measure of market access is constant after 1920, and so we use this value for subsequent decades.

First, false positive matches may introduce systematic bias in the analysis. Second, individuals in the linked samples are not fully representative of the overall population, leading to questions about the generalizability of results. We address these concerns by using the full count 1940 US Census, which asked individuals where they lived 5 years before. As in Hornbeck (2023), we can thus derive a county-pair migration matrix from the full count Census that does not rely on linking algorithms.

In column 6 of Table 2, we first replicate our baseline specification (column 5) using the linked sample, restricting attention to the 1930-40 decade. Coefficients on climate distance become somewhat lower (in absolute value) than in the entire sample period—a pattern that we discuss when exploring the mechanisms in Section 4 below. However, they remain negative and statistically significant. In column 7, we turn to the county-pair migration flows obtained from the 1940 full count Census. Coefficients remain close to those obtained in the linked sample.[38] Even though we could not replicate this exercise for Norwegian immigration, the similarity of the findings derived using the (US) linked sample and the full count Census is reassuring. This indicates that false positive matches are unlikely to introduce bias in the regressions, and increases confidence in the generalizability of results obtained from the linked sample.

Additional robustness checks. We conclude this section by presenting additional robustness checks to bolster the confidence in our main findings. First, in Table A18 we verify that our estimates are unchanged when measuring climate distance in different ways. In column 1, we replicate our baseline specification, which defines temperature and precipitation as yearly averages. Then, we define temperature and precipitation distance by considering the yearly maximum temperature (column 2) and summer averages (column 3).[39] In column 4, we again use average yearly temperature and precipitation, but we attribute to each decade: *i*) before 1910 the climate computed over the 1895-1910 period; *ii*) after 1910 the climate measured as a rolling average of the 15 previous years (e.g., we assign average climate over the period 1905-1920 to 1910-1920 migration flows). Second, in Table A19 we test the robustness of our findings to: *i*) excluding county-pairs that are at the top 1% and 5% of the distribution of bilateral migration (columns 2-3); *ii*) considering only counties that existed already in 1860 (column 4); *iii*) including only county-pairs that are present in the dataset for the entire period (column 5); and, *iv*) restricting attention to observations with strictly positive migration flows (column 6). Finally, we address concerns about the spatial correlation of climate discussed above by estimating standard errors in different ways. In Table A20, we show that the statistical significance of our estimates is unaffected by: clustering standard errors at the state of origin

[38] Because the sample in columns 6 and 7 is slightly different, in unreported analyses, we verified that results are unchanged when focusing on county-pairs that are present in both the linked and the full count samples.

[39] Note that in column 2 we do not use the yearly maximum of precipitation since precipitation is far more variable day-to-day than temperature, and thus excessively sensitive to outliers.

by state of destination level (column 2); and applying the methodology in Conley (1999) using different lag parameters (columns 3 to 10).

3.4 Long-Run Changes in Climate Distance and Migration

Thus far, we have exploited cross-sectional variation in bilateral climate distance. We showed that our findings are robust to the inclusion of a large set of county-pair controls—such as differences in geographic features (Table A16), soil type (Table A17), railroad connection, frontier exposure as well as many other economic, social, and demographic characteristics (Figure A12). However, one may still be worried that our estimates reflect the influence of other origin-destination specific variables also correlated with climate distance. In this section, we tackle this concern by leveraging variation that arises from *changes* in average climate. Specifically, we test whether the *change* in climate distances between US counties predicts the *change* in migration patterns throughout the 20$^{\text{th}}$ century.

We focus on two periods: a historical window in the early 20$^{\text{th}}$ century and a modern window in the early 21$^{\text{st}}$ century. We measure historical migration using the 1935-1940 migration matrix derived from the 1940 US Census (see also Table 2, column 7). To measure modern migration, we leverage IRS tables that track the number of migrants between each US county-pair from 2011 onwards. In our baseline specification, we consider migration flows recorded between 2018 and 2019.[40] Historical and modern climate are defined by taking the annual average temperature and precipitation over the 1895-1920 and the 1990-2020 periods, respectively.[41] Figure A14 illustrates visually the change in climate distance from two selected US counties—one in the Delta region in the South, and the other one in the Northeast. The map highlights two notable features. First, changes in climate distances vary substantially across the country. Second, the spatial distribution of the change in average climate distances does not follow identifiable geographic patterns.

We replicate our baseline domestic gravity model, specified in equation (2), but allowing climate distances to vary between the early and the modern periods. We report results in Table 3. In column 1, we only include origin county-by-period and destination county-by-period fixed effects as well as physical distance.[42] The negative and statistically significant coefficients indicate that county-pairs that became climatically more distant experienced a larger reduction in migration over time. In column 2, we add county-of-origin by county-of-destination fixed effects, which absorb any county-pair (time-invariant) characteristic, including physical distance.

[40] We describe this data in more detail in Section 4.3 and Appendix B.4.
[41] As for historical climate, data on modern climate come from Vose et al. (2014).
[42] Results are unchanged when using different measures of geographic distance.

We also allow geographic distance to have a heterogeneous effect by interacting it with period dummies. Note that this specification only exploits the *change* in climate distances between county-pairs over the 100-year period we consider. Coefficients are barely affected, and remain negative and statistically significant at the 1% level. According to our estimates, a 20% increase in temperature distance over time—about 1 °C, or the average annual temperature distance between and Boston and Detroit—reduces migration by about 3.5%. Coefficients on the change in precipitation distance are similar in magnitude. To put these numbers in perspective, consider that global temperatures have risen by about 1°C since the late 19$^{\text{th}}$ century due to anthropogenic climate change.

Subsequent columns of Table 3 explore the robustness of these results. In column 3, we interact period dummies with the battery of historical county-pair controls measured in 1930 (see also Figure A12). This specification addresses the possibility that climate distance changed differentially across county-pairs in a way that was correlated with historical origin-destination variables that may have persistent effects on the change in migration. In columns 4 and 5, we measure modern migration flows using 2011-2012 and 2014-2015, respectively. In column 6, we measure historical and modern migration using average migration flows calculated over different averaging windows: from 1900-1910 to 1930-1940 in the early period (from the linked sample) and from 2011-2012 to 2018-2019 for the modern period (from IRS data). In all cases, coefficients on temperature distance are stable and highly robust. Consistent with our previous evidence, the estimates for the effects of precipitation distance are somewhat less robust, and become smaller and imprecisely estimated in column 6.

To more intuitively illustrate the variation exploited in Table 3, we replicate this analysis by taking variables in long differences, as in Burke and Emerick (2016). Using OLS, we estimate:

$$\Delta_t \log(M_{odt}) = \alpha_o + \gamma_d + \beta_1 \Delta_t \log(\text{Dist}_{odt}^{\text{Temp}}) + \beta_2 \Delta_t \log(\text{Dist}_{odt}^{\text{Precip}}) + \epsilon_{od} \tag{3}$$

where $\Delta_t \log(M_{odt})$, $\Delta_t \log(\text{Dist}_{odt}^{\text{Temp}})$, and $\Delta_t \log(\text{Dist}_{odt}^{\text{Precip}})$ are the long-run change in the log of migration and climate distances. Besides its intuitive appeal, this model has a further advantage: since it considers the change in the log of the number of migrants, it effectively exploits only variation between county-pairs with non-zero migration flows in both the historical and the modern periods. This reduces the potential concern that our estimates might pick up compositional changes in the geography of US migration throughout the 20$^{\text{th}}$ century (Molloy et al., 2011; Boustan et al., 2013; Zimran, 2023).

Results are reported in Table A21, which follows the same structure as Table 3. In column 1, we only include county-of-origin and county-of-destination fixed effects, which account for county specific trends. In column 2, we present our preferred specification, further control-

ling for geographic distance. As before, coefficients are negative and statistically significant at the 1% level. Perhaps not surprisingly given that we are only leveraging variation along the intensive margin of migration, the implied magnitudes are somewhat smaller (in absolute value) than in Table 3. The remainder of the table verifies that results are highly robust. In column 3, we include the vector of bilateral controls, allowing county-pairs to be on differential trends according to several historical origin-destination specific variables. In columns 4 to 6, we replicate column 2 measuring migration over different time periods. In all cases, coefficients remain negative, stable, and statistically significant.

4 Mechanisms

In this section, we first document that results cannot be explained by past migration or the persistence of ethnic enclaves alone (Section 4.1). Then, we consider two complementary mechanisms for our findings: climate-specific human capital (Section 4.2) and climate-as-amenity (Section 4.3).

4.1 Ethnic Enclaves and Migration Persistence

Given the persistence of ethnic enclaves and migration patterns documented in the literature (Altonji and Card, 1991; Card, 2001), one may wonder whether this mechanism is, at least in part, responsible for the relationship between migration and climate distance. For instance, early settlers might select destinations because of climate similarity, and latecomers may then follow the same routes because of the presence of familiar people, rather than a familiar climate. Climate similarity and the persistence of ethnic enclaves are not mutually exclusive forces, and they may well coexist with each other. However, understanding whether climate has an independent effect on migration even once ethnic enclaves are established might have important policy implications.

Norwegian immigration. In Figure 5, we focus on Norwegian immigration. We first estimate our preferred specification (Table 1, column 4) on linked samples obtained by matching individuals across different Census years.[43] We focus first on the black dots, which depict the coefficients for the effect of temperature distance (Panel A) and precipitation distance (Panel B) on migration for each separate time-period. The first two dots from the left refer to individuals linked between the 1865 Norwegian Census and the US Censuses of 1870 and 1880 (our baseline). In the third and fourth dots, we consider individuals linked between the 1865

[43] See Appendix B.2 for more details. To ease comparisons, climate distances are standardized in each sample to have zero mean and standard deviation equal to one.

Norwegian Census and the 1900 US Census, regardless of the year of immigration and those moving after 1880, respectively. In the remaining three dots, we focus on the second wave of Norwegian migration, which occurred in the early 20th century. The grey triangles in Figure 5 plot these same coefficients after controlling for the previous period's number of migrants between the Norwegian municipality of origin and the US county of destination.[44] The grey triangles are indistinguishable from the black dots, suggesting that climate matching influences migration independently of historical enclaves.[45]

Interestingly, all coefficients become somewhat smaller in absolute value after 1900, indicating a weakened relationship between climate similarity and migration over time. One interpretation is that climate distance became less relevant as ethnic networks grew, but this is unlikely since estimates are unchanged when controlling for lagged migration. An alternative interpretation is that Norwegian pioneers settled land that was a better climate match first, while late-comers were more constrained in terms of land availability. Furthermore, to the extent that farmers valued climate more than non-farmers (as shown in Section 4.2), climate matching may have become less important as the economy shifted away from agriculture.

US internal migration. While ethnic networks may be particularly important for international migrants, they can favor the persistence of domestic migration routes as well (Stuart and Taylor, 2021). In Figure 6, we replicate the analysis presented in Figure 5 focusing on US internal migration. We plot coefficients on temperature (Panel A) and precipitation (Panel B) distance separately for each Census decade, both for the baseline specification (black dots) and for regressions that control for the number of migrants between each county-pair in the previous decade (grey triangles). As with the Norwegians, results are virtually unchanged when controlling for past migration, suggesting that climate distance has an independent effect on US internal migration.[46]

In Table A22, we provide additional evidence that migrant networks are not driving the relationship between climate similarity and migration. In column 2, we confirm the patterns displayed in Figure 6 by replicating the preferred specification for US internal migration (Table 2, column 5) while controlling for lagged migration flows.[47] To reduce concerns that lagged migration flows over the previous decade may be measured with noise, we control for the stock of migrants (cumulated starting from 1850) in column 3 and for the average of past flows in

[44] For 1865-1870, we cannot perform this exercise, since we have no data on past migration.

[45] As expected, past migration is positively correlated with subsequent migration (Figure A13). In unreported results, we compared the standardized beta coefficients associated with climate distance to those on past migration. The former were systematically larger (in absolute value) than the latter, indicating that the effect of climate distance may be stronger than that of past migration.

[46] Similarly, coefficients become smaller (in absolute value) over time—a pattern that we discuss in more detail in Section 4.2.

[47] As expected, lagged migration has a positive and strong effect on subsequent migration flows. However, coefficients on climate distance remain negative and precisely estimated. Moreover, as noted before, standardized beta coefficients indicate that the effects of climate distance are as large as (if not larger than) those of past migration.

column 4. In columns 5 and 6, we consider only migration from 1900 onwards, separately controlling for pre-1900 (stock and average flows) migration. In all cases, coefficients on climate distance remain negative, precisely estimated, and quantitatively large.

Finally, we consider a setting where past networks should have limited scope to influence migration, namely US counties with very low population density that had been only recently settled by white individuals (who account for almost 95% of migrants in our sample, as shown in Table A10). To this end, we restrict destinations to the set of counties that between 1850 and 1890 were on the American frontier, as defined by Bazzi et al. (2020).[48] Then in column 2 of Table A23, we replicate the baseline specification for this selected set of origin-destination pairs. Despite the drastic reduction in sample size, the effect of climate distance remains in line with estimates derived using the full sample over the the same time period (reported in column 1 to ease comparisons). Subsequent columns verify this when considering each decade separately, especially for temperature distance.

Taken together, the evidence presented in this section suggests that past networks are unlikely to explain the strong, positive relationship between climate similarity and migration documented above. We now turn to two separate sets of mechanisms: climate-specific human capital and climate-amenity.

4.2 Climate-Specific Skills

Weather, and extreme heat in particular, has well-documented impacts on labor productivity and economic outcomes (Dell et al., 2012; Heal and Park, 2016; Deryugina and Hsiang, 2017). In this section, we examine an alternate nexus, exploring the possibility that climate-specific skills explain why migrants seek a familiar climate. Climate-specific skills include farming expertise *vis-a-vis* crop suitability and pest management as well as techniques for food preservation, survival of extreme heat or cold, and house construction (insulation, ventilation, and hurricane resilience).[49] Immunity and familiarity with climate-specific illnesses, such as malarial vectors, may also influence migration decisions. In pioneering work on this topic, Steckel (1982) discussed the relationship between climate-specific human capital and the horizontal migration patterns that prevailed in the US during the 19th century.

One interesting example involves Vietnamese refugees settling the US Gulf Coast after the

[48] Bazzi et al. (2020) trace out the evolution of the American frontier line—defined as the boundary at which population density falls below two people per square mile—over time. For each decade, they then assign to the frontier all counties that are in close proximity (100 km) to the frontier line and that have population density below six people per square mile (a threshold first indicated by Porter et al., 1890). Since we cannot compute migration for the 1880-1890 decade, we consider migration between 1880-1900, and use frontier status in 1890 (the last year considered in Bazzi et al., 2020).

[49] Such knowledge helps explain why mortality from exposure to the same level of extreme heat is lower, on average, in hot places than cold places (Barreca et al., 2015; Heutel et al., 2021; Carleton et al., 2022).

Vietnam War. The region's hot and humid climate is somewhat similar to the climate of the Mekong Delta region, which is known for its fishing and aquaculture. Climate similarity—and the corresponding knowledge about aquaculture techniques in warm-water wetlands—likely eased the transition of Vietnamese fishermen, especially in comparison to a counterfactual of moving to other prominent fishing states like Alaska or Maine. A compelling non-US case is the Indonesian government's resettlement program of two million people from Java and Bali to the Outer Islands. Bazzi et al. (2016) showed that subsequent productivity was higher in rural villages that received migrants from regions with a more similar climate, suggesting the importance of climate-specific human capital and skill transferability.

To test the relevance of this channel in our context, we explore the heterogeneity of results by occupation. Intuitively, skill transferability should be more relevant for individuals working in more weather-exposed occupations, such as farming. We focus on domestic migration to leverage rich variation across US counties and over time. We replicate our preferred specification separately for different subsample of migrants, based on the industry reported in the US Census. We plot the estimated coefficients, with corresponding 95% confidence intervals, in Figure 7. To ease comparisons, we standardize temperature (Panel A) and precipitation (Panel B) distances so that they have zero mean and unit variance in each sub-sample. The first dot encompasses the full sample (i.e., Table 2, column 5), and the second dot restricts the sample to men 15+ in the labor force. From the third dot onwards, we cut the sample by occupational sector. In line with our hypothesis, the estimated climate-migration elasticity is twice as large (in absolute value) for farmers than for non-farmers—though consistent with our other findings, the pattern is evident only for temperature.

To further explore the climate-as-skills mechanism, we return to Figure 6, where we examined the evolution of the climate-migration elasticity from 1850-1860 to 1930-1940 (see also Section 4.1). During this period, the US labor force in farming declined from 55% to 17% (Lebergott, 1966), whereas technological innovation may have reduced the value of specific skills.[50] This suggests that the elasticity of migration with respect to climate distance should decline (in absolute value) over time. This is precisely what we observe in Figure 6: coefficients remain relatively stable (negative and precisely estimated) until the turn of the 20th century, and then gradually become smaller (in absolute value). By 1940, the coefficient on temperature (resp., precipitation) distance is about 25% (resp., 50%) smaller than that in 1860. This trend also resonates with the climate change adaptation literature that finds a dampening in the temperature-mortality relationship over time (Barreca et al., 2016), attributable to adaptive technologies that insulated people from climates extremes.

[50] For instance, the adoption of cold storage in the late 19th century rendered hot or cold weather-specific crop and food preservation techniques less valuable.

4.3 Climate as Amenity

Climate might also act as an amenity driving migration decisions. Evidence indicates that climate stability favors the transmission of culture across generations (Giuliano and Nunn, 2021). To the extent that climate facilitates the persistence of cultural practices—from hobbies to socializing norms to food preferences to religious rituals—migrants may choose locations with climates that align with such personal and cultural preferences. Consistent with the climate-as-amenity channel, we detect a statistically significant and economically relevant relationship between climate distance and migration among non-agricultural workers, including those employed in the service sector (Figure 7). Because these sectors are less tied to climate, climate-specific human capital should be less relevant than for farmers.

Moreover, the negative association between climate distance and migration persisted through 1940 (Figure 6), when agriculture accounted for less than 20% of the US labor force. Using IRS county-to-county migration tables between 2011 and 2019, we test whether climate distance correlates with migration even today—when farmers represent a negligible share of US employment. Table 4 replicates the preferred specification (Table 2, column 5) for the modern period. Coefficients on climate distance remain negative and statistically significant, even though smaller in absolute value than those obtained for the historical period.[51] In particular, standardizing regressors produces coefficients that are between 30% and 50% smaller in absolute value than those estimated over the pre-1900 sample (not reported for brevity). These trends indicate that, even if climate became somewhat less relevant for migrants over as time passed, it still plays an important role in driving location decisions today—consistent with a climate-as-amenity mechanism.

Findings in this section resonate with work by Albouy et al. (2021), who provide evidence of a positive correlation between the characteristics (e.g., number of winter days with mild temperature, distance from coast, safety) of the countries of origin and those of the cities where immigrants settled in the US after 2000. Note, though, that the human capital and the amenity channels are likely to be intertwined, particularly in this context. Take a simple example: immigrants' food preferences are shaped by the foods available and cooking practices in their country of origin. In turn, farming practices and food preservation techniques—a form of climate-specific human capital—determine which foods are produced. Likewise, in the aforementioned case of Vietnamese fishermen in Louisiana, the enjoyment of the Gulf Coast's familiar seafood, wetlands, and weather is difficult to untangle from the climate-specific fishing skills that they brought to the region.

[51] See Appendix B.4 for more details on the IRS data. In our preferred specification, we exclude post-2019 migration data, which may be unduly influenced by COVID-19 (Ramani and Bloom, 2021). In column 2 of Table 4, we verify that results are unchanged when including IRS migration data up until 2021. Columns 3 and 4 replicate column 1 and document that results hold when controlling for historical migration flows.

5 Implications of Climate Similarity and Migration

In this section, we discuss the implications of our findings. In Section 5.1, we estimate the value of climate similarity for migrants in two ways. First, we leverage temporal and spatial variation generated by the Homestead Act of 1862, which altered the price of land for migrants and reduced climate matching for farmers. Second, we provide novel evidence of a negative relationship between climate mismatch and life expectancy, which we combine with the EPA's value of statistical life to derive an estimate for the marginal value of climate. In Section 5.2, we use climate projections to illustrate the impact that, by altering climate similarity between counties, climate change might have on US domestic migration.

5.1 Estimating the Marginal Value of Climate

5.1.1 The Homestead Act

The Homestead Act of 1862 was the largest land distribution program in US history. It provided up to 160 acres of essentially free land to farmers, conditional on five years of residency and cultivation. Under the Act, 10% of US land was transferred from the federal government to 1.6 million farmers (Edwards et al., 2017).[52] We obtain information on the universe of land patents from the Bureau of Land Management General Office Land records, which report the year a parcel was contracted, the area, and the name of the homesteader. We aggregate the data from the individual land patent to the county level, and compute the share of the county area that was homesteaded in any given decade, relative to the overall area of the county. Because a Homestead patent was granted five years after the arrival of the homesteader on the land, we attribute patents signed in a given 10-year period to migrants arrived in a place five years before.[53] As a concrete example, we assign contracts signed between 1915 and 1924 to migrants arriving in a county between 1910 and 1919 (i.e., the 1910-1920 decade). As the vast majority of contracts were signed between 1870 and 1920, we focus our attention on this time period.

Figure A15 plots the distribution of the share of county area that was homesteaded in each decade from 1870 to 1920. It shows that, during this period, between 10 and 30% of available land was homesteaded in the Midwest and West, but that the Homestead Act also made land available across southern states, such as Florida, Alabama, and Mississippi. We calculate the

[52] Under the Homestead Act, individuals would acquire land in a three-step process. First, applicants would register the land by filing an application and an affidavit to a local land office and by paying a $12 fee. Second, they had to settle on the land within six months from the application and provide evidence of permanent and continuous residence as well as of cultivation. Third, homesteaders would file for a deed of title and pay an additional $6 fee. See Mattheis and Raz (2019) and Smith (2020) for more details.

[53] Remember that we are able to measure migration only across decades.

80^{th} percentile of county homesteaded shares, considering data from all US counties and all years spanning from 1860 and 1920. We then construct a dummy variable that takes the value of one if the homesteaded share of the destination county in a given decade surpasses this calculated 80^{th} percentile.[54] Figure A17 maps the counties that meet this threshold across decades, and documents the substantial variation across both time and space.

US internal migration. We augment the baseline specification for domestic migration in equation (2) by interacting the Homestead dummy with both climate and geographic distances.[55] As in the baseline specification, county of origin by decade and county of destination by decade fixed effects absorb any temporal and location-specific push and pull factors that might be correlated with county-pair climate distances. Moreover, the latter set of fixed effects takes into account the possibility that places with a large share of homesteaded land in a given decade also became more (or less) attractive for migrants from other parts of the US that had a more (or less) similar climate.

We present results in Table 5, column 1. Consistent with our baseline estimates, coefficients on temperature and precipitation distance remain negative and statistically significant—and, again, larger for temperature than for precipitation. The coefficient on the interaction between temperature distance and the Homestead dummy is positive and statistically significant. That is, the degree of temperature similarity between origin and destination counties is lower for migrants moving to homesteaded counties as compared to other migrants from the same county (and in the same decade) who selected a non-homesteaded destination. In line with earlier results, the interaction between the Homestead dummy and precipitation distance is quantitatively small and imprecisely estimated.

Since the Homestead Act was explicitly designed to attract farmers, one would expect a stronger pull effect for this population. On the other hand, climate matching is more important to farmer migrants than non-farmer migrants (Figure 7). Hence, whether the Homestead Act reduced the importance of climate distance more for farmers than for non-farmers is *ex-ante* ambiguous. In columns 2 and 3, we explore these ideas, focusing on farmer and non-farmer migrants, respectively. Results suggest that farmers' preferences for temperature were relatively less distorted by the prospect of homesteading than those of non-farmers.[56] In particular, the temperature distance elasticity for farmers settling in a Homestead county is 53% smaller than

[54] Figure A16 shows graphically the definition of the threshold.

[55] The Homestead dummy is absorbed by the county of destination by decade fixed effects. In our preferred specification, we impose two sample restrictions. First, because most Homestead contracts were signed between 1870 and 1920, we restrict attention to the decades from 1860-1870 to 1910-1920. Second, we define the sample of origin counties those that are never homesteaded, while leaving the set of destinations unrestricted. All results are robust to using alternative time windows or origin-destination samples, and different thresholds to define the Homestead dummy (see also Tables A24 and A25).

[56] The fact that we find effects for both farmers and non-farmers is also consistent with the possibility that the Homestead Act generated economic spillovers outside the agricultural sector.

that of farmers settling in a non-Homestead county. Meanwhile, the temperature distance elasticity for non-farmers settling in a Homestead county is 37% of the elasticity observed for the same migrants settling outside of a Homestead county.[57]

Norwegian immigration. An interesting feature of the Homestead Act is that its introduction coincided temporally with the first wave of Norwegian immigration. Furthermore, anecdotal accounts indicate that the prospects of free land under the Homestead Act compelled many Norwegians to embark for the US (Semmingsen, 1978). Indeed, comparing the settlements of Norwegian immigrants and the availability of homesteaded land between 1865 and 1880, one can observe a significant degree of overlap (Figures A9 and A17). In column 4 of Table 5, we explore whether the Homestead Act reduced climate distance's influence on location decisions of Norwegians moving to the US. We replicate the specification in Table 1, column 4, by interacting climate (and geographic) distance with the Homestead dummy.[58]

Because the Homestead dummy is not absorbed by the fixed effects in this case, we can explicitly control for it. The corresponding coefficient is positive, large, and precisely estimated. That is, Norwegian immigrants were more likely to settle in counties that offered Homestead land, aligning with the historical narrative. Turning to climate distance, the patterns mirror those in Table 1: both temperature and precipitation distances reduce migration, with the effect being larger for the former than for the latter. Finally, the positive and statistically significant coefficient on the interaction between temperature distance and the Homestead dummy indicates that—similar to what we found for domestic migration—Norwegians moving to homesteaded counties had a lower degree of temperature similarity.

Discussion. Results in Table 5 suggest that the Homestead Act caused migrants to match *less* on climate than they would have otherwise. This finding relates to recent work by Mattheis and Raz (2019) and Allen and Leonard (2021), who show that the Homestead Act slowed down economic development in the western US.[59] Allen and Leonard (2021) stress the role of agricultural path dependence due to the Homestead Act's farming obligation, while Mattheis and Raz (2019) posit that homesteaders were negatively selected. Our findings offer an additional, possibly complementary, explanation: the Homestead Act may have reduced farmers' productivity by increasing their climate mismatch.

Relatedly, our results are relevant for the literature on migration and labor misallocation

[57] We obtain these numbers by dividing the total elasticity of climate distance (e.g., $\beta_3 + \beta_4$) by the elasticity of climate distance net of Homestead (e.g., β_3) for each sample.

[58] Since we consider individuals that migrated to the United States between 1865 and 1880, we set the Homestead dummy equal to one if the homesteaded share of the county area was above the 80^{th} percentile during this specific period.

[59] Findings in Smith (2020) indicate that these negative effects were partly offset by the Homestead Act's promotion of smaller-scale land cultivation, which was put at more productive use compared to large and concentrated plots distributed through railroads grants.

(Young, 2013; Bryan et al., 2014; Munshi and Rosenzweig, 2016), particularly Bazzi et al. (2016), who document the importance of migrants' climate-specific human capital for economic development in Indonesia. We complement this literature by showing that skill transferability may shape migration patterns by inducing people to sort based on climate. Our results also showcase the potential labor (mis)allocation impacts of public programs, and speak to the literature on place-based policies (Glaeser and Gottlieb, 2008; Kline and Moretti, 2014), with relevance for discussions of managed retreat and climate change adaptation.

Estimating the marginal value of climate. The Homestead Act provides a unique setting to estimate how individuals value climate similarity. The main result of this paper—robust across time periods, geography, and migrant groups—is that migrants seek to minimize the climate distance between their origin and their destination. However, when offered cheap land through the Homestead Act, farmers deviated from the baseline—that is, climate matching was less important when the land was free. Comparing the size of this deviation to the difference in land prices offered under the Homestead Act versus direct government purchase allows us to estimate farmers' valuation of climate similarity. If migrants deviated greatly from their baseline climate preferences when offered Homestead land, this would imply a low valuation of climate similarity. By contrast, a small deviation of migration flows from the baseline implies a smaller valuation.[60]

To estimate the marginal value of climate similarity, we proceed as follows. First, we compute counterfactual migration flows under two alternative scenarios. In the baseline case, we predict farmer migration based on climate and geographic distances, taking into account the roll out of the Homestead Act. In a second counterfactual scenario, we predict migration based on distance only, assuming that the Homestead Act was never passed. We use the coefficient from column 2 in Table 5 to derive counterfactual migration flows. These are, in turn, used to obtain measures of migration-weighed average climate and geographic distance between counties. Under the baseline Homestead Act scenario, a migrant would travel to counties that were, on average, 3.74°C apart from their origin. Under the "No Homestead Act" scenario, the average travelled temperature distance drops to 3.66°C. Let us denote the difference between the two counterfactual travelled distances by ΔDist Temp. The .08°C difference between the two scenarios represents the resulting increase in climate mismatch attributable to free land under the Homestead Act.

Then, we assume that all farmers migrating to a Homestead county paid the administrative fee of $345 (in 2020 dollars), whereas those migrating to non-Homestead counties paid the price of

[60] As noted above, the Homestead Act reduced climate matching also for non-farmers. However, we focus on farmers since it is easier to calculate the value of land for this group.

a direct purchase (about $4,763 in 2020 dollars), for the same amount of land.[61] As we did for average travelled distance, we use migration flows predicted under the two scenarios to obtain the average price paid for land. In the "No Homestead Act" scenario, the average price paid for land is $4,763, no matter where a migrant settled. In the baseline scenario, instead, a migrant would pay an average of $4,373 for 160 acres of land.[62] We denote this price difference with Δp. Finally, we recover an estimate for the marginal valuation of climate and geographic distance for farmers using the following formula:

$$MWTP(\text{Temp}) = abs(\frac{\Delta p}{\Delta \text{Dist Temp}})$$

The marginal valuation of temperature is estimated at $4,506 in current dollars per 1°C. For comparison, farmers would be willing to pay this amount to avoid the cost of travelling an extra 483 km. This estimate may be an upper bound, since homesteaded land was likely of lower quality than nearby alternatives Mattheis and Raz (2019). Below, we study a different setting—mortality increases due to climate mismatch—which allows us to compare magnitudes of marginal value of climate.

5.1.2 Mortality and Climate Mismatch

In this section, we use a different approach, dataset, and population sample to estimate the marginal value of climate similarity. We rely on US mortality data used in Section 2, where we documented a positive relationship between the average temperature in the country of birth and the temperature in the US county where people died (Figure 2).[63] We leverage the fact that this dataset reports information on the age at death to test whether climate mismatch among immigrants in the US is associated with differences in life expectancy. We estimate individual-level regressions of age at death (in years) on climate mismatch among the over 700,000 foreign-born individuals who died in the US between 1959 and 1961. We define climate mismatch as the absolute value of the difference between the temperature (or precipitation) in the capital city of the country of birth and the US county of death.

We present results in Table 6. Columns 1 and 3 indicate that climate mismatch is negatively correlated with age at death. That is, migrants who settled in US counties with a climate similar to their country of origin tend to live longer. Subsequent columns document that

[61] For Homestead land, we assume a cost of $22 for 160 acres that includes average filing fees and commission. Converting this value from 1865 to 2020 dollars, we obtain $345. For direct purchase, we assume $300 for 160 acres, which is the midpoint of $1.25/acre, the price for unclaimed federal land, and $2.50/acre, the price for railroad grant land (Allen and Leonard, 2021). Adjusted for inflation, this $300 equals $4,763.

