NBER WORKING PAPER SERIES

MIGRATION, CLIMATE SIMILARITY, AND THE CONSEQUENCES OF CLIMATE MISMATCH

Marguerite Obolensky
Marco Tabellini
Charles Taylor

Working Paper 32035
http://www.nber.org/papers/w32035

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2024, revised November 2024

Previously circulated as 'Homeward Bound: How Migrants Seek Out Familiar Climates.' We thank Ran Abramitzky, David Albouy, Marcella Alsan, Sam Bazzi, Chris Boulos, Tamma Carleton, Paola Conconi, Leonardo D'Amico, Catherine De Vries, Tatyana Deryugina, Christian Dustmann, Silvia Farina, James Feigenbaum, Martin Fiszbein, Nicola Gennaioli, Paola Giuliano, Edward Glaeser, James Hammitt, Gordon Hanson, Joe Henrich, Andrew Hultgren, Peter Huybers, Kelsey Jack, Namrata Kala, William Kerr, Eliana La Ferrara, Ross Mattheis, Eoin McGuirk, Tom Nicholas, Nathan Nunn, Elias Papaioannou, Maria Petrova, Jonathan Proctor, Paola Sapienza, Wolfram Schlenker, Ambra Seck, Marlon Seror, Jesse Shapiro, Andrei Shleifer, Felipe Valencia, Nico Voigtlaender, Matt Weinzierl, Noam Yuchtman, 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Migration, Climate Similarity, and the Consequences of Climate Mismatch
Marguerite Obolensky, Marco Tabellini, and Charles Taylor
NBER Working Paper No. 32035
January 2024, revised November 2024
JEL No. J15,J61,N31,N32,Q54,R11

ABSTRACT

This paper examines the concept of "climate matching" in migration—the idea that migrants seek out destinations with familiar climates. Focusing on the US, we document that temperature distance between origin and destination predicts the distribution of migrants across counties. This pattern holds for internal and international migration in the past (1850-1940) and today (2011-2019), and is not explained by the spatial correlation of climate or the persistence of ethnic networks. We provide suggestive evidence for two mechanisms driving climate matching: climate-specific skills and climate-as-amenity. Then, we study the implications of climate matching for migrants. Leveraging plausibly exogenous variation in climate mismatch, we document that climate distance reduces life expectancy among immigrants, and increases mortality rates for their US-born children. We calculate an individual-level mortality cost of a 1° C change in climate to be $5,250.

Marguerite Obolensky
m.obolensky@columbia.edu

Marco Tabellini
Harvard Business School
279 Morgan Hall
Soldiers Field
Boston, MA 02163
and CEPR
and also NBER
mtabellini@hbs.edu

Charles Taylor
Harvard Kennedy School
79 John F Kennedy St, B-409
Cambridge, MA 02138
USA
ctaylor@hks.harvard.edu

A data appendix is available at http://www.nber.org/data-appendix/w32035

"[New England is] more suitable to the nature of our people, who neither finde content in the colder Climates, nor health in the hotter; but (as hearbs and plants) affect their native temperature, and prosper kindly no where else." – Sir Ferdinando Gorges, English founder of the Province of Maine, 1622 (Kupperman, 1984)

1 Introduction

Climate change is an increasingly studied driver of migration (Stern, 2007; Missirian and Schlenker, 2017; Mahajan and Yang, 2020). However, little is known about how *relative* climate influences migration. Do migrants care about the similarity in climate between origin and destination? If so, what happens when migrants cannot "match" on climate due to economic (e.g., travel costs) or political (e.g., immigration restrictions) barriers?

Historical accounts frequently reference the idea of migrants seeking out familiar climates and the importance of climate matching on a migrant's ultimate success. In 1925, US President Calvin Coolidge observed that "the newcomers from Europe commonly sought climatic conditions here [in the United States] 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 European settlement of the Americas (Fischer, 1989; Crosby, 2004) and US internal migration during the 19$^{\text{th}}$ century (Steckel, 1983).[1] Indeed, English migrants to North America and the West Indies expressed "profound anxiety" about the health consequences of climates that diverged from England's temperate climate as early as the 1600s (Kupperman, 1984).

Despite its intuitive appeal, little systematic evidence exists on climate matching in migration and its implications for migrants. In this paper, we seek to fill this gap. We study the relationship between climate distance and migration, and examine the consequences of climate mismatch on the mortality of immigrants and their foreign-born children. We focus on the US and consider international and internal migration from both historical and modern periods. The US is an ideal setting to study the role of climate preferences on migration, given its relative lack of internal and, historically, international restrictions on mobility and the wide range of climate zones, thus implying that migrants have many climatic options to choose from.[2] Rich datasets also allow us to measure the relationship between climate similarity and migration precisely.

[1] Steckel (1983) noted that a "farmer contemplating a move sought, other things being equal, a location that maximized the return on previous investments in human capital; namely, a place where the climate, soil, and terrain were familiar."

[2] Although the movement of white individuals within the US was largely unregulated until the 1920s, formal and informal restrictions severely limited the movement of African Americans and Native Americans (Alston and Ferrie, 1999; Nichols, 2014).

The first part of the paper documents a strong, positive relationship between climate similarity and migration, summarized in Figure 1. The figure shows that immigrants from warmer (colder) countries tend to settle in similarly warmer (colder) parts of the US. The correlation holds *across* and *within* countries over different time periods, and does not merely reflect factors such as geographic distance, income levels, or past migration.

We formalize this result by applying gravity equations common in the trade literature (Anderson and Van Wincoop, 2003) to three different settings. We define climate distance as the absolute value of the difference in average annual temperature at origin and destination. First, we use a linked dataset of individuals who moved from Norway to the US between 1865 and 1880. This setting allows us to measure climate within the country of origin at a highly granular level (Norwegian municipality), and at a unique time early in Age of Mass Migration (1850-1920) when more than 30 million Europeans moved to the US. Second, we rely on repeated cross-sections of international immigrants from any country of origin across US counties, both historically (1880-1920) and in more modern times (1970-2010). This setting complements the Norwegian one by encompassing the two main migration eras in US history and including many more immigrant-sending countries, and thus, more variation in climate distance. Yet, due to data limitations, it uses less granular climate information on the origin country. Third, we consider US internal migration both during the 1850-1940 period, characterized by the US westward expansion (Bazzi et al., 2020), and modern times since 2011.

In each setting, climate similarity strongly predicts migration flows. We find that reducing temperature distance by the corresponding mean sample distance (which ranges from 5°C to 10°C, depending on the setting) increases migration between 0.3% and 1.4%.[3] To gauge the implied magnitudes, consider that our estimated coefficients are of the same order as those associated with reducing physical distance by 500 to 1,500 km, depending on the context. They are also in line with the results obtained in the literature for the elasticity of migration with respect to wages (Tombe and Zhu, 2019; Caliendo et al., 2021; Morten and Oliveira, 2024). Consistent with the environmental economics literature, the effect of temperature is stronger and more robust than that of precipitation.[4] We perform several robustness checks, such as controlling for origin-destination pair differences in economic activity and market access, accounting for past ethnic networks, and using alternative linking methods. Further, we show that results are robust to using alternate measures of climate (e.g., max and min temperature;

[3] For reference, 5.3°C is comparable to the average annual temperature distance between Chicago (9.8°C) and Washington DC (14.6°C); Seattle (11.3°C) and San Jose, CA (16.4°C); and Berlin (10.3°C) and Rome (15.2°C).

[4] While temperature is a consistently important factor influencing health and economic outcomes (Carleton and Hsiang, 2016; Heal and Park, 2016; Deryugina and Hsiang, 2017), the link with precipitation is less clear. In the agricultural context, temperature is a better predictor of crop yields than precipitation (Lobell and Burke, 2008; Schlenker and Roberts, 2009), though this is partly mediated through irrigation and soil moisture (Taylor, 2022b; Proctor et al., 2022). Further, precipitation is spatially heterogeneous, less precisely measured than temperature, and subject to spatial aggregation bias (Fezzi and Bateman, 2015).

seasonality measures).

A key concern in analyzing the relationship between climate and migration is that any association between the two variables may reflect the effect of geographic (rather than climate) distance. Our estimates for Norwegian and international migration to the US are unlikely to suffer from this problem because the transatlantic component of migration breaks the spatial correlation of climate. However, such concerns may be more relevant for US internal migration, since one might expect climate and geographic distances to be correlated. For this reason, all our analyses control for county-pair geographic distance (which in our setting explains at most 22% of the variation in temperature 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 like California, *vi*) use travel cost measures that vary over time due to rail expansion; and, *vii*) adjust standard errors to account for spatial correlation.

Finally, we leverage shifts in average climate throughout the 20$^{\text{th}}$ century, driven by multidecadal oceanic oscillations (e.g., North Atlantic Oscillation) and anthropogenic climate change, to isolate the impact of *changes* in climate distance on changes in migration patterns. This addresses the concern that, even after controlling for a large set of county-pair covariates, our estimates may still reflect the persistence of cultural, institutional, or economic factors correlated with both climate and migration. We find that changes in climate distance from 1900 to 2020 predict changes in both internal and international migration over the same period. Our estimates suggest that a 1°C reduction in temperature distance throughout the 20$^{\text{th}}$ century increases internal and international migration by approximately 0.3%. Beyond reducing identification concerns, these results have significant policy implications in light of projected climate change, which we discuss below.

The granularity of historical US Census data, combined with the substantial variation across space and over time, allows us to explore the mechanisms driving climate matching in migration. First, we document that the estimates are larger for farmers and individuals working in climate-exposed occupations. This supports the notion that climate matching is driven—at least in part—by climate-specific skills (Steckel, 1983; Michalopoulos, 2012; Bazzi et al., 2016), such as farming expertise and techniques for surviving extreme heat or cold events. Second, we show that climate matching continues to hold even for individuals working in sectors other than agriculture (e.g., services) and indoor occupations. This suggests that migrants also value climate similarity as an amenity. The finding that climate similarity predicts migration even today, when workers are far more insulated from climate than in the past, further corroborates this interpretation.

What are the implications of climate mismatch? Even though migrants seek out familiar climates, they are unlikely to eliminate the origin-destination climate distance. Take, for example, Somali immigrants in the US: no US county is hot enough to match the average temperature of their home country. Moreover, economic pull factors, as well as the persistence of ethnic networks that might reflect "historical accidents" (e.g., the California Gold Rush, a sudden change in immigration or refugee resettlement policy) remain important factors in a migrant's location choice (Card, 2001; Burchardi et al., 2019).[5] In the second part of the paper, we examine the implications of climate mismatch for both immigrants and their US-born children.