[62] This is the average of Homestead and non Homestead land prices, weighed using predicted migration flows.

[63] As noted above, US mortality records typically report whether an individual was born in the US or the "rest of the world." However, for the years 1959-1961, the individual's country of birth is reported for all locations outside the continental US.

this relationship holds after controlling for US county of death, country of origin, and year of death fixed effects as well as for individual characteristics, such as gender and marital status.[64] These more saturated specifications isolate the impact of climate distance on mortality for an individual after controlling for demographic characteristics as well as for the average age at death among all immigrants in the US: from the same origin country, in the same county where the individual died (which could be tied to local economic conditions), and who died in the same year.

Different mechanisms might explain results in Table 6. However, these patterns are consistent with either the climate-specific human capital or the climate-as-amenity channels (or both) discussed in Section 4. For instance, climate mismatch might reduce returns to climate-specific skills and thus earnings. This may lead to a deterioration of health outcomes. A mismatch in climate-specific skills may also have a more direct impact on mortality through increased accidental deaths, e.g., driving on icy roads for someone not accustomed to such conditions. Alternatively, hotter temperatures may increase crime and violence (Ranson, 2014; Mukherjee and Sanders, 2021), which could in turn expose people to premature death—more so those coming from colder climates. Climate mismatch may lower life expectancy also through an amenity channel, for example if mismatched migrants are less happy and less socially connected (Holt-Lunstad et al., 2010).

In Figure A18, we test whether the mortality effect documented in Table 6 varies with the direction of climate mismatch (i.e., moving to a hotter or a colder location). To this end, we regress age at death on 1°C temperature distance bins between the US county of death and birth country capital, partialling out the same fixed effects and demographic controls as in our most stringent specification above (column 6). We find an inverted U-shaped curve that resonates with the climate change literature that has documented increased mortality in response to both hot and cold shocks (Barreca et al., 2015; Heutel et al., 2021; Carleton et al., 2022).

Finally, as for the Homestead Act, we use our results to estimate the marginal value of climate. Coefficients from our preferred specification (Table 6, column 6) imply that a 1°C of temperature distance reduces life expectancy by .12 years, or approximately one month of life. We implement an admittedly crude back-of-the-envelope by taking the EPA's value of a statistical life (VSL) of $9.6 million (2020 dollars), dividing it by the average death age in our sample (73 years), and multiplying this number by .12, i.e., the coefficient from column 6 in Table 6. This implies that 1°C of additional climate similarity is roughly equal to $15,800 per person (in today's currency)—a number three times larger than the $4,506 value we found using the Homestead Act.

[64] Consistent with other results in this paper, the effect of precipitation distance shows a similar sign but is less robust and statistically insignificant when including a more stringent set of controls.

5.2 Climate Change and Migration Projections

In Section 3.4, we leveraged climate change over the 20th century to identify the impact of climate distance on migration. In this section, we use these estimates to explore what climate projections over the 21st century imply for future patterns of county connectivity and bilateral migration flows. We rely on daily projections under a moderate warming path (RCP4.5) to compute projected climate averages at the county level for the period 2080-2100.[65] We then calculate the percent change in average temperature distance between today and this future period for all contiguous US county-pairs. Using the gravity models estimated in Sections 3.3 and 3.4, we obtain the predicted change in migration between any two counties.

Figure A19 illustrates the spatial distribution of such predicted changes for two example counties in the US.[66] Relative to present day climate distances, counties shaded in green are projected to get closer to the temperature in the example county over time (i.e., converge), while red counties are projected to diverge (range of -4 to 2.5°C in magnitude). The projected temperature divergence between the Northeast and the Midwest implies that, *ceteris paribus*, there will be less migration between these two regions in the future relative to today. On the other hand, in the second example we find that more migrants will travel between the Midwest and the Delta region in the US South, as the climates of these areas are expected to converge.

What does climate change imply for a county's overall connectivity? To answer this question, we first compute the matrix of percent changes in bilateral migration predicted by the change in temperature distance, ΔMig_{od}. Then, we derive the change in a county climate-connectivity as:

$$CC_o = \sum_d (\text{Pop}_d^{2010} \times \Delta Mig_{od})$$

where Pop_d^{2010} is the 2010 population share of county d, relative to total US population. Similar in spirit to the "market access" measure from Donaldson and Hornbeck (2016), CC_o captures the degree of connectivity that each origin county o has to all possible destinations d. As in Donaldson and Hornbeck (2016), this measure is higher for counties that will become more connected (climatically) to more populous counties.

Figure 8 depicts the change in climate-induced connectivity between the beginning and the end of the 21st century. It shows that Florida, and the US South more generally, will experience an increase in connectivity as a result of climate change. Intuitively, since climate models predict more warming in colder places than in hotter places, Florida's temperature is expected to converge (in a relative sense) to the rest of the US—increasing its climate connection to

[65] See https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp.

[66] We use an elasticity of migration with respect to temperature distance of -.2, as per results in Tables 2 and 3.

populous places like the Northeast.

6 Discussion and Conclusions

In this paper, we provide robust evidence on the relationship between climate similarity and migration. Across time periods, geography, and migrant groups, we find that individuals tend to settle in places with climates similar to where they came from. These patterns cannot be explained by the spatial correlation of climate or pre-existing migrant networks, and do not reflect geographic, economic, and demographic differences between origins and destinations. Exploring mechanisms, we document that both climate-specific skills and climate-as-amenity likely explain our results. To our knowledge, we are the first to estimate the effects of climate distance on historical migration patterns. We also advance and complement different strands of the literature.

First, with relevance for climate change economics, we provide two novel approaches to estimating the marginal value of climate—or, more specifically individuals' willingness to hold climate constant. Using historical data from the Homestead Act and US mortality records, we find that a temperature mismatch of 1°C among migrants is associated with a cost of \$4,500-\$15,800 in current dollars. Moreover, we derive a measure of climate-induced connectivity that can be used to predict migration flows and how climate change may alter future migration patterns. In a simple and largely illustrative example, we apply this connectivity measure to project future US internal migration. Future work can build on our approach to compute more precise measures of climate connectivity for both the US and the entire world.

Second, the systematic relationship between climate similarity and migration uncovered in our work can be exploited as a source of identifying variation to study the impact of migration across several domains. Our analysis suggests that climate similarity, if combined with appropriate push shocks, can be used to refine the standard shift-share instrument used in the literature that leverages variation in the pre-determined distribution of settlers (Altonji and Card, 1991; Card, 2001). Our approach has the potential to offer a micro-foundation for such initial settlements, possibly addressing the critiques that this class of instruments has received recently (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022).

By extension, our findings open the door to several fascinating questions on economic history and political economy. Combining climate's predictive power on settlement patterns with push shocks like the Dust Bowl (Hornbeck, 2012), the boll weevil (Lange et al., 2009), or severe frosts in mid-19th century Scandinavia (Karadja and Prawitz, 2019) can provide insights into how climate-driven migration contributed to the structural transformation of the US economy.

The relationship between climate similarity and migration documented in our work can also be used to study how migrants affect the political and cultural landscape of receiving areas (Alesina and Tabellini, 2023; Bazzi et al., 2023a,b). This margin may be especially important in the context of climate-driven migration, given the well-documented relationship between agricultural practices and cultural norms (Alesina et al., 2013; Cao et al., 2021; Becker, 2023).

Finally, our results are relevant for the emerging literature that seeks to understand the spatial effects of climate shocks on the global economy through general equilibrium models. Such models assume that migration costs are exogenous to climate change (Cruz and Rossi-Hansberg, 2021; Desmet and Rossi-Hansberg, 2023; Bilal and Rossi-Hansberg, 2023), but our findings indicate that migrants place a large weight on climate similarity between origin and destination. Climate distance can thus be viewed as a travel cost that depends on the spatial distribution of climate change, and our results can be used to enrich these models, improving the precision of welfare estimates and the assessment of the macroeconomic effects of climate change. This is also relevant for climate migration and refugee resettlement policies. Since 2008, over 20 million people annually, on average, have been forcibly displaced by weather-related events (UNHCR)—a number that is estimated to increase with further global warming. For economic migrants, climate change often acts as a push shock and has a disproportionate impact on farmers. Our findings suggest that climate similarity should be taken into account when resettling existing climate refugees, as well as when designing "managed retreat" policies in anticipation of climate change.

References

Abatzoglou, J. T., S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch (2018): "TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015," *Scientific data*, 5, 1–12.

Abramitzky, R. and L. Boustan (2017): "Immigration in American economic history," *Journal of economic literature*, 55, 1311–1345.

——— (2022): *Streets of gold: America's untold story of immigrant success*, PublicAffairs.

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum, and S. Pérez (2021): "Automated linking of historical data," *Journal of Economic Literature*, 59, 865–918.

Abramitzky, R., L. Boustan, K. Eriksson, S. Perez, and M. Rashid (2020): "Census Linking Project: Version 2.0 [dataset]." *Data retrieved from, https://censuslinkingproject. org*.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102, 1832–56.

Albouy, D., H. Cho, and M. Shappo (2021): "Immigration and the pursuit of amenities," *Journal of Regional Science*, 61, 5–29.

Albouy, D., W. Graf, R. Kellogg, and H. Wolff (2016): "Climate amenities, climate change, and American quality of life," *Journal of the Association of Environmental and Resource Economists*, 3, 205–246.

Alesina, A., P. Giuliano, and N. Nunn (2013): "On the origins of gender roles: Women and the plough," *The quarterly journal of economics*, 128, 469–530.

Alesina, A. and M. Tabellini (2023): "The Political Effects of Immigration: Culture or Economics?" *Journal of Economic Literature*, Forthcoming.

Allen, D. W. and B. Leonard (2021): "Property Right Acquisition and Path Dependence: Nineteenth-Century Land Policy and Modern Economic Outcomes," *The Economic Journal*, 131, 3073–3102.

Alston, L. J. and J. P. Ferrie (1999): *Southern Paternalism and the American Welfare State: Economics, Politics, and Institutions in the South, 1865-1965*, Cambridge University Press New York.

Althoff, L., H. B. Gray, and H. Reichardt (2022): "Intergenerational mobility and assortative mating," *Working Paper*.

Altonji, J. G. and D. Card (1991): "The effects of immigration on the labor market outcomes of less-skilled natives," in *Immigration, trade, and the labor market*, University of Chicago Press, 201–234.

Anderson, J. E. and E. Van Wincoop (2003): "Gravity with gravitas: A solution to the border puzzle," *American economic review*, 93, 170–192.

Angrist, J. D. and J.-S. Pischke (2009): *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Atack, J. and R. A. Margo (2011): "The impact of access to rail transportation on agricultural improvement: The American Midwest as a test case, 1850–1860," *Journal of Transport and Land Use*, 4, 5–18.

Bailey, M. J., C. Cole, M. Henderson, and C. Massey (2020): "How well do automated linking methods perform? Lessons from US historical data," *Journal of Economic Literature*, 58, 997–1044.

Barreca, A., K. Clay, O. Deschênes, M. Greenstone, and J. S. Shapiro (2015): "Convergence in adaptation to climate change: Evidence from high temperatures and mortality, 1900–2004," *American Economic Review*, 105, 247–251.

Barreca, A., K. Clay, O. Deschenes, M. Greenstone, and J. S. Shapiro (2016): "Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the twentieth century," *Journal of Political Economy*, 124, 105–159.

Bazzi, S., A. Ferrara, M. Fiszbein, T. Pearson, and P. A. Testa (2023a): "The Other Great Migration: Southern Whites and the New Right," *The Quarterly Journal of Economics*, 138, 1577–1647.

Bazzi, S., A. Ferrara, M. Fiszbein, T. P. Pearson, and P. A. Testa (2023b): "The Confederate Diaspora," Tech. rep., National Bureau of Economic Research.

Bazzi, S., M. Fiszbein, and M. Gebresilasse (2020): "Frontier culture: The roots and persistence of "rugged individualism" in the United States," *Econometrica*, 88, 2329–2368.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2016): "Skill transferability, migration, and development: Evidence from population resettlement in Indonesia," *American Economic Review*, 106, 2658–2698.

Becker, A. (2023): "On the Origins of Restricting Women's Promiscuity," *Review of Economic Studies*, Forthcoming.

Berkes, E., E. Karger, and P. Nencka (2023): "The census place project: A method for geolocating unstructured place names," *Explorations in Economic History*, 87, 101477.

Bertoli, S., F. Docquier, H. Rapoport, and I. Ruyssen (2022): "Weather shocks and migration intentions in Western Africa: Insights from a multilevel analysis," *Journal of Economic Geography*, 22, 289–323.

Bilal, A. and E. Rossi-Hansberg (2023): "Anticipating Climate Change Across the United States," Working Paper 31323, National Bureau of Economic Research.

Blegen, T. C. (1955): *Land of their choice: The immigrants write home*, U of Minnesota Press.

Bolt, J., R. Inklaar, H. de Jong, and J. L. van Zanden (2018): "Rebasing 'Maddison': new income comparisons and the shape of long-run economic development," Maddison Project Database, version 2018.

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 89, 181–213.

Bosetti, V., C. Cattaneo, and G. Peri (2021): "Should they stay or should they go? Climate migrants and local conflicts," *Journal of Economic Geography*, 21, 619–651.

Boustan, L. P., D. M. Bunten, and O. Hearey (2013): "Urbanization in the United States, 1800-2000," Tech. rep., National Bureau of Economic Research.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014): "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh," *Econometrica*, 82, 1671–1748.

Burke, M. and K. Emerick (2016): "Adaptation to climate change: Evidence from US agriculture," *American Economic Journal: Economic Policy*, 8, 106–140.

Cao, Y., B. Enke, A. Falk, P. Giuliano, and N. Nunn (2021): "Herding, Warfare, and a Culture of Honor: Global Evidence," Tech. rep., National Bureau of Economic Research.

Card, D. (2001): "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 19, 22–64.

Carleton, T., A. Jina, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. Nath, et al. (2022): "Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits," *The Quarterly Journal of Economics*, 137, 2037–2105.

Carleton, T. A. and S. M. Hsiang (2016): "Social and economic impacts of climate," *Science*, 353, aad9837.

Collins, W. J. and A. Zimran (2023): "Working their way up? us immigrants' changing labor market assimilation in the age of mass migration," *American Economic Journal: Applied Economics*, 15, 238–269.

Conley, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of econometrics*, 92, 1–45.

Coolidge, C. (1926): *Foundations of the Republic; Speeches and Addresses*, New York, Scribner.

Cruz, J.-L. and E. Rossi-Hansberg (2021): "The economic geography of global warming," Tech. rep., National Bureau of Economic Research.

De Juan, A., F. Haass, C. Koos, S. Riaz, and T. Tichelbaecker (2021): "War and nationalism: Evidence from world war I and the rise of the Nazi party," .

Dell, M., B. F. Jones, and B. A. Olken (2012): "Temperature shocks and economic growth: Evidence from the last half century," *American Economic Journal: Macroeconomics*, 4, 66–95.

Deryugina, T. and S. Hsiang (2017): "The marginal product of climate," Tech. rep., National Bureau of Economic Research.

Desmet, K. and E. Rossi-Hansberg (2023): "Climate Change Economics over Time and Space," .

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

Eckert, F., A. Gvirtz, J. Liang, and M. Peters (2020): "A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790," Tech. rep., National Bureau of Economic Research.

Eckert, F., J. Juneau, and M. Peters (2023): "Sprouting Cities: How Rural America Industrialized," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 113, 87–92.

Eckert, F. and M. Peters (2022): "Spatial structural change," Tech. rep., National Bureau of Economic Research.

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

Eriksson, K. (2020): "Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration," *European Review of Economic History*, 24, 427–446.

Fezzi, C. and I. Bateman (2015): "The impact of climate change on agriculture: nonlinear effects and aggregation bias in Ricardian models of farmland values," *Journal of the Association of Environmental and Resource Economists*, 2, 57–92.

Fischer, D. H. (1989): *Albion's seed: Four British folkways in America*, vol. 1, America: A Cultural History.

Fishback, P. V., W. C. Horrace, and S. Kantor (2006): "The impact of New Deal expenditures on mobility during the Great Depression," *Explorations in Economic History*, 43, 179–222.

Fogel, R. W. (1964): *Railroads and American economic growth*, Johns Hopkins Press Baltimore.

Giuliano, P. and N. Nunn (2021): "Understanding cultural persistence and change," *The Review of Economic Studies*, 88, 1541–1581.

Glaeser, E. L. and J. D. Gottlieb (2008): "The economics of place-making policies," Tech. rep., National Bureau of Economic Research.

Goldin, C. (1994): "The political economy of immigration restriction in the United States, 1890 to 1921," in *The regulated economy: A historical approach to political economy*, University of Chicago Press, 223–258.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik instruments: What, when, why, and how," *American Economic Review*, 110, 2586–2624.

Gregory, J. N. (2006): *The southern diaspora: How the great migrations of black and white southerners transformed America*, Univ of North Carolina Press.

Heal, G. and J. Park (2016): "Reflections—temperature stress and the direct impact of climate change: a review of an emerging literature," *Review of Environmental Economics and Policy*.

Heutel, G., N. H. Miller, and D. Molitor (2021): "Adaptation and the mortality effects of temperature across US climate regions," *Review of Economics and Statistics*, 103, 740–753.

Holt-Lunstad, J., T. B. Smith, and J. B. Layton (2010): "Social relationships and mortality risk: a meta-analytic review," *PLoS medicine*, 7, e1000316.

Hornbeck, R. (2012): "The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe," *American Economic Review*, 102, 1477–1507.

——— (2023): "Dust bowl migrants: Environmental refugees and economic adaptation," *Journal of the Economic History*, Forthcoming.

Karadja, M. and E. Prawitz (2019): "Exit, voice, and political change: Evidence from Swedish mass migration to the United States," *Journal of Political Economy*, 127, 1864–1925.

Kirchner, N. (2021): "Warming stripes — visualisations of annual temperature series from the Nordic countries," Dataset version 3.

Kline, P. and E. Moretti (2014): "People, places, and public policy: Some simple welfare economics of local economic development programs," *Annu. Rev. Econ.*, 6, 629–662.

Lange, F., A. L. Olmstead, and P. W. Rhode (2009): "The impact of the boll weevil, 1892–1932," *The Journal of Economic History*, 69, 685–718.

Lebergott, S. (1966): "Labor force and employment, 1800–1960," in *Output, employment, and productivity in the United States after 1800*, NBER, 117–204.

Lobell, D. B. and M. B. Burke (2008): "Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation," *Environmental Research Letters*, 3, 034007.

Mahajan, P. and D. Yang (2020): "Taken by storm: Hurricanes, migrant networks, and US immigration," *American Economic Journal: Applied Economics*, 12, 250–277.

Mattheis, R. and I. T. Raz (2019): "There's no such thing as free land: The homestead act and economic development," *Unpublished Manuscript*.

McGuirk, E. F. and N. Nunn (2023): "Transhumant pastoralism, climate change, and conflict in africa," *Review of Economic Studies*, Forthcoming.

Missirian, A. and W. Schlenker (2017): "Asylum applications respond to temperature fluctuations," *Science*, 358, 1610–1614.

Molloy, R., C. L. Smith, and A. Wozniak (2011): "Internal migration in the United States," *Journal of Economic perspectives*, 25, 173–196.

Mukherjee, A. and N. J. Sanders (2021): "The causal effect of heat on violence: Social implications of unmitigated heat among the incarcerated," Tech. rep., National Bureau of Economic Research.

Munshi, K. and M. Rosenzweig (2016): "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 106, 46–98.

National Center for Health Statistics (2021): "Detailed Mortality Records, Multiple Cause-of-Death (1959-1961)," National Center for Health Statistics, Hyattsville, Maryland. Accessed via NBER, https://www.nber.org/research/data/mortality-data-vital-statistics-nchs-multiple-cause-death-data.

Nichols, R. L. (2014): *American Indians in US history*, vol. 248, University of Oklahoma Press.

Olivetti, C. and M. D. Paserman (2015): "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940," *American Economic Review*, 105, 2695–2724.

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

Proctor, J., A. Rigden, D. Chan, and P. Huybers (2022): "More accurate specification of water supply shows its importance for global crop production," *Nature Food*, 3, 753–763.

Ramani, A. and N. Bloom (2021): "The Donut effect of COVID-19 on cities," Tech. rep., National Bureau of Economic Research.

Ranson, M. (2014): "Crime, weather, and climate change," *Journal of environmental economics and management*, 67, 274–302.

Rosenbloom, J. L. (2002): *Looking for work, searching for workers: American labor markets during industrialization*, Cambridge University Press.

Ruggles, S., C. A. Fitch, G. Ronald, H. J. David, M. A. Nelson, E. Roberts, M. Schouweiler, and M. Sobek (2021): "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]," *Minneapolis, MN: IPUMS*.

Sayers, A. (2014): "NYSIIS: Stata module to calculate nysiis codes from string variables," Statistical Software Components, Boston College Department of Economics.

Schlenker, W. and M. J. Roberts (2009): "Nonlinear temperature effects indicate severe damages to US crop yields under climate change," *Proceedings of the National Academy of sciences*, 106, 15594–15598.

Semmingsen, I. (1960): "Norwegian emigration in the nineteenth century," *Scandinavian Economic History Review*, 8, 150–160.

——— (1978): *Norway to America: a history of the migration*, U of Minnesota Press.

Silva, J. S. and S. Tenreyro (2006): "The log of gravity," *The Review of Economics and statistics*, 88, 641–658.

Sinha, P., M. Caulkins, and M. Cropper (2021): "The value of climate amenities: A comparison of hedonic and discrete choice approaches," *Journal of Urban Economics*, 126, 103371.

Smith, C. (2020): "Land Concentration and Long-Run Development in the Frontier United States," Tech. rep., Mimeo.

Steckel, R. H. (1982): "The economic foundations of east-west migration during the nineteenth century," Tech. rep., National Bureau of Economic Research.

Stern, N. H. (2007): *The economics of climate change: the Stern review*, Cambridge University Press.

Stuart, B. A. and E. J. Taylor (2021): "Migration networks and location decisions: Evidence from US mass migration," *American Economic Journal: Applied Economics*, 13, 134–175.

Taylor, C. A. (2022): "Irrigation and Climate Change: Long-run Adaptation and its Externalities," *Working Paper*.

The Digital Archive, The National Archive; Norwegian Historical Data Centre, U. o. T. t. M. P. C. (2008): "National Sample of the 1865 Census of Norway, Version 2.0," .

Tobler, W. R. (1970): "A computer movie simulating urban growth in the Detroit region," *Economic geography*, 46, 234–240.

Turner, F. J. (2017): "The significance of the frontier in American history," in *The Structure of Political Geography*, Routledge, 132–139.

Verein für Computergenealogie e.V. (2014): "Verlustlisten Erster Weltkrieg." Accessed via https://www.compgen.de/portfolio-item/db-verlustlisten/.

Vose, R. S., S. Applequist, M. Squires, I. Durre, M. J. Menne, J. Williams, Claude N., C. Fenimore, K. Gleason, and D. Arndt (2014): "NOAA Monthly U.S. Climate Divisional Database (NClimDiv)," .

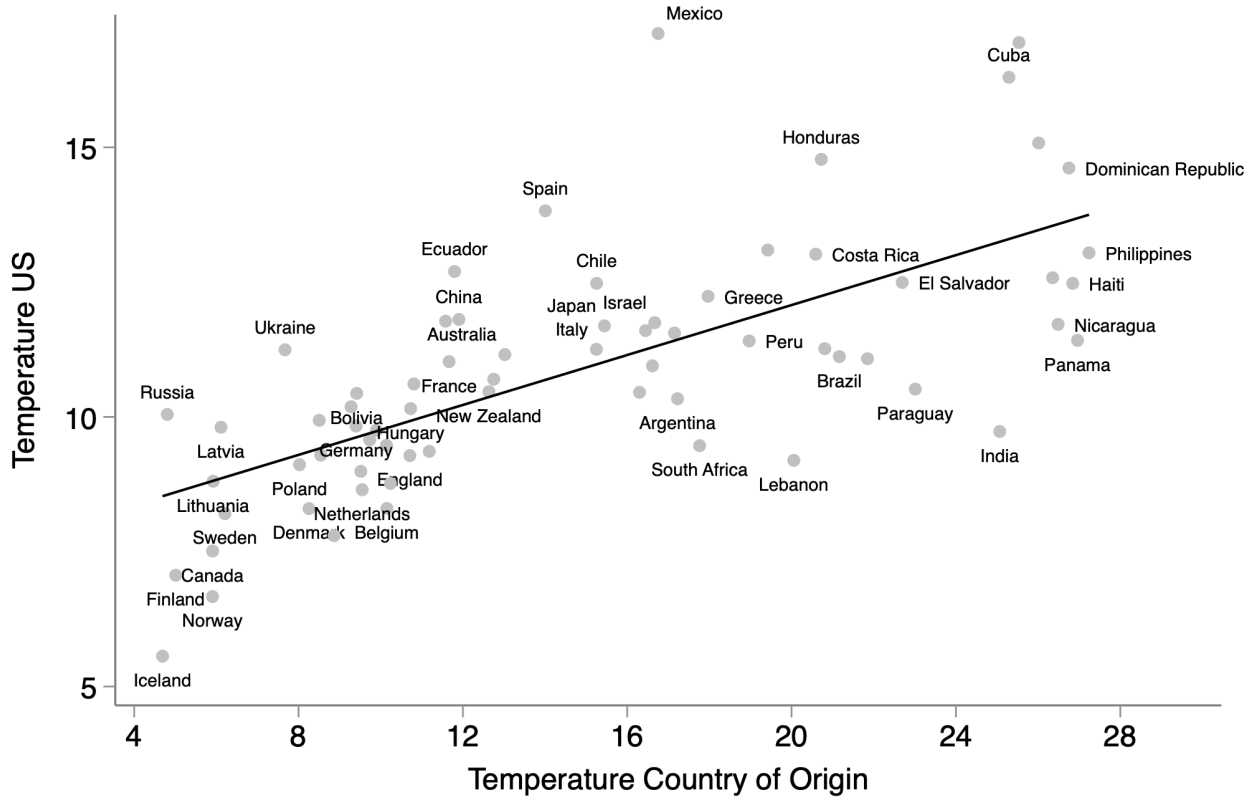
Young, A. (2013): "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics*, 128, 1727–1785.

Zedlitz, J. (2016): "Web-based collaborative system for transcription of serial historic sources to structured data," .

Zimran, A. (2023): "Internal Migration in the United States: Rates, Selection, and Destination Choice, 1850-1940," *Journal of the Economic History*, Forthcoming.

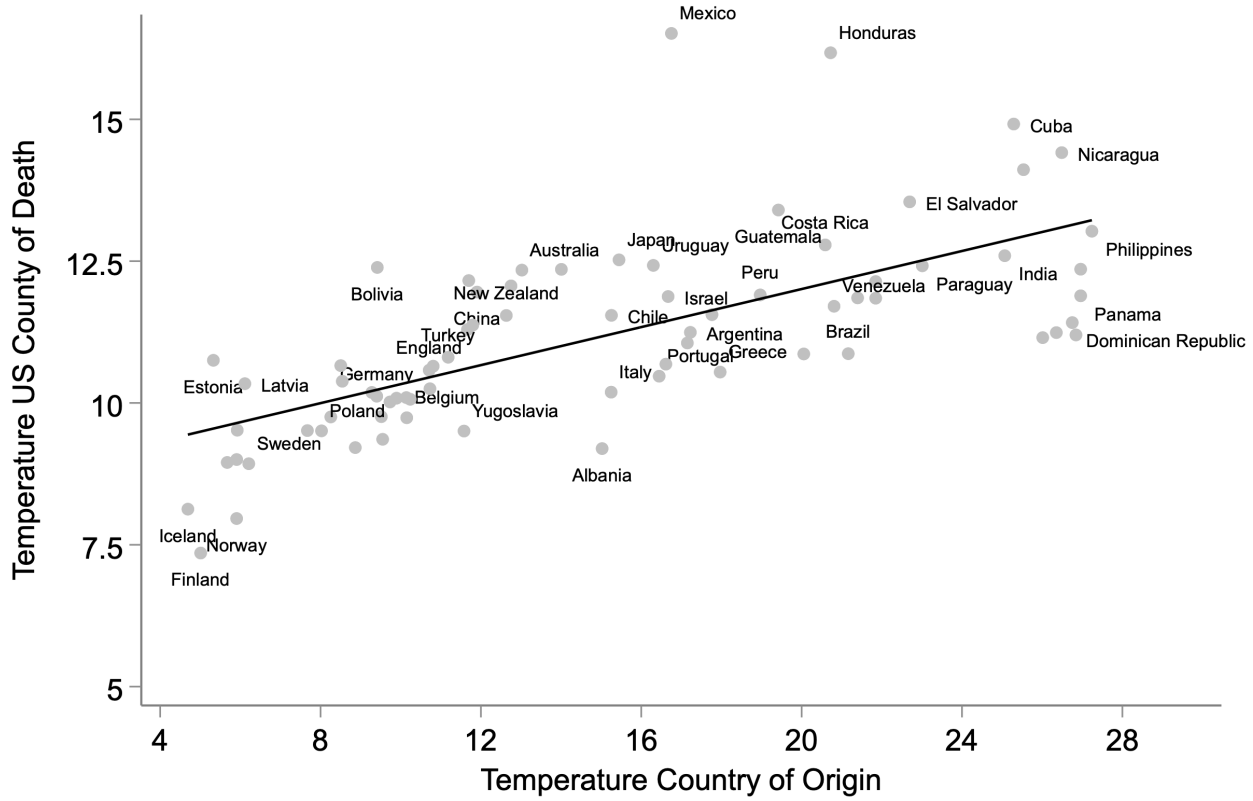
Figures and Tables

Figure 1. Temperature Matching of Immigrants in the US (1880)



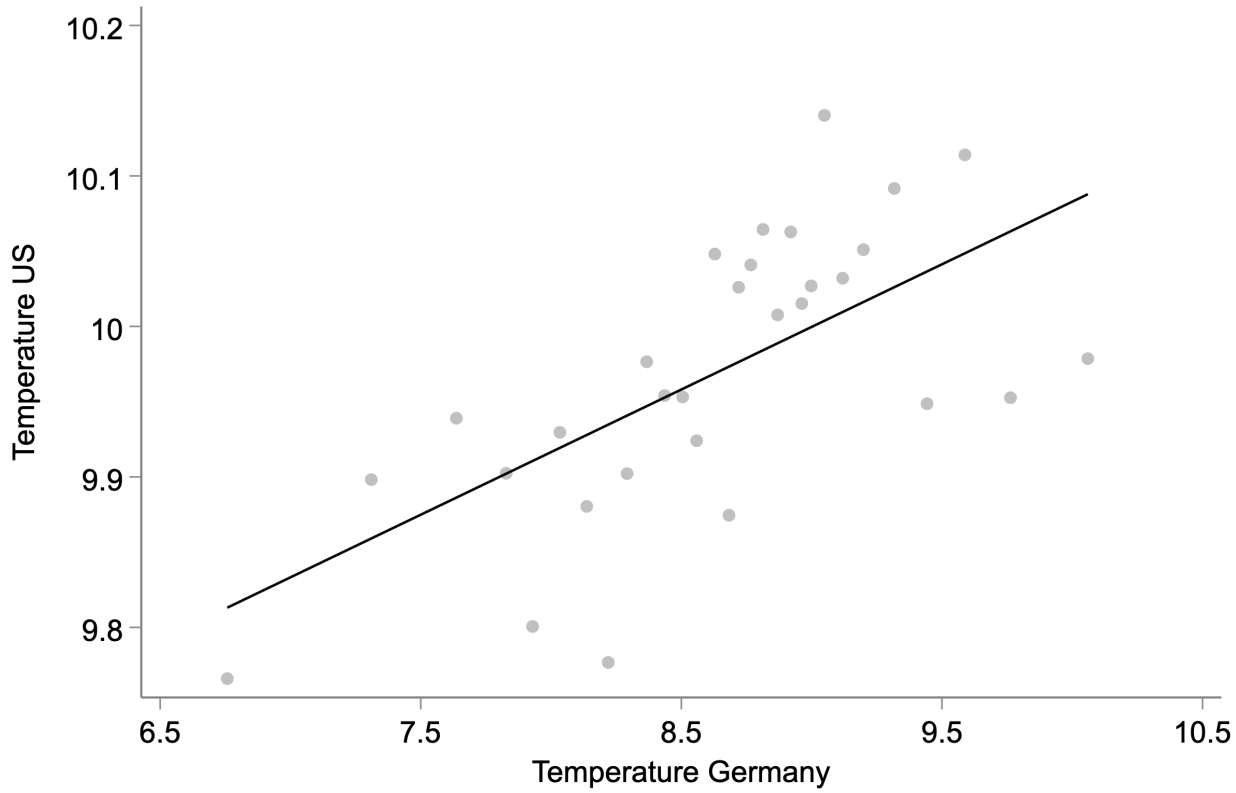
Notes: The figure displays the relationship between average temperature in degrees Celsius across US counties where immigrants from each origin were living in 1880 (y-axis) and the average temperature in the capital city of their country of origin (x-axis). The regression coefficient and the corresponding robust standard errors are, respectively, .231 and .033.