First, we use US death records of non-European immigrants from 1988 to 2005 to study whether climate distance reduces life expectancy. Second, we rely on US birth records from 2005 to 2021 and test whether a foreign-born mother's climate distance affects infant mortality and other infant health outcomes. Climate mismatch might negatively impact health and longevity, either directly through climate-specific human capital, including "indigenous knowledge" and adaptation techniques, or indirectly through lower happiness and higher social isolation, with consequences for health outcomes where migrants value climate similarity as an amenity.

To address potential concerns about selection (e.g., individuals with better health outcomes being better able to match on climate), we derive an instrument for climate mismatch between the temperature in the country of origin and that in the county of death. We adapt the methodology developed by Burchardi et al. (2019) and Terry et al. (2024) to our context. This approach effectively randomizes the climate distance assigned to immigrants from each country of origin. We proceed in three steps. First, starting from 1880, we predict the number of non-European immigrants from a given origin country to a given US county by combining two variables: the number of immigrants who arrived in the US from other countries within the same continent—a proxy for push shocks that induced individuals to leave their country of origin; and, the share of European immigrants who settled in each US county—a proxy for pull shocks shifting the relative attractiveness of each US destination. We iterate this procedure over 100 years from 1880 to 1970 to isolate quasi-random variation in the distribution of non-European ancestry across counties in 1970. Second, we isolate exogenous non-European immigration shocks by interacting predicted ancestry in a given county with contemporaneous nationwide immigration from that origin.[6] Finally, we use these matrices of country-county immigration flows to construct plausibly exogenous weights to obtain predicted origin-specific average temperature distance.

Estimating 2SLS individual-level regressions, we find a strong, negative relationship between

[5] Data from the American Community Survey indicate that, as of 2020, more than 60,000 individuals with Somali ancestry (out of the 160,000 in the entire US) lived in Minnesota—one of the coldest US states.

[6] As in Terry et al., 2024, we apply a leave-out methodology, omitting individuals from that specific country who settled in the US Census division of a given destination county.

climate distance and age at death. According to our preferred specification—which controls for individual characteristics and fixed effects for birth cohort, US county of death, and continent of origin—increasing climate mismatch by 1°C reduces life expectancy by about six days (on a population with an average age at death of 80 years old). These results are robust to excluding specific countries of origin (e.g., Cuba, Mexico, or China) and US states, calculating the predicted number of immigrants in different ways, and estimating alternative specifications.

Then, we ask whether the next generation also bears the costs of climate mismatch by looking at infant mortality and other infant health outcomes. We use vital statistics data from the CDC's National Center for Health Statistics to create a panel of the 11 million births in a given county from mothers from a given country in each year from 2005 to 2021. Applying the same procedure described above to instrument for mothers' climate distance, we find that each additional degree Celsius of climate mismatch increases infant mortality by 0.27 deaths per thousand births—a 6% increase relative to the average infant mortality rate in our sample. These effects are mirrored by an increase in the probability of low birth weight and premature birth—two predictors of subsequent infant mortality. This suggests that climate distance impacts maternal and fetal health, and not just the newborns' environment after birth.

We use the estimates of the impact of climate distance on mortality to back out the value of climate for migrants. Using the EPA's value of a statistical life, we find that one additional degree Celsius of distance between a migrant origin and destination costs $5,250 in terms of lost life. This number is similar to the valuation of climate similarity we obtain when studying infant mortality: in that context, each additional degree of distance between a mother's origin and destination costs $2,700.

Our findings provide systematic evidence in support of the long-standing idea that historical migration patterns were influenced by climate similarity (Coolidge, 1926; Steckel, 1983; Fischer, 1989; Crosby, 2004),[7] and are 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, 2023; Desmet and Rossi-Hansberg, 2023; Bilal and Rossi-Hansberg, 2023). Our evidence indicates that migrants place a large weight on climate distance—effectively 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.

[7] By shedding light on this driver of migration 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.

Our paper also speaks to recent work that has used hedonic approaches to 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, finding a positive correlation between selected features in the country of origin and city of destination—from safety to coastal proximity to the 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 and migration using mean annual temperature, a straightforward and widely applicable measure. This elasticity is easy to track and incorporate 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, climate statistics, time periods, and geographic areas. Third, we explore the channels behind the relationship between climate similarity and migration. Fourth, we assess the implications of climate matching in migration, documenting that climate mismatch has a large, negative effect on mortality for migrants and their offspring.

By showing that climate mismatch reduces life expectancy and increases infant mortality, we complement the large literature on the mortality impacts of climate change (Deschênes and Greenstone, 2011; Carleton and Hsiang, 2016; Heal and Park, 2016; Deryugina and Hsiang, 2017). Focusing on a very different setting, we also find that temperature plays an important role in influencing health outcomes. Our results resonate with the U-shaped mortality response to hot and cold shocks in the literature (Barreca et al., 2015; Heutel et al., 2021; Carleton et al., 2022; Bressler et al., 2024): we find that *relative* temperature matters, not just the absolute level of hot or cold. The notion of relative climate might help explain the large heterogeneity in climate-mortality impacts, including why mortality from exposure to extreme heat is lower, on average, in hot places than in cold places (Barreca et al., 2015; Heutel et al., 2021), as well as the related intuition behind estimating climate adaptation potential (Carleton et al., 2022).

Finally, our findings on the impact of climate mismatch may have important implications. Consider that there are 281 million international migrants, and upwards of one billion internal migrants.[8] In the US, nearly one-third of Americans have moved across state lines in their lifetime (Molloy et al., 2011). Effectively, all migrants have experienced a form of "climate change." Our paper shows that the climate distance of a move can affect individual welfare

[8] International migration from UN IOM, World Migration Report 2022 (https://worldmigrationreport.iom.int/wmr-2022-interactive); global counts of internal movers are harder to come by, but a UN study estimates that as of 2005 there were 763 million people living within their own country but outside their region of birth (Bell and Charles-Edwards, 2013).

and, potentially, the migrant-receiving region's capacity to handle climate shocks (and future adaptation). In addition, by linking climate mismatch to mortality—not only for migrants themselves but also for their offspring via infant mortality—the paper expands the scope of research on the inter-generational effects of climate change.

2 Data and Motivating Evidence

2.1 Data Sources

Throughout the paper, we use data from a variety of sources. In this section, we briefly summarize the main datasets, which are also reported in Table A.1.

Climate. We use yearly temperature and precipitation averages as our main proxy for a location's climate. We obtain measures of climate for US counties by taking annual averages from the NOAA Monthly US Climate Divisional Database (Vose et al., 2014). Throughout the analysis, unless otherwise noted, we use a sufficiently long period from 1960 to 2000 to capture the climatological average of a location. However, we verify that all results are robust to using alternative time windows that are more closely aligned with a given migration sample's time period. When estimating regressions that exploit changes in climate over the long term from the early 20th century to the early 21st century, we measure historical (circa 1900) and modern (circa 2020) climate using 1901-1930 and 1991-2020, respectively.

The main data source used to measure climate outside the US is a global monthly gridded data product from TerraClimate (Abatzoglou et al., 2018), a commonly used dataset in the climate literature. As for US counties, we take the average of yearly temperature and precipitation between 1960 and 2000 (TerraClimate starts in 1958). We then take the spatial climate average in grid cells within a 5km radius of each country's capital city. For robustness, we also replicate historical analyses using the CRU (Climatic Research Unit, University of East Anglia, Harris et al., 2020) gridded climate dataset, which goes back to 1901. This data product is interpolated from weather observations without any underlying atmospheric or hydrological model, which can create spatial or temporal data gaps reflecting concentrations of weather stations. Nevertheless, when examining the change in climate distances over the 20th century for both the US and international locations, we use the CRU product to measure historical (circa 1900) and modern (circa 2020) climate normals using 1901-1930 and 1991-2020, respectively.

The variation in climate distance used in some of our analyses can be visualized in Figure 2.[9]

[9] Additional maps showing baseline climate and climate distances are contained in Figure 4, as well as in Appendix Figures C.2, E.2, and E.3.

Migration. To measure Norwegian migration to the US, we create a linked dataset of Norwegian immigrants that reports their municipality of origin in 1865 and their US county of residence in 1880, using full count censuses made available by IPUMS (Ruggles et al., 2021; The Digital Archive, 2008) and automated linking algorithms (Abramitzky et al., 2020, 2021). Using the same methodology, we replicate the analysis linking censuses over alternative years (e.g., 1865-1900; 1900-1910; 1900-1920). More details on the linking algorithms and the resulting samples are presented in Appendix C.2.

Beyond the Norwegian experience, we measure international immigration to the US from all countries. We use US full count censuses from 1880 to 1920, as well as micro-samples from the US Census between 1970 and 2000 and from the American Community Survey (ACS) between 2005 and 2010 to capture international immigration patterns over time.[10] We follow Burchardi et al. (2019), who harmonize countries and counties to their 1990 boundaries, ensuring consistency across different datasets and time periods.[11]

Finally, we assemble data on US internal migration—both historical and modern. We measure historical internal migration in the US using linked samples from 1850 to 1940 assembled by Abramitzky et al. (2021) and available through the Census Linking Project. We obtain the harmonized addresses of more than 90% Americans using the Census Place Project (Berkes et al., 2023). We discuss the details of the US internal migration linked samples in Appendix E.2.[12]

For modern US internal migration data, we use the IRS Change-of-Address Tables dataset, which provides migration data based on year-to-year address changes reported on individual income tax returns.[13] This dataset includes county-level inflows and outflows of residents and is available for filing years 1991 through 2022, but we focus on migration flows measured after 2011, since data collection methods changed in that year. To avoid any distortions created by the COVID-19 pandemic, we further exclude post-2019 observations (though results are robust to including 2020 and 2021).

Mortality. We consider mortality data for both immigrants and their US-born children. First, we rely on the Berkeley Unified Numident Mortality Database (BUNMD)—a publicly available dataset containing over 49 million individual-level mortality records from 1988 to 2005—to

[10] While our main analysis with historical data spans the 1880-1920 period, we also use the 1870, 1930, and 1940 US full count censuses when assessing the robustness of our results. The 1890 full count census was destroyed in a fire and is thus unavailable. For all census years except 1880, we consider immigrants who arrived in the US in the previous decade. Since no information on year of arrival or time spent in the US is provided in the 1880 Census, we consider all immigrants for this census year.

[11] Since we have to match the migration data with many other datasets, for convenience, we apply the procedure from Burchardi et al. (2019) to map county boundaries to their 2010 definition.

[12] In a robustness exercise, we also derive the number of internal migrants using information from the 1940 US full count census, which records the county of residence in both 1940 and 1935.

[13] See Statistics of Income (SOI) Division Tax Stats—Migration Data, https://www.irs.gov/statistics/soi-tax-stats-migration-data.

obtain detailed information on place of birth and death and age at death. Throughout the paper, we restrict the sample to migrants who arrived in the US after 1970, and who died after age 65.[14] Second, we rely on the CDC's National Center for Health Statistics (NCHS) Birth Records—a dataset that provides comprehensive demographic and health data for births in the United States based on information from birth certificates between 2005 and 2021. This dataset is part of the National Vital Statistics System (NVSS) and includes key metrics such as birthweight, gestation weeks, and infant death.

2.2 Motivating Evidence

In this section, we present suggestive evidence on the relationship between temperature at origin and temperature at destination, which motivates the formal analysis conducted in the rest of the paper. Figure 1, already presented in the introduction, documents that immigrants coming from warmer (colder) countries were living in similarly warmer (colder) parts of the US. In this section, we briefly describe the wide variety of settings where this correlation holds.[15]

Cross-country evidence from full count US censuses. Figure 1, Panel A, presents a scatterplot for the relationship between the average temperature in the country of origin and the average temperature in the US county of residence for immigrants living in the US in 1880. In Figure B.1, we use full count censuses to show that the positive relationship between origin-destination temperature holds in other years as well (1900, 1920, and 1940). In Figure B.2, we document that results also hold when controlling for factors that might correlate with both temperature distance and migration: origin country GDP per capita (Panel B), geographic distance from the origin country (Panel C), and number of immigrants from a given country living in a given US county in 1870 (Panel D).

Evidence from US death certificates. Census data records where a person is living at a point in time. However, this may not correspond with where they permanently settle. In Panel B of Figure 1, we thus use death records of immigrants who arrived in the US after 1970 and died between 1988 and 2005. The figure shows a strong, positive relationship between the temperature in the country of birth and the temperature in the US county of death. It also indicates that the relationship between climate at origin and climate at destination holds for the modern era, when the composition of immigrants moving to the US differs starkly from that prevailing in the early 20th century (Abramitzky and Boustan, 2017). Figure B.3 uses

[14] For robustness, in Section 2.2, we also rely on mortality records for years 1959-1961, obtained from National Center for Health Statistics (2021). See Table A.1 for more details.

[15] Both in the headline Figure 1 and in the additional figures presented in this section, we restrict the sample to immigrant groups that accounted for at least 0.1% of the US foreign-born population in the corresponding year. Results are robust to dropping this restriction.

mortality records for years 1959-1961, and documents that the patterns in Panel B of Figure 1 also hold for this earlier period, when most immigrants came from Europe.

***Within* country-of-origin climate variation.** The relationship between origin and destination climates continues to hold *within* countries. First, we use a linked dataset of Norwegian immigrants that reports their municipality of origin in 1865 and their US county of residence in 1880. In Figure B.4, Panel A, we show that the temperature in the Norwegian municipality of residence in 1865 is strongly correlated with the temperature in the US county of residence in 1880. That is, Norwegians living in colder parts of Norway systematically moved to colder US destinations. Next, we consider German immigrants in the US during the late 19$^{\text{th}}$ century, when they accounted for 30% of the US foreign-born population. In Appendix J.1, we detail our surname-level matching procedure to calculate the average temperature associated with the same German surname in Germany and the US, respectively. Figure B.4, Panel B, shows that German surnames more prevalent in colder (warmer) parts of Germany were also more likely to appear in colder (warmer) parts of the US.

US internal migration. Finally, we document that climate matching also holds for internal migrants in the US. In Figure B.5, we use a linked sample of men moving across US counties in each decade from 1850 to 1940, and show that people originating from colder counties systematically moved to colder destinations. This relationship is robust to controlling for geographic distance (Panel B).[16]

3 Part I: Documenting Climate Matching in Migration

In this section, we formally examine the correlations between climate distance and migration presented above. We focus on three different settings: (1) a linked sample of Norwegian immigrants to the US between 1865 and 1880; (2) repeated cross-sections of international immigrants to the US from all countries, both historically (1880-1920) and in the modern era (1970-2010); and (3) US internal migrants, using linked samples that follow the same individual in each decade from 1850 to 1940, and relying on IRS county-to-county migration tables from 2011 to 2019. Following the trade literature, we estimate gravity models (Anderson and Van Wincoop, 2003), and express migration flows as a function of the climate distance between origins and destinations.

[16] Concerns about the spatial correlation of climate in the context of internal migration are addressed in detail in Section 3.

3.1 Empirical Approach

3.1.1 Norwegian Immigration to the US

We begin by focusing on Norwegian immigrants to the US. This setting offers at least two advantages. First, the availability of historical censuses in both countries allows us to link the same individual over time and measure the origin climate at a highly granular level (the municipality). Second, this represents a large migration episode: more than 800,000 Norwegians—or one-third of Norway's 1900 population—moved to the US between the 1850s and 1920 (Ulvestad, 2010). Our main analysis links individuals across the 1865 Norwegian and the 1880 US Censuses—a period that coincides with the first wave of the Norwegian mass migration to the US—but results hold when considering alternative census years.

We collapse the data at the Norwegian municipality of origin (o) by US county of destination (d) level. We restrict attention to US counties that, as of 1880, had at least one European immigrant, to proxy for the set of destinations available to Norwegians.[17] The climate variation we use can be visualized in Figure C.2.

Since the migration matrix is sparse, we follow the literature (Silva and Tenreyro, 2006) and use Poisson Pseudo Maximum Likelihood (PPML). We estimate:

$$M_{od} = \exp[\alpha_o + \gamma_d + \beta \text{Dist}_{od}^{\text{Climate}}]\epsilon_{od} \tag{1}$$

where M_{od} is the number of immigrants from Norwegian municipality o to US county d between 1865 and 1880; $Dist_{od}^{Climate}$ is a vector containing the absolute value of the difference in the climate between origin o and destination d. Temperature distance is our primary climate measure, but we also control for precipitation distance. α_o and γ_d are Norwegian municipality and US county fixed effects. We cluster standard errors at the Norwegian province by US state-level.

Additional details on the Norwegian migration analysis are contained in Appendix C, where we describe the historical context (Appendix C.1), provide details on the linking procedure (Appendix C.2), present additional exercises and robustness checks (Appendix C.3), and report all the relevant figures and tables (Appendix C.4). In particular, we replicate the analysis using varying sets of geographic fixed effects, including additional geographic controls (e.g., elevation and terrain matching) and alternate samples based on different census years and linking methods.

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

3.1.2 International Immigration to the US

Next, we replicate the analysis considering all immigrants in the US during the two main eras of immigration in American history (Abramitzky and Boustan, 2017). First, we use repeated cross-sections from full count censuses from 1880 to 1920. Second, we rely on micro-samples from the US Census and the ACS to measure migration from 1970 to 2010. For each period (1880-1920; 1970-2010), we stack the data at the country of origin by US county of residence by decade-level. The climate variation we use can be visualized for four major immigrant-sending countries in Figure 2.

Using PPML, we estimate:

$$M_{odt} = \exp[\alpha_{ot} + \gamma_{dt} + \beta \text{Dist}_{od}^{\text{Climate}}]\epsilon_{odt} \tag{2}$$

where M_{odt} is the number of immigrants from country o living in US county d in decade t. Similar to equation (1), $Dist_{od}^{Climate}$ includes the absolute temperature and precipitation differences between o and d, and α_{ot} and γ_{dt} are country of origin by decade and US county of destination by decade fixed effects. We cluster standard errors at the US state by country of origin by decade-level.

One caveat to this analysis is that, since we lack full count censuses for most countries, we measure the origin climate in the capital city of the country.[18] Despite the coarser definition of origin climate, this analysis complements the one performed with spatially granular Norwegian data in three ways. First, it allows us to test the hypothesis of climate matching in migration more broadly, both across space and over time. Second, it does not rely on linked data, which may be subject to potential limitations (Bailey et al., 2020). Third, it leverages substantial variation across time and climate zones, relative to the Norwegians who were clustered geographically within the US and came from a single, relatively small country.

Additional details on the international migration analysis are contained in Appendix D, where we describe the historical context (Appendix D.1), present additional exercises and robustness checks (Appendix D.2), and report all the relevant figures and tables (Appendix D.3). In particular, we replicate the analysis controlling for varying sets of geographic fixed effects and controls, using different samples of years and countries, and measuring climate at origin in different ways.

[18] Results are robust to using population-weighed climate measures (see Appendix D.2).

3.1.3 US Internal Migration

Next, we analyze the internal movements of individuals within the US. While this setting presents identification concerns around the spatial correlation of climate, which we address to the best of our ability, it provides advantages in terms of the large sample size, long time frame, and large geographic scope. Similar to the international immigration context, we consider both historical and modern periods. First, we use the linked sample of individuals who moved across counties between 1850 and 1940 assembled by Abramitzky et al. (2020), which coincided with the westward expansion of the US population (Bazzi et al., 2020; Zimran, 2023). Second, we rely on IRS migration tables that count the number of people moving from each origin county to each destination county in each year between 2011 and 2019. The climate variation we use can be visualized for two example locations within the US in Figure E.3.

We stack the data at the county-pair by decade-level, from 1850-1860 to 1930-1940, and at the county-pair by calendar year-level, from 2011 to 2019. We derive the number of migrants between any origin and destination county in each period (the decade or the year). Then using PPML, we estimate:

$$M_{odt} = \exp[\alpha_{ot} + \gamma_{dt} + \beta \text{Dist}_{od}^{\text{Climate}} + \theta \text{Dist}_{od}^{\text{Physical}}]\epsilon_{odt} \quad (3)$$

where migration flows from county o to county d are measured between decade $t-10$ and t (or, between calendar year t and calendar year $t-1$), climate distances are defined as before, $Dist_{od}^{Physical}$ is the physical distance between counties o and d, and α_{ot} and γ_{dt} are county of origin by period and county of destination by period fixed effects. We cluster standard errors at the state of origin by state of destination by period-level.[19]

Additional details on the internal migration analysis are contained in Appendix E, where we describe the historical context (Appendix E.1), provide details on the linking procedure and the other samples used (Appendix E.2), present additional exercises and robustness checks (Appendix E.3), and report all the relevant figures and tables (Appendix E.4). In particular, we replicate our analysis including different sets of geographic fixed effects, using alternative samples (including the 1940 US full count census, which allows us to address concerns related to linking procedures, Bailey et al., 2020), measuring of climate in different ways (e.g., max and min temperature as opposed to mean), and controlling for seasonality. We are especially careful to deal with the issue of spatial correlation of climate in the historical internal migration context: we replicate the analysis including more flexible controls for geographic distance, considering each movement direction separately, measuring geographic distance in a time-varying way to

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

account for the spread of the railroad network, and controlling for a large set of county-pair variables—such as differences in geographic features, soil type, frontier exposure, and other economic, social, and demographic characteristics.