Figure 2. Climate Matching from US Mortality Data (1959-1961)



Notes: The figure displays the relationship between average temperature across US counties where foreign-born individuals from each country of origin died from 1959 to 1961 (y-axis) and average temperature in the capital city of their country of origin (x-axis). The regression coefficient and the corresponding robust standard errors are, respectively, .168 and .022.

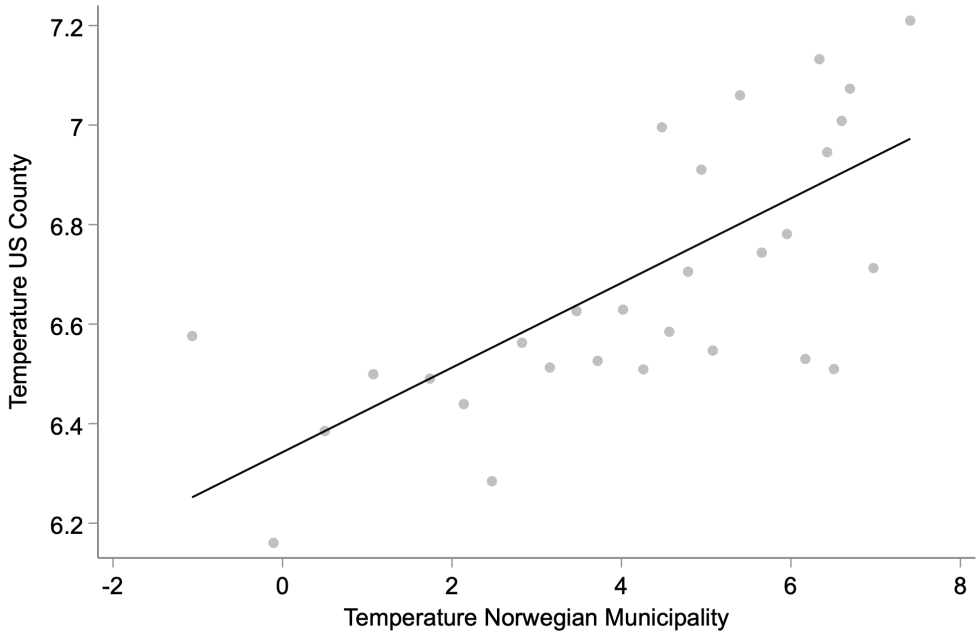
Figure 3. Temperature Matching from German Surnames



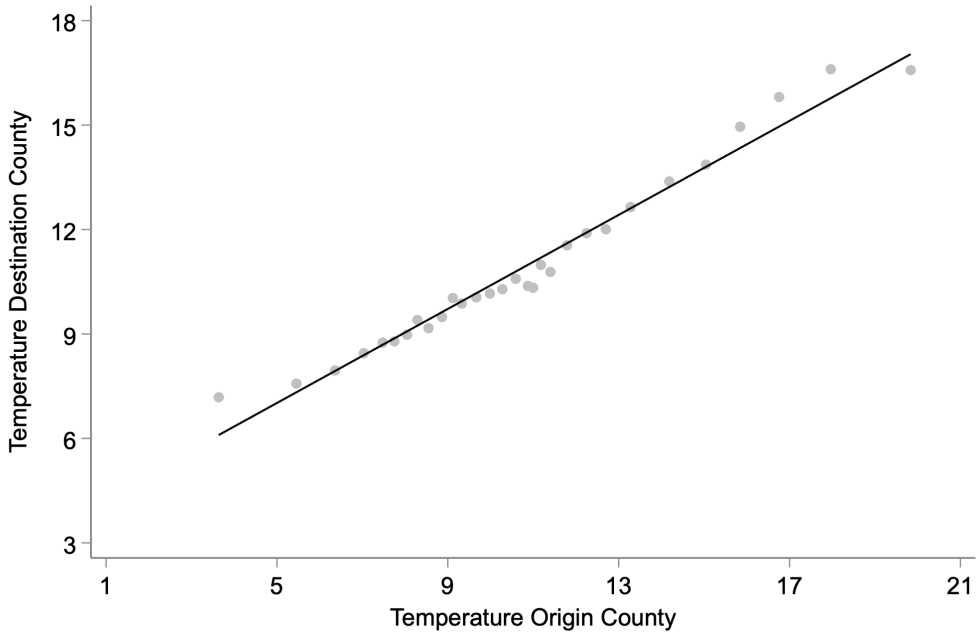
Notes: The figure displays a binned scatterplot obtained from a surname-level regression of average temperature at destination against average temperature at origin. The figure considers the sample of German immigrants' surnames linked between the WWI German casualties list and the 1880 US full count Censuses. The regression coefficient and the corresponding robust standard errors are, respectively, .083 and .011.

Figure 4. Temperature Matching and Migration at the Individual-Level

(a) Norwegian Immigrants



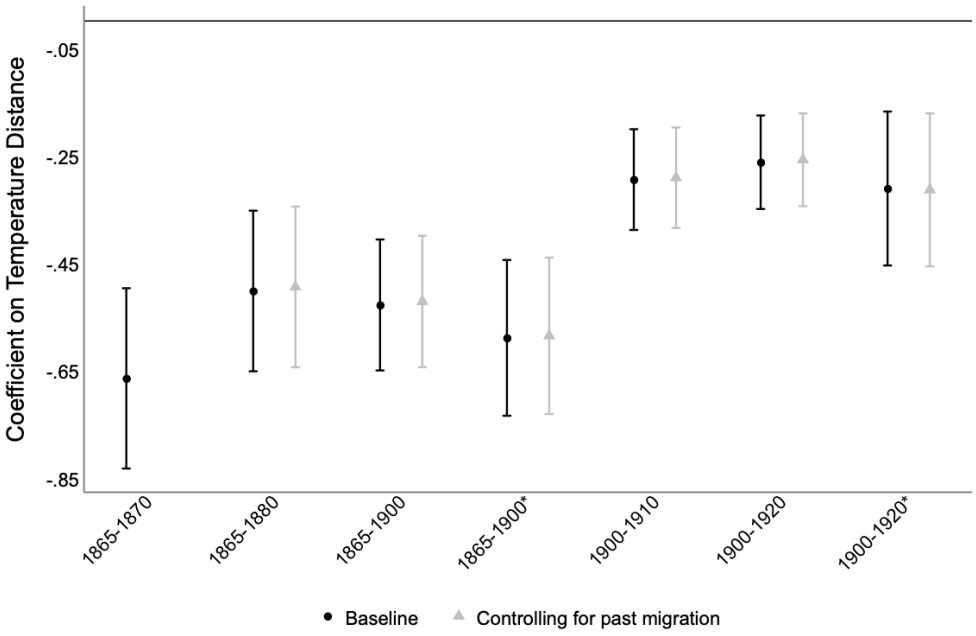
(b) US Internal Migrants



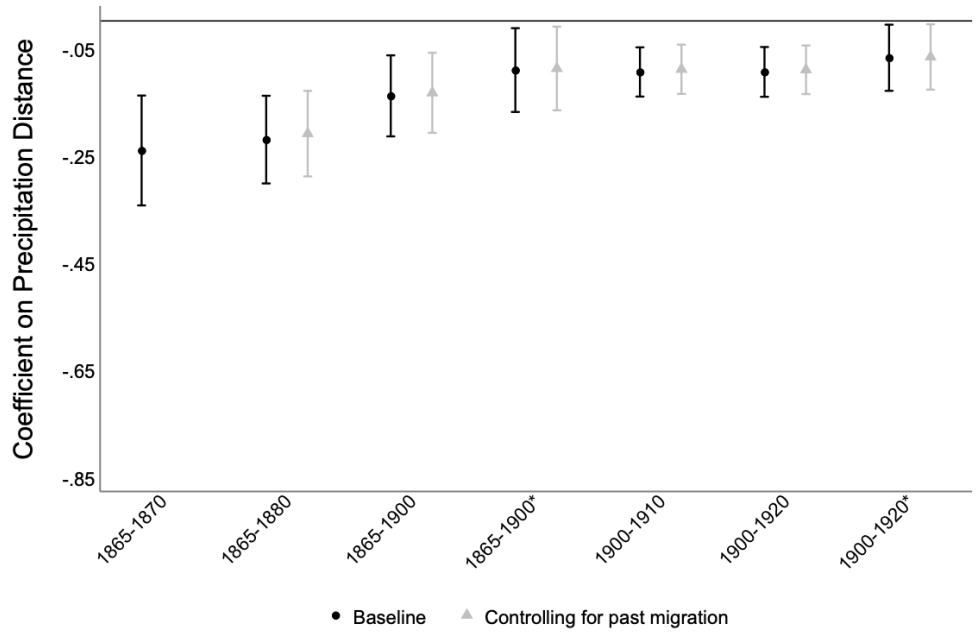
Notes: The figure displays binned scatterplots obtained from individual-level regressions of average temperature at destination against average temperature at origin. Panel A considers the sample of Norwegian immigrants linked between the 1865 Norwegian and the 1880 US full count Censuses. Origins and destinations are, respectively, Norwegian municipalities and counties in the contiguous US. Panel B considers the sample of individual migrants (within the contiguous US) linked across US full count Censuses over the period 1850-1860 to 1930-1940. Origins and destinations are counties in the contiguous US. In Panel A, the regression coefficient and the corresponding standard errors, clustered at the Norwegian province by US state level, are .085 and .038. In Panel B, the regression coefficient and the corresponding standard errors, clustered at the US state of origin by US state of destination by decade level, are .675 and .015.

Figure 5. Climate Distance and Norwegian Immigration, by Period

(a) Temperature



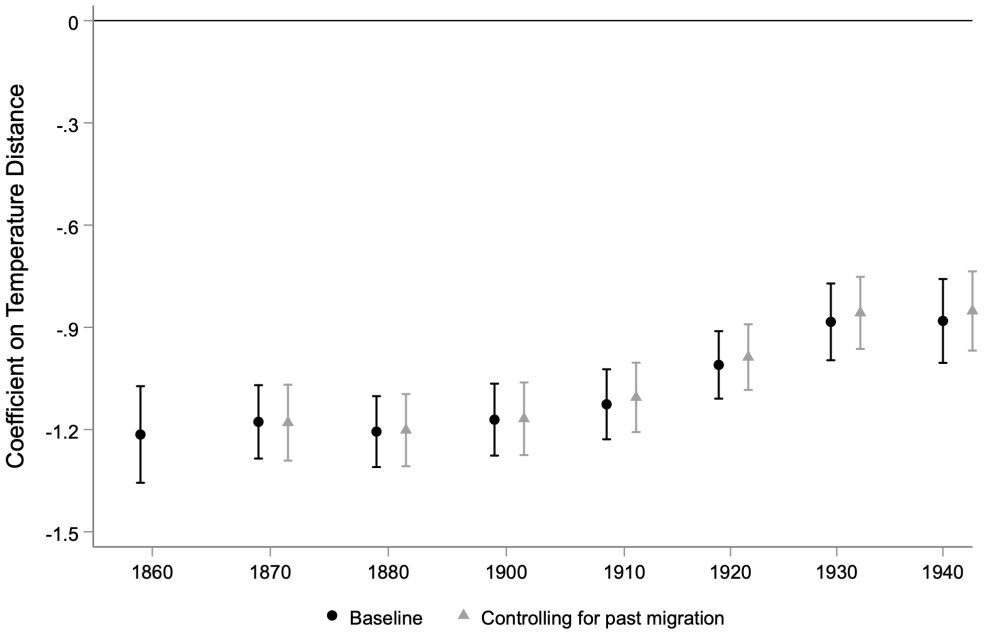
(b) Precipitation



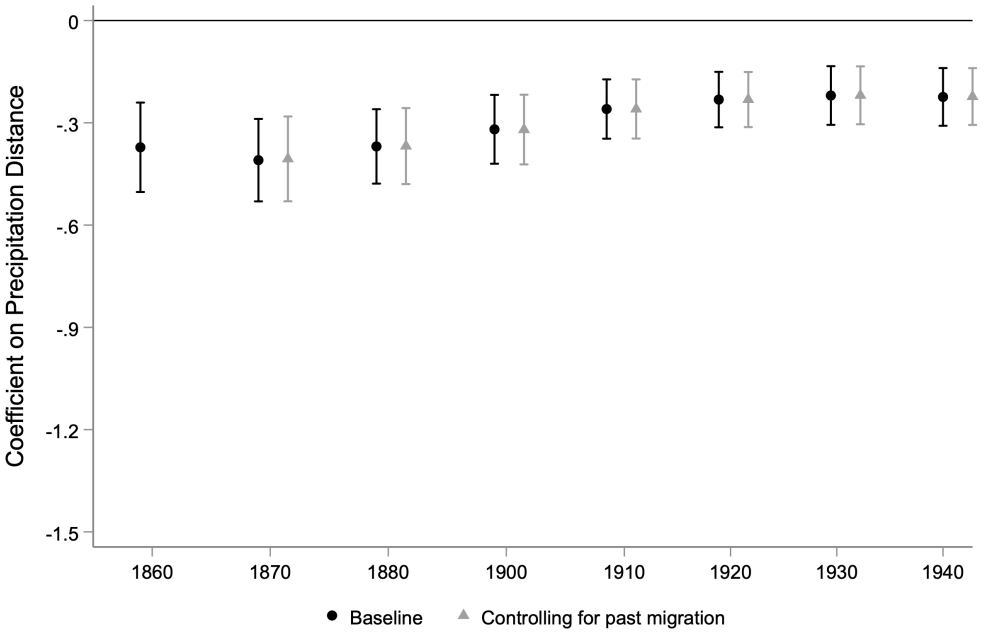
Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature (Panel A) and precipitation (Panel B) between Norwegian municipality of origin and US county of destination. Dots refer to the baseline specification, described in the notes of Table 1. Triangles refer to specifications that further controls for lagged number of immigrants from the municipality to the US county. The number of immigrants is defined over the window reported on the x-axis. The fourth (resp., seventh) set of coefficients from the left refers to individuals linked between the 1865 Norwegian and the 1900 US (resp., the 1900 Norwegian and the 1920 US) Censuses and who migrated after 1880 (resp., after 1910). Lagged migration refers to number of immigrants from 1865 to 1870 (resp., from 1865 to 1900) in the second to fourth triangles (in the last three triangles). Temperature and precipitation distances are standardized within the relevant sample to have zero mean and standard deviation equal to one. Standard errors are clustered at the US state by Norwegian province level.

Figure 6. Climate Distance and US Internal Migration, by Period

(a) Temperature



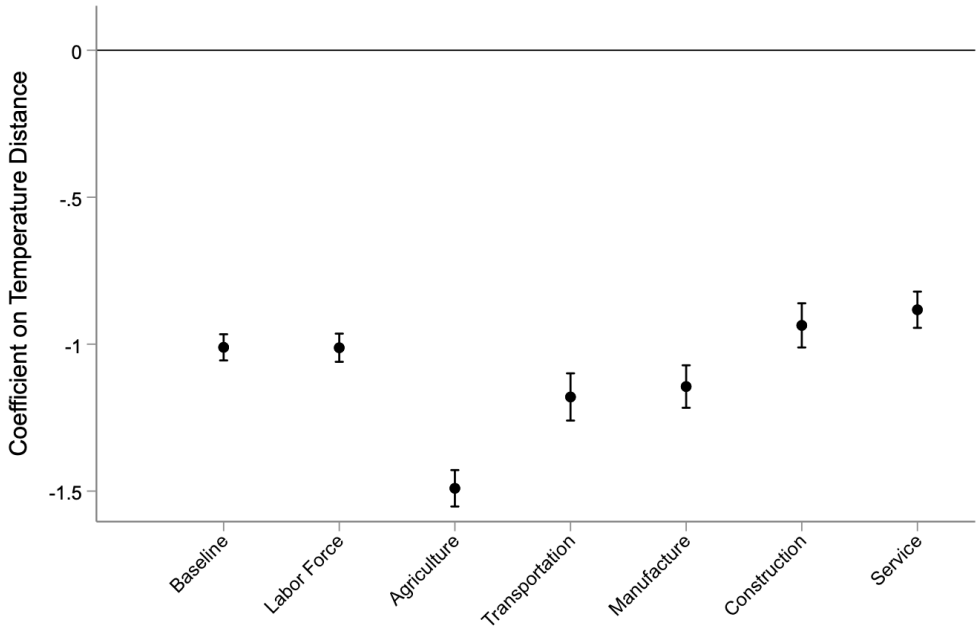
(b) Precipitation



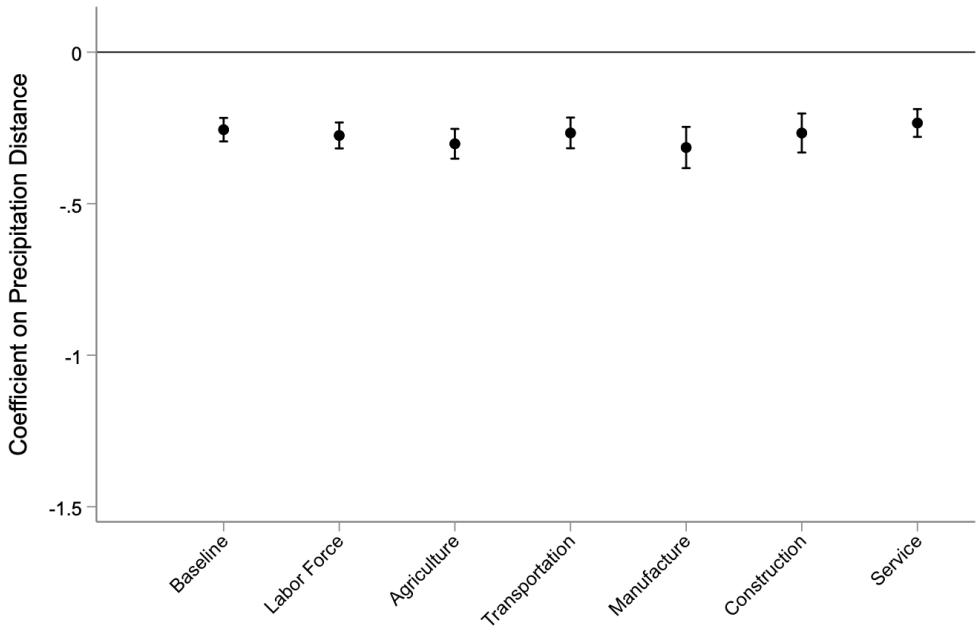
Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature (Panel A) and precipitation (Panel B) between origin and destination counties for each decade reported on the x-axis. Dots refer to the baseline specification, described in the notes of Table 2. Triangles refer to specifications that further controls for cumulated lagged number of immigrants. The year reported refers to the end of the decade. The sample includes all US county-pairs in the contiguous US from 1850-1860 to 1930-1940. Temperature and precipitation distances are standardized within the relevant sample to have zero mean and standard deviation equal to one. Standard errors are clustered at the state of origin by state of destination level.

Figure 7. Climate Distance and US Internal Migration: Heterogeneity by Sector

(a) Temperature

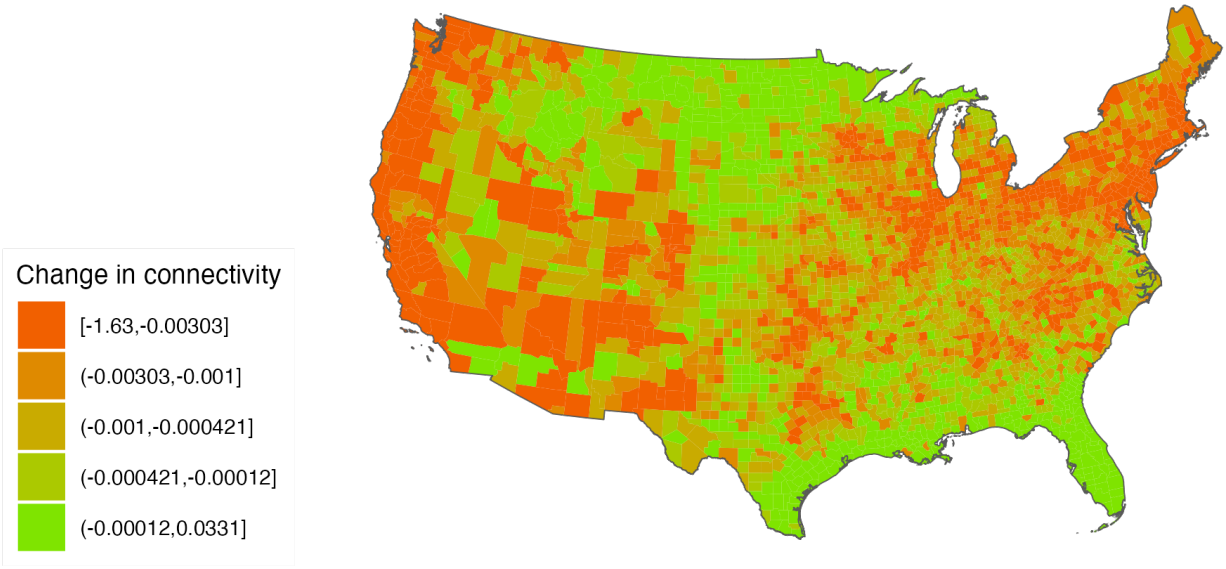


(b) Precipitation



Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature (Panel A) and precipitation (Panel B) between origin and destination counties for all migrants (resp., all migrants in the labor force) in the first (resp., second) dot from the left. From the third dot onward, the number of migrants refers to migrant men in the sector reported on the x-axis (defined according to 1-digit IND1950 code from IPUMS). The sample includes all US county-pairs in the contiguous US from 1850-1860 to 1930-1940. Temperature and precipitation distances are standardized within the relevant sample to have zero mean and standard deviation equal to one. Standard errors are clustered at the state of origin by state of destination by decade level.

Figure 8. Change in Temperature Induced Migration Connectivity over the 21st Century



Notes: The figure depicts the change in temperature -induced connectivity between each contiguous US county pairs between the beginning and the end of the 21st century. Temperature-induced connectivity is given by: $CC_o = \sum_d (\text{Pop}_d^{2010} \times \Delta Mig_{od}^{2000-2100})$ where $\Delta Mig_{od}^{2000-2100}$ is the percent change in bilateral migration flows between any two counties o and d, driven by changes in temperature distance, between 1990-2010 and 2080-2100. Pop_d^{2010} is the 2010 population share of county d, relative to total US population. Florida's climate is expected to converge to the rest of the US—increasing its climate connection to populous place.

Table 1. Climate Distance and Norwegian Immigration (1865-1880)

Dep. var.:	Number of Migrants			
	(1)	(2)	(3)	(4)
Temperature Distance	-0.373***	-0.375***	-0.143***	-0.107***
	(0.019)	(0.019)	(0.024)	(0.016)
	[-16.377]	[-16.474]	[-5.683]	[-4.257]
Precipitation Distance	-0.007***	-0.007***	-0.007***	-0.006***
	(0.002)	(0.003)	(0.001)	(0.001)
	[-2.188]	[-2.456]	[-2.135]	[-1.967]
Distance from NYC (in 100km)		0.009**	-0.136***	-0.051**
		(0.004)	(0.035)	(0.025)
		[0.728]	[-10.937]	[-4.130]
Observations	1,002,822	1,002,822	847,894	847,894
Pseudo R-squared	0.121	0.122	0.287	0.401
Mean Temp. Dist.	7.954	7.954	7.429	7.429
SD Temp. Dist.	4.691	4.691	4.609	4.609
Mean Precip. Dist.	40.04	40.04	40.17	40.17
SD Precip. Dist.	35.06	35.06	35.16	35.16
US State FE			Yes	Yes
Norwegian Province FE			Yes	Yes
US Controls				Yes
Norwegian Controls				Yes

Notes: The sample includes pairs formed by all Norwegian municipalities (origins) and all US counties (destinations) with at least one European immigrant in 1880. *Number of Migrants* is the number of 10+ Norwegian male immigrants linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the Norwegian municipality and the US county. *Distance from NYC* refers to the distance between a US county and New York City (and is expressed in 100 km). US and Norwegian controls include: population density, sex ratios (defined as the ratio between women 18-33 and men 20-35), the urban population share, the share of men (15-64) in the labor force, and the share of male employment in agriculture and manufacturing. US controls also include the following variables, which cannot be computed for Norwegian municipalities: the immigrant population share, the Norwegian population share, the Black population share, total frontier experience from Bazzi et al. (2020), and market access from Donaldson and Hornbeck (2016). All Norwegian controls are measured in 1865; all US controls are measured in 1880, except for market access, which is measured in 1870. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the US state by Norwegian province level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 2. Climate Distance and US Internal Migration

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.375***	-0.302***	-0.264***	-0.262***	-0.260***	-0.224***	-0.231***
	(0.011)	(0.008)	(0.006)	(0.006)	(0.006)	(0.016)	(0.020)
	[-0.111]	[-0.089]	[-0.078]	[-0.076]	[-0.070]	[-0.042]	[-0.033]
Precipitation Distance	-0.039***	-0.014***	-0.012***	-0.011***	-0.011***	-0.009***	-0.010***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
	[-0.070]	[-0.025]	[-0.021]	[-0.020]	[-0.018]	[-0.011]	[-0.009]
Distance (in 100km)		-0.138***	-0.149***	-0.148***	-0.149***	-0.176***	-0.237***
		(0.012)	(0.004)	(0.004)	(0.005)	(0.014)	(0.018)
		[-0.085]	[-0.091]	[-0.090]	[-0.085]	[-0.067]	[-0.070]
Observations	35,634,556	35,634,556	35,332,029	34,504,246	30,479,809	4,684,923	4,818,958
Pseudo R-squared	0.169	0.206	0.682	0.691	0.705	0.763	0.746
Mean Temp. Dist.	5.426	5.426	5.418	5.417	5.321	5.461	5.479
SD Temp. Dist.	3.884	3.884	3.879	3.881	3.829	3.917	3.931
Mean Precip. Dist.	31.16	31.16	31.11	30.96	29.61	31.15	31.27
SD Precip. Dist.	23.62	23.62	23.61	23.60	23.21	23.78	23.83
County o FE			Yes	Yes		Yes	Yes
County d FE			Yes	Yes		Yes	Yes
Decade FE			Yes				
State o × Decade FE				Yes			
State d × Decade FE				Yes			
County o × Decade FE					Yes		
County d × Decade FE					Yes		
Sample	Linked	Linked	Linked	Linked	Linked	Linked 30-40	Full Count 35-40

Notes: The sample includes all county-pairs in the contiguous US: for each decade from 1850-1860 to 1930-1940 (columns 1 to 5), for 1930-40 (column 6), and for 1935-1940 (column 7). *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period (decade in columns 1 to 6; 5-year period in column 7). The migration matrix in columns 1 to 6 (resp., column 7) is derived from the linked sample (resp., the 1940 full count US Population Census). See Appendix B for more details. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). County (resp., State) o refers to county (resp., state) of origin; county (resp., State) d refers to county (state) of destination. Standard errors are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Climate Distance and Migration in the Long Run

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.227***	-0.166***	-0.222***	-0.134***	-0.192***	-0.298***
	(0.024)	(0.047)	(0.045)	(0.037)	(0.049)	(0.074)
	[-0.015]	[-0.005]	[-0.007]	[-0.005]	[-0.008]	[-0.009]
Precipitation Distance	-0.015***	-0.018***	-0.020***	-0.018***	-0.019***	-0.007
	(0.003)	(0.005)	(0.004)	(0.005)	(0.006)	(0.006)
	[-0.006]	[-0.004]	[-0.005]	[-0.004]	[-0.005]	[-0.001]
Distance (in 100km)	-0.226***					
	(0.016)					
	[-0.032]					
Observations	7,703,638	2,188,200	2,184,658	2,670,748	2,049,054	1,959,248
Pseudo R-squared	0.771	0.964	0.965	0.963	0.961	0.952
Mean Temp. Dist.	5.302	4.103	4.105	4.052	4.118	3.570
SD Temp. Dist.	3.825	3.391	3.392	3.364	3.403	3.088
Mean Precip. Dist.	31.78	25.61	25.62	25.78	25.58	25.04
SD Precip. Dist.	24.45	23.95	23.96	23.93	23.98	23.22
County o × Period FE	Yes	Yes	Yes	Yes	Yes	Yes
County d × Period FE	Yes	Yes	Yes	Yes	Yes	Yes
County o × County d FE		Yes	Yes	Yes	Yes	Yes
Distance × Period FE		Yes	Yes	Yes	Yes	Yes
Controls × Period FE			Yes			
Modern Period	2018-2019	2018-2019	2018-2019	2011-2012	2014-2015	Avg. 2011-2019
Historical Period	1935-1940	1935-1940	1935-1940	1935-1940	1935-1940	Avg. 1910-1940

Notes: The sample includes all county-pairs in the contiguous US. *Number of Migrants* in historical period is the number of men 15+ in the baseline year who moved from the origin to the destination county. *Number of Migrants* in the modern period is the number of people who changed address from the origin to the destination county in each year, and is constructed from IRS Migration data. The migration matrix is derived from the linked sample. See Appendix B for more details. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). Controls include absolute difference in employment share, manufacturing employment share, agriculture employment share, black share, urban share, immigrant share, sex ratio, population density, exposure to the frontier, access to market distance, and a dummy equals to 1 if the two counties are both connected to the railroad. County o (resp., county d) refers to county of origin (resp., of destination). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by period level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 4. Climate Distance and US Internal Migration (2011-2019)

Dep. var.:	Number of Migrants			
	(1)	(2)	(3)	(4)
Temperature Distance	-0.215***	-0.212***	-0.207***	-0.207***
	(0.014)	(0.013)	(0.014)	(0.014)
	[-0.010]	[-0.009]	[-0.009]	[-0.009]
Precipitation Distance	-0.017***	-0.017***	-0.017***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
	[-0.005]	[-0.005]	[-0.005]	[-0.005]
Distance (in 100km)	-0.215***	-0.215***	-0.207***	-0.207***
	(0.009)	(0.008)	(0.009)	(0.009)
	[-0.022]	[-0.021]	[-0.021]	[-0.021]
Lagged Migration Period			3.932***	0.482***
			(0.414)	(0.050)
			[0.000]	[0.000]
Observations	25,071,430	31,024,472	25,071,430	25,071,430
Pseudo R-squared	0.781	0.781	0.783	0.783
Mean Temp. Dist.	5.125	5.123	5.125	5.125
SD Temp. Dist.	3.722	3.721	3.722	3.722
Mean Precip. Dist.	31.35	31.30	31.35	31.35
SD Precip. Dist.	25.22	25.21	25.22	25.22
County o × Decade FE	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes
Definition Lagged Mig.			Avg. Flows 1850-1940	Stock 1850-1940

Notes: The sample includes all county-pairs in the contiguous US for each year from 2011-2012 to 2018-2019, except for column 2, which includes moves occurring in 2019-2020 and 2020-2021. *Number of Migrants* is the number of people who changed address from the origin to the destination county in each year and is constructed from IRS Migration data. These data are based on year-to-year address changes reported on individual income tax returns filed with the IRS. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). Stock 1850-1940 (resp., Avg. Flows 1850-1940) is the total (resp., the average) number of migrants from the origin to the destination county in each decade between 1850-1860 and 1930-1940 obtained from the linked sample (and is expressed in 10 thousand). All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Climate, Migration, and the Homestead Act

Dep. var.:	Number of Migrants			
	(1)	(2)	(3)	(4)
Temperature Distance	-0.299***	-0.385***	-0.256***	-0.131***
	(0.007)	(0.009)	(0.007)	(0.017)
	[-0.181]	[-2.833]	[-0.163]	[-5.213]
Temperature Distance × Homestead	0.147***	0.179***	0.159***	0.081***
	(0.011)	(0.015)	(0.011)	(0.021)
	[0.042]	[0.616]	[0.046]	[2.193]
Precipitation Distance	-0.014***	-0.024***	-0.011***	-0.006***
	(0.001)	(0.002)	(0.001)	(0.001)
	[-0.050]	[-1.035]	[-0.042]	[-1.743]
Precipitation Distance × Homestead	0.002	0.013***	0.000	0.000
	(0.002)	(0.004)	(0.003)	(0.001)
	[0.005]	[0.290]	[0.000]	[0.031]
Distance (in 100km)	-0.185***	-0.322***	-0.147***	-0.065**
	(0.009)	(0.013)	(0.009)	(0.026)
	[-0.245]	[-5.206]	[-0.206]	[-5.241]
Distance × Homestead	0.065***	0.066***	0.073***	-0.062***
	(0.007)	(0.011)	(0.007)	(0.014)
	[0.049]	[0.615]	[0.058]	[-4.722]
Homestead				0.975***
				(0.224)
				[3.297]
Observations	27,678,550	26,956,182	26,937,930	847,894
Pseudo R-squared	0.600	0.387	0.719	0.402
Mean Temp. Dist.	5.113	5.071	5.091	7.429
SD Temp. Dist.	3.699	3.668	3.686	4.609
Mean Precip. Dist.	27.22	26.94	27.10	40.17
SD Precip. Dist.	22.06	21.85	22.03	36.16
County o × Decade FE	Yes	Yes	Yes	
County d × Decade FE	Yes	Yes	Yes	
US Controls				Yes
Norwegian Controls				Yes
US State FE				Yes
Norwegian Province FE				Yes
Occupation	Any	Farmer	Non-Farmer	Any
Sample	US	US	US	Norwegian

Notes: The sample includes: for each decade from 1860-1870 to 1910-1920 all county-pairs in the contiguous US, such that origin counties never offered any Homestead land, in columns 1 to 3; pairs formed by all Norwegian municipalities (origins) and all US counties (destinations) with at least one European immigrant in 1880, in column 4. *Number of Migrants* is the number of men: aged 15+ in the baseline year who moved from the origin to the destination county in each period, derived from the US linked sample (columns 1 to 3); age 10+ in 1865 linked between the 1865 Norwegian and the 1880 US Censuses (column 4). See Appendix B for more details. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the temperature (resp., precipitation) difference between origin and destination. *Distance* is the physical distance between counties (resp., between a county and New York City) in columns 1 to 3 (resp., column 4), and is expressed in 100 km. In columns 1 to 3 (resp., column 4) *Homestead* is a dummy equal to one if the homesteaded share of the area in the county of destination in a given decade (resp., between 1865 and 1880) is above the 80^{th} percentile of the homesteaded share of county area calculated over the entire US for the 1860-1920. Column 1 includes all migrants, while columns 2 and 3 restrict attention to men 15+ (in the baseline year) who were farmers and non-farmers, respectively. See notes to Table 2 (resp., Table 1) for the detailed list of controls included in columns 1 to 3 (resp., column 4). County o (resp., county d) refers to county of origin (resp., of destination). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by decade (resp., at the US state by Norwegian province) level in columns 1 to 3 (resp., column 4). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Climate Mismatch and Age at Death

Dep. var.:	Age at Death					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.277***	-0.120***			-0.271***	-0.121***
	(0.049)	(0.028)			(0.047)	(0.028)
	[-0.073]	[-0.032]			[-0.071]	[-0.032]
Precipitation Distance			-0.031***	-0.001	-0.029***	-0.003
			(0.010)	(0.006)	(0.009)	(0.006)
			[-0.052]	[-0.002]	[-0.050]	[-0.005]
Observations	727,818	727,570	727,818	727,570	727,818	727,570
R-squared	0.005	0.187	0.003	0.186	0.008	0.187
Mean Temp. Dist.	3.239	3.238			3.239	3.238
SD Temp. Dist.	3.022	3.022			3.022	3.022
Mean Precip. Dist.			31.30	31.30	31.30	31.30
SD Precip. Dist.			19.41	19.41	19.41	19.41
US County FE		Yes		Yes		Yes
Country FE		Yes		Yes		Yes
Death Year FE		Yes		Yes		Yes
Demographic Controls		Yes		Yes		Yes

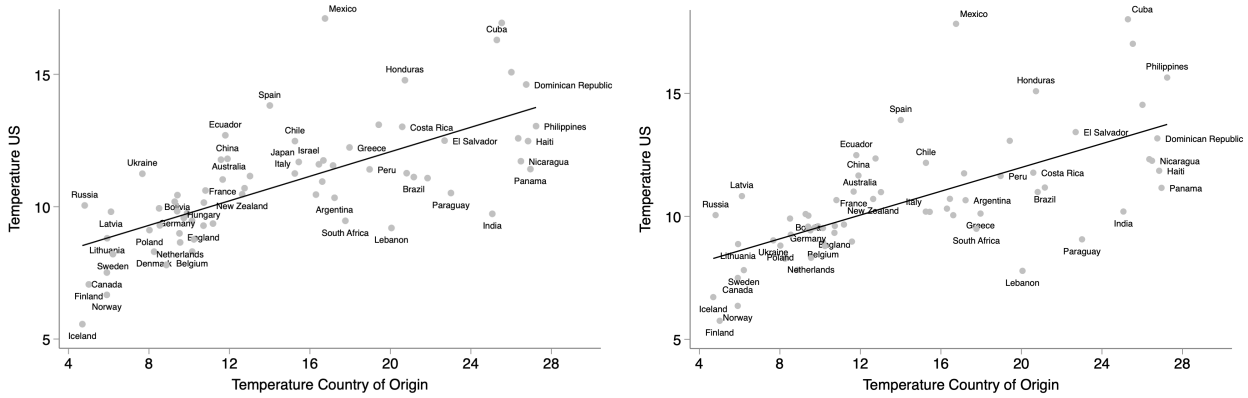
Notes: The sample includes all foreign-born individuals died from 1959 to 1961 in the continental US. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between birth country capital and US county of death. Columns 1, 3, and 5 show the raw correlations. Columns 2, 4, and 6 include: US county of death, country of birth, and year of death fixed effects; and, individual controls (gender and marital status). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the US state by country of birth level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

A Additional Figures and Tables

Figure A1. Temperature Matching of Immigrants in the US, by Decade

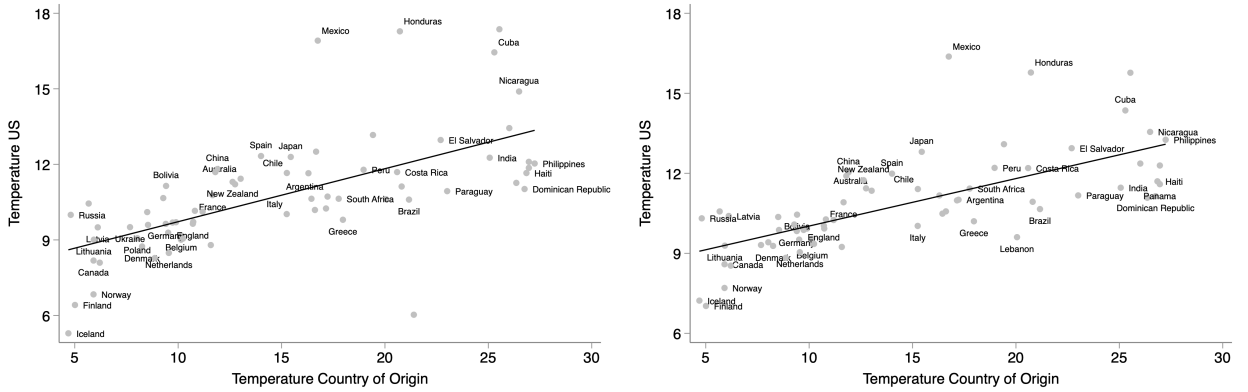
(a) 1880

(b) 1900



(c) 1920

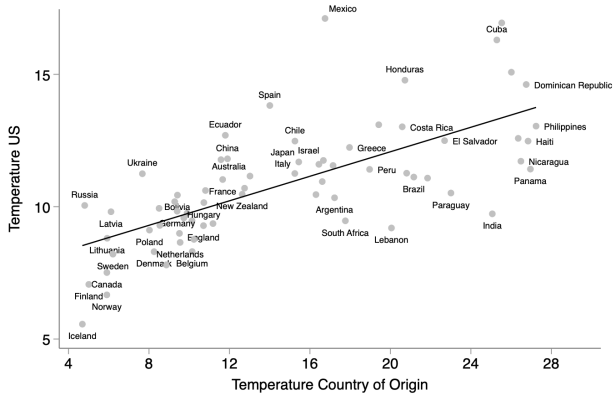
(d) 1940



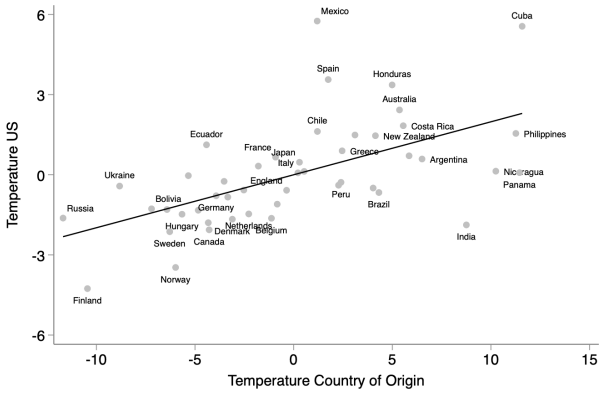
Notes: The figure displays the relationship between average temperature in degrees Celsius across US counties where immigrants from each origin lived in a given decade (y-axis) and the average temperature in the capital city of their country of origin (x-axis). Panel A reports results for 1880, replicating Figure 1. Panels B, C, and D consider, respectively: 1900, 1920, and 1940. The regression coefficients (robust standard errors) are: .231 (.033) in Panel A, .242 (.038) in Panel B, .210 (.034) in Panel C, and .178 (.024) in Panel D.

Figure A2. Temperature Matching of Immigrants in the US (1880): Robustness

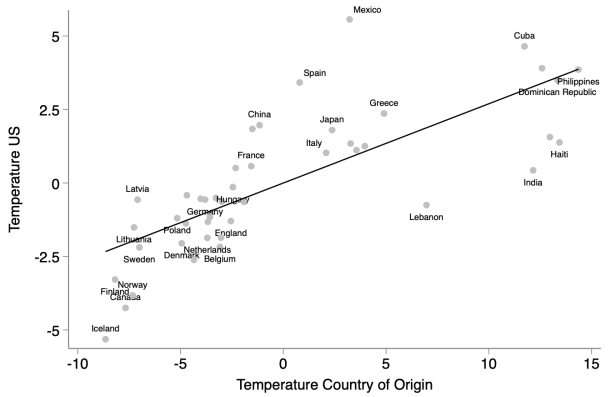
(a) Baseline



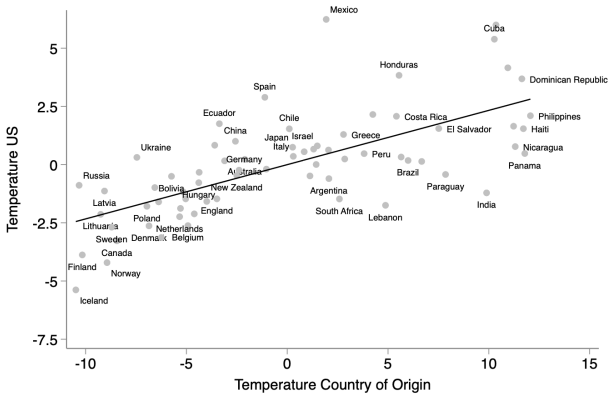
(b) GDP per capita



(c) Distance



(d) Immigrants in 1870

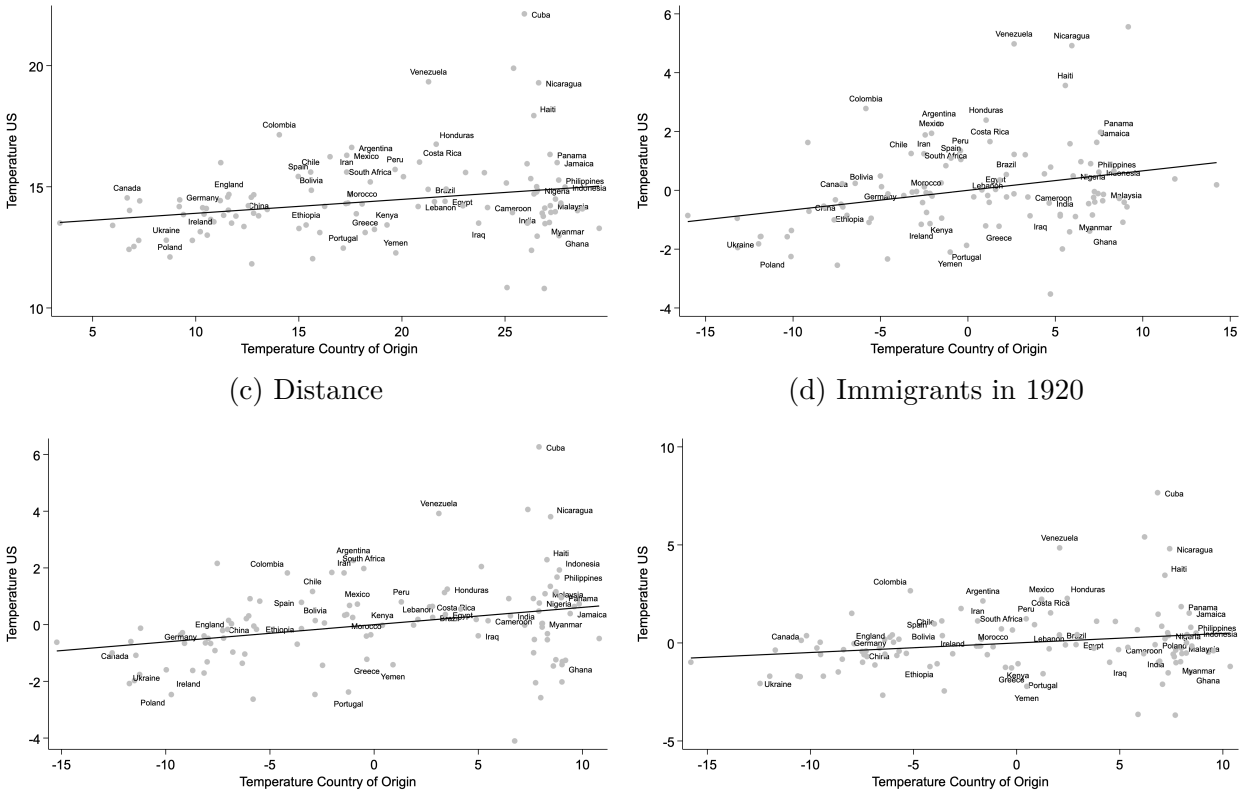


Notes: The figure displays the relationship between average temperature in degrees Celsius across US counties where immigrants from each origin were living in 1880 (y-axis) and the average temperature in the capital city of their country of origin (x-axis), after controlling for additional factors. Panel A replicates Figure 1. Panels B, C, and D control for: country of origin GDP per capita in 1920; the distance between the US county of residence and the foreign capital city; the number of immigrants from each country of origin who were living in each US county in 1870. The regression coefficients (robust standard errors) are: .231 (.033) in Panel A, .198 (.047) in Panel B, .270 (.038) in Panel C, and .233 (.034) in Panel D.

Figure A3. Temperature Matching of Immigrants in the US: 2015

(a) Baseline

(b) GDP per capita



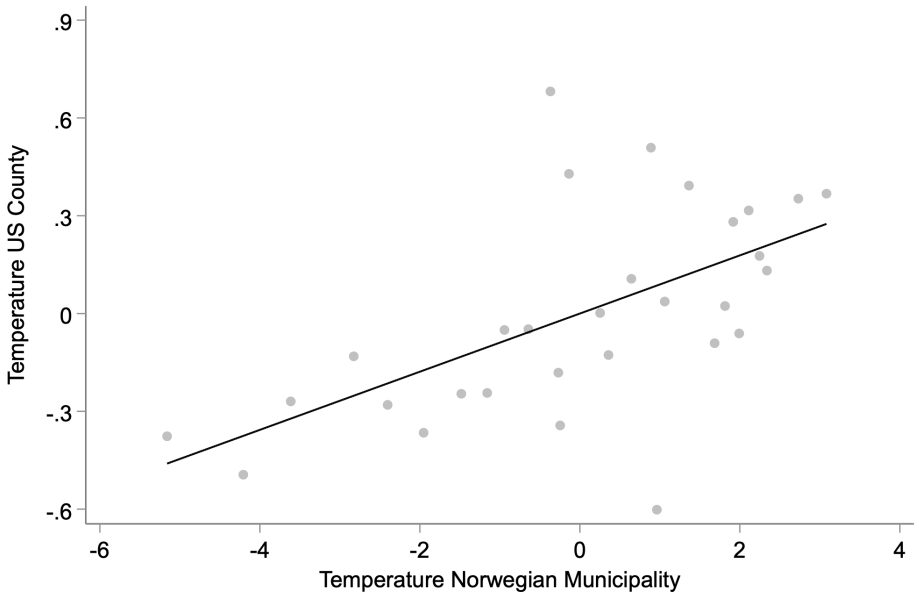
(c) Distance

(d) Immigrants in 1920

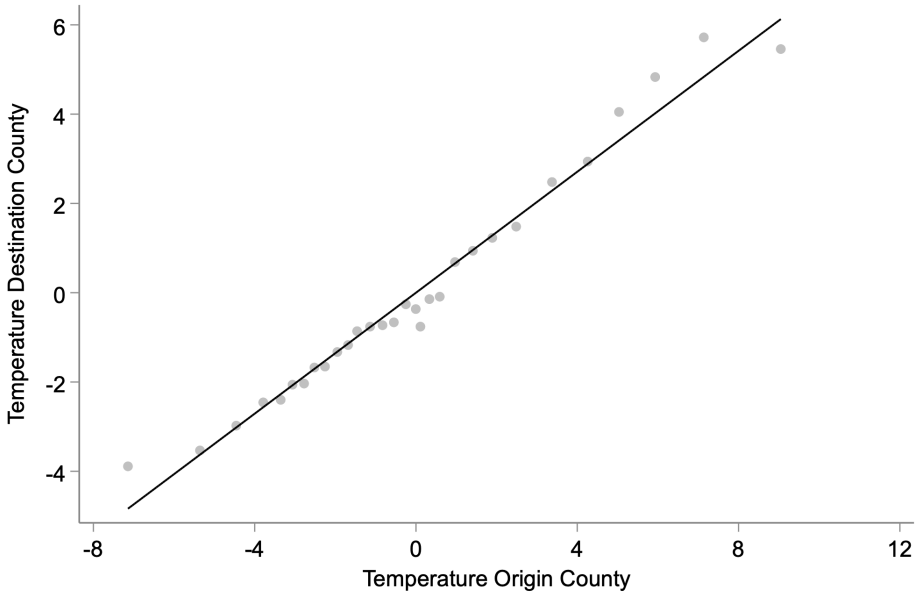
Notes: The figure displays the relationship between average temperature in degrees Celsius across US counties where immigrants from each origin were living in 2011-2015 (y-axis) and the average temperature in the capital city of their country of origin (x-axis), after controlling for additional factors. Panel A displays the raw correlation. Panels B, C, and D control for: country of origin GDP per capita in 2010; the distance between the US county of residence and the foreign capital city; the number of immigrants from each country of origin who were living in each US county in 1920. The regression coefficients (robust standard errors) are: .057 (.020) in Panel A, .066 (.020) in Panel B, .061 (.019) in Panel C, and .049 (.021) in Panel D.

Figure A4. Temperature Matching and Migration, Controlling for Geography

(a) Norwegian Immigrants



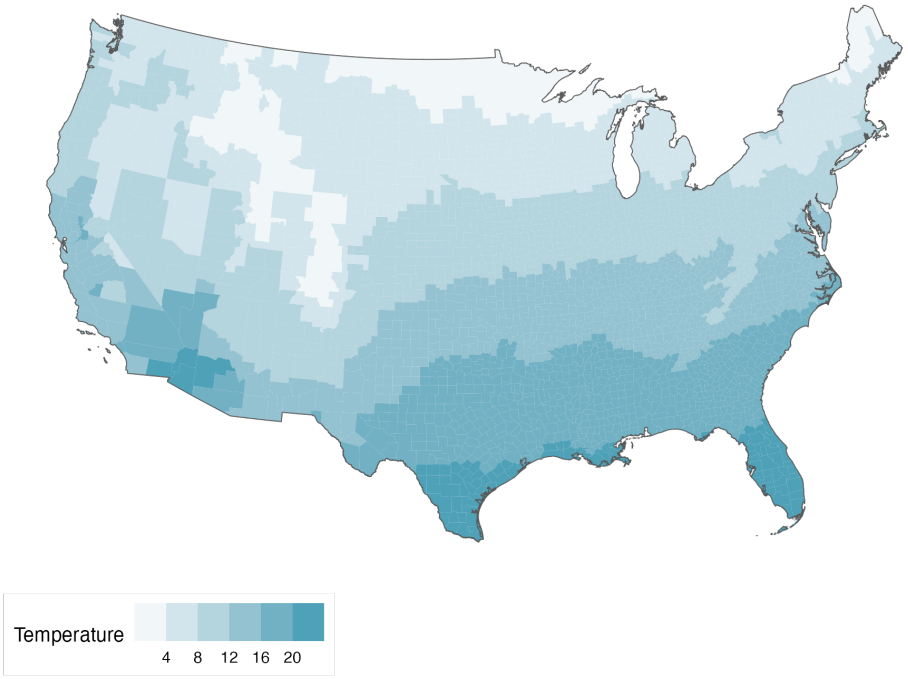
(b) US Internal Migrants



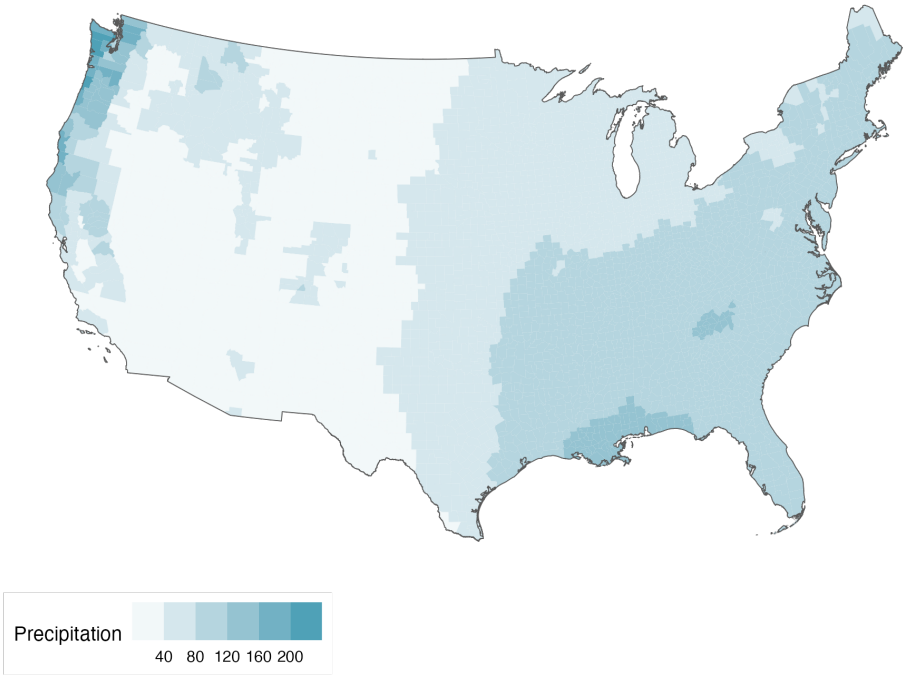
Notes: The figure displays binned scatterplots obtained from individual-level regressions of average temperature at destination against average temperature at origin when other controls are included. Panel A considers the sample of Norwegian immigrants linked between the 1865 Norwegian and the 1880 US full count Censuses. Origins and destinations are, respectively, Norwegian municipalities and counties in the contiguous US. In this panel, we control for the distance of the destination county from New York and for the distance of the origin municipality from the closest Atlantic port (either Oslo or Kristiansand). Panel B considers the sample of individual migrants (within the contiguous US) linked across US full count Censuses over the period 1850-1860 to 1930-1940. Origins and destinations are counties in the contiguous US. In this panel, we control for the distance between the origin county and the destination county (expressed in 100 km). In Panel A, the regression coefficient and the corresponding standard errors, clustered at the Norwegian province by US state level, are .089 and .039. In Panel B, the regression coefficient and the corresponding standard errors, clustered at the US state of origin by US state of destination by decade level, are .677 and .016.

Figure A5. US County Historical Climate (1895-1920)

(a) Temperature



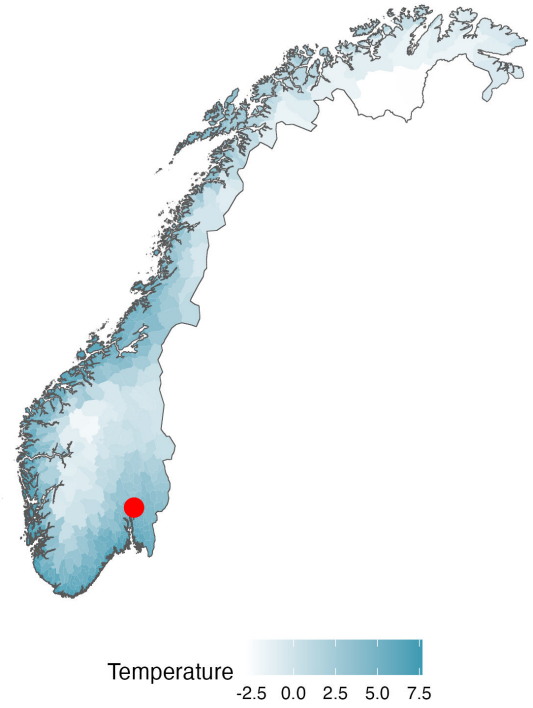
(b) Precipitation



Notes: All year averages at the US county level for the period 1895-1920 based on NOAA monthly gridded historical data (Vose et al., 2014). Temperature and precipitation are expressed in Celsius degrees and mm/month respectively.

Figure A6. Norwegian Municipalities Climate (1958-1980)

(a) Temperature



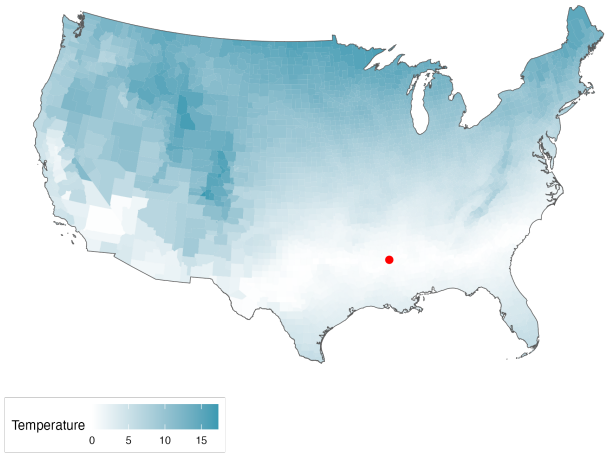
(b) Precipitation



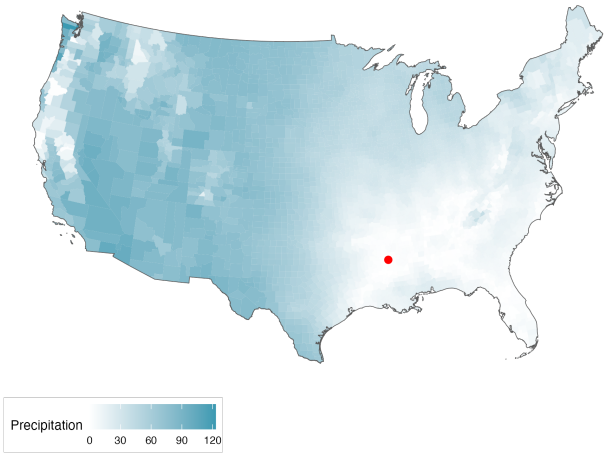
Notes: All year averages at the Norwegian municipality level for the period 1958-1980 based on TerraClimate monthly gridded historical data (Abatzoglou et al., 2018). Temperature and precipitation are expressed in Celsius degrees and mm/month respectively.