3.2 Consolidated Results

Table 1 summarizes our main results. Column 1 estimates equation (1) using the linked dataset of Norwegian immigrants between 1865 and 1880. Columns 2 and 3 estimate international migration to the US using equation (2) for the historical (1880-1920) and modern (1970-2010) periods. Columns 4 and 5 estimate the internal migration regressions from equation (3), for the historical (1850-1940) and modern (2011-2019) periods.

In all cases, results confirm the correlations presented in Section 2.2: reducing climate distance increases migration flows between each origin and destination. While the negative relationship between climate distance and migration holds across geographies and time periods, the implied magnitudes vary somewhat depending on the specific context. In the Norwegian case (column 1), reducing climate distance by the sample mean (4.7°C) increases migration by 0.4%. The implied magnitudes for historical international migration (column 2) as well as for historical and modern internal migration (columns 4 and 5) are similar: lowering climate distance by the corresponding sample mean (10°C, 5°C, and 5.1°C) increases migration by 1.4%, 1.3%, 1.1%, respectively. Results for international modern migration (column 3) are an order of magnitude smaller in absolute value. In this case, reducing climate distance by 10°C (the sample mean) increases migration by 0.3%. Details about additional exercises, robustness checks, and results specific to each setting are presented in Appendix C, D, and E.

One may wonder the extent to which climate matching varies by baseline climate. For example, migrants from mild places may seek out similar climates, but people from hot or cold places may not (i.e., they are happy to move to a milder climate). Figure 3 examines the heterogeneity of results depending on the temperature at origin for international (Panels A and B) and internal US (Panels C and D) migration. We place locations into three buckets based on their mean annual temperature: cold (5-10°C), mild (10-15°C), and warm (15+°C). The distribution of countries across temperature buckets can be visualized in Figure 1, which reports cold ones (e.g., Norway, Germany, and Poland), mild ones (e.g., France, China, and UK), and warm ones (e.g., Italy, Greece, and Mexico). Interestingly, coefficients in Figure 3 are rather similar across temperature bins—implying that climate matching occurs regardless of a migrant's baseline climate. In the case of the international historical sample (Panel A), estimates become smaller and no longer statistically significant for the hottest origin countries, perhaps reflecting the relatively small share of migrants to the US during the historical time period (1880-1920).

3.3 Long-Run Changes in Climate Distance and Migration

Thus far, we have exploited cross-sectional variation in bilateral climate distance computed from the average climatology of each location over a sufficiently long period (1960-2000). Despite the robustness tests and the consistency of the results across diverse settings (see Appendix C, D, and E), one may still be worried that our estimates reflect the influence of other origin-destination specific factors also correlated with climate distance. In this section, we tackle this concern by leveraging variation that arises from *changes* in average climate occurring differentially across space. Specifically, we test whether, over the 20th century, the *change* in climate distance between countries of origin and US counties of destination predicts the *change* in international migration patterns. We also repeat this in the context of internal migration.[20]

We focus on two periods: a historical window in the early 20th century and a modern window in the early 21st century.[21] We visually illustrate the change in climate distance for two selected US counties relative to each other US county (Figure 4): one in the Delta region in the South (Panel A), and the other in the Northeast (Panel B). The map makes it clear that changes in climate distances vary substantially across the US without systematic geographic patterns. Similar trends, not reported for brevity, are evident for international climate distances.

We replicate our baseline international and internal migration gravity models estimated with equations (2) and (3), respectively, allowing climate distances to vary between the early and the modern periods. In addition to the set of controls and fixed effects included, we add origin-by-destination fixed effects, which absorb any pair-specific (time-invariant) variable. To allow the effects of the *change* in climate distance on migration to vary depending on geographic distance, we further control for the interaction between geographic distance and period fixed effects.

Formally, we estimate:

$$M_{od\tau} = \exp\left[\alpha_{o\tau} + \gamma_{d\tau} + \gamma_{od} + \beta \text{Dist}^{Climate}_{od\tau} + \theta_\tau \text{Dist}^{Physical}_{od}\right] \varepsilon_{od\tau} \tag{4}$$

where everything is as before, but τ refers to the period (historical and modern), γ_{od} are origin-by-destination fixed effects, and $\theta_\tau \text{Dist}^{Physical}_{od}$ are period dummies interacted with geographic distance.

Consolidated results are presented in Table 2. In columns 1 and 2, we examine international migration, considering different time windows to measure migration flows (1900 and 2010;

[20] While since 1980, anthropogenic emissions have driven the overall warming of the climate, such spatiotemporal variation is a normal part of the climate system, influenced by multi-decadal oceanic circulation patterns (e.g., North Atlantic Oscillation), solar variability, volcanic activity, and land use change.

[21] The climate over the historical and modern periods are measured over the 1901-1930 and 1991-2020, respectively.

average over 1900-1920 and over 1990-2010, respectively). In columns 3 and 4, we focus on internal migration, using the 1900-1910 linked sample (resp., the 1900 to 1940 averages from the linked sample) and the 2018-2019 IRS migration data (resp., the average over the 2011-2019 period). We find that a 1°C increase in temperature distance over time—the average annual temperature distance between Boston and Detroit—reduces migration by about 0.3%. To put these numbers in perspective, consider that global temperatures have risen by about 1°C since the late 19th century due to anthropogenic climate change.

To illustrate our long-run variation in another way, we replicate this analysis by taking variables in long differences, as in Burke and Emerick (2016). Besides its intuitive appeal, this approach 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 20th century for both internal and international migrants (Molloy et al., 2011; Boustan et al., 2013; Abramitzky and Boustan, 2017). Results in Table F.2 are qualitatively similar to those in Table 2, though the coefficient becomes smaller (in absolute value) for internal migration. Additional details related to the long-run analysis are included in Appendix F.

3.4 Mechanisms Behind Climate Matching

What explains climate matching in migration? We consider two complementary mechanisms: climate-specific human capital and climate-as-amenity. In Figure 5, we replicate the analysis presented in Table 1, separately for each decade, or migration period.[22] Across the board, patterns are similar: the elasticity of migration with respect to climate distance declines over time—with the decline being more pronounced for international, than for internal, migration. In other words, while remaining important, climate matching becomes a less significant factor in migration decisions over time.

In terms of drivers, the persistence of ethnic enclaves (Card, 2001; Munshi and Rosenzweig, 2016)—whereby the first pioneers match on climate and subsequent immigrant waves follow their networks—is unlikely to fully drive these results. First, as explained in more detail in Appendix G.1, Figure 5 also replicates the analysis by controlling for lagged origin-destination migrants. The coefficients (grey triangles) remain almost identical to those from the baseline specification (black dots). Second, in Appendix G.2, we exploit variation from the expansion of the US frontier (Turner, 2017; Bazzi et al., 2020) to show that climate matching also holds

[22] For modern internal migration, we plot coefficients obtained when focusing on the calendar year 2018-2019, but results are similar when using other years (or, the average over the entire 2011-2019 period).

in contexts preceding the establishment of migrant networks that could act as a pull factor.

An alternative interpretation for results in Figure 5 is that migrants value climate similarity—at least in part—due to the importance of climate-specific skills. Between 1850 and 1940, the share of employment in agriculture in the US dropped from 55% to 17% (Lebergott, 1966). Since climate-specific human capital—such as farming expertise *vis-a-vis* crop suitability and pest management—should matter more for farmers, this trend is consistent with climate-specific skills being an important driver of our results.[23] We provide further evidence for the relevance of this channel in Figure 6 and Figure G.1, where we explore the heterogeneity of results by occupation among internal movers. Our estimates show that for both international and internal migration, the effects of temperature distance are larger in agriculture than in other sectors and in outdoor and climate-intensive occupations. More details are included in Appendix G.3.

At the same time, Figure 6 and Figure G.1 reveal that temperature distance reduces migration even for individuals working in occupations where climate-specific human capital is unlikely to matter, such as services. Moreover, even in modern times, climate distance continues to have a negative effect on migration (Figure 5)—which is particularly pronounced for US internal migrants (Panel C). Together, these results suggest that climate similarity also acts as an amenity driving migration decisions. For instance, if 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.[24]

It is important to note that the evidence presented here should be interpreted as suggestive, and that we were unable to tease out the relative importance of the climate-specific skills vs. the climate-as-amenity channel. Yet, our results indicate that both mechanisms are likely to be at play. We also suspect that the two forces are complementary and reinforce each other. 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.

[23] Notably, Michalopoulos (2012) finds evidence that location-specific human capital is a factor explaining the formation and maintenance of distinct ethnic groups in Africa.

[24] This idea resonates with findings in 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.

4 Part II: Mortality Impacts of Climate Mismatch

Even though migrants attempt to match on climate, they are unlikely to fully eliminate the difference in temperature between their origin and their destination. One notable example, already mentioned in the introduction, is that of Somali immigrants: no US county is sufficiently hot to match the average temperature in Somalia. More generally, economic pull factors and the persistence of ethnic enclaves that might reflect historical events unrelated to climate distance, such as changes in refugee resettlement or immigration policies, may continue to shape migrants' location choice over time. Continuing with the previous example, a large share of the Somali community in the US lives in Minneapolis—the coldest large city in the continental US.

In this section, we examine the implications of climate matching for immigrants' well-being using an instrumental variable approach. First, we provide novel evidence of a negative relationship between climate mismatch and the life expectancy of migrants (Section 4.1). Next, we show that climate mismatch also negatively affects health outcomes of infants born in the US to immigrant mothers (Section 4.2). We conclude by combining our findings with the EPA's value of statistical life to derive an estimate of the value of climate (Section 4.3).

4.1 Climate Mismatch and Adult Mortality

Preliminary evidence. We rely on US mortality data used in Figure 1, Panel B, where we documented a positive relationship between an immigrant's temperature in the country of birth and the temperature in the US county of death. We now test whether climate mismatch among immigrants in the US is associated with differences in lifespan. We consider the 730,000 foreign-born individuals who arrived after 1970 and died in the US between 1988 and 2005 at age 65 or older. Summary statistics by country of origin and climate and demographic characteristics can be viewed in Panel A of Tables H.1 and H.2.

As in the rest of this analysis, we define climate distance as the absolute value of the difference between the temperature in the capital city of the country of birth and the US destination, which in this case is the county of death. We begin by plotting simple correlations from the raw data, aggregated to the country-level in Figure H.1. Results suggest that age at death is lower among immigrant groups for which climate distance is higher.