Figure A7. Climate Distances Between US Counties

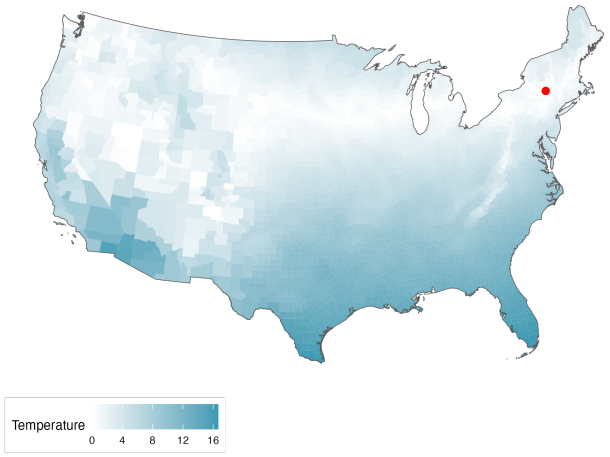
(a) Temp. Distance, Delta region



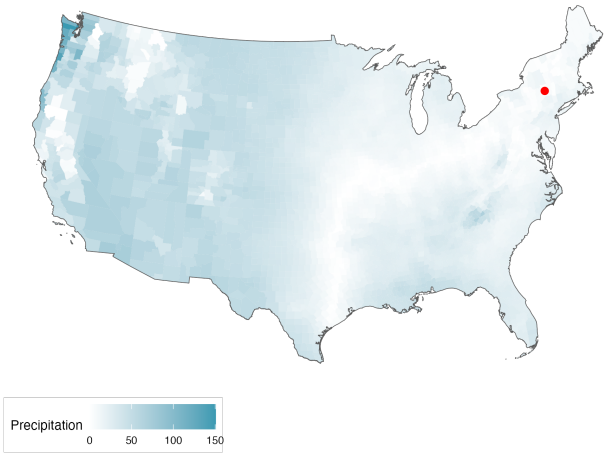
(b) Precip. Distance, Delta region



(c) Temp. Distance, US Northeast



(d) Precip. Distance, US Northeast

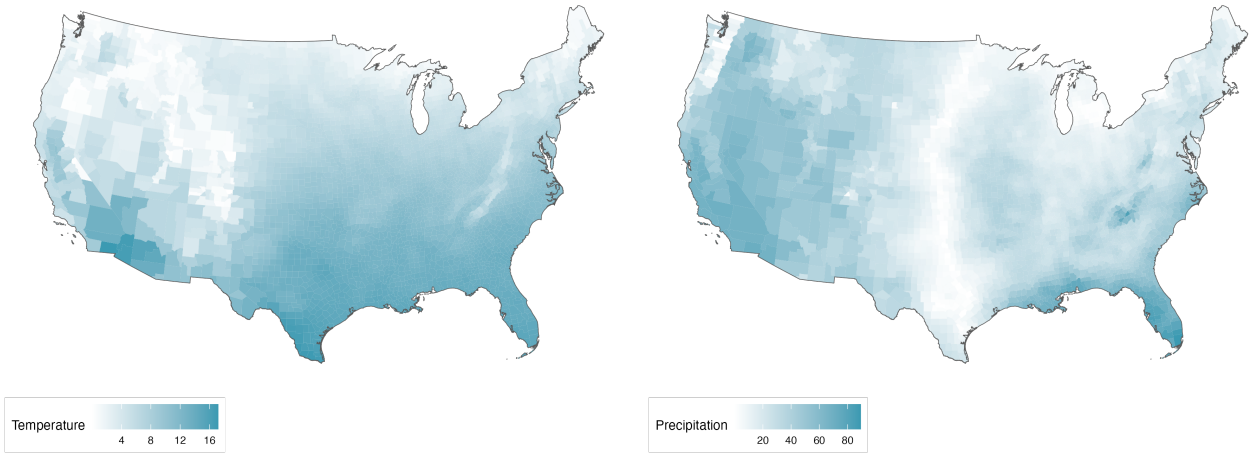


Notes: Red dot is the centroid of the considered region. Temperature and precipitation are averages of all-year climate in degrees C and mm/month, respectively. These averages are taken over the period 1895-1920 using NOAA data (Vose et al., 2014). To interpret the figure, note that the distance in temperature between the Delta region and New York City is +12° C, even though New York City is 12° C colder than the Delta region.

Figure A8. Climate Distances Between Oslo and US Counties

(a) Temp. Distance, Oslo

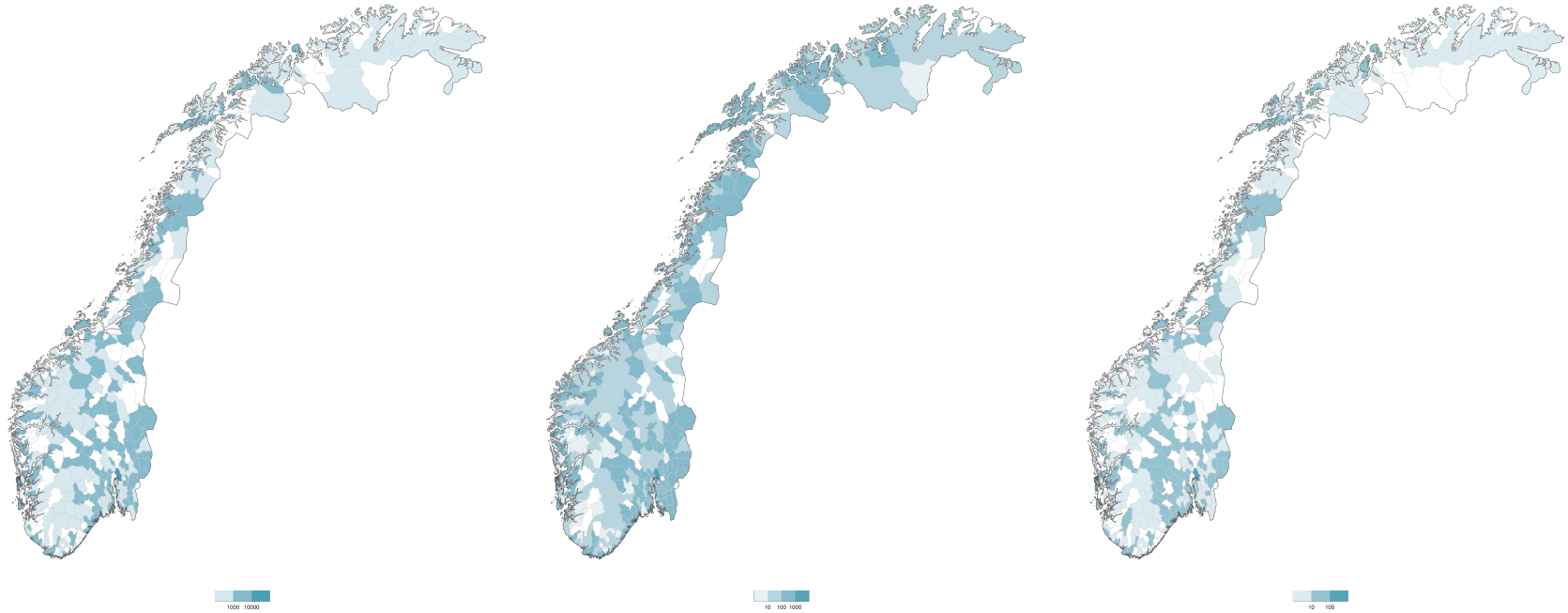
(b) Precip. Distance, Oslo



Notes: Temperature and precipitation are averages of all-year climate in degrees C and mm/month, respectively. For the US, these averages are taken over the period 1895-1920 using NOAA data (Vose et al., 2014). For Norway, TerraClimate data are averaged over the period 1958-1980 (Abatzoglou et al., 2018).

Figure A9. Distribution of Population in Norway (1865)

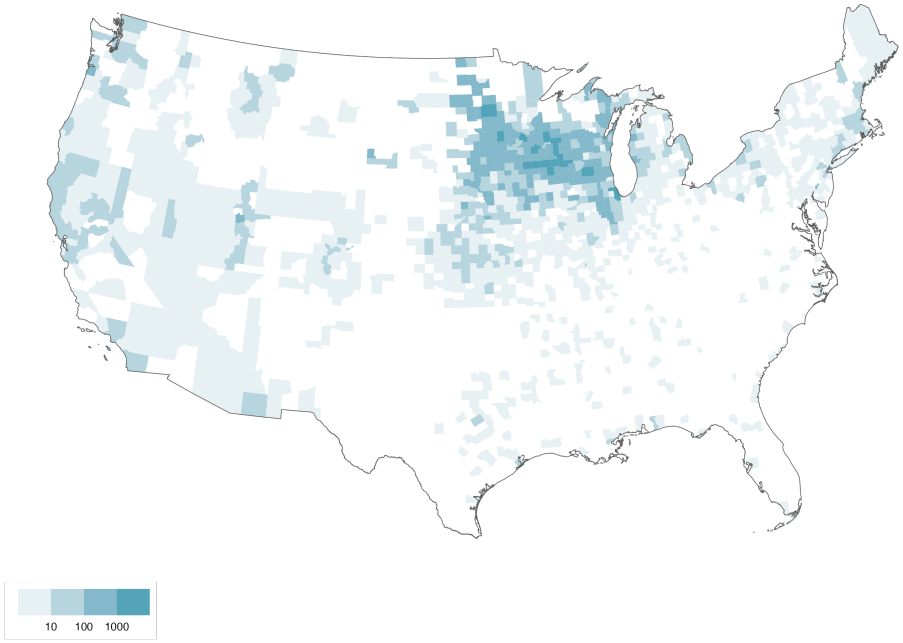
(a) Norwegians in 1865... (b) ...Matched to 1900 Norwegian Census (c) ...Matched to 1880 US Census



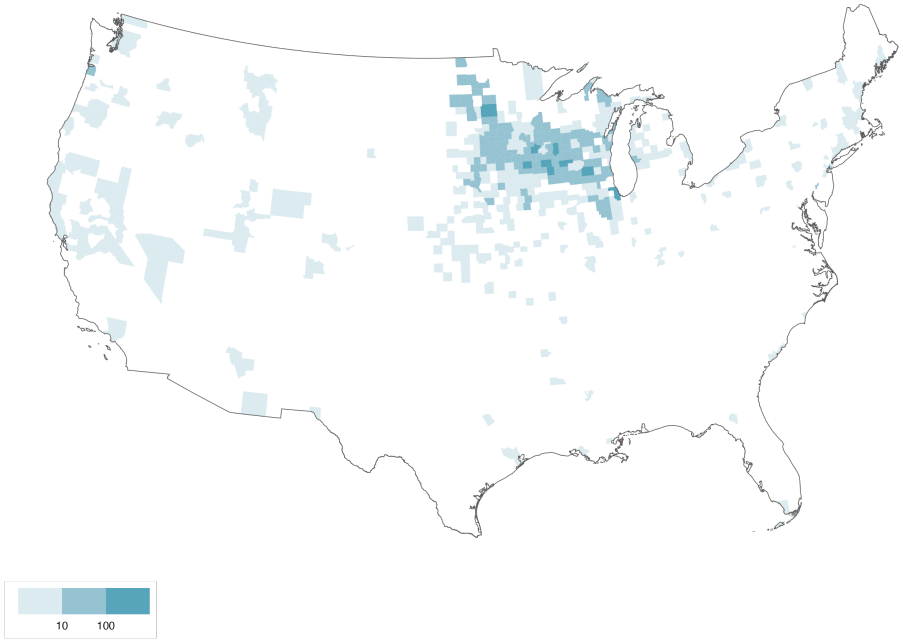
Notes: Panel A plots the distribution of men 15+ in the full count 1865 Norwegian Census. Panels B and C plot the distribution of men 15+ in the full count 1865 Norwegian Census who are matched, respectively to the 1900 Norwegian Census and the 1880 US Census.

Figure A10. Distribution of Norwegian Immigrants in the US (1880)

(a) Norwegian Immigrants in 1880 Full Count US Census



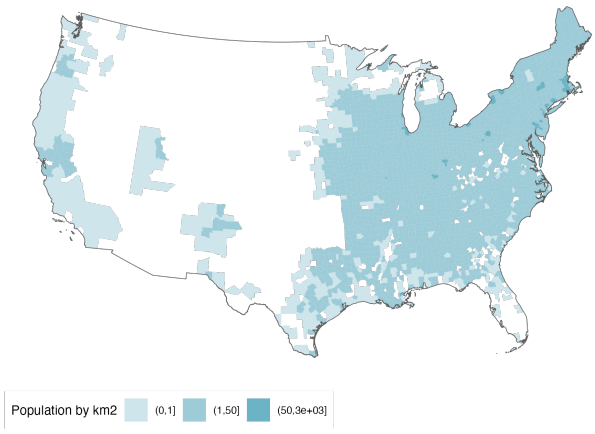
(b) Norwegian Immigrants Matched to 1865 Norwegian Census



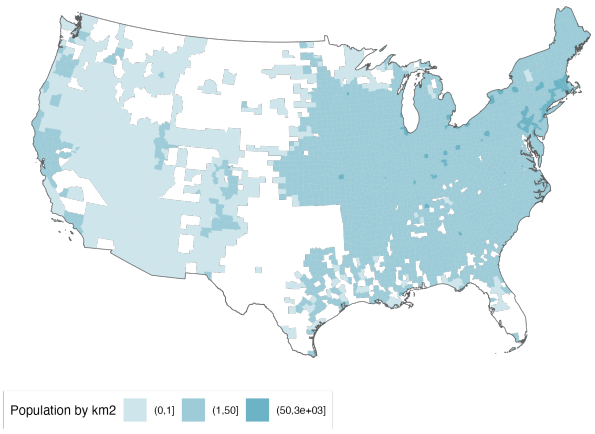
Notes: Panel A (resp., Panel b) plots the distribution of Norwegian male immigrants of age 15+ in the full count 1880 US Census (resp., who are linked between the 1865 Norwegian and the 1880 US full count Censuses).

Figure A11. Distribution of US Population, by Decade

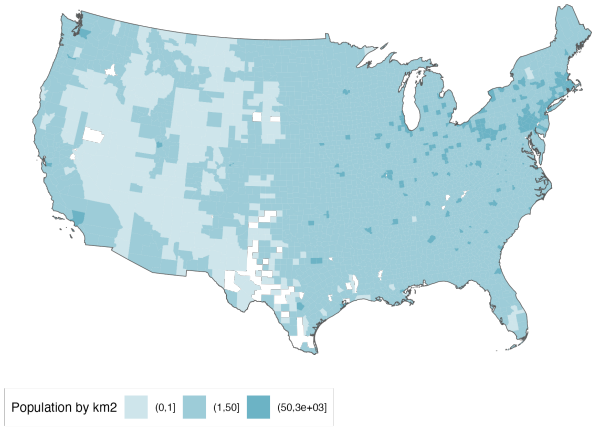
(a) 1860



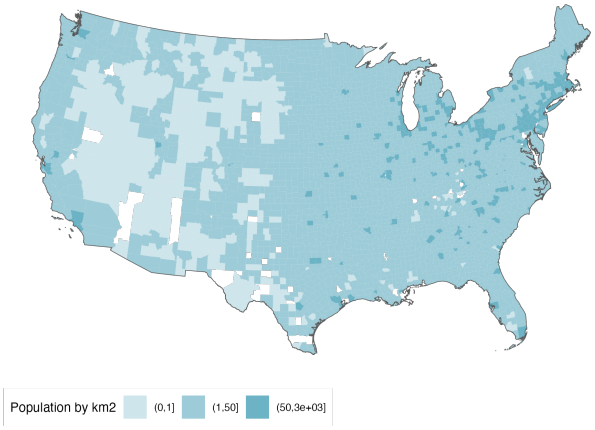
(b) 1880



(c) 1920



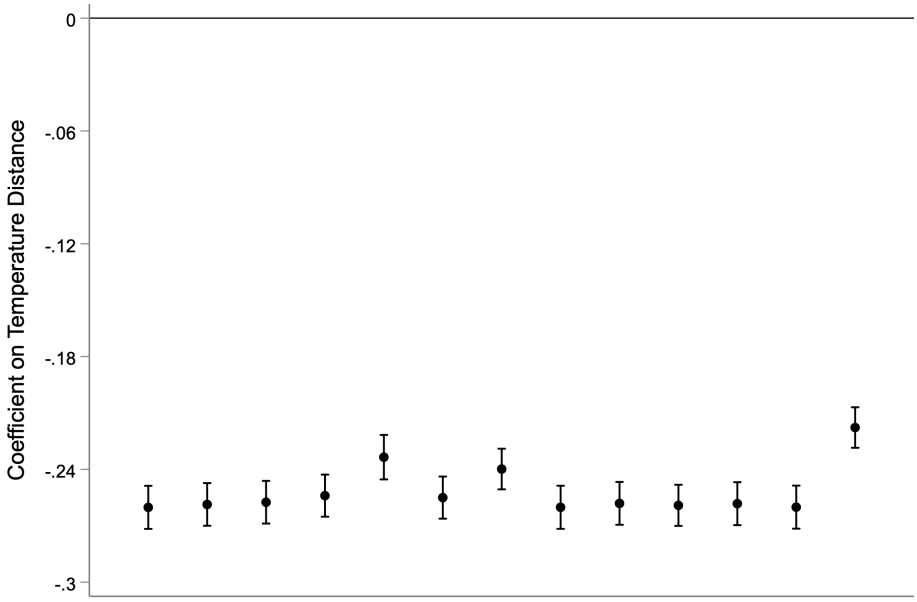
(d) 1940



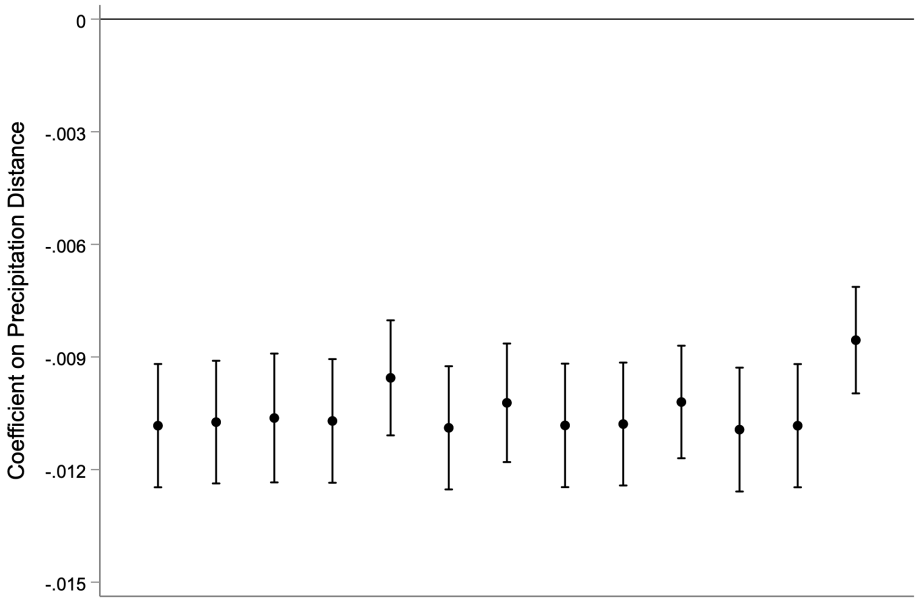
Notes: Each map plots county population density (per square km) in a given decade.

Figure A12. Climate Distance and Migration: Including County-Pair Controls

(a) Temperature

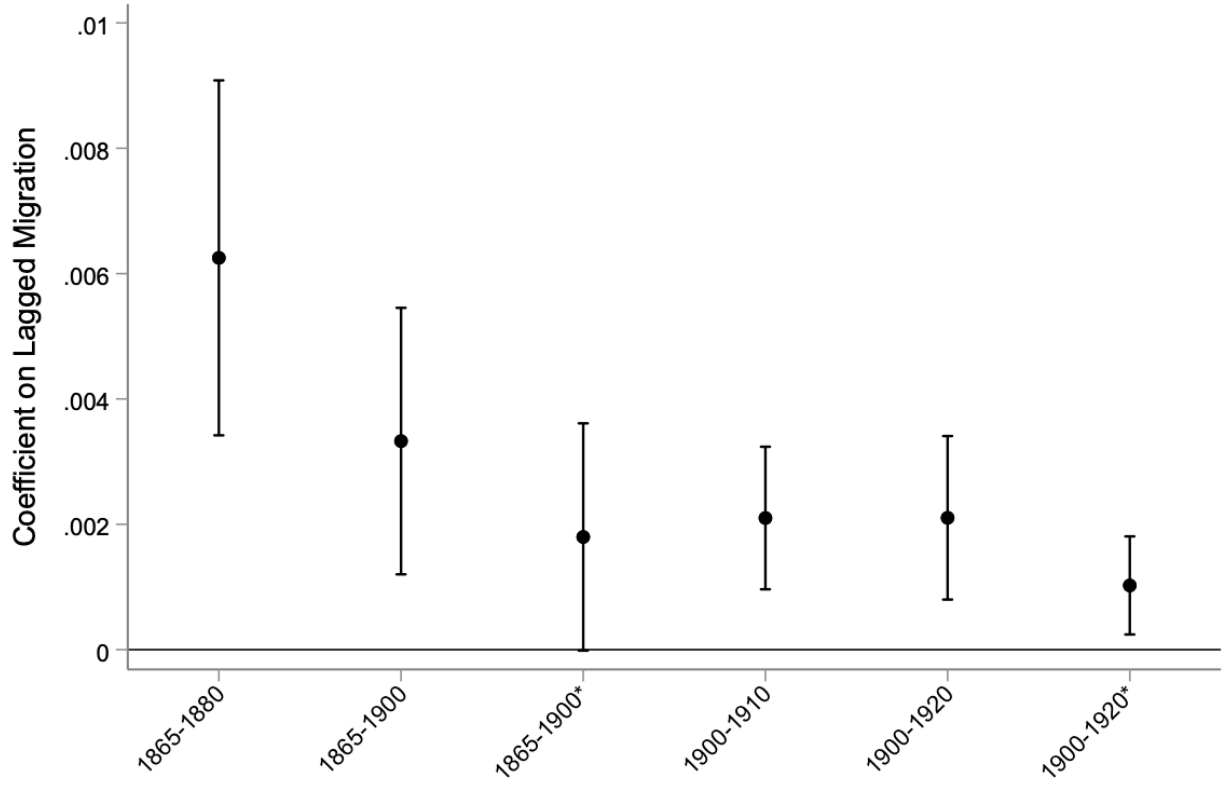


(b) Precipitation



Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature (Panel A) and precipitation (Panel B) between origin and destination counties. The first dot replicates the baseline specification (Table 2, column 5). Each subsequent dot reports results obtained when also including, one at the time, the absolute value of the difference in the following county-pair variables (measured in the baseline decade): employment share; manufacturing employment share; agriculture employment share; Black population share; urban population share; immigrant population share; sex ratios; population density; frontier exposure from Bazzi et al. (2020); market access from Donaldson and Hornbeck (2016); and a dummy equal to one for being connected to railroads. The last dot reports results obtained including all variables simultaneously. Frontier exposure is time invariant, and so the (absolute value of the) county-pair difference of this variable is interacted with decade dummies. Standard errors are clustered at the state of origin by state of destination by decade level.

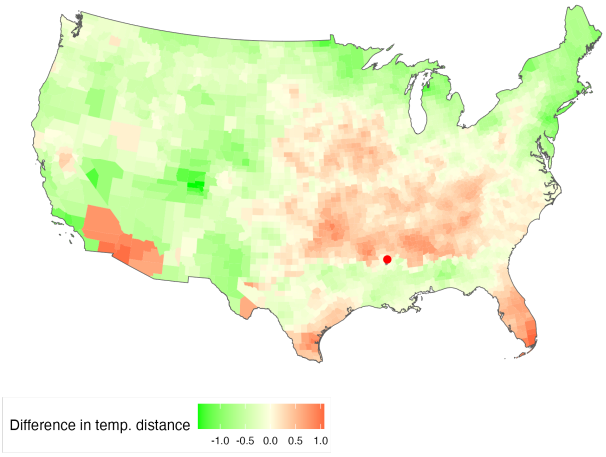
Figure A13. Lagged Number of Immigrants and Subsequent Norwegian Immigration



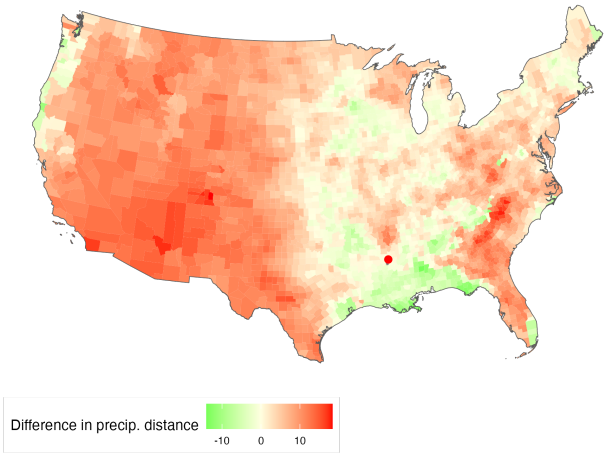
Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the lagged number of migrants in regressions, where the dependent variable is the number of Norwegian immigrants linked between the Norwegian and the US Censuses of the year reported on the x-axis. All regressions also include the set of controls and fixed effects from the baseline specification (see notes to Table 1). The third (resp., sixth) dot from the left refers to individuals linked between the 1865 Norwegian and the 1900 US (resp., the 1900 Norwegian and the 1920 US) Censuses and who migrated after 1880 (resp., after 1910). Lagged migration refers to number of immigrants from 1865 to 1870 (resp., from 1865 to 1900) in the second to fourth triangles (in the last three triangles). The lagged number of migrants is standardized within the relevant sample to have zero mean and standard deviation equal to one. Standard errors clustered at the US state by Norwegian province level.

Figure A14. Change in Climate Distances and Migration Between US Counties

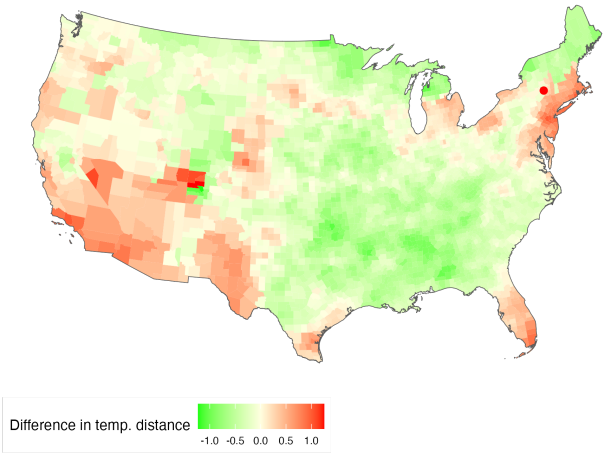
(a) Change in Temp. Distance, Delta Region



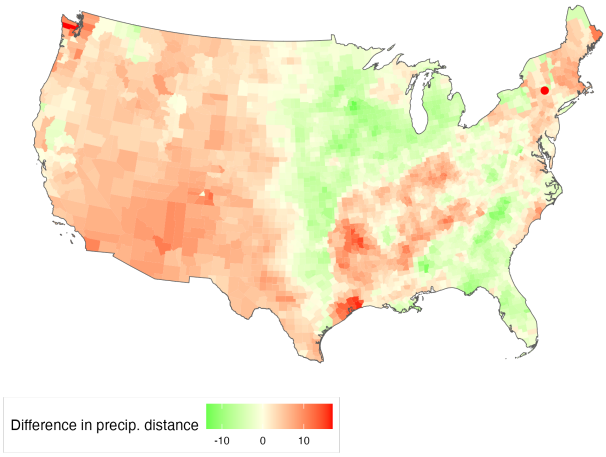
(b) Change in Precip. Distance, Delta Region



(c) Change in Temp. Distance, US Northeast



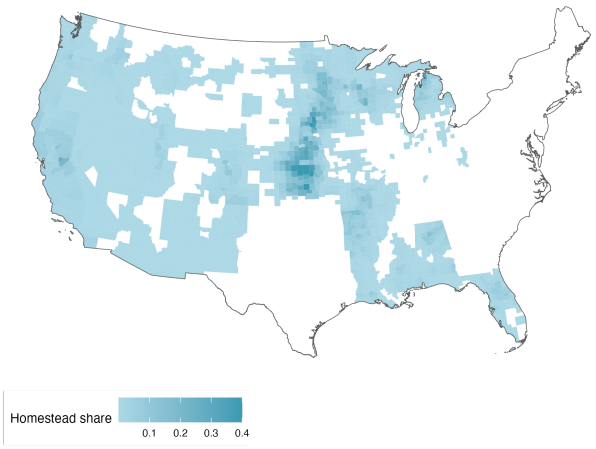
(d) Change in Precip. Distance, US Northeast



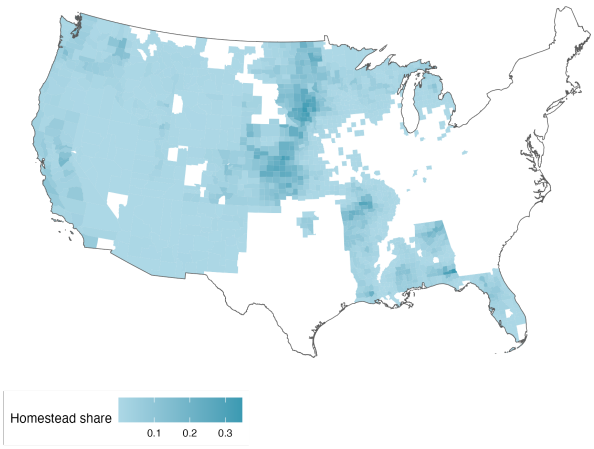
Notes: Red dot is the centroid of the considered region. Temperature and precipitation are averages of all-year climate in degrees C and mm/month, respectively. These averages are taken over the period 1895-1920 and 1990-2020 using NOAA data (Vose et al., 2014). To interpret the figure, note that the *distance* in temperature between the Delta region and New York City decreased by 1.5° C: it went from 12° C to 10.5° C ; New York City today is colder on average than the Delta region but less so on average than 100 years ago.