OLS regressions. Next, we turn to the formal regression analysis. We restrict our main analysis to non-European immigrants, which, as explained below, is required to derive an instrument for climate mismatch at the country of origin level. We estimate individual-level regressions of the form:

$$y_{ibod} = \alpha_{c(o)b} + \alpha_{db} + \beta_1 \text{Dist}_{od}^{Climate} + X'\gamma + \varepsilon_{ibod} \tag{5}$$

where y_{iodb} is age at death (in months) of individual i, belonging to birth cohort b, who was born in country o and died in US county d. As in Section 3, the main regressor of interest is climate distance, $Dist_{od}^{Climate}$, defined as the absolute value of temperature difference between o and d, but regressions also separately account for precipitation distance. We also include a vector of controls X (geographic distance between o and d and individual characteristics).[25] In our preferred specification, we also include fixed effects for birth cohort by continent of origin ($\alpha_{c(o)b}$) and birth cohort by US county of death (α_{db}). Standard errors are clustered at the country of origin by US state of death-level.

We present OLS results from equation (5) in Table 3, Panel A. In column 1, besides including climate and geographic distance as well as individual characteristics, we also control for continent of origin, US county of death, and birth cohort (grouped into 10-year periods) fixed effects. This set of controls isolates the relationship between climate distance and mortality, holding constant average age at death at the continent of origin level, at the US county-level, and at the birth cohort-level (all of which could be tied to local economic conditions), and accounting for individual-level demographic characteristics. The estimates confirm the patterns of Figure H.1, and indicate that climate distance is negatively correlated with age at death. That is, migrants who settled in US counties with a climate similar to that of their home country tend to live longer.

Column 2 further controls for the total number of individuals from the same country of origin who died between 1988 and 2005 in the individual's county of death—a proxy for immigrant network effects. Columns 3 and 4 document that this relationship holds after controlling for more stringent sets of fixed effects: US county-by-birth cohort and continent of origin-by-birth cohort. In Table H.3, Panel A, we verify that results hold when replacing continent fixed effects with country of origin fixed effects (columns 1 and 2), when interacting birth cohort fixed effects with country and county fixed effects (columns 3 and 4), and even when replacing county by cohort fixed effects with zip-code by cohort fixed effects. Regarding sample selection, results are robust to considering non-European arrivals post 1960 rather than 1970 (Panel B), and to dropping all restrictions and using the full sample of European and non-European migrants (Panel C).

We also analyze heterogeneity by direction of climate distance, asking whether the increase in

[25] Individual controls include year of migration (proxied for by the year in which the individual applied to Social Security), a dummy for being male, and a dummy for dying in a state other than that where the person applied for Social Security.

mortality occurs among people moving both to hotter and colder places. Figure 7 shows the OLS results by binned temperature distance (non-absolute value). While less precise, results suggest a symmetric mortality response across cold and hot moves, although the effect is amplified among people moving to much hotter destinations.

Instrumental variable analysis. Even with this rich set of controls, country-to-county-specific unobservable factors might influence the degree of climate matching for migrants. To assuage this concern, we implement an instrumental variable approach adapted from Burchardi et al. (2019) and Terry et al. (2024), effectively randomizing climate distance assigned to immigrants from each origin country. We present the formal derivation of the instrument in Appendix H.1, and only discuss the general intuition here.

We predict the number of (non-European) immigrants from a given origin country to a given US county by combining *i*) the number of immigrants who arrived in the US from other countries within the same continent, a proxy for push shocks from origin, and *ii*) the share of European immigrants who settled in each US county, a proxy for pull shocks to destination. We iterate this procedure over 100 years from 1880 to 1970 to isolate quasi-random variation in the distribution of non-European ancestry across counties in 1970. Second, we interact the observed national-level flows between 1970 and 1980 to the US from each non-European country of origin, with the predicted ancestry obtained in the first step, omitting individuals from that specific country who settled in the US Census division of a given county. This allows us to get the predicted number of immigrants who arrived after 1970 from each non-European origin in each US county. Third, we calculate the country-specific average climate distance, using as weights the predicted share of immigrants from a given country in each US county, relative to the predicted number of immigrants from that country in the US.[26]

Table 3, Panel B, replicates the OLS regressions reported in Panel A, instrumenting climate distance with the measure just described.[27] Since the instrument is effectively defined at the country-level, we cannot include country of origin fixed effects. For this reason, we use continent fixed effects, as in the corresponding OLS specification. As before, climate mismatch decreases the age at death of migrants. According to our preferred specification, every additional degree Celsius of climate mismatch is associated with death occurring 7 days earlier (relative to a population where the average age at death is 80 years). The similarity of OLS and 2SLS coefficients and their stability across specifications suggests that selection is unlikely to explain

[26] For illustration, take country of origin A, with an average annual temperature of 15°C. Say there are three counties in the US with average temperatures of 12°C, 15°C, 22°C, respectively, and our instrument predicts flows in 1990 from country A of 10, 20, and 100 people, respectively. Thus, the country A-specific average climate distance in 1990 would be: $|12-15| \cdot \frac{10}{130} + |15-15| \cdot \frac{20}{130} + |22-15| \cdot \frac{100}{130} = 5.6$°C.

[27] Panel C reports first stage results, and documents that the instrument is strong, as confirmed by the F-statistic above 100, displayed in Panel B.

OLS results.[28]

Results are robust to excluding the top five immigrant-sending countries to the US and the top five US state destinations (Figure H.2), calculating the predicted number of immigrants in different ways, and estimating alternative specifications, such as replacing county with zip-code of death fixed effects or using more granular birth-cohort groups (Table H.4).

Discussion. The patterns in Table 3 can be reconciled with different mechanisms, including the climate-specific human capital channel and climate-as-amenity channel discussed in Section 3.4. For one, people moving from a different climate may lack the "indigenous knowledge" or other adaptations to maximize their survival probability in the destination area (Crosby, 2004; Henrich, 2016). Another example entails accidental deaths, such as driving on icy roads for someone unfamiliar with such conditions. Moreover, if climate mismatch reduces returns to climate-specific skills and, thus, earnings, it may lead to a deterioration of health outcomes. Additionally, hotter temperatures may increase crime and violence (Ranson, 2014; Mukherjee and Sanders, 2021), which could, in turn, expose people to premature death—perhaps more so from those coming from colder climates who are less acclimated. Finally, climate distance may lower life expectancy through an amenity channel if mismatched migrants become less happy and less socially connected over time (Holt-Lunstad et al., 2010).

4.2 Climate Mismatch and Infant Mortality

Having documented that climate mismatch reduces life expectancy for first-generation immigrants, we now ask whether the next generation also bears the cost of their parents' climate mismatch. We use US birth records from 2005 to 2021 to study whether a foreign-born non-European mother's climate distance affects infant mortality and other infant health impacts. Summary statistics by mother's country of origin and climate and demographic characteristics can be viewed in Panel B of Table H.1 and Table H.2, respectively. We aggregate the data to the county-country-birth year-level and instrument mothers' climate distance using the same methodology described above. We estimate:

$$imr_{odb} = \alpha_{c(o)b} + \alpha_{db} + \beta_1 \text{Dist}_{od}^{Climate} + X'\gamma + \varepsilon_{ibod} \tag{6}$$

where imr_{odb} is the infant mortality rate for children born in US county d in year b from immigrant mothers born in country o. As before, the main regressor of interest is the absolute value of temperature difference between o and d, $Dist_{od}^{Climate}$. Regressions also include precipitation

[28] If anything, the 2SLS coefficients are larger (in absolute value) than that of OLS. This suggests that, at least in this context, selection may lead to an underestimation of the mortality impact of climate mismatch.

distance and a vector of controls, X (mother-level characteristics aggregated to the o-d level and geographic distance between o and d).[29] In the preferred specification, described below, we also include fixed effects for birth cohort by continent of origin ($\alpha_{c(o)b}$) and birth cohort by US county of birth (α_{db}), and control for the predicted number of immigrants in 2000. Standard errors are clustered at the country of origin by US state of birth by year-level.

We present 2SLS results in Table 4, Panel A. In column 1, we control for continent of the mother, county of birth, and year of birth fixed effects as well as for a vector of mothers' controls (age, marital status, and number of live births had until that point). In column 2, we further control for the predicted number of immigrants from the same origin of the mother in the same county of birth of the child, in 2000. In columns 3 and 4, we add interactions between continent and year fixed effects and county and year fixed effects. We take column 4 as our preferred specification.

We find that, among the 11 million births occurring to foreign-born mothers between 2005 and 2021, each additional degree of climate mismatch increases infant mortality by 0.27 deaths per thousand live births—a 6% increase relative to the average infant mortality rate. Coefficients are very stable across specifications. In Panels B and C, we also examine the probability of low birth weight and of premature birth, both of which are predictors of subsequent infant mortality. Consistent with patterns in Panel A, we find that one additional degree of climate mismatch increases the probability of low birth weight (4.5 per thousand live births or 6% relative to the mean) and premature birth (3.5 per thousand live births or 3% relative to the mean). Together, these results suggest that climate distance affects maternal and fetal health—and not just the newborns' post-birth environment.[30]

As for the age at death analysis, results are robust to excluding the top five immigrant-sending countries and US state destinations in this dataset (Figure H.3). They are also robust to calculating the predicted number of immigrants in different ways and estimating alternative specifications, such as using Poisson or unweighted regression (Table H.5).

4.3 Estimating the Cost of Climate Mismatch

We now use the mortality results obtained before to estimate the value of climate similarity. For adult mortality in Section 4.1, the coefficient from our preferred specification (Table 3, Panel B, column 4) implies that 1°C of temperature distance reduces life expectancy by 0.021 years,

[29] Mother controls are age, marital status, as well as the number of live births she had at the time of birth of the child.

[30] These effect sizes are within the range of those found in the literature on infant mortality response to environmental hazards, including extreme heat (Banerjee and Maharaj, 2020), air pollution (Chay and Greenstone, 2003; Currie and Neidell, 2005; Currie et al., 2009), and pesticide exposure (Frank, 2016; Taylor, 2022a).

or approximately 7.5 days of life. We implement an admittedly crude back-of-the-envelope by taking the EPA's value of a statistical life year (VSLY) of $250,000 per year (2020 dollars), which is based on a VSL of $10 million divided by a standardized remaining lifespan of 40 years (Masterman and Viscusi, 2020; Hammitt, 2023). We then multiply this $250,000 by 0.021, i.e., the coefficient from column 4 in Table 3, implying that 1°C of additional climate distance costs migrants $5,250.[31]

We proceed in a similar way using results on infant mortality from Section 4.2. We use our estimates from Table 4, Panel A (column 4), to compute a back-of-the-envelope value of climate similarity. Combining our estimate of 0.27 deaths per 1,000 live births with the $10 million VSL (0.27 / 1,000 x $10,000,000), we find that each additional degree Celsius of climate mismatch costs $2,700. If we combine this infant mortality cost with the adult mortality cost (with the caveat that these estimates are derived from different populations), this would equal $7,950. These are sizable numbers, especially for the poorest migrants whose travels to the US cost thousands of dollars. Illustratively, getting from Central American countries to Texas is estimated to cost between $9,000 and $12,000.[32]

To ground these numbers, we leverage our rich historical census data in order to estimate the costs of climate mismatch derived from a completely different empirical setting: the Homestead Act of 1862. The largest land distribution program in US history, the Homestead Act 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). This policy provides a unique setting to estimate the value of climate for migrants.

In the interest of space, all the details related to the historical context of the Homestead Act, the empirical strategy and results, and the approach to calculating the value of climate similarity are presented in Appendix I. Here, we only summarize the main takeaways. Consistent with our findings in the first part of the paper, migrants minimize the climate distance between origin and destination during the most active years of the Homestead Act (1870-1920). However, we find that, when offered cheap land through the Act, farmers deviate from the baseline—that is, they seek destinations with a more dissimilar climate to their origin. 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. A large deviation from baseline climate preferences when offered Homestead land implies a low valuation of climate similarity. By contrast, a small deviation of migration flows from the

[31] This value can be scaled by one's preferred VSLY: for example, using a life-year value of $100,000, as in Deryugina et al. (2019), the mortality cost is $2,100. A similar number is derived by dividing the VSL of $10 million by the average life expectancy in our sample (80 years). This yields a VSLY of $125,000.

[32] See https://www.telegraph.co.uk/news/african-migrants-in-america/

baseline implies a large valuation of climate similarity. We estimate the marginal valuation of temperature at $4,765 in current dollars per 1°C. For comparison, farmers were willing to pay this amount to avoid the cost of traveling an extra 180 km when migrating.[33] These estimates are broadly in line with those obtained using more recent adult and infant mortality data.

5 Discussion and Conclusions

In this paper, we obtain two main results. First, we provide evidence that individuals tend to settle in places with climates similar to those of their origins. These patterns—which hold across periods, geography, and migrant groups—are not explained by the spatial correlation of climate or other known drivers of migration, including pre-existing migrant networks (Altonji and Card, 1991; Stuart and Taylor, 2021), geographic distance (Schwartz, 1973; Steckel, 1983), or differences in economic opportunities (Borjas, 2001; Cadena and Kovak, 2016) between origins and destinations. Exploring the mechanisms, we document that both climate-specific skills and climate-as-amenity likely explain our results. Second, we implement an instrumental variable strategy to estimate the cost of climate distance for migrants. We find that climate mismatch reduces the life expectancy of first-generation immigrants and increases infant mortality of their US-born children. Through the EPA's value of a statistical life framework, we find that a temperature mismatch of 1°C among adult migrants to be associated with a cost of $5,250 in current dollars due to increased mortality.

We advance and complement different strands of the literature. For urban economics, other papers have examined how local amenities, including a destination's climate, influence where people settle (Glaeser and Tobio, 2007; Albouy et al., 2016). Our paper highlights the importance of *relative* climate as amenity. With relevance to climate change economics, while the estimates obtained here may be specific to the US context and migrants may not be representative of the whole population, we document the novel fact that maintaining a stable climate for individuals leads to mortality gains—both among adults and infants.

For the migration literature, 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), offering a micro-foundation for such initial

[33] This estimate may be an upper bound since homesteaded land was likely of lower quality than nearby alternatives (Mattheis and Raz, 2019).

settlements. This may help address the critiques that this class of instruments has recently received (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022).

Our findings also open the door to fascinating questions in 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-19$^{\text{th}}$ 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 (Bazzi et al., 2023a,b; Alesina and Tabellini, 2024). This margin may be especially important in climate-driven migration, given the well-documented relationship between agricultural practices and cultural norms (Alesina et al., 2013; Cao et al., 2021; Becker, 2024).

Finally, our results are relevant to 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, 2023; Desmet and Rossi-Hansberg, 2023; Bilal and Rossi-Hansberg, 2023), but our findings indicate that migrants place significant 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 warming. We hope that 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.

By extension, our results are relevant to climate refugee policy. On average, over 20 million people annually have been forcibly displaced by weather-related events since 2008 (UNHCR)—a number that is estimated to increase with further global warming. Even among economic migrants, climate change often acts as a push shock. Given the significant welfare implications explored in this paper, our findings suggest that climate similarity should be considered when resettling existing climate refugees and 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 (2024): "The Political Effects of Immigration: Culture or Economics?" *Journal of Economic Literature*, 62, 739–760.

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.

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.

Banerjee, R. and R. Maharaj (2020): "Heat, infant mortality, and adaptation: Evidence from India," *Journal of Development Economics*, 143, 102378.

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.

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. (2024): "On the Economic Origins of Concerns Over Women's Chastity," *Review of Economic Studies*, Forthcoming.

Bell, M. and E. Charles-Edwards (2013): "Cross-national comparisons of internal migration: an update of global patterns and trends," Tech. rep., United Nations, Department of Economic and Social Affairs, Population Division.

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.

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.

Borjas, G. J. (2001): "Does immigration grease the wheels of the labor market?" *Brookings papers on economic activity*, 2001, 69–133.

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

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

Bressler, R. D., A. J. Wilson, C. Ivanovich, C. Tuholske, C. Raymond, R. M. Horton, A. Sobel, P. Kinney, T. Cavazos, and J. Shrader (2024): "Heat Disproportionately Kills Young People: Evidence From Wet-Bulb," *Working Paper*, temperature in Mexico [Draft available upon request].

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.

Burchardi, K. B., T. Chaney, and T. A. Hassan (2019): "Migrants, ancestors, and foreign investments," *The Review of Economic Studies*, 86, 1448–1486.

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

Cadena, B. C. and B. K. Kovak (2016): "Immigrants equilibrate local labor markets: Evidence from the Great Recession," *American Economic Journal: Applied Economics*, 8, 257–290.

Caliendo, L., L. D. Opromolla, F. Parro, and A. Sforza (2021): "Goods and factor market integration: A quantitative assessment of the EU enlargement," *Journal of Political Economy*, 129, 3491–3545.

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.

Chay, K. and M. Greenstone (2003): "Air quality, infant mortality, and the Clean Air Act of 1970," .

Ciesin, I. (2016): "Gridded population of the world, version 4 (GPWv4): Population count," *Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).*

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.

Crosby, A. W. (2004): *Ecological imperialism: the biological expansion of Europe, 900-1900*, Cambridge University Press.

Cruz, J.-L. and E. Rossi-Hansberg (2023): "The Economic Geography of Global Warming," *Review of Economic Studies*, rdad042.

Currie, J. and M. Neidell (2005): "Air pollution and infant health: what can we learn from California's recent experience?" *The quarterly journal of economics*, 120, 1003–1030.

Currie, J., M. Neidell, and J. F. Schmieder (2009): "Air pollution and infant health: Lessons from New Jersey," *Journal of health economics*, 28, 688–703.

De Juan, A., F. Haass, C. Koos, S. Riaz, and T. Tichelbaecker (2021): "War and Nationalism: How WW1 Battle Deaths Fueled Civilians' Support for the Nazi Party," *American Political Science Review*, 118, 144–162.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019): "The mortality and medical costs of air pollution: Evidence from changes in wind direction," *American Economic Review*, 109, 4178–4219.

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

Deschênes, O. and M. Greenstone (2011): "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US," *American Economic Journal: Applied Economics*, 3, 152–185.

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.

Frank, E. (2016): "The Effects of Bat Population Losses on Infant Mortality through Pesticide Use in the US Working Paper," *CDEP*.

Glaeser, E. L. and J. D. Gottlieb (2008): "The economics of place-making policies," *Brookings Papers on Economic Activity*, 2008, 155–239.

Glaeser, E. L. and K. Tobio (2007): *The Rise of the Sunbelt: Edward L. Glaeser, Kristina Tobio*, 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.

Goldstein, J. R., M. Alexander, A. M. G. Casey Breen, F. Menares, M. Osborne, M. Snyder, and U. Yildirim. (2023): "CenSoc Mortality File: Version 3.0." Berkeley: University of California.

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

Hammitt, J. K. (2023): "Consistent valuation of a reduction in mortality risk using values per life, life year, and quality-adjusted life year," *Health economics*, 32, 1964–1981.

Harris, I., T. J. Osborn, P. Jones, and D. Lister (2020): "Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset," *Scientific Data*, 7.

Hatton, T. J. and J. G. Williamson (1998): *The age of mass migration: Causes and economic impact*, Oxford University Press on Demand.

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*.

Hengl, T. (2018a): "Soil bulk density (fine earth) 10 x kg / m-cubic at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.2)," .

——— (2018b): "Soil pH in H2O at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.2)," .

Hengl, T. and S. Gupta (2019): "Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.1)," .

Hengl, T. and I. Wheeler (2018): "Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.2)," .

Henrich, J. (2016): *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*, princeton University press.

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.

Higham, J. (1955): *Strangers in The Land: Patterns of American Nativism, 1860-1925*, Rutgers University Press.

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.

King, D. (2002): *Making Americans: Immigration, race, and the origins of the diverse democracy*, Harvard University Press.

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.

Kupperman, K. O. (1984): "Fear of Hot Climates in the Anglo-American Colonial Experience," *The William and Mary Quarterly*, 41, 213–240.

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.

Lee, E. (2003): *At America's gates: Chinese immigration during the exclusion era, 1882-1943*, Univ of North Carolina Press.

Leonhardt, D. (2023): *Ours was the Shining Future: The Story of the American Dream*, Hachette UK.

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.

Masterman, C. J. and W. K. Viscusi (2020): "Publication selection biases in stated preference estimates of the value of a statistical life," *Journal of Benefit-Cost Analysis*, 11, 357–379.

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

Michalopoulos, S. (2012): "The origins of ethnolinguistic diversity," *American Economic Review*, 102, 1508–1539.

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.

Morten, M. and J. Oliveira (2024): "The effects of roads on trade and migration: Evidence from a planned capital city," *American Economic Journal: Applied Economics*, 16, 389–421.

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.