Figure A15. Homestead Shares, by County and Decade

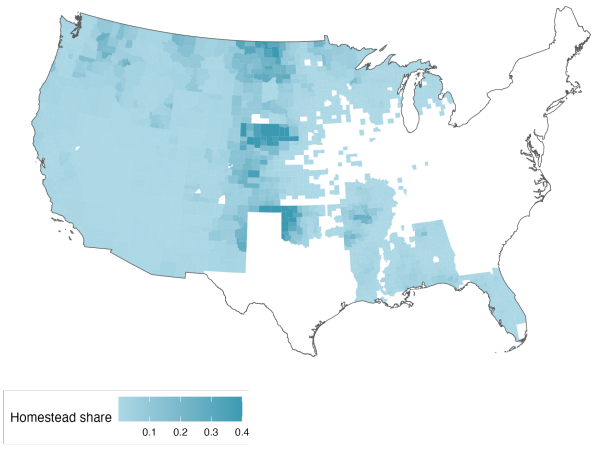
(a) 1870



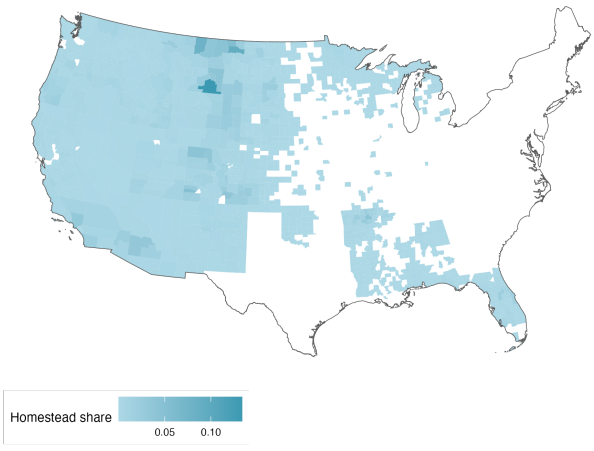
(b) 1880



(c) 1900

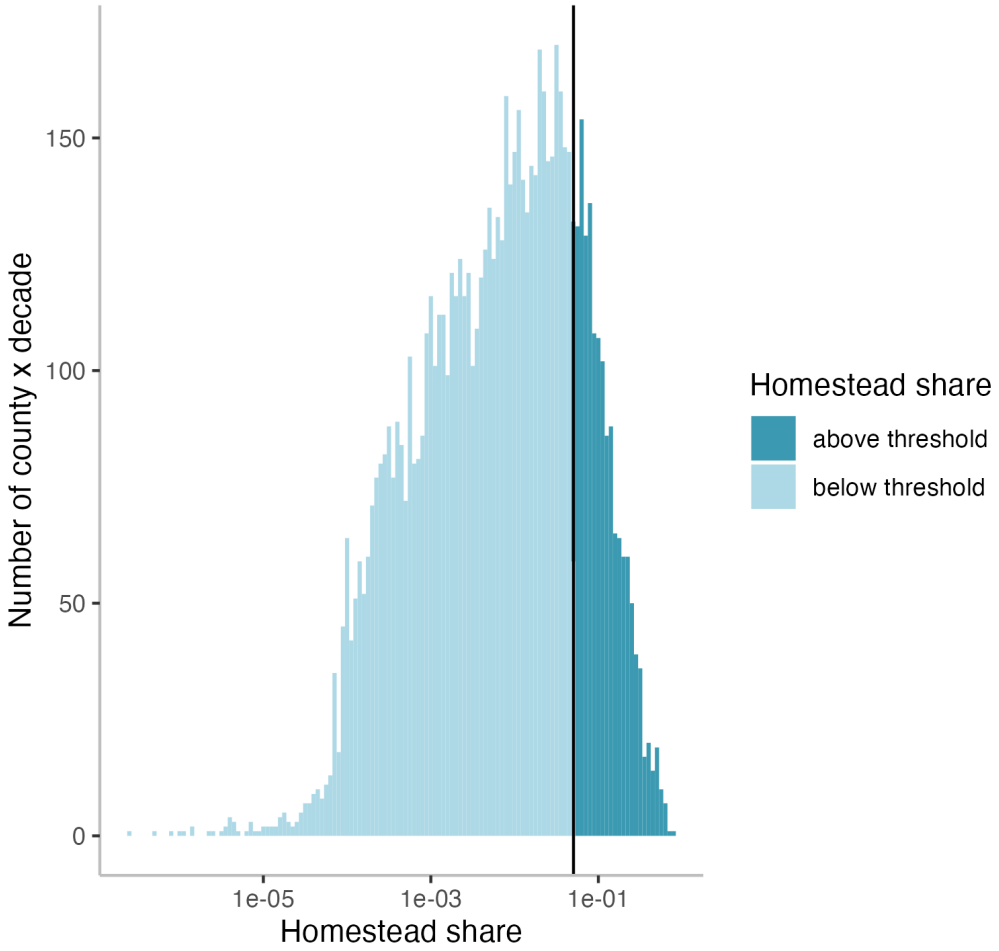


(d) 1920



Notes: Each map plots the share of county area that is distributed through Homestead in each county and decade. Because a Homestead patent is signed five years after the arrival of the homesteader on the land, we attribute patents signed between, e.g., 1915 and 1924 to the 1910 decade.

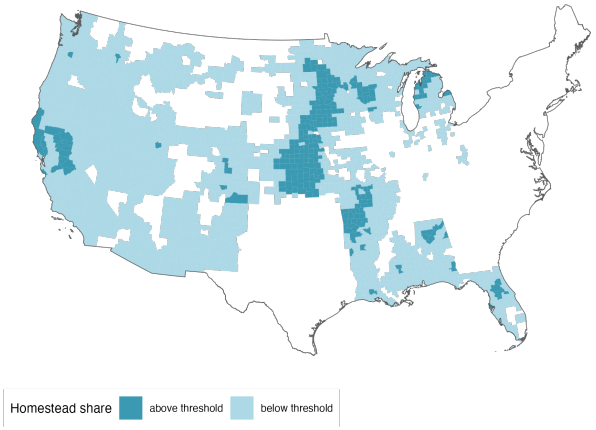
Figure A16. Distribution of Homestead Shares (1860-1920)



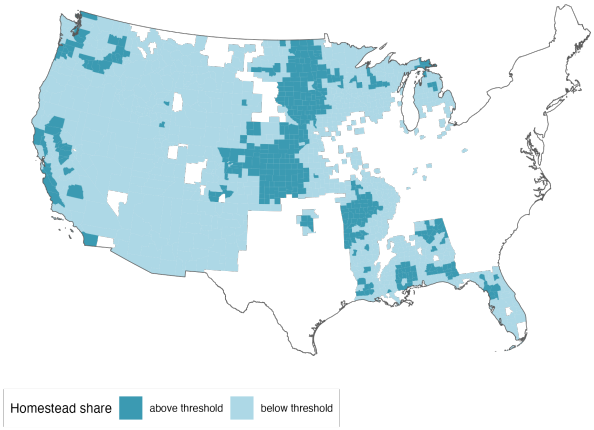
Notes: The figure plots the distribution of county level Homestead shares for all years from 1860 to 1920. Because a Homestead patent was signed five years after the arrival of the homesteader on the land, we attribute patents signed over a given decade to the previous 5 years. For instance, patents signed between 1915 and 1924 are assigned to migrants moving to a county between 1910 and 1919 (i.e., the 1910-1920 decade). The vertical bar corresponds to the 80^{th} percentile of the distribution, and is the threshold chosen to define the Homestead county dummy used in the analysis.

Figure A17. Homestead Counties, by Decade

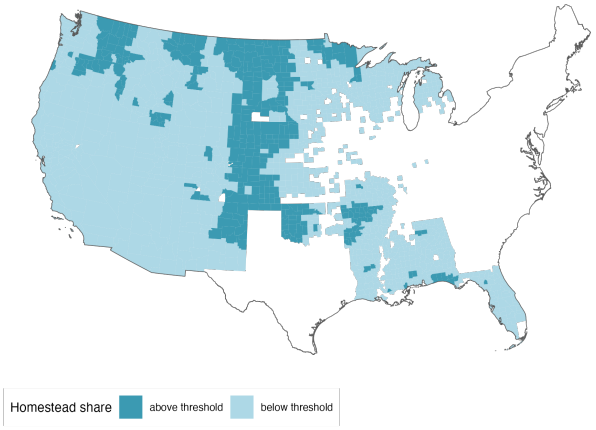
(a) 1870



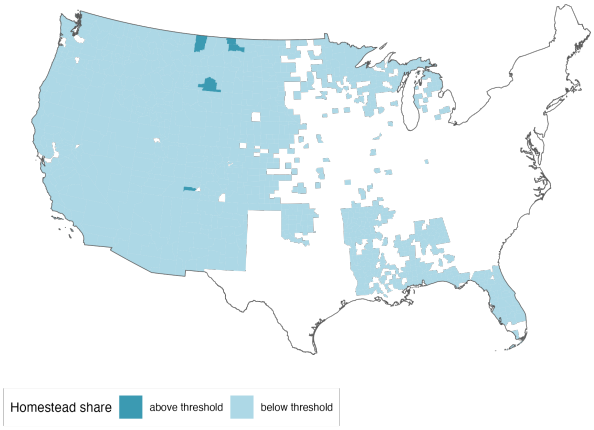
(b) 1880



(c) 1900

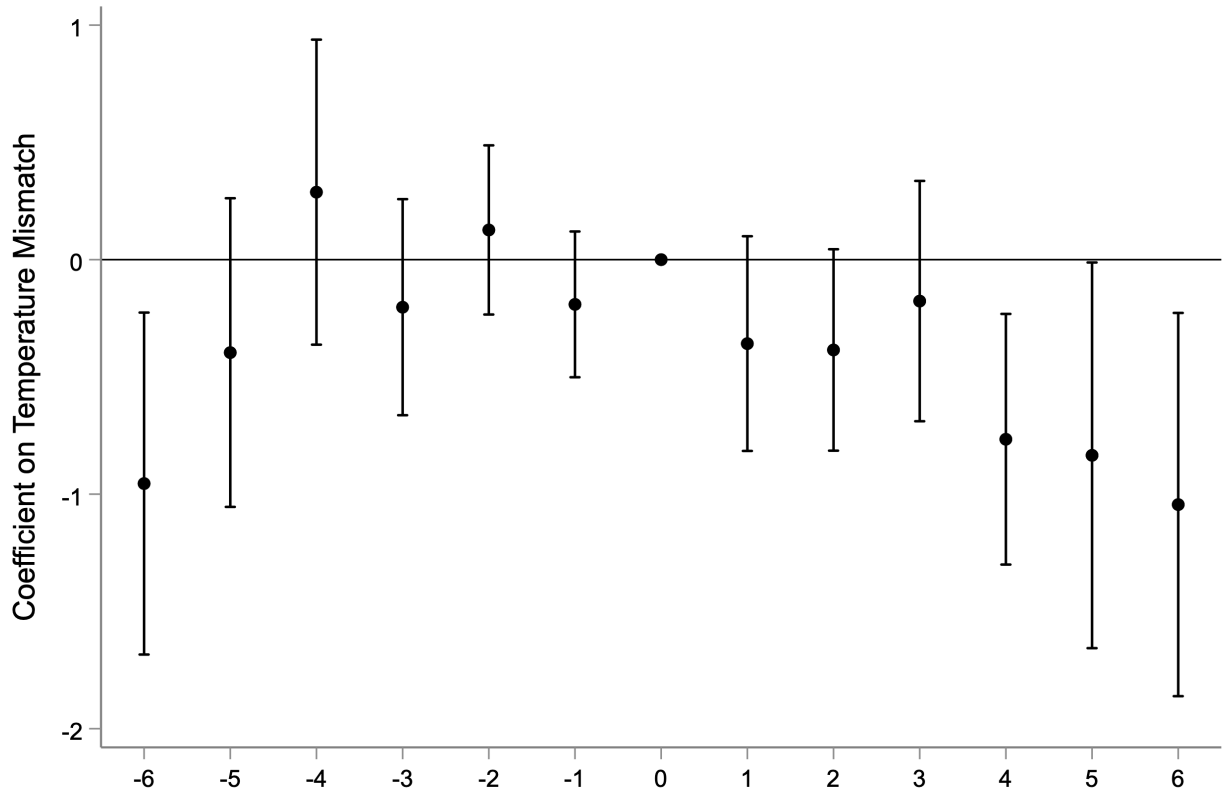


(d) 1920



Notes: The figure plots the counties that are classified as "Homestead counties" in each decade. Homestead counties are those where the share of land that is homesteaded in a given decade is above the 80^{th} percentile of the distribution of Homestead shares for the entire US over the 1860-1920 period. Because a Homestead patent is signed five years after the arrival of the homesteader on the land, we attribute patents signed between, e.g., 1915 and 1924 to the 1910 decade.

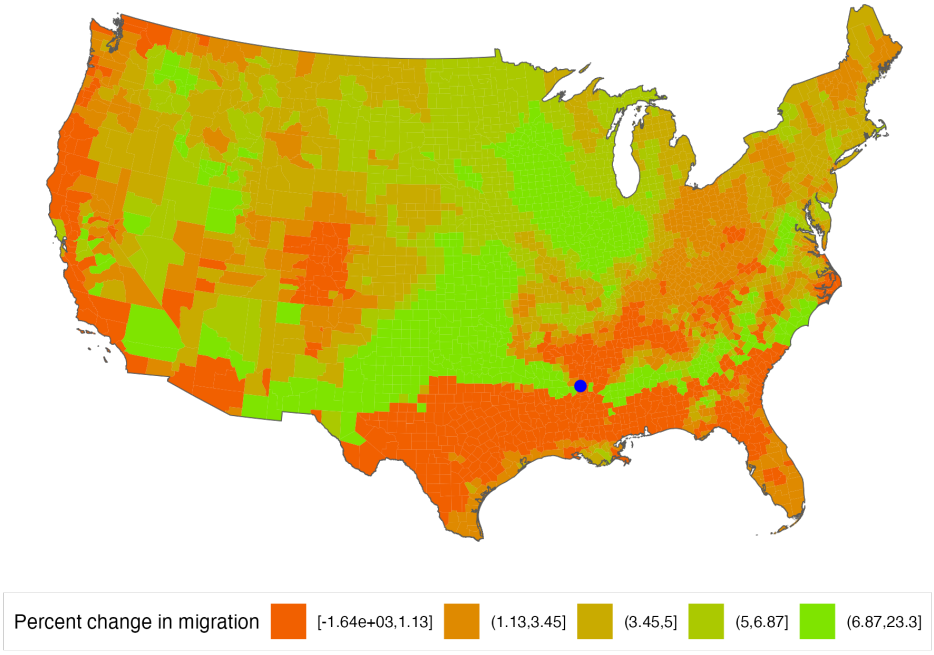
Figure A18. Temperature Mismatch and Age at Death of Immigrants in the US (1959-1961)



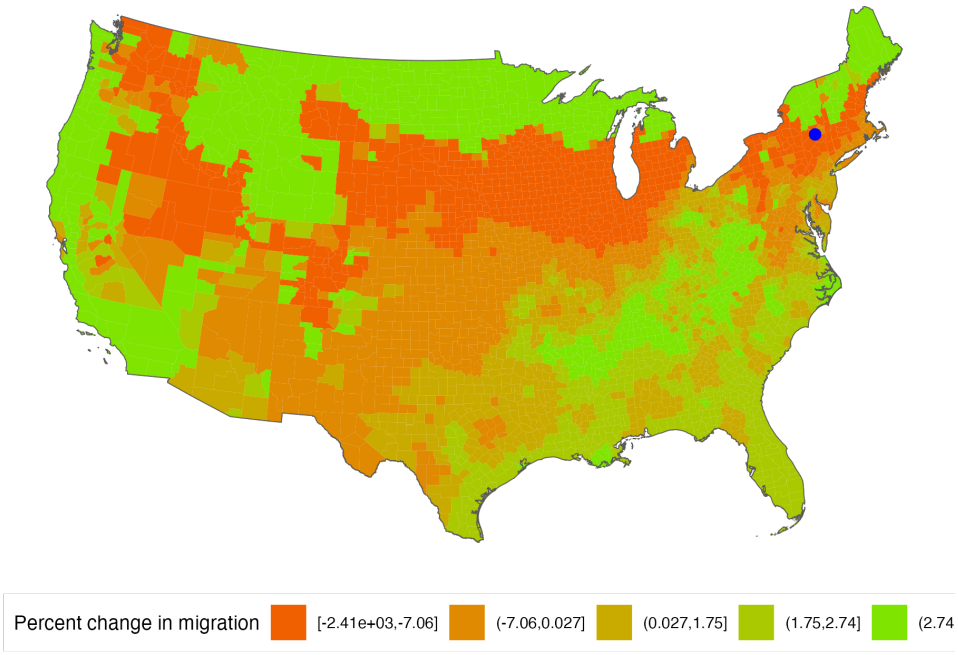
Notes: The figure plots the coefficients, with corresponding 95% confidence intervals, of the regression of age of death on 1°C temperature distance bins for the difference in temperature between the US county of death and birth country capital. The 0°C bin is omitted, whereas the -6°C and 6°C bins include observations with differences beyond these thresholds. The sample includes all foreign-born individuals died from 1959 to 1961 in the continental US. All models include fixed effects for US county of death, year of death, and country of birth; and, individual controls (gender and marital status). Standard errors clustered at the US state by country of birth level.

Figure A19. Change in Climate Distances and Migration Between US Counties

(a) Change in Temp. Distance Induced Migration, Delta Region



(b) Change in Temp. Distance Induced Migration, US Northeast



Notes: The blue dot is the centroid of the considered region. Changes in temperature-induced migration are computed using averages of all-year temperature from 1990-2020 to 2080-2100. To interpret the figure, note that the distance in temperature between the Delta region and New York City is projected to decline, increasing migration flows by 1.13-3.45%.

Table A1. 1865 Norwegian Census: Full Count and Linked Sample

	Full Count			Linked		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Individual-level Characteristics						
Age	33.822	18.103	608,157	25.470	11.164	82,274
Married	0.532	0.499	512,548	0.446	0.497	64,690
Urban	0.153	0.360	519,934	0.160	0.367	65,807
In Labor Force	0.866	0.340	471,522	0.842	0.365	65,725
Farmer	0.497	0.500	408,431	0.451	0.498	55,315
In Manufacturing	0.133	0.340	408,431	0.138	0.345	55,315
Municipality-level Characteristics						
Population Density	13.591	36.068	610,271	14.190	36.984	82,274
Distance from Closest Port (in 100km)	2330.311	2706.986	588,891	2473.883	2926.971	80,269
Average Temperature	4.093	2.254	588,891	4.292	2.150	80,269
Average Precipitation	99.142	44.702	588,891	98.357	44.262	80,269

Notes: Common restriction is on individuals to be men and at least 10 years old as of 1865. Linked refers to both individuals linked between the Norwegian 1865 Census and the US 1880 Census (migrants), and to individuals linked between the Norwegian 1865 Census and the Norwegian 1900 Census (stayers). *i) Individual-level Characteristics*: the dummies for married and urban status are defined only for men 15+. The dummy for labor force participation is defined for men in the age range 15-64. The dummies for being farmer and working in manufacturing are defined for men in the age range 15-64 and being in the labor force. *ii) Municipality-level Characteristics*: the mean and standard deviation of population density are rescaled by 100,000. Distance from closest international port is defined as the shortest between the following distances: Municipality-Port of Kristiansand; Municipality-Port of Oslo.

Table A2. 1865 Norwegian Census (Linked Sample): Stayers and Movers

	Linked Stayers			Linked Movers		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Individual-level Characteristics						
Age	25.254	10.917	76,308	28.232	13.652	5,966
Married	0.444	0.497	59,893	0.474	0.499	4,797
Urban	0.153	0.360	60,878	0.252	0.434	4,929
In Labor Force	0.840	0.367	60,876	0.862	0.345	4,849
Farmer	0.455	0.498	51,134	0.399	0.490	4,181
In Manufacturing	0.134	0.341	51,134	0.185	0.389	4,181
Municipality-level Characteristics						
Population Density	13.609	36.349	76,308	21.625	43.633	5,966
Distance from Closest Port (in 100km)	2490.500	2943.744	74,521	2258.451	2691.065	5,748
Average Temperature	4.300	2.147	74,521	4.200	2.192	5,748
Average Precipitation	98.480	44.414	74,521	96.757	42.206	5,748

Notes: Common restriction is on individuals to be men and at least 10 years old as of 1865. *i) Individual-level Characteristics*: the dummies for married and urban status are defined only for men 15+. The dummy for labor force participation is defined for men in the age range 15-64. The dummies for being farmer and working in manufacturing are defined for men in the age range 15-64 and being in the labor force. *ii) Municipality-level Characteristics*: the mean and standard deviation of population density are rescaled by 100,000. Distance from closest international port is defined as the shortest between the following distances: Municipality-Port of Kristiansand; Municipality-Port of Oslo.

Table A3. Norwegian Immigrants in the US, 1880: Full Count and Linked Sample

	Full Count			Linked		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Individual-level Characteristics						
Age	41.693	12.981	69,440	43.187	13.695	5,747
Married	0.728	0.445	69,440	0.733	0.442	5,747
Urban	0.177	0.381	69,440	0.175	0.380	5,747
Literate	0.865	0.342	69,440	0.848	0.359	5,747
In Labor Force	0.964	0.185	65,224	0.963	0.190	5,280
Farmer	0.631	0.482	62,906	0.640	0.480	5,082
In Manufacturing	0.079	0.270	62,906	0.077	0.266	5,082
Occupational Score	17.536	7.176	62,906	17.463	7.102	5,082
County-level Characteristics						
Population Density	0.000	0.001	69,440	0.000	0.001	5,747
Norwegian Population Share	0.103	0.089	69,440	0.106	0.089	5,747
Immigrant Population Share	0.328	0.095	69,440	0.332	0.093	5,747
Distance from NYC (in 100km)	1593.613	548.586	69,440	1575.773	510.642	5,747
Connected to Railroad	0.957	0.202	69,440	0.954	0.209	5,747
Average Temperature	6.753	2.053	69,440	6.698	1.977	5,747
Average Precipitation	64.834	13.556	69,440	64.423	12.316	5,747

Notes: Common restriction is on individuals to be men. The full count is restricted to men being at least 25 as of 1880 (col.1,2,3); the linked sample is restricted to men being at least 10 as of 1865 (col.4,5,6). *i) Individual-level Characteristics*: the dummy for labor force participation is defined for men younger than 64. The dummies for being farmer and working in manufacturing, along with the occupational score, are defined for men younger than 64 and being in the labor force. *ii) County-level Characteristics*: mean and standard deviation of population density are rescaled by 100,000. Norwegian and immigrant population shares are computed as a share of total county population.

Table A4. Climate Distance and Norwegian Immigration: Additional Controls

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.107***	-0.116***	-0.106***	-0.136***	-0.064***	-0.076***	-0.060***
	(0.016)	(0.015)	(0.016)	(0.018)	(0.018)	(0.019)	(0.019)
	[-4.257]	[-4.600]	[-4.195]	[-5.335]	[-0.945]	[-1.112]	[-0.880]
Precipitation Distance	-0.006***	-0.007***	-0.006***	-0.012***	-0.003***	-0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)
	[-1.967]	[-2.225]	[-1.836]	[-3.737]	[-0.545]	[-0.215]	[-0.271]
Observations	847,894	847,894	847,894	839,838	180,188	178,476	178,476
Pseudo R-squared	0.401	0.398	0.406	0.430	0.293	0.330	0.334
Mean Temp. Dist.	7.429	7.429	7.429	7.387	4.606	4.566	4.566
SD Temp. Dist.	4.609	4.609	4.609	4.591	3.689	3.667	3.667
Mean Precip. Dist.	40.17	40.17	40.17	40.22	42.06	42.19	42.19
SD Precip. Dist.	36.16	36.16	36.16	36.27	40.21	40.33	40.33
US Controls	Yes	Yes	Yes	Yes			
Norwegian Controls	Yes		Yes		Yes		
US State FE	Yes	Yes	Yes	Yes			
Norwegian Province FE	Yes	Yes	Yes		Yes		
Bilateral Controls		Yes	Yes				Yes
Norwegian Municipality FE				Yes		Yes	Yes
US County FE					Yes	Yes	Yes

Notes: The sample includes pairs formed by all Norwegian municipalities (origins) and all US counties (destinations) with at least one European immigrant in 1880. *Number of Migrants* is the number of 10+ Norwegian male immigrants linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the Norwegian municipality and the US county. The table replicates column 4 of Table 1, reported in column 1 (see notes to Table 1 for the list of baseline controls). Column 2 replaces the vectors of US county and Norwegian controls with a vector that includes their difference (X_{od}). Column 3 includes both the separate controls and their difference. Column 4 (resp., column 5) replaces Norwegian (resp., US) controls with municipality (resp., US county) fixed effects. Column 6 includes simultaneously Norwegian municipality and US county fixed effects. Column 7 augments column 6 with the vector of bilateral controls, X_{od}. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the US state by Norwegian province level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A5. Climate Distance and Norwegian Immigration: Controlling for Geography

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.107***	-0.105***	-0.119***	-0.111***	-0.125***
	(0.016)	(0.016)	(0.015)	(0.017)	(0.016)
	[-4.257]	[-4.167]	[-4.739]	[-4.417]	[-4.942]
Precipitation Distance	-0.006***	-0.006***	-0.007***	-0.006***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-1.967]	[-1.990]	[-2.026]	[-1.997]	[-2.103]
Observations	847,894	847,894	847,894	847,894	847,894
Pseudo R-squared	0.401	0.401	0.401	0.401	0.402
Mean Temp. Dist.	7.429	7.429	7.429	7.429	7.429
SD Temp. Dist.	4.609	4.609	4.609	4.609	4.609
Mean Precip. Dist.	40.17	40.17	40.17	40.17	40.17
SD Precip. Dist.	36.16	36.16	36.16	36.16	36.16
US State FE	Yes	Yes	Yes	Yes	Yes
Norwegian Province FE	Yes	Yes	Yes	Yes	Yes
US Controls	Yes	Yes	Yes	Yes	Yes
Norwegian Controls	Yes	Yes	Yes	Yes	Yes
Geographic Control	None	Elevation	Ruggedness	Coastal	All

Notes: The sample includes pairs formed by all Norwegian municipalities and all US counties with at least one European immigrant in 1880. *Number of Migrants* is the number of 10+ Norwegian male immigrants linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the Norwegian municipality and the US county. US and Norwegian controls include: distance between a US county and New York City, population density, sex ratios (ratio between women 18-33 and men 20-35), the urban population share, share of men (15-64) in the labor force, and share of male employment in agriculture and manufacturing. US controls also include the following variables: immigrant population share, the Norwegian population share, the Black population share, total frontier experience Bazzi et al. (2020), and market access Donaldson and Hornbeck (2016). All Norwegian controls are measured in 1865; all US controls are measured in 1880, or 1870 for market access. Columns 2 and 3 also add the absolute value of the difference in, respectively, elevation and ruggedness, between the Norwegian municipality and the US county. Column 4 includes a dummy if the US county has coastal access. Column 5 includes all variables in columns 2 to 4 simultaneously. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the US state by Norwegian province level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A6. Climate Distance and Norwegian Immigration: Alternative Samples

Dep. var.:	Number of Migrants							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temperature Dist.	-0.107***	-0.110***	-0.101***	-0.075***	-0.105***	-0.093***	-0.104***	-0.106***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.016)	(0.015)	(0.015)
	[-4.257]	[-5.413]	[-2.746]	[-1.101]	[-5.635]	[-4.872]	[-5.505]	[-4.834]
Precipitation Dist.	-0.006***	-0.007***	-0.006***	-0.005***	-0.006***	-0.004***	-0.005***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-1.967]	[-2.438]	[-1.516]	[-0.782]	[-2.539]	[-1.786]	[-2.244]	[-1.342]
Observations	847,894	1,095,442	467,731	181,451	845,789	809,536	847,828	817,161
Pseudo R-squared	0.401	0.423	0.344	0.244	0.359	0.358	0.368	0.402
Mean Temp. Dist.	7.429	7.675	5.834	4.601	7.439	7.234	7.429	7.620
SD Temp. Dist.	4.609	5.007	4.256	3.684	4.609	4.521	4.609	4.572
Mean Precip. Dist.	40.17	43.09	42.29	42.15	40.17	40.21	40.17	40.02
SD Precip. Dist.	36.16	37.94	39.48	40.25	36.15	36.46	36.16	35.90
US State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Norwegian Prov. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
US Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Norwegian Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table replicates column 4 of Table 1, reported in column 1 (see notes to Table 1 for the list of baseline controls), using different samples. *Number of Migrants* is the number of 10+ Norwegian male immigrants linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the Norwegian municipality and the US county. The sample includes: *i*) all US counties in column 2; *ii*) counties with at least one Norwegian immigrant in the 1880 full count US Census (resp., in the linked sample) in column 3 (resp., in column 4). Column 5 (resp., column 6) drops US counties (resp., Norwegian municipalities) at the top 1% of the distribution of the linked sample population. Column 7 drops municipality-county-pairs at the top 1% of the distribution of bilateral migration flows. Column 8 excludes Minnesota. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the US state by Norwegian province level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A7. Climate Distance and Norwegian Immigration: Alternative Linking Methods

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.107***	-0.097***	-0.108***	-0.115***	-0.099***	-0.087***
	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)	(0.022)
	[-4.257]	[-5.009]	[-4.431]	[-3.871]	[-4.208]	[-5.552]
Precipitation Distance	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-1.967]	[-2.533]	[-2.054]	[-1.661]	[-1.998]	[-4.030]
Observations	847,894	847,894	847,894	847,894	822,213	759,484
Pseudo R-squared	0.401	0.382	0.399	0.416	0.401	0.367
Mean Temp. Dist.	7.429	7.429	7.429	7.429	7.267	7.416
SD Temp. Dist.	4.609	4.609	4.609	4.609	4.553	4.754
Mean Precip. Dist.	40.17	40.17	40.17	40.17	40.02	40.66
SD Precip. Dist.	36.16	36.16	36.16	36.16	36.17	36.60
US State FE	Yes	Yes	Yes	Yes	Yes	Yes
Norwegian Province FE	Yes	Yes	Yes	Yes	Yes	Yes
US Controls	Yes	Yes	Yes	Yes	Yes	Yes
Norwegian Controls	Yes	Yes	Yes	Yes	Yes	Yes
Restriction	Older than 10 in 1865	15-40 in 1865 (Nor)	18-65 in 1880 (US)	None	ABE-JW	Conservative ABE-NYSIIS

Notes: The sample includes pairs formed by all Norwegian municipalities (origins) and all US counties (destinations) with at least one European immigrant in 1880. *Number of Migrants* is the number of 10+ Norwegian male immigrants linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the Norwegian municipality and the US county. Column 1 replicates the baseline specification in column 4 of Table 1, where the linked sample used to derive the number of immigrants is restricted to men 10+ in 1865. In columns 2, 3, and 4, the sample is restricted to men: 15-40 in 1865, 18-65 in 1880, and without age restrictions. Columns 5 and 6 derive the linked sample using, respectively, the Jaro-Winkler and the conservative version of the NYSIIS algorithm (i.e., individuals who are unique within a 5-year window around their birth date in terms of the matching variables) from Abramitzky et al. (2012). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the US state by Norwegian province level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A8. US Population: Full Count vs Linked Samples, by Decade

	Full Count			Linked		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Panel A: 1860						
Age	33.31	14.73	6,077,761	33.34	13.85	985,763
Foreign Born	0.178	0.382	6,077,761	0.128	0.334	985,763
White	0.980	0.141	6,077,761	0.989	0.105	985,763
Urban	0.167	0.373	6,077,761	0.151	0.358	985,763
Literate	0.757	0.429	6,077,761	0.787	0.410	985,763
In Labor Force	0.865	0.342	5,819,994	0.884	0.320	959,596
Occupational Score	20.24	9.937	5,032,160	20.21	9.989	848,616
Farmer	0.525	0.499	5,032,160	0.551	0.497	848,616
In Manufacturing	0.126	0.332	5,032,160	0.140	0.347	848,616
Panel B: 1900						
Age	35.15	15.32	15,183,543	30.66	11.55	2,217,092
Foreign Born	0.218	0.413	15,183,543	0.158	0.365	2,217,092
White	0.882	0.323	15,183,543	0.942	0.234	2,217,092
Urban	0.279	0.449	15,183,543	0.262	0.440	2,217,092
Literate	0.847	0.360	15,183,543	0.912	0.284	2,217,092
In Labor Force	0.925	0.264	14,394,375	0.912	0.284	2,209,530
Occupational Score	19.86	10.26	13,311,350	19.93	10.39	2,014,568
Farmer	0.486	0.500	13,311,350	0.510	0.500	2,014,568
In Manufacturing	0.125	0.330	13,311,350	0.133	0.340	2,014,568
Panel C: 1940						
Age	38.64	14.70	38,246,367	35.08	12.10	8,467,821
Foreign Born	0.180	0.384	38,246,367	0.124	0.330	8,467,821
White	0.901	0.299	38,246,367	0.950	0.218	8,467,821
Urban	0.577	0.494	38,246,367	0.595	0.491	8,467,821
Literate	0.951	0.216	38,246,367	0.979	0.143	8,467,821
In Labor Force	0.982	0.132	36,076,601	0.987	0.112	8,375,759
Occupational Score	23.76	10.77	34,149,990	24.76	10.99	7,966,512
Farmer	0.251	0.433	35,436,485	0.233	0.423	8,270,197
In Manufacturing	0.233	0.423	35,436,485	0.247	0.431	8,270,197

Notes: The table reports summary statistics for men in the full count Censuses and linked samples by decade. All characteristics are measured in the year of origin. Since Censuses 1850 and 1860 are missing the variable *nativity*, we use variable *bpl* instead. The dummy for labor force participation is defined for men in the age range 15-64. The dummies for employment-related variables are defined for men in the age range 15-64 and being in the labor force. Remaining dummies are defined for men 15+.