Obolensky, M., M. Tabellini, and C. Taylor (2024): "Homeward Bound: How Migrants Seek Out Familiar Climates," Tech. rep., National Bureau of Economic Research.

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.

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.

Schwartz, A. (1973): "Interpreting the effect of distance on migration," *Journal of political economy*, 81, 1153–1169.

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. (1983): "The economic foundations of East-West migration during the 19th century," *Explorations in Economic History*, 20, 14–36.

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. (2022a): "Cicadian rhythm: Insecticides, infant health and long-term outcomes," *Center for Environmental Economics and Policy Working Paper*, 1–45.

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

Terry, S. J., T. Chaney, K. B. Burchardi, L. Tarquinio, and T. A. Hassan (2024): "Immigration, innovation, and growth," Tech. rep., National Bureau of Economic Research.

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," .

Tombe, T. and X. Zhu (2019): "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 109, 1843–1872.

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

Ulvestad, M. (2010): *Norwegians in America, their History and Record*, vol. 1, Astri My Astri Publishing.

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.

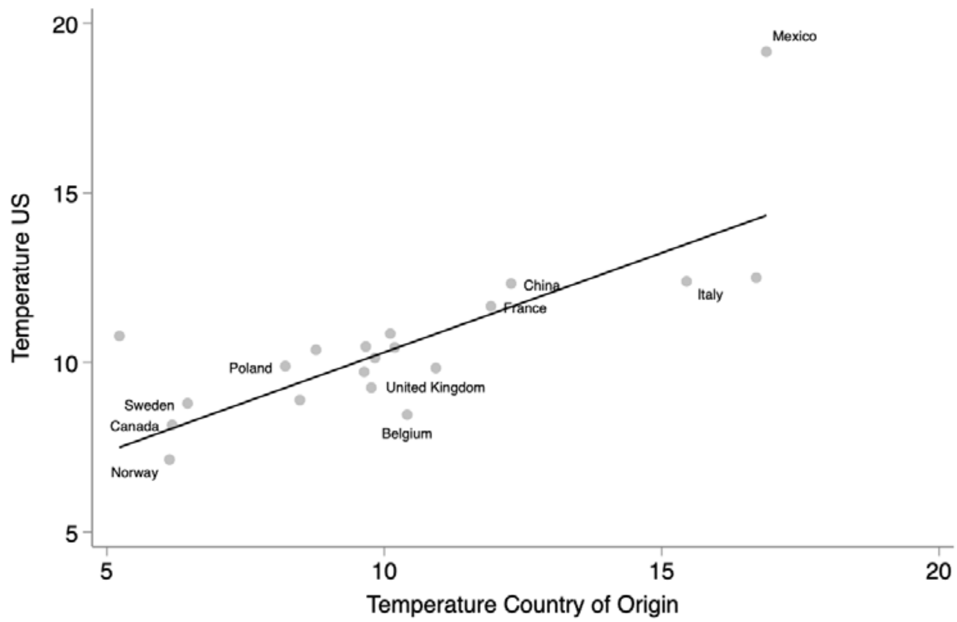
Zedlitz, J. (2016): "Web-based collaborative system for transcription of serial historic sources to structured data," .

Zimran, A. (2023): "Internal Migration in the United States: Rates, Selection, and Destination Choice, 1850-1940," *Journal of the Economic History*, 84, 727–766.

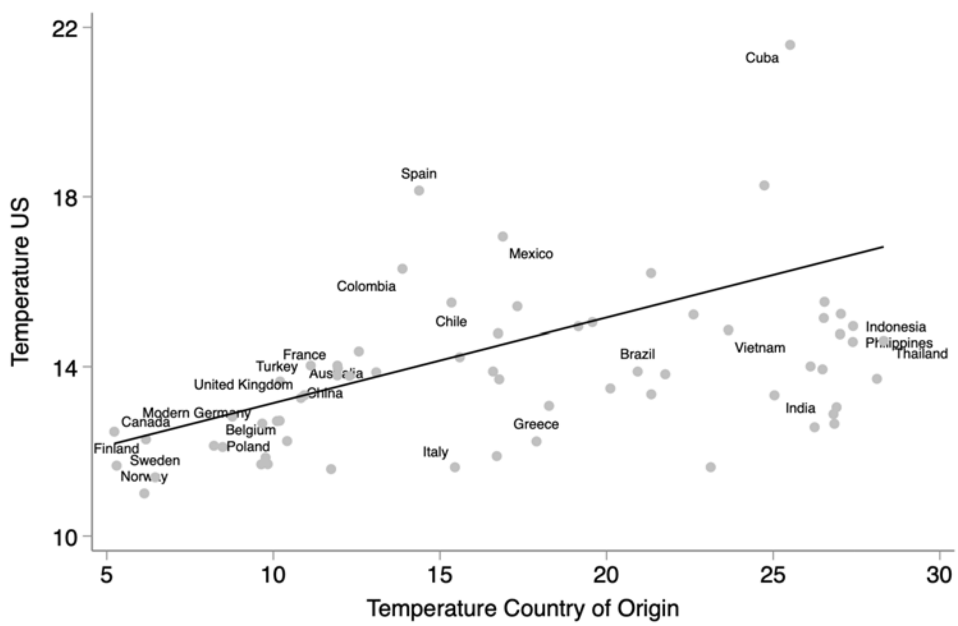
Figures and Tables

Figure 1. Temperature Matching for Immigrants in Historical and Modern Era

(A) Historical census (1880)



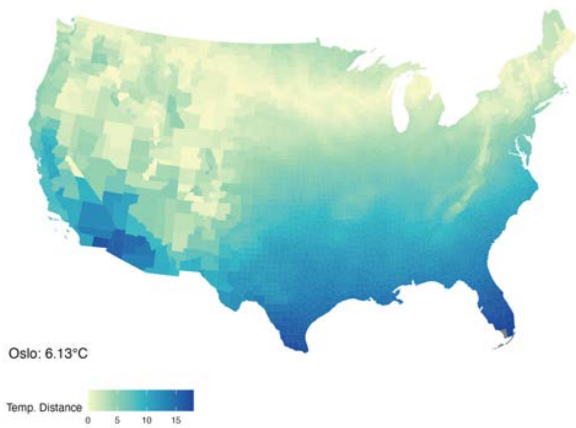
(B) US mortality data (1988-2005)



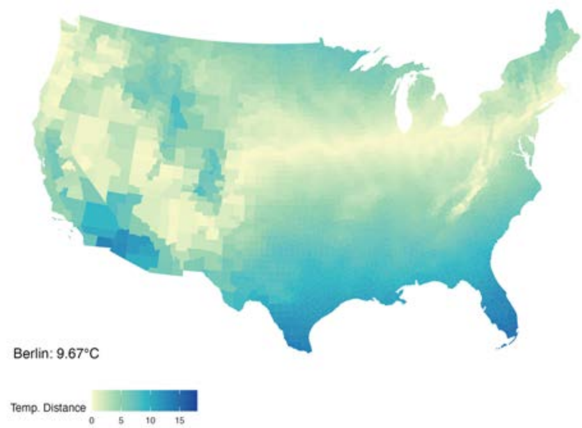
Notes: Figures display the relationship between the average temperature in degrees Celsius in the capital city of an immigrant's country of origin (x-axis) and the average temperature across US counties on the y-axis based on census records of where immigrants from each origin were living in 1880 (Panel A) and BUNMD mortality records of where immigrants died between 1988 and 2005 (Panel B). The latter includes foreign-born individuals who migrated after 1970. Regressions are weighed by the number of individuals from an origin country in the US. Panels only display points for countries comprising at least 0.1% of the foreign-born share at the time. The regression coefficient and the corresponding robust standard errors are, respectively, 0.586(0.140) and 0.201(0.085).