Table A9. US Population: Full Count vs Linked Samples, by Period

	Full Count			Linked		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Panel A: 1860-1940						
Age	36.59	14.95	166,172,765	34.15	12.82	32,943,677
Foreign Born	0.209	0.407	166,172,765	0.139	0.346	32,943,677
White	0.902	0.297	166,172,765	0.949	0.220	32,943,677
Urban	0.447	0.497	166,172,765	0.461	0.498	32,943,677
Literate	0.897	0.304	166,172,765	0.944	0.231	32,943,677
In Labor Force	0.936	0.245	157,632,799	0.943	0.232	32,442,995
Occupational Score	22.02	10.62	144,194,339	22.85	11.06	29,851,951
Farmer	0.359	0.480	147,492,877	0.353	0.478	30,586,862
In Manufacturing	0.188	0.391	147,492,877	0.198	0.399	30,586,862
Panel B: 1860-1900						
Age	34.47	15.13	41,581,297	32.80	13.45	6,762,309
Foreign Born	0.219	0.414	41,581,297	0.155	0.362	6,762,309
White	0.916	0.277	41,581,297	0.958	0.201	6,762,309
Urban	0.254	0.435	41,581,297	0.236	0.425	6,762,309
Literate	0.808	0.394	41,581,297	0.855	0.352	6,762,309
In Labor Force	0.849	0.358	39,596,861	0.849	0.358	6,627,222
Occupational Score	19.80	10.27	33,612,787	19.82	10.41	5,629,415
Farmer	0.510	0.500	33,612,787	0.543	0.498	5,629,415
In Manufacturing	0.120	0.325	33,612,787	0.127	0.333	5,629,415
Panel C: 1910-1940						
Age	37.29	14.82	124,591,468	34.50	12.63	26,181,368
Foreign Born	0.206	0.405	124,591,468	0.134	0.341	26,181,368
White	0.898	0.303	124,591,468	0.947	0.224	26,181,368
Urban	0.511	0.500	124,591,468	0.519	0.500	26,181,368
Literate	0.926	0.261	124,591,468	0.967	0.180	26,181,368
In Labor Force	0.965	0.184	118,035,938	0.967	0.179	25,815,773
Occupational Score	22.70	10.64	110,581,552	23.56	11.08	24,222,536
Farmer	0.315	0.464	113,880,090	0.310	0.463	24,957,447
In Manufacturing	0.208	0.406	113,880,090	0.214	0.410	24,957,447

Notes: The table reports summary statistics for men in the full count Censuses and linked samples by period. All characteristics are measured in the year of origin. Since Censuses 1850 and 1860 are missing the variable *nativity*, we use variable *bpl* instead. The dummy for labor force participation is defined for men in the age range 15-64. The dummies for employment-related variables are defined for men in the age range 15-64 and being in the labor force. Remaining dummies are defined for men 15+.

Table A10. Linked Sample: Stayers vs Domestic Migrants, by Period

	Stayers			Migrants		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Panel A: 1860-1940						
Age	35.38	13.11	20,073,837	32.24	12.11	12,869,840
Foreign Born	0.125	0.330	20,073,837	0.161	0.367	12,869,840
White	0.965	0.184	20,073,837	0.925	0.264	12,869,840
Urban	0.473	0.499	20,073,837	0.442	0.497	12,869,840
Literate	0.957	0.203	20,073,837	0.923	0.267	12,869,840
In Labor Force	0.952	0.214	19,699,172	0.929	0.257	12,743,823
Occupational Score	23.14	11.21	18,289,799	22.40	10.79	11,562,152
Farmer	0.358	0.479	18,750,778	0.346	0.476	11,836,084
In Manufacturing	0.204	0.403	18,750,778	0.189	0.391	11,836,084
Panel B: 1860-1900						
Age	34.76	14.15	3,597,638	30.56	12.23	3,164,671
Foreign Born	0.111	0.314	3,597,638	0.206	0.404	3,164,671
White	0.971	0.169	3,597,638	0.943	0.232	3,164,671
Urban	0.221	0.415	3,597,638	0.253	0.435	3,164,671
Literate	0.879	0.326	3,597,638	0.827	0.378	3,164,671
In Labor Force	0.861	0.346	3,498,306	0.837	0.370	3,128,916
Occupational Score	19.65	10.50	3,011,825	20.01	10.32	2,617,590
Farmer	0.584	0.493	3,011,825	0.496	0.500	2,617,590
In Manufacturing	0.122	0.328	3,011,825	0.132	0.338	2,617,590
Panel C: 1910-1940						
Age	35.51	12.87	16,476,199	32.78	12.02	9,705,169
Foreign Born	0.128	0.334	16,476,199	0.146	0.353	9,705,169
White	0.964	0.187	16,476,199	0.919	0.273	9,705,169
Urban	0.528	0.499	16,476,199	0.504	0.500	9,705,169
Literate	0.974	0.159	16,476,199	0.954	0.210	9,705,169
In Labor Force	0.971	0.166	16,200,866	0.959	0.199	9,614,907
Occupational Score	23.83	11.22	15,277,974	23.10	10.83	8,944,562
Farmer	0.314	0.464	15,738,953	0.304	0.460	9,218,494
In Manufacturing	0.219	0.414	15,738,953	0.205	0.404	9,218,494

Notes: The table reports summary statistics for men stayers and migrants in the linked samples by period. All characteristics are measured in the year of origin. Since Censuses 1850 and 1860 are missing the variable *nativity*, we use variable *bpl* instead. The dummy for labor force participation is defined for men in the age range 15-64. The dummies for employment-related variables are defined for men in the age range 15-64 and being in the labor force. Remaining dummies are defined for men 15+.

Table A11. Climate Distance and Migration: Controlling Non-Linearly for Geography

Dep. var.:	Number of Migrants			
	(1)	(2)	(3)	(4)
Temperature Distance	-0.260***	-0.325***	-0.342***	-0.347***
	(0.006)	(0.007)	(0.007)	(0.007)
	[-0.070]	[-0.088]	[-0.092]	[-0.094]
Precipitation Distance	-0.011***	-0.027***	-0.034***	-0.037***
	(0.001)	(0.001)	(0.002)	(0.002)
	[-0.018]	[-0.044]	[-0.056]	[-0.061]
Distance (in 100km)	-0.149***	-0.002***	-0.000***	-0.000***
	(0.005)	(0.000)	(0.000)	(0.000)
	[-0.085]	[-0.051]	[-0.027]	[-0.015]
Observations	30,479,809	30,479,809	30,479,809	30,479,809
Pseudo R-squared	0.705	0.672	0.662	0.659
Mean Temp. Dist.	5.321	5.321	5.321	5.321
SD Temp. Dist.	3.829	3.829	3.829	3.829
Mean Precip. Dist.	29.61	29.61	29.61	29.61
SD Precip. Dist.	23.21	23.21	23.21	23.21
County o × Decade FE	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes
Distance	Linear	Quadratic	Cubic	Quartic

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). Column 1 controls for physical distance linearly. Columns 2, 3, and 4 control, respectively, for the second, third, and fourth order polynomials of physical distance, always including lower order terms. The coefficient refers to the sum of distance terms. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A12. Climate Distance and Migration, by Direction of Move

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.260***	-0.216***	-0.210***	-0.223***	-0.139***	-0.152***	-0.122***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.011)
	[-0.070]	[-0.053]	[-0.111]	[-0.082]	[-0.036]	[-0.085]	[-0.050]
Precipitation Distance	-0.011***	-0.010***	-0.009***	-0.012***	-0.008***	-0.006***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
	[-0.018]	[-0.019]	[-0.036]	[-0.032]	[-0.010]	[-0.017]	[-0.019]
Distance (in 100km)	-0.149***	-0.142***	-0.131***	-0.157***	-0.425***	-0.437***	-0.420***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.013)	(0.014)	(0.014)
	[-0.085]	[-0.089]	[-0.179]	[-0.149]	[-0.146]	[-0.325]	[-0.230]
Observations	30,479,809	20,085,470	19,826,771	19,114,830	10,136,584	9,676,434	9,640,674
Pseudo R-squared	0.705	0.687	0.628	0.674	0.789	0.755	0.754
Mean Temp. Dist.	5.321	4.278	4.269	4.254	7.320	7.270	7.237
SD Temp. Dist.	3.829	3.388	3.380	3.378	3.818	3.809	3.783
Mean Precip. Dist.	29.61	31.23	31.16	30.76	26.29	26.33	25.90
SD Precip. Dist.	23.21	24.93	24.85	24.75	18.90	18.88	18.61
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Direction	All	Horizontal	East-West	West-East	Vertical	South-North	North-South

Notes: The sample in column 1 includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. Subsequent columns consider county-pairs within the direction of move reported at the bottom of the table. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). All regressions control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A13. Climate Distance and Migration: Controlling for Latitude and Longitude

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.260***	-0.130***	-0.157***	-0.127***	-0.133***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
	[-0.070]	[-0.035]	[-0.043]	[-0.034]	[-0.036]
Precipitation Distance	-0.011***	-0.010***	-0.009***	-0.009***	-0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-0.018]	[-0.017]	[-0.014]	[-0.015]	[-0.021]
Distance (in 100km)	-0.149***	-0.142***	-0.369***	-0.288***	
	(0.005)	(0.004)	(0.013)	(0.018)	
	[-0.085]	[-0.081]	[-0.211]	[-0.165]	
Observations	30,479,809	30,479,809	30,479,809	30,479,809	30,479,809
Pseudo R-squared	0.705	0.708	0.709	0.709	0.706
Mean Temp. Dist.	5.321	5.321	5.321	5.321	5.321
SD Temp. Dist.	3.829	3.829	3.829	3.829	3.829
Mean Precip. Dist.	29.61	29.61	29.61	29.61	29.61
SD Precip. Dist.	23.21	23.21	23.21	23.21	23.21
County o × Decade FE	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes
Latitude Distance		Yes		Yes	Yes
Longitude Distance			Yes	Yes	Yes

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). Columns 2 and 3 replicate the baseline specification of Table 2, column 5 (also reported in column 1), by controlling for latitude and longitude distance, respectively. Column 4 includes both measures simultaneously. Column 5 replicates column 4 without controlling for physical distance. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A14. Climate Distance and Migration: Dropping Selected Counties

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.260***	-0.207***	-0.189***	-0.189***	-0.180***	-0.257***	-0.254***
	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)	(0.006)
	[-0.070]	[-0.113]	[-0.128]	[-0.128]	[-0.126]	[-0.068]	[-0.070]
Precipitation Distance	-0.011***	-0.007***	-0.007***	-0.006***	-0.007***	-0.010***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-0.018]	[-0.024]	[-0.027]	[-0.026]	[-0.028]	[-0.016]	[-0.016]
Distance (in 100km)	-0.149***	-0.122***	-0.109***	-0.109***	-0.090***	-0.153***	-0.190***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)
	[-0.085]	[-0.141]	[-0.157]	[-0.157]	[-0.132]	[-0.083]	[-0.102]
Observations	30,479,809	30,391,967	30,244,609	30,255,171	29,356,823	26,596,850	29,274,530
Pseudo R-squared	0.705	0.683	0.664	0.664	0.649	0.725	0.711
Mean Temp. Dist.	5.321	5.330	5.348	5.350	5.428	5.723	5.349
SD Temp. Dist.	3.829	3.826	3.822	3.822	3.817	3.865	3.847
Mean Precip. Dist.	29.61	29.65	29.75	29.77	30.16	31.63	29.22
SD Precip. Dist.	23.21	23.19	23.18	23.18	23.17	23.45	22.94
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Drop Adjacent Counties	Drop Adjacent and Next to Adjacent Counties	Drop Moves < 100km	Between States Only	Drop Adjacent States	Drop California

Notes: The sample in column 1 includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. Subsequent columns exclude counties (or county-pairs) as specified at the bottom of the table. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A15. Climate Distance and Migration: Controlling for Time-Varying Travel Distance

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.260***	-0.260***	-0.252***	-0.271***	-0.253***	-0.296***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)
	[-0.070]	[-0.070]	[-0.074]	[-0.079]	[-0.074]	[-0.086]
Precipitation Distance	-0.011***	-0.011***	-0.010***	-0.011***	-0.014***	-0.020***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
	[-0.018]	[-0.017]	[-0.017]	[-0.019]	-[0.025]	[-0.034]
Distance (in 100km)	-0.149***		-0.068***	-0.136***		
	(0.005)		(0.006)	(0.005)		
	[-0.085]		[-0.042]	[-0.084]		
Transportation Cost			-0.018***	-0.005***	-0.028***	-0.061***
			(0.001)	(0.002)	(0.001)	(0.003)
			[-0.110]	[-0.009]	[-0.176]	[-0.118]
Observations	30,479,809	30,479,809	23,119,668	23,119,668	23,119,668	23,119,668
Pseudo R-squared	0.705	0.706	0.700	0.693	0.696	0.679
Mean Temp. Dist.	5.321	5.321	5.348	5.348	5.348	5.348
SD Temp. Dist.	3.829	3.829	3.845	3.845	3.845	3.845
Mean Precip. Dist.	29.61	29.61	29.28	29.28	29.28	29.28
SD Precip. Dist.	23.21	23.21	22.98	22.98	22.98	22.98
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). Column 2 replicates column 1, which estimates the baseline specification (Table 2, column 5), by interacting physical distance with decade dummies. Columns 3 and 5 (resp., columns 4 and 6) control for time-varying county-pair level travel costs measured in hours (resp., time). See Appendix B.6 for more details on the construction of these variables. Columns 5 and 6 do not include (time invariant) physical distance. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A16. Climate Distance and Migration: Controlling for Geography

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.260***	-0.255***	-0.259***	-0.258***	-0.251***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
	[-0.070]	[-0.069]	[-0.070]	[-0.070]	[-0.068]
Precipitation Distance	-0.011***	-0.010***	-0.011***	-0.011***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-0.018]	[-0.016]	[-0.018]	[-0.017]	[-0.016]
Distance (in 100km)	-0.149***	-0.145***	-0.152***	-0.145***	-0.146***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
	[-0.085]	[-0.083]	[-0.087]	[-0.083]	[-0.084]
Observations	30,479,809	30,479,809	30,479,809	30,479,809	30,479,809
Pseudo R-squared	0.705	0.706	0.706	0.706	0.707
Mean Temp. Dist.	5.321	5.321	5.321	5.321	5.321
SD Temp. Dist.	3.829	3.829	3.829	3.829	3.829
Mean Precip. Dist.	29.61	29.61	29.61	29.61	29.61
SD Precip. Dist.	23.21	23.21	23.21	23.21	23.21
County o × Decade FE	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes
Elevation		Yes			Yes
Ruggedness			Yes		Yes
Coastal				Yes	Yes

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period. The migration matrix is derived from the linked sample. See Appendix B for more details. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). All regressions control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. In column 2, we control for the distance in elevation between county o and d. The ruggedness control is the distance in standard deviation in elevation between county o and d. Coastal control is a dummy equal to one if both county o and d are coastal. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A17. Climate Distance and Migration: Controlling for Soil Types

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.260***	-0.251***	-0.248***	-0.260***	-0.256***	-0.244***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)
	[-0.070]	[-0.068]	[-0.067]	[-0.070]	[-0.069]	[-0.066]
Precipitation Distance	-0.011***	-0.010***	-0.010***	-0.007***	-0.009***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-0.018]	[-0.016]	[-0.016]	[-0.011]	[-0.015]	[-0.008]
Distance (in 100km)	-0.149***	-0.139***	-0.146***	-0.146***	-0.147***	-0.137***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
	[-0.085]	[-0.080]	[-0.084]	[-0.084]	[-0.084]	[-0.078]
Bulk Distance		-0.030***				-0.019***
		(0.002)				(0.003)
		[-0.020]				[-0.013]
Concentration Distance			-0.168***			-0.115***
			(0.017)			(0.018)
			[-0.025]			[-0.017]
pH Distance				-0.024***		-0.021***
				(0.002)		(0.002)
				[-0.012]		[-0.010]
Water Distance					-0.078***	-0.052***
					(0.005)	(0.005)
					[-0.017]	[-0.011]
Observations	30,479,809	30,479,809	30,479,809	30,479,809	30,479,809	30,479,809
Pseudo R-squared	0.705	0.707	0.707	0.706	0.707	0.709
Mean Temp. Dist.	5.321	5.321	5.321	5.321	5.321	5.321
SD Temp. Dist.	3.829	3.829	3.829	3.829	3.829	3.829
Mean Precip. Dist.	29.61	29.61	29.61	29.61	29.61	29.61
SD Precip. Dist.	23.21	23.21	23.21	23.21	23.21	23.21
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). County o (resp., county d) refers to county of origin (resp., of destination). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A18. Climate Distance and Migration: Alternative Climate Definitions

Dep. var.:	Number of Migrants			
	(1)	(2)	(3)	(4)
Temperature Distance	-0.260***	-0.253***	-0.277***	-0.245***
	(0.006)	(0.006)	(0.007)	(0.007)
	[-0.070]	[-0.070]	[-0.062]	[-0.054]
Precipitation Distance	-0.011***	-0.012***	-0.003**	-0.010***
	(0.001)	-0.001	(0.001)	(0.001)
	[-0.018]	[-0.02]	[-0.005]	[-0.014]
Distance (in 100km)	-0.149***	-0.146***	-0.161***	-0.155***
	(0.005)	(0.004)	(0.005)	(0.006)
	[-0.085]	[-0.083]	[-0.092]	[-0.070]
Observations	30,479,809	30,479,809	30,479,809	18,594,032
Pseudo R-squared	0.705	0.706	0.698	0.730
Mean Temp. Dist.	5.321	5.455	4.375	31.65
SD Temp. Dist.	3.829	3.899	3.197	23.88
Mean Precip. Dist.	29.61	29.61	27.90	5.435
SD Precip. Dist.	23.21	23.21	24.85	3.899
County o × Decade FE	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes
Temperature Measurement	Yearly Avg.	Yearly Max	Summer	Rolling Avg.
Precipitation Measurement	Yearly Avg.	Yearly Avg.	Summer	Rolling Avg.

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). *Yearly Avg.* is the baseline measure of climate, obtained by taking the 1895-1920 mean of yearly average climate. *Yearly Max* refers to the 1895-1920 mean of yearly maximum temperature. *Summer* corresponds to the 1895-1920 mean of summer average climate. In the *Rolling Avg.* specification, we use average yearly climate, but we attribute to each decade: *i*) before 1910, the climate computed over the 1895-1910 period; *ii*) after 1910, the climate measured as a rolling average of the 15 previous years (e.g., we assign average climate over the period 1905-1920 to 1910-1920 migration flows). All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by decade level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A19. Climate Distance and Migration: Alternative Samples

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.260***	-0.190***	-0.160***	-0.284***	-0.285***	-0.175***
	(0.006)	(0.003)	(0.002)	(0.007)	(0.007)	(0.006)
	[-0.070]	[-1.334]	[-2.500]	[-0.047]	[-0.046]	[-0.011]
Precipitation Distance	-0.011***	-0.008***	-0.007***	-0.017***	-0.017***	-0.009***
	(0.001)	(0.000)	(0.000)	(0.002)	(0.002)	(0.001)
	[-0.018]	[-0.355]	[-0.627]	[-0.015]	[-0.014]	[-0.004]
Distance (in 100km)	-0.149***	-0.090***	-0.063***	-0.159***	-0.159***	-0.106***
	(0.005)	(0.002)	(0.001)	(0.006)	(0.006)	(0.003)
	[-0.085]	[-1.334]	[-2.047]	[-0.059]	[-0.058]	[-0.020]
Observations	30,479,809	30,161,185	29,229,535	14,005,934	13,389,801	3,093,539
Pseudo R-squared	0.705	0.346	0.195	0.748	0.748	0.607
Mean Temp. Dist.	5.321	5.357	5.443	4.763	4.711	3.002
SD Temp. Dist.	3.829	3.827	3.828	3.425	3.391	2.787
Mean Precip. Dist.	29.61	29.78	30.13	21.67	21.43	20.05
SD Precip. Dist.	23.21	23.19	23.19	17.82	17.68	20.57
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940, in column 1. Column 2 (resp., column 3) drops observations in the top 1% (resp., 5%) of the distribution of the number of migrants. Column 4 includes only counties already existing in 1860. Column 5 (resp., column 6) includes county-pairs that existed in all decades (resp., observations with a strictly positive number of migrants). *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by decade level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A20. Climate Distance and Migration: Alternative Standard Errors

Dep. var.:	Number of Migrants									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Temperature Distance	-0.260***	-0.260***	-0.260***	-0.260***	-0.260***	-0.260***	-0.260***	-0.260***	-0.260***	-0.260***
	(0.006)	(0.013)	(0.008)	(0.011)	(0.017)	(0.024)	(0.009)	(0.012)	(0.019)	(0.027)
	[-0.077]	[-0.077]	[-0.077]	[-0.077]	[-0.077]	[-0.077]	[-0.077]	[-0.077]	[-0.077]	[-0.077]
Precipitation Distance	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.003)
	[-0.019]	[-0.019]	[-0.019]	[-0.019]	[-0.019]	[-0.019]	[-0.019]	[-0.019]	[-0.019]	[-0.019]
Distance (in 100km)	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***
	(0.005)	(0.010)	(0.008)	(0.011)	(0.019)	(0.025)	(0.008)	(0.010)	(0.005)	(0.018)
	[-0.091]	[-0.091]	[-0.091]	[-0.091]	[-0.091]	[-0.091]	[-0.091]	[-0.091]	[-0.091]	[-0.091]
Observations	30,485,232	30,485,232	30,485,232	30,485,232	30,485,232	30,485,232	30,485,232	30,485,232	30,485,232	30,485,232
Pseudo R-squared	0.70528	0.70528	0.70528	0.70528	0.70528	0.70528	0.70528	0.70528	0.70528	0.70528
Mean Temp. Dist.	5.426	5.426	5.426	5.426	5.426	5.426	5.426	5.426	5.426	5.426
SD Temp. Dist.	3.884	3.884	3.884	3.884	3.884	3.884	3.884	3.884	3.884	3.884
Mean Precip. Dist.	31.16	31.16	31.16	31.16	31.16	31.16	31.16	31.16	31.16	31.16
SD Precip. Dist.	23.62	23.62	23.62	23.62	23.62	23.62	23.62	23.62	23.62	23.62
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conley Cutoff (in km)			50	100	250	500	50	100	250	500
Lat & Long			Origin	Origin	Origin	Origin	Destination	Destination	Destination	Destination

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, and is derived from the linked sample described in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors are reported in parentheses, and are clustered at the state of origin by state of destination by decade level in column 1, and at the state of origin by state of destination level in column 2. Columns 3 to 10 estimate standard errors using the procedure in Conley (1999) with the spatial cutoffs reported at the bottom of the table. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A21. Climate Distance and Migration: Long Differences

Dep. var.:	Change in Log Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Change Log Temp. Distance	-0.073***	-0.047***	-0.048**	-0.035**	-0.045***	-0.043***
	(0.020)	(0.016)	(0.023)	(0.016)	(0.016)	(0.012)
	[-0.017]	[-0.011]	[-0.011]	[-0.007]	[-0.012]	[-0.017]
Change Log Precip. Distance	-0.041**	-0.038**	-0.058**	-0.039***	-0.040***	-0.028***
	(0.020)	(0.015)	(0.024)	(0.015)	(0.015)	(0.009)
	[-0.012]	[-0.011]	[-0.015]	[-0.010]	[-0.013]	[-0.014]
Observations	26,331	26,331	17,866	49,973	19,377	53,456
R-squared	0.572	0.618	0.641	0.583	0.628	0.640
County *o* FE	Yes	Yes	Yes	Yes	Yes	Yes
County *d* FE	Yes	Yes	Yes	Yes	Yes	Yes
Log Distance		Yes	Yes	Yes	Yes	Yes
Log Bilateral Controls			Yes			
Modern Period	2018-2019	2018-2019	2018-2019	2011-2012	2014-2015	Avg. 2011-2019
Historical Period	1935-1940	1935-1940	1935-1940	1935-1940	1935-1940	Avg. 1910-1940

Notes: Change in Log Migrants is the difference between the log number of modern and historical migrants. *Change Log Temp. Distance* (resp., *Change Log Precip. Distance*) is the difference between the log temperature (resp., precipitation) of modern and historical period. Log Distance refers to the the log of physical distance (expressed in 100 km) between counties. Bilateral Controls include absolute difference in employment share, manufacturing employment share, agriculture employment share, black share, urban share, immigrant share, sex ratio, population density, exposure to the frontier, access to market distance, and a dummy equals to 1 if the two counties are both connected to the railroad. County *o* (resp., county *d*) refers to county of origin (resp., of destination). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A22. Climate Distance and Migration: Controlling for Lagged Migration

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.260***	-0.255***	-0.255***	-0.255***	-0.241***	-0.241***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
	[-0.070]	[-0.067]	[-0.069]	[-0.069]	[-0.053]	[-0.053]
Precipitation Distance	-0.011***	-0.011***	-0.011***	-0.011***	-0.010***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-0.018]	[-0.017]	[-0.018]	[-0.018]	[-0.013]	[-0.013]
Distance (in 100km)	-0.149***	-0.146***	-0.146***	-0.146***	-0.153***	-0.153***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
	[-0.085]	[-0.082]	[-0.084]	[-0.083]	[-0.069]	[-0.069]
Lagged Migration		1.362***	0.549***	3.706***	1.389***	5.596***
		(0.274)	(0.106)	(0.643)	(0.343)	(1.378)
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Observations	30,479,809	29,328,452	30,479,809	30,479,809	18,594,032	18,594,032
Pseudo R-squared	0.705	0.708	0.707	0.707	0.731	0.731
Mean Temp. Dist.	5.321	5.306	5.321	5.321	5.454	5.454
SD Temp. Dist.	3.829	3.813	3.829	3.829	3.912	3.912
Mean Precip. Dist.	29.61	29.53	29.61	29.61	31.07	31.07
SD Precip. Dist.	23.21	23.08	23.21	23.21	23.77	23.77
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Definition Lagged Mig.		Previous Decade	Lagged Stock	Avg. Lagged Flows	1900 Stock	Pre-1900 Avg. Flows
Sample Period	Full	Full	Full	Full	1910-1940	1910-1940

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940 in columns 1 to 4, and from 1900-1910 to 1930-1940 in columns 5 and 6. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period. The migration matrix is described in detail in Appendix B. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). *Lagged Migration* is: *i*) the total number of migrants in the last decade (resp., in all previous decades) in column 2 (resp., column 3); *ii*) the average number of migrants from 1850-1860 to the previous decade in column 4; *iii*) the total (resp., the average) number of migrants from 1850-1860 to 1880-1900 in column 5 (resp., column 6). Historical immigration is expressed in 10,000. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A23. Climate Distance and Migration: Evidence from the US Frontier

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.323***	-0.206***	-0.251***	-0.151***	-0.188***	-0.206***
	(0.008)	(0.010)	(0.019)	(0.018)	(0.019)	(0.018)
	[-0.434]	[-1.683]	[-1.815]	[-1.581]	[-1.278]	[-2.034]
Precipitation Distance	-0.018***	-0.024***	-0.025***	-0.008	-0.048***	0.005
	(0.001)	(0.004)	(0.005)	(0.007)	(0.008)	(0.006)
	[-0.145]	[-0.993]	[-0.746]	[-0.380]	[-1.568]	[0.257]
Distance (in 100km)	-0.129***	-0.229***	-0.267***	-0.303***	-0.199***	-0.201***
	(0.008)	(0.017)	(0.029)	(0.038)	(0.032)	(0.028)
	[-0.389]	[-2.778]	[-2.937]	[-4.754]	[-1.948]	[-2.749]
Observations	19,994,393	1,227,965	309,380	305,085	307,040	306,460
Pseudo R-squared	0.527	0.366	0.378	0.321	0.386	0.349
Mean Temp. Dist.	5.001	5.534	5.558	5.372	5.461	5.746
SD Temp. Dist.	3.615	3.715	3.720	3.618	3.660	3.846
Mean Precip. Dist.	26.28	30.50	21.11	29.17	32.68	39.10
SD Precip. Dist.	21.50	18.65	15.07	16.98	17.36	20.14
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Origins	Any	Never Frontier	Never Frontier	Never Frontier	Never Frontier	Never Frontier
Destinations	Any	Frontier	Frontier	Frontier	Frontier	Frontier
Endline Decade	1860-1900	1860-1900	1860	1870	1880	1900

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1880-1900. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period. The migration matrix is derived from the linked sample. See Appendix B for more details. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). County o (resp., county d) refers to county of origin (resp., of destination). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by year of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A24. Climate, Migration, and the Homestead Act: Robustness to Sample

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.299***	-0.279***	-0.299***	-0.225***	-0.186***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
	[-0.181]	[-0.137]	[-0.206]	[-0.344]	[-0.363]
Temperature Distance	0.147***	0.165***	0.124***	0.072***	0.035***
× Homestead	(0.011)	(0.013)	(0.011)	(0.011)	(0.010)
	[0.042]	[0.039]	[0.038]	[0.068]	[0.057]
Precipitation Distance	-0.014***	-0.017***	-0.012***	-0.013***	-0.009***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
	[-0.050]	[-0.048]	[-0.049]	[-0.121]	[-0.108]
Precipitation Distance	0.002	-0.005*	0.002	0.004*	0.001
× Homestead	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
	[0.005]	[-0.007]	[0.005]	[0.025]	[0.009]
Distance (in 100km)	-0.185***	-0.267***	-0.153***	-0.142***	-0.140***
	(0.009)	(0.013)	(0.008)	(0.009)	(0.008)
	[-0.245]	[-0.297]	[-0.229]	[-0.471]	[-0.563]
Distance × Homestead	0.065***	0.135***	0.049***	0.020***	0.010*
	(0.007)	(0.011)	(0.007)	(0.005)	(0.006)
	[0.049]	[0.093]	[0.040]	[0.051]	[0.044]
Observations	27,678,550	19,546,989	35,164,852	13,337,957	6,345,932
Pseudo R-squared	0.600	0.639	0.581	0.536	0.501
Mean Temp. Dist.	5.113	4.863	5.229	5.634	5.730
SD Temp. Dist.	3.699	3.514	3.773	3.951	3.971
Mean Precip. Dist.	27.22	25.10	28.92	34.07	37.52
SD Precip. Dist.	22.06	20.31	22.92	24.05	23.33
County o × Decade FE	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes
Origins	Never Htd	Never Htd States	No Htd t-10, t	Never Htd	Never Htd
Destinations	Any	Any	Any	Some Htd 1862-1920	Some Htd t-10, t

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1860-1870 to 1910-1920 such that: *i*) origin counties never offered any Homestead land (column 1); *ii*) origin counties belong to US states that never offered any Homestead land (column 2); *iii*) origin counties did not offer any Homestead land in the previous decade (column 3); *iv*) origin counties never offered any Homestead land and destination counties offered at least some Homestead land, respectively, between 1862 and 1920 and during the previous decade (columns 4 and 5). *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, derived from the US linked sample. See Appendix B for more details. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). *Homestead* is a dummy equal to one if the homesteaded share of the area in the county of destination in a given decade is above the 80^{th} percentile of the homesteaded share of county area calculated over the entire US for the 1860-1920. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by decade level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A25. Climate, Migration, and the Homestead Act: Robustness to Area Threshold

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.293***	-0.300***	-0.298***	-0.299***	-0.297***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
	[-0.158]	[-0.166]	[-0.171]	[-0.181]	[-0.191]
Temperature Distance × Homestead	0.171***	0.177***	0.173***	0.147***	0.103***
	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
	[0.070]	[0.065]	[0.057]	[0.042]	[0.022]
Precipitation Distance	-0.019***	-0.020***	-0.017***	-0.014***	-0.012***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
	[-0.058]	[-0.062]	[-0.057]	[-0.050]	[-0.047]
Precipitation Distance × Homestead	-0.001	-0.001	0.001	0.002	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	[-0.004]	[-0.002]	[0.003]	[0.005]	[0.009]
Distance (in 100km)	-0.259***	-0.223***	-0.208***	-0.185***	-0.155***
	(0.011)	(0.011)	(0.010)	(0.009)	(0.008)
	[-0.303]	[-0.267]	[-0.260]	[-0.245]	[-0.219]
Distance × Homestead	0.141***	0.112***	0.090***	0.065***	0.016**
	(0.010)	(0.008)	(0.008)	(0.007)	(0.007)
	[0.162]	[0.116]	[0.081]	[0.049]	[0.009]
Observations	23,031,689	24,027,360	25,577,474	27,678,550	30,637,624
Pseudo R-squared	0.626	0.617	0.609	0.600	0.589
Mean Temp. Dist.	4.921	4.968	5.033	5.113	5.187
SD Temp. Dist.	3.561	3.597	3.646	3.699	3.747
Mean Precip. Dist.	25.46	25.82	26.48	27.22	27.98
SD Precip. Dist.	20.39	20.69	21.48	22.06	22.60
County o × Decade FE	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes
Threshold	P50	P60	P70	P80	P90

Notes: The sample includes all county-pairs in the contiguous US for each decade from 1860-1870 to 1910-1920 such that origin counties never offered any Homestead land. *Number of Migrants* is the number of men 15+ in the baseline year who moved from the origin to the destination county in each period, derived from the US linked sample. See Appendix B for more details. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). *Homestead* is a dummy equal to one if the homesteaded share of the area in the county of destination in a given decade is above the percentile of the homesteaded share of county area calculated over the entire US for the 1860-1920, and reported at the bottom of the table. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the state of origin by state of destination by decade level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

B Data Appendix

B.1 Climate Spatial Correlation

As discussed in the main paper (Section 3), a potential concern with our analysis is that climate is spatially correlated. For this reason, it is possible that our estimates pick up the effects of geographic—rather than climate—distance. Such concern is especially relevant for domestic migration, where the "Atlantic reset" does not create a clear break between the origins and the destinations, as instead happens for Norwegian immigrants. Temperature and precipitation are determined by continuous physical phenomenon, and discontinuity can only be introduced by topography: mountain ranges or large bodies of water (such as the Atlantic Ocean) can interrupt climate correlation.

The higher the correlation between physical and climate distance, the more concrete the possibility that our results might capture migrants' distaste for travelling long distances. To explore this possibility, we measure the extent of spatial correlation of climate within the US. In Table B1, we flexibly regress temperature and precipitation distance on non-linear functions of physical distance. We regress temperature and precipitation distance between the universe of US counties on distance polynomials of increasing degrees, from 1 to 6. Interestingly, we find that physical distance can explain at most 22% (31%) of the variation in bilateral temperature (precipitation) distance between US counties. The low predictive power of these models suggests that spatial correlation does not absorb all the variation in our regressors of interest.

In Table B2, we explore the possibility that climate distance might be highly correlated within short physical distances. This might pose a threat to our analysis, since migrants in the domestic linked sample tend to travel relatively short distances.[67] Column 1 replicates column 2 of Table B1. In subsequent columns, we regress temperature and precipitation distances on polynomials of degree 2 of geographic distance, restricting the sample to county-pairs that are apart no more than the number of km reported at the bottom of each panel. Column 3 of Table B2 shows that, even for short-range migration (e.g., 0 to 500km), a quadratic function of physical distance can only explain 19% (resp., 7%) of the temperature (resp., precipitation) variation between counties.

Taken together, these results indicate that, while climate and physical exhibit some degree of correlation, there remains large variation in climate distance even after controlling for physical distance. In the main text, we leverage such variation to identify the relationship between climate distance and migration.

[67] For instance, the median distance traveled by migrants in our sample is 580km.

Table B1. Spatial Correlation of Climate: Flexibly Predicting Climate Distances

Dep. var.:	Temperature Distance					
	(1)	(2)	(3)	(4)	(5)	(6)
Distance (in 100km)	4,202***	4,202***	4,202***	4,202***	4,202***	4,202***
	(3.690)	(3.485)	(3.482)	(3.477)	(3.477)	(3.477)
Distance (in 100km)^2		-3,768***	-3,768***	-3,768***	-3,768***	-3,768***
		(3.485)	(3.482)	(3.477)	(3.477)	(3.477)
Distance (in 100km)^3			429.1***	429.1***	429.1***	429.1***
			(3.482)	(3.477)	(3.477)	(3.477)
Distance (in 100km)^4				582.4***	582.4***	582.4***
				(3.477)	(3.477)	(3.477)
Distance (in 100km)^5					128.3***	128.3***
					(3.477)	(3.477)
Distance (in 100km)^6						122.8***
						(3.477)
Observations	9,644,130	9,644,130	9,644,130	9,644,130	9,644,130	9,644,130
R-squared	0.11852	0.21385	0.21509	0.21736	0.21747	0.21758
Dep. var.:	Precipitation Distance					
	(1)	(2)	(3)	(4)	(5)	(6)
Distance (in 100km)	39,624***	39,624***	39,624***	39,624***	39,624***	39,624***
	(20.13)	(19.82)	(19.77)	(19.77)	(19.77)	(19.77)
Distance (in 100km)^2		-10,950***	-10,950***	-10,950***	-10,950***	-10,950***
		(19.82)	(19.77)	(19.77)	(19.77)	(19.77)
Distance (in 100km)^3			-4,033***	-4,033***	-4,033***	-4,033***
			(19.77)	(19.77)	(19.77)	(19.77)
Distance (in 100km)^4				-184.4***	-184.4***	-184.4***
				(19.77)	(19.77)	(19.77)
Distance (in 100km)^5					1,637***	1,637***
					(19.77)	(19.77)
Distance (in 100km)^6						339.9***
						(19.77)
Observations	9,644,130	9,644,130	9,644,130	9,644,130	9,644,130	9,644,130
R-squared	0.28664	0.30853	0.31150	0.31151	0.31200	0.31202

Notes: The sample includes all county-pairs in the contiguous US. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table B2. Spatial Correlation of Climate: Flexibly Predicting Temperature Distances within Distance Bins

Dep. var.:	Temperature Distance							
Distance Bins	All	(0,2]	(0,5]	(0,10]	(0,15]	(0,20]	(0,30]	(0,40]
Distance (in 100km)	4,202***	1,288***	685.1***	2,285***	4,006***	5,016***	5,103***	4,235***
	(3.485)	(0.7930)	(1.267)	(2.053)	(2.730)	(3.178)	(3.444)	(3.479)
Distance (in 100km)^2	-3,769***	-6.466***	-24.23***	-65.52***	-343.8***	-973.3***	-2,494***	-3,771***
	(3.485)	(0.7930)	(1.267)	(2.053)	(2.730)	(3.178)	(3.444)	(3.479)
Observations	9,644,130	214,464	1,171,944	3,651,476	6,084,924	7,776,054	9,182,151	9,624,016
R-squared	0.21385	0.10813	0.19986	0.25342	0.26283	0.24949	0.22847	0.21637
Dep. var.:	Precipitation Distance							
Distance Bins	All	(0,2]	(0,5]	(0,10]	(0,15]	(0,20]	(0,30]	(0,40]
Distance (in 100km)	39,624***	708.2***	3,068***	10,234***	19,072***	27,498***	38,492***	39,774***
	(19.82)	(7.369)	(10.07)	(13.62)	(16.20)	(17.97)	(19.31)	(19.79)
Distance (in 100km)^2	-10,950***	-100.4***	-296.9***	-113.2***	-695.5***	-1,024***	-3,228***	-10,207***
	(19.82)	(7.369)	(10.07)	(13.62)	(16.20)	(17.97)	(19.31)	(19.79)
Observations	9,644,130	214,464	1,171,944	3,651,476	6,084,924	7,776,054	9,182,151	9,624,016
R-squared	0.30853	0.04208	0.07405	0.13390	0.18571	0.23161	0.30355	0.30911

Notes: The sample includes all county-pairs in the contiguous US. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin and the destination county. *Distance* refers to the physical distance between counties (and is expressed in 100 km). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

B.2 Norwegian Immigration to the US

B.2.1 Linked Sample over 1865-1880

In the paper, we rely on a dataset that measures the number of immigrants who were living in a given Norwegian municipality in 1865 and who moved to a given US county in 1880. To derive this dataset, we proceed as follows. We adapt the matching algorithm described in Abramitzky et al. (2021) to match the Norwegian Census of 1865 to the population of Norwegian immigrants enumerated in the US Census of 1880. As is standard in the literature and in order to circumvent changes from maiden to married name, common for women at the time, we restrict attention to men in both Censuses (Abramitzky et al., 2021). More specifically, we consider men who were at least 10 years old in the Norwegian Census, and at least 25 in the 1880 US Census. This ensures that we perform the Census linking on samples of men with plausible birth dates, and who were old enough (i.e., at least 15 years old) in the 10 years prior to the 1880 US Census (when we observe them in the new location). We clean first and last names by removing special characters, accents, middle names, and name extensions, such as "Jr." or numbers. Then, we standardize nicknames using a crosswalk provided by the Minnesota Population Center and made available by Abramitzky et al. (2012). Finally, both first and last names are standardized using the NYSIIS algorithm implemented through the code in Sayers (2014). This procedure translates all names to their English phonetic form, ensuring that they are transcribed in the closest possible way to what a Census enumerator would have done in 1880.

Next, we search for *unique exact* matches of observations between the two Censuses. Matches are performed on first names, last names, and age in 1880. As in Abramitzky et al. (2012), we augment the matched dataset by relaxing the exact matching constraint on the age of the individual by accepting a unique match if the first and last names agree and if the $\text{age}^{NOR,1865} + 15 = \text{age}^{US,1880} \pm 2$. We discard all ambiguous matches. That is, we drop from the dataset all individuals who can be matched to two or more Norwegian immigrants in the US based on their names and date of birth. The resulting dataset is a panel dataset with two observations for each individual—one in 1865 and one in 1880. As shown in Table B3, we are able to match 5,966 individuals, corresponding to a match rate of 8.24% (very close to the 8.52% match rate obtained for the same period by Eriksson, 2020).[68]

We locate Norwegian immigrants in the US using data from the Census Place Project (Berkes et al., 2023). Information on Norwegians municipality of residence in Norway is obtained from the Norwegian Census.[69] Additional individual characteristics, such as labor force participation,

[68] The slight discrepancy between our match rate and that in Eriksson (2020) is likely due to the different sets of individuals included in the linking procedure: Eriksson (2020) focuses on men aged 18-65 in 1880 and 1910, and aged 21-65 in 1920.

[69] Few municipalities change boundaries over time. In the event of a split, we assign population to the largest of the two resulting

occupation, and urban status, are retrieved from the two Censuses. Table A3 in Appendix A presents the characteristics (reported in the US Census of 1880) of individuals linked between the 1865 Norwegian Census and the 1880 US Census, and compares them to those of Norwegian immigrant men (of comparable age) in the 1880 US full count Census.[70]

Finally, we collapse the data to obtain the number of individuals who migrated from Norwegian municipality o to US county d. The resulting dataset consists of all possible pairs of Norwegian municipalities in 1865 and US counties in 1880, provided that at least one European immigrant—irrespective of age and gender—was recorded in the US county in the US full count Census of 1880.[71]

B.2.2 Linked Samples over Alternative Periods

In the paper, we also present results considering Norwegian immigrants linked over alternative time periods. Besides the 1865-1880 sample, we also consider: 1865-1870, 1865-1900, 1900-1910, and 1900-1920. To this end, we adopt a procedure very similar to the one described in Appendix B.2.1 for 1865-1880. Specifically, we: *i*) select only Norwegian men older than a given age threshold—10 years at the time in Norway; *ii*) clean and standardize their first and last names via the aforementioned procedure; and, *iii*) link them based on their standardized names and reported age in the US, always allowing for some misreporting in age. The sample of individuals in the US (our target population) is constructed by selecting individuals who are: *i*) older than 10 in the Norwegian Census; *ii*) old enough to appear in the Norwegian Census (either of 1865 or of 1900).[72]

Concretely, take Peter—a Norwegian immigrant living in the US who was 45 in 1900. According to his reported age in 1900, Peter: *i*) is older than 10 in 1865; and, *ii*) is old enough to appear in the 1865 Norwegian Census. Thus, he is part of our sample. On the other hand, Eric—a Norwegian immigrant living in the US in 1900 who reports an age of 37—is not part of our sample. This is because, even if Eric is old enough to appear in the 1865 Norwegian Census, he was younger than 10 at the time of that Census, making him ineligible for our sample.

Similarly to Eriksson (2020), since from 1900 onward the US Census reports also the year of

geographies.

[70] Using a procedure very similar to the one described above, we also create a sample of Norwegian men linked between the 1865 and the 1900 Norwegian Censuses. We compare their characteristics with those of linked migrants in Table A2. Table A1 compares the 1865 characteristics of Norwegian men in either linked sample (migrants and stayers) to those of men 10+ in the full count Norwegian Census. Appendix A also presents the geographic distribution of Norwegian men, both in Norway and in the US (Figures A9 and A10).

[71] As discussed in the main text, results are robust to using alternative sample definitions.

[72] The actual constraint on age that we impose is slightly less stringent, so as to allow for potential mis-reporting in the Census. For example, if the age restriction requires men to be older than 45, the constraint we impose is that men have to be 40 or older. Results are robust to using exact age.

immigration (while this piece of information is not available in 1880), we augment the linking procedure with this variable too, whenever possible. That is, for the linked samples of 1865-1900, 1900-1910, and 1900-1920, in addition to the parameters mentioned above, we further exploit the year of immigration reported in the US Census, discarding individuals whose immigration year predates the year of the Norwegian Census of origin.[73]

Table B3 reports the number of Norwegian men linked in each time window, together with the corresponding match rate (expressed relative to the number of Norwegian-born men in the US who satisfy our aforementioned constraints). Overall, the resulting match rates are in line with those in the literature (Abramitzky et al., 2021), including specifically those in Eriksson (2020) for Norwegian immigrants.

Table B3. Linked Individuals and Match Rates, by Period

Sample Period	Number of Individuals	Match Rate (%)
1865-1870	3,874	6.86
1865-1880	5,966	8.24
1865-1900	4,486	8.30
1900-1910	8,262	11.22
1900-1920	8,292	11.57

Notes: Number of individuals is the number of matched individuals in that sample period. *Match Rate* is computed as the ratio between the number of matched individuals and the number of individuals in the US who meet our eligibility criteria, namely Norwegian-born men older than a given threshold.

B.3 Historical US Internal Migration

Linked samples. To construct the county-to-county migration dataset used in the paper, we proceed as follows. We use the Census Linking Project to create longitudinal datasets from historical US Census records that span each decade from 1850-1860 to 1930-1940 (Abramitzky et al., 2020).[74] The Census Linking Project relies on automated algorithms to link individuals across Census pairs using first and last name, year of birth and state or country of birth. Matches must be unique. Following Abramitzky et al. (2021) and the approach described in Appendix B.2.1, we focus only on men. We also exclude any individual who was younger than

[73] Results are virtually unchanged when dropping this additional restriction.

[74] Since the 1890 Census was destroyed in a fire, this year cannot be used in the analysis; we thus consider the 1880-1900 period.

15 in the baseline year.[75] We then rely on the Census Place Project (Berkes et al., 2023) to assign individuals in the linked sample to the county of residence in a given decade. We exclude individuals reporting to live in either Hawaii or Alaska. This process yielded a dataset with 32,943,677 observations. Tables A8 and A9 in Appendix A present the characteristics of the individuals included in the linked sample (separately by decade) and compare them to those of individuals in the full count Censuses.[76]

Similar to what we did for Norwegian immigration, we collapse the data at the county-pair by decade level to obtain a dataset with the number of county-to-county migrants in each decade. Whenever the full count US Census of the baseline year reported no men 15+ in the origin county, we set the number of migrants to missing (since there could be no individual within the age range considered in our analysis migrating from that origin).

Full count 1940 US Census. As discussed in Section 3.3 of the paper, we complement the linked sample with a dataset obtained from the 1940 full count US Census. We take advantage of the fact that in 1940, for the first time, the US Census asked individuals their place of residence 5 years before. We use this piece of information to construct a migration matrix that tracks moves of individuals between 1935 and 1940. As for the linked sample, we adjust county boundaries at destination using the Census Place Project in 1940 (Berkes et al., 2023). Instead, since the Census Place Project cannot be used for the (1935) county of origin, we rely on the cross-walk made available by Eckert et al. (2020). For consistency with the linked sample, we further restrict attention to individuals who, in 1940, were 20 years or older (who, thus, were at least 15 years old in 1935, i.e., in their location of origin), and who in either 1935 or 1940 did not live in Alaska or Hawaii. Finally, as before, we collapse the data at the county of origin by county of destination level to measure the county-to-county moves between 1935 and 1940.

B.4 Modern US Internal Migration: IRS Data

In Sections 3 and 4, we leverage county-to-county migration data from 2011 to 2019. To construct this dataset, we build on the County-to-County Migration files produced by the Internal Revenue Service (IRS) Statistics of Income (SOI) Division.[77] These data are based on year-to-year address changes reported on individual income tax returns filed with the IRS, and present migration patterns by state or by county. In particular, they record both the number of new residents who moved to a state or county (from each origin) and the number

[75] As in the Norwegian case, this restriction is imposed to increase confidence that we focus on individuals who are old enough to make the decision as to whether and where to move. All results are unchanged when imposing alternative age thresholds (including capping age in the endline year, as in Zimran, 2023) or when leaving the sample completely unrestricted in terms of age.

[76] Table A10 also compares the characteristics of individuals in the linked sample who did not move to those of domestic migrants.

[77] See SOI Tax Stats—Migration Data, https://www.irs.gov/statistics/soi-tax-stats-migration-data.

of residents leaving a state or county (to each destination). Data are available for tax filing years from 1991 to 2021. However, because data collection methods changed in 2011, we focus on migration flows between 2011 and 2019, further excluding 2020 and 2021 to avoid potential COVID-related migration anomalies.[78] For robustness, we verify that results are unchanged when including 2020 and 2021.

B.5 Surname-Level Climate Distance from German Data

In Section 2 of the paper, we explore the relationship between the temperature in Germany and the temperature in the US counties where German immigrants lived in 1880. Since full count Census data for late 19th century Germany are not available, we cannot proceed as we did for the Norwegians, by matching individuals between the US and Germany. Instead, we match spatially clustered surnames in Germany to surnames of German immigrants in the US. For each surname, we compute its associated (weighed) average German temperature, based on the distribution of that surname in Germany. Likewise, leveraging the distribution of German immigrants across US counties, we compute the (weighed) average temperature associated with that surname in the US. In what follows, we explain in detail our procedure.

We build on the lists of German WWI casualties compiled by the Association for Computer Genealogy (Verein für Computergenealogie e.V., 2014). The lists, published between 1914 and 1919 by the Prussian State Gazette, encompass the governmental notifications on the military losses sustained by the entire army of the German Empire (Zedlitz, 2016), and include wounded, missing, captured, and dead soldiers. We acquire information on approximately 3.3 million entries through data scraping.[79] Each entry contains information about the soldier's first and last name, and town of birth. The lists allow us to infer the geographical distribution of German surnames around the time of WWI. We acknowledge that the lists are only partially representative of the German population, most notably because the draft was not uniform across the Empire (De Juan et al., 2021). Moreover, the lists represent a rare source of information on the geographical distribution of the population in a country that experienced mass migration to the US at the turn of the 20th century.

In a first step, we derive a measure of average temperature at the surname-level in Germany. We geolocate the soldiers' birthplaces using ArcGIS. To ensure consistency with the definition of Germany used by US Census enumerators, we restrict attention to individuals born within the 1880 borders of the German Empire. Then, we standardize last names by removing special

[78] As for the historical analysis, we exclude records from Alaska and Hawaii. Results are unchanged when including also years prior to 2011.

[79] This represents 38.8% of the total 8.5 million entries in the official lists.

characters, accents, and *umlauts*. These steps leave us with a list of 132,910 surnames. Next, we select geographically clustered surnames, in order to ensures that the associated climate is measured with precision. We compute the centroid of birthplaces of each soldier bearing a given surname. We consider a surname as clustered if 85% of the soldiers were born within 30km of the centroid.[80] Finally, we discard surnames with less than three associated casualties. This is done for two reasons. First, very rare surnames are more likely to be the result of misspelling. Second, surnames associated with only one individual would automatically get selected based on our definition of geographically clustered surnames. Given that our goal is to select surnames that are both informative of their region of origin and common enough to be matched to German immigrants in the US, we decided to exclude them.[81] This leaves us with 9,633 surnames.

The average German temperature associated to each surname is given by:

$$\text{Temp}^{GER}\left(s^{GER}\right) = \sum_{o} \mathbf{P}(k \in o \mid s(k) = s^{GER}) \times \text{Temp}_{o}^{GER}$$

where $\mathbf{P}(k \in o \mid s(k) = s^{GER}) := \frac{\sum_k 1(k \in (o,s))}{\sum_o 1(k \in s)}$ is the probability that individual k was born in county o, conditional on having surname s. Intuitively, $\text{Temp}^{GER}(s)$ can be interpreted as the average temperature experienced in Germany by an individual with surname s.

To derive a similar surname-level average temperature for German immigrants living in the US, we turn to the 1880 full count US Census, restricting attention to German-born men 15+. We clean last names by removing special characters, accents, and numbers. This leaves us with 275,856 surnames. Next, we compute the Jaro-Winkler (JW) spelling distance between each of the previously selected German surname and the surnames of German immigrants in the US. We keep a matched surname pair if the associated JW distance is below .1.[82] Out of the 275,856 surnames of eligible German immigrants in the US, we match 95,721 (34.7%) to at least one of the 9,633 selected German surnames. Finally, we assign German climate to German immigrants in the US as follows:

$$\text{Temp}^{GER}\left(s^{US}\right) = \sum_{s^{GER} \text{ s.t. } d(s^{GER}, s^{US}) \leq 0.1} w\left(s^{GER}, s^{US}\right) \times \text{Temp}\left(s^{GER}\right)$$

[80] The threshold was selected based on the 10% of surnames within the lowest 85$^{\text{th}}$ percentile to define an implicit threshold of 30km. The 85$^{\text{th}}$ percentile of the distance for the average surname is 338km.

[81] Results, not reported for brevity, are robust to using alternative percentile thresholds and considering different radii. They are also unchanged when changing the minimum number of casualties for each surname to values higher than three.

[82] This is the same threshold used in the literature (e.g., Abramitzky et al., 2021) and in our own analysis with Norwegian immigrants (see also Appendix B.2).

where $d(s, z)$ is the JW distance between s and z and:

$$w\left(s^{GER}, s^{US}\right) = \frac{1/d\left(s^{GER}, s^{US}\right)}{\sum_{s^{GER} \text{ s.t. } d(s^{GER}, s^{US}) \leq 0.1} 1/d\left(s^{GER}, s^{US}\right)}$$

$\text{Temp}^{GER}\left(s^{US}\right)$ can thus be interpreted as the expected temperature experienced *in Germany* by a German-born US immigrant.[83] To see this, note that $\text{Temp}^{GER}\left(s^{US}\right)$ is the average of the temperature associated to each matched surname weighed by the spelling similarity between the immigrant's surname and the matched German surname. Finally, we collapse the dataset of matched German immigrants in the US Census at the surname-level. The resulting dataset consists of $95,721$ surnames of German-born men aged 15+ living in a US county in 1880 who are matched to at least one of the geographically concentrated German surname defined above.

In the last step, similar to what we did above for $\text{Temp}^{GER}(s^{GER})$, we define the US temperature associated to each German immigrant's surname, $\text{Temp}^{US}(s^{US})$ as:

$$\text{Temp}^{US}\left(s^{US}\right) = \sum_{d} \mathbf{P}(k \in d \mid s(k) = s^{US}) \times \text{Temp}_{d}^{US}$$

where $\mathbf{P}(k \in d \mid s(k) = s^{US}) := \frac{\sum_k 1(k \in (d,s))}{\sum_d 1(k \in s)}$ is the probability that individual k lives in US county d, conditional on having surname s. Intuitively, $\text{Temp}^{US}(s)$ can be interpreted as the average temperature experienced in US by a German immigrant with surname s.

Figure 3 in the main text plots the raw correlation between the two temperature measures, reporting German and US climate on the x-axis and on the y-axis respectively.

[83] Note that when two surnames are identical, $d\left(s^{GER}, s^{US}\right) = 0$, and $w\left(s^{GER}, s^{US}\right)$ is not defined. Therefore, to allow for perfect matches, whenever $d\left(s^{GER}, s^{US}\right) = 0$ we artificially set it to $d\left(s^{GER}, s^{US}\right) = 10^{-8}$.

B.6 Railroads and Transportation Costs

Data on the US railroad networks expansion are taken from Donaldson and Hornbeck (2016). Figure B1 plots the evolution of the rail lines between 1860 and 1900. Most of the rail construction was finished by 1920. We use these data to compute bilateral time-varying transportation costs for each US county-pair. To this end, we augment the railroad network with straight connections within and between county centroids, which are meant to proxy stage coach connections (Figure B2). Then, we assign a weight to each edge of the resulting network. Transportation costs can be expressed either in hours or in dollars. In the former case, we use as weights 25 miles per hour for a railroad edge and 8 miles per hour for all other edges. In the latter case, we use \$1.5 per mile for a train ride and \$.03 for all other edges.[84] Then, with this weighed network at hand, we compute the least-cost paths between each county pair, for each decade between 1860 and 1920. As the railroad network grew, we observe substantial over time as well as cross-sectional variation in both measures of transportation costs.

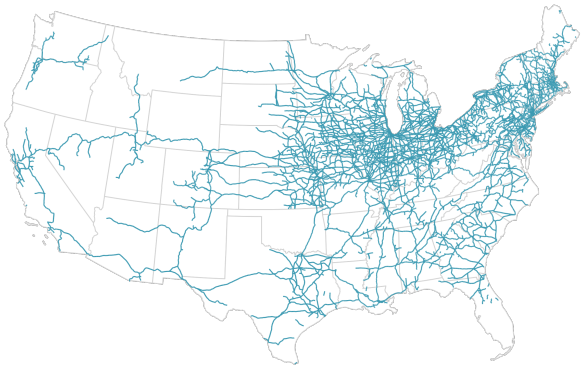
[84] These numbers were gathered from various sources, including the Poor's Railroad Manuals available at http://www.pacificng.com/template.php?page=/ref/rrmanuals/index.htm and the Annual report on the railroads of New York.

Figure B1. Railroads Expansion over Time

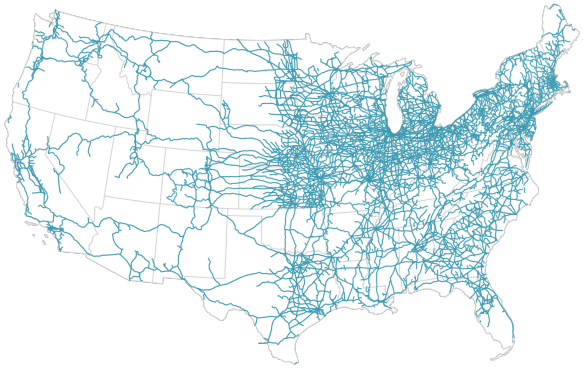
(a) 1860



(b) 1880

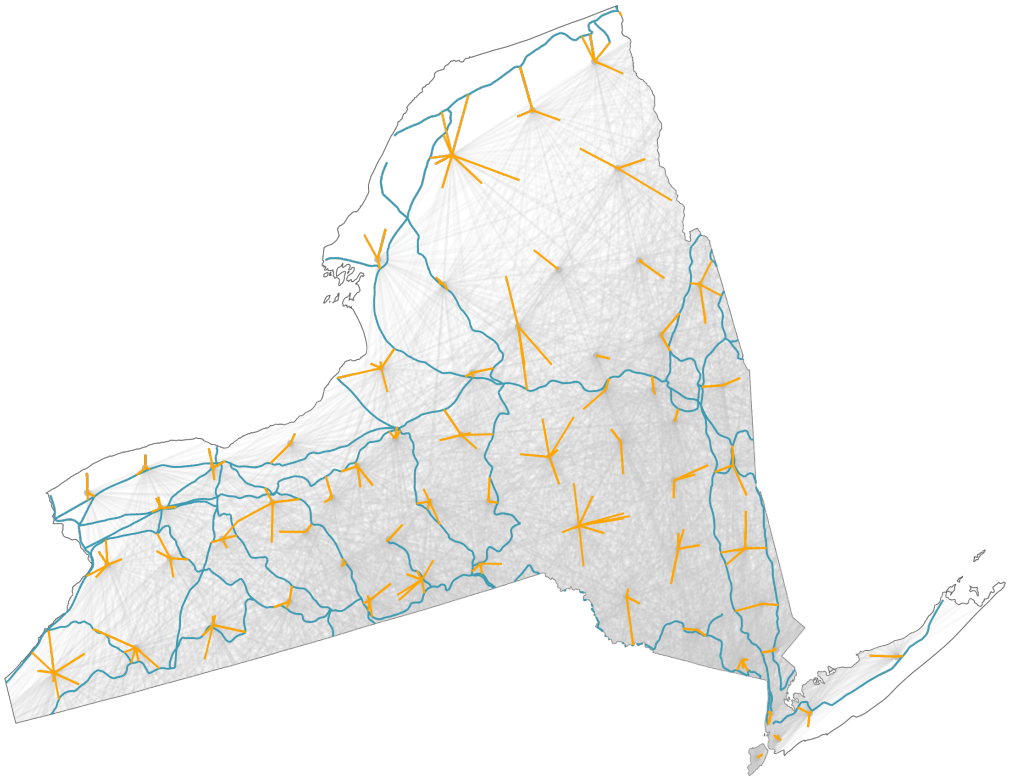


(c) 1900



Notes: The maps display the railroad network across US counties in 1860 (Panel A), 1880 (Panel B), and 1900 (Panel C).

Figure B2. Network Used to Compute Transportation Costs (1860)



Notes: The figure displays the map of New York state county centroids (grey dots), showing: (in blue) the railroad network as of 1860, (in orange) within county connection, (in grey) out of county connections.