Figure 2. Temperature Distances Between Selected Foreign Cities and US Counties

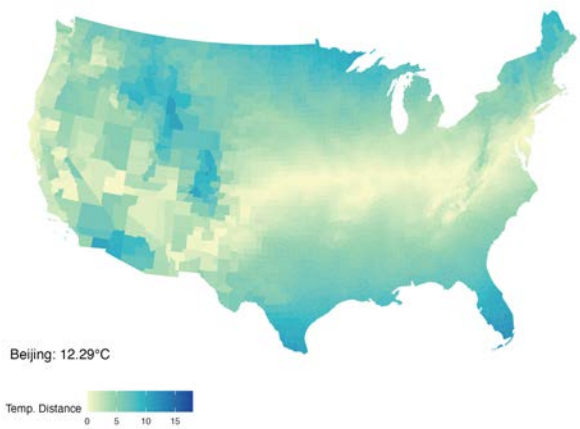
(A) Oslo



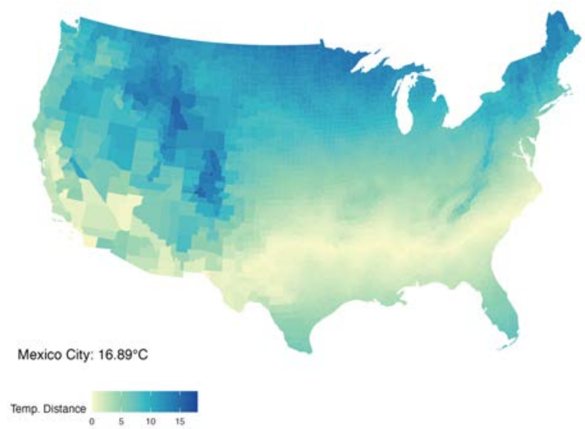
(B) Berlin



(C) Beijing

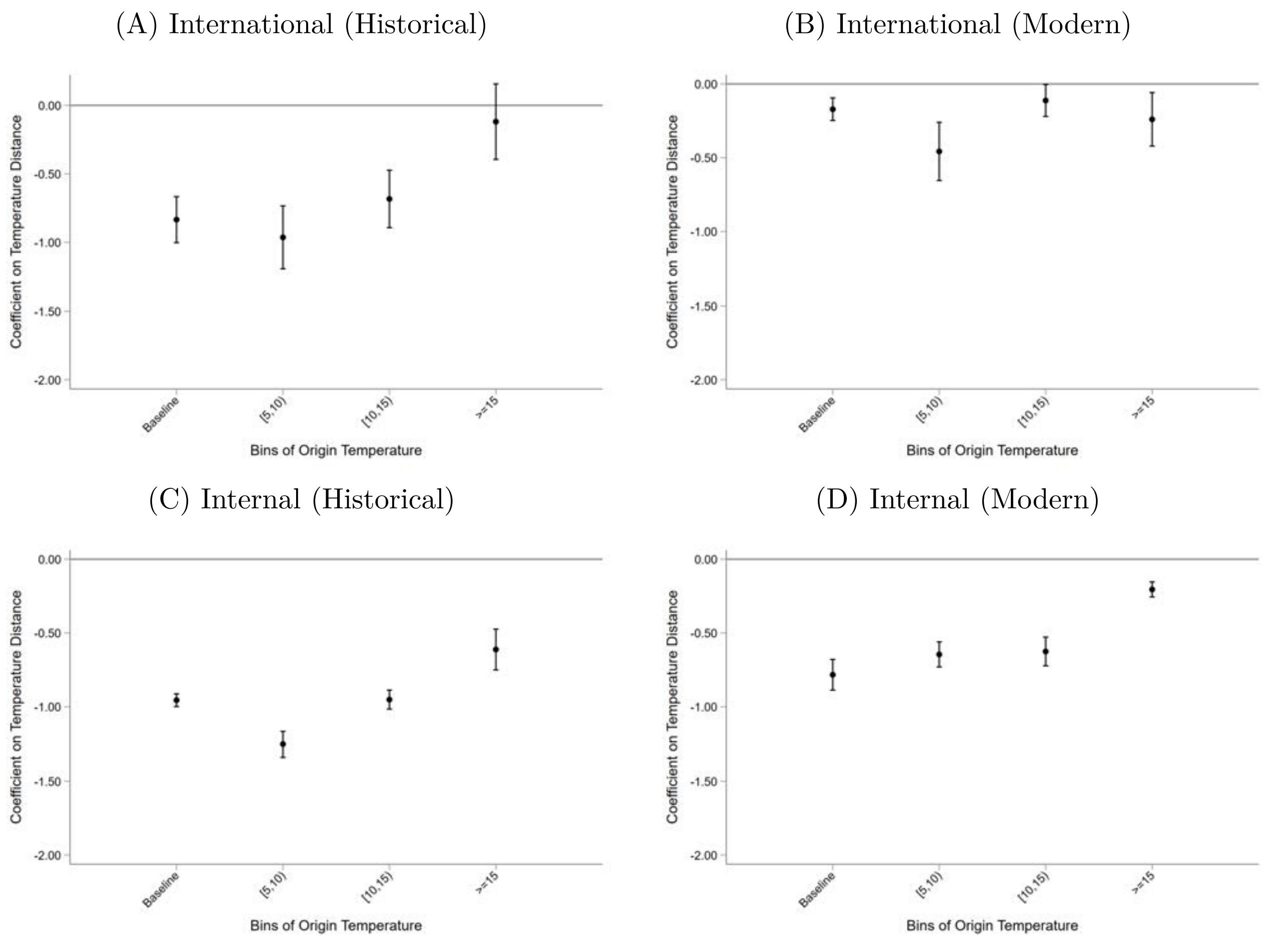


(D) Mexico City



Notes: The figure shows the climate distances between each US county and the cities considered (Oslo, Berlin, Beijing, Mexico City). Temperature is the average across the year in degrees Celsius, calculated over the period 1960-2000 using NOAA data for the US (Vose et al., 2014), and TerraClimate data for foreign cities (Abatzoglou et al., 2018). Climate distances are defined as the absolute difference in average temperature between each US county and the respective city.

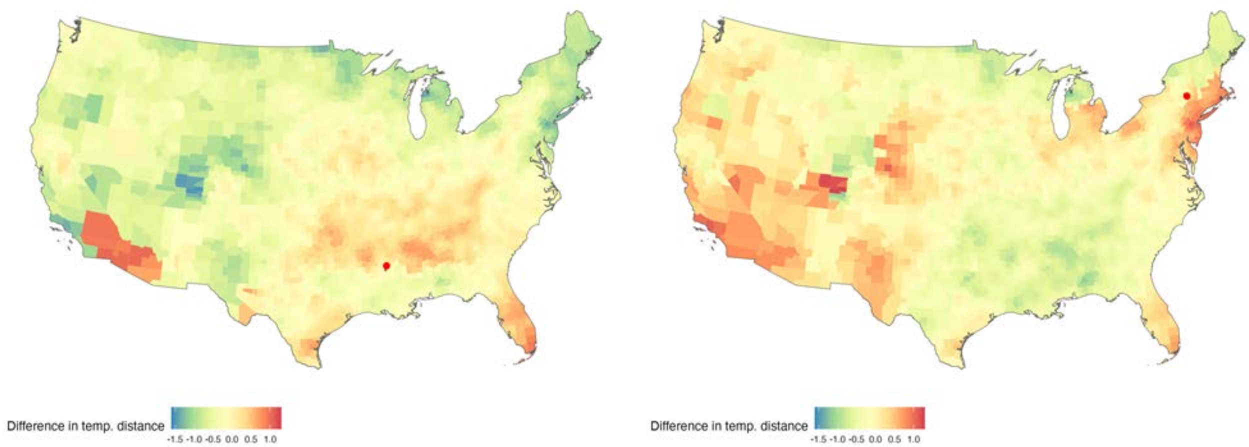
Figure 3. Heterogeneity by Origin Temperature Bins



Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the temperature difference for the international country of birth vs. US destination county (Panels A and B), and for US origin counties vs. US destination counties for internal movers (Panels C and D). Standard errors are clustered at the country (resp., state) of origin by state of destination by decade-level.

Figure 4. Historical Change in Climate Distances Between US Counties

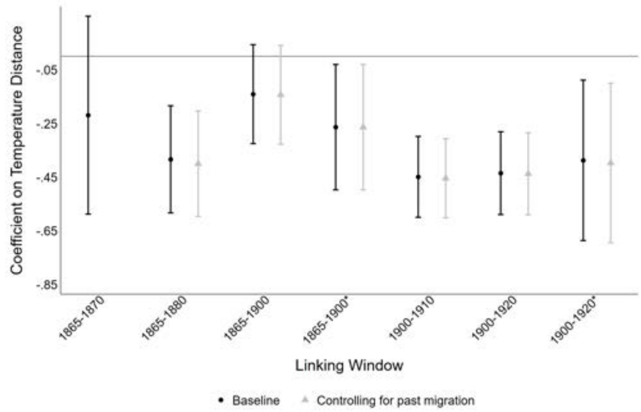
(A) Change in Temp. Distance, Delta Region (B) Change in Temp. Distance, US Northeast



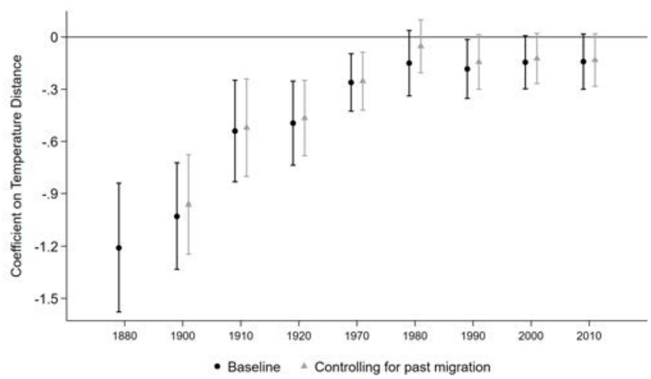
Notes: The figure plots the change in temperature distances between each US county and the centroid of the considered region. The red dot is the centroid of the considered region. Temperature is an average of all-year climate in degrees C. These averages are taken over the period 1901-1930 and 1991-2020 using NOAA data (Vose et al., 2014). To interpret the figure, note that in Panel A, the *distance* in temperature between the Delta region and New York City decreased by 1.3° C: it went from 6.9° C to 5.6° C; New York City today is colder on average than the Delta region but less so on average than 100 years ago. As a result New York City appears bright green on the map.

Figure 5. Climate Distance and Migration, by Period and Sample

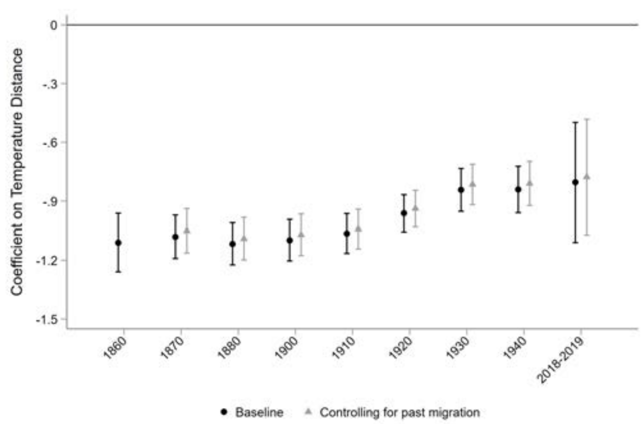
(A) Norwegian Immigration



(B) International Migration



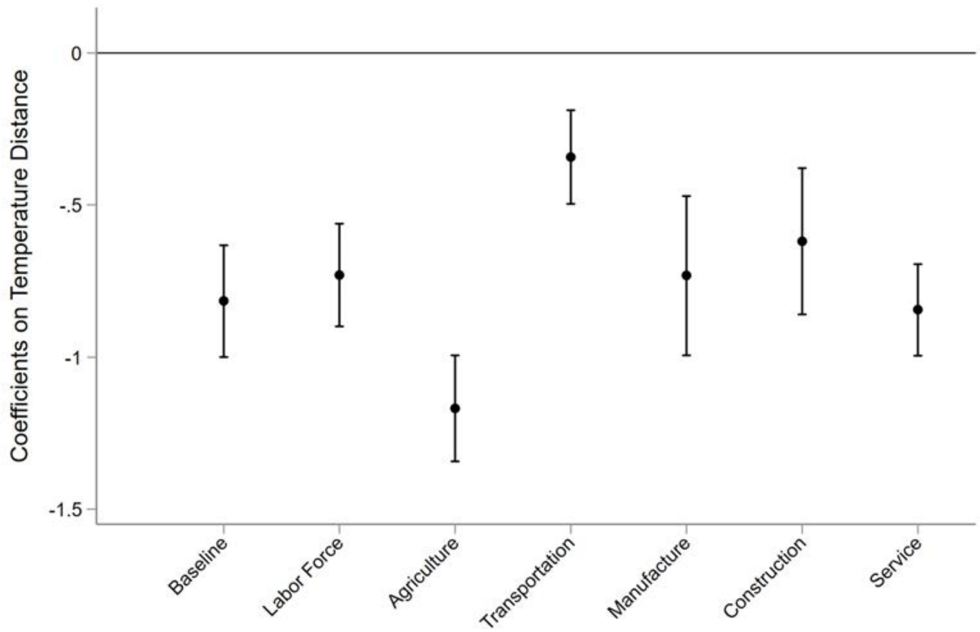
(C) Internal Migration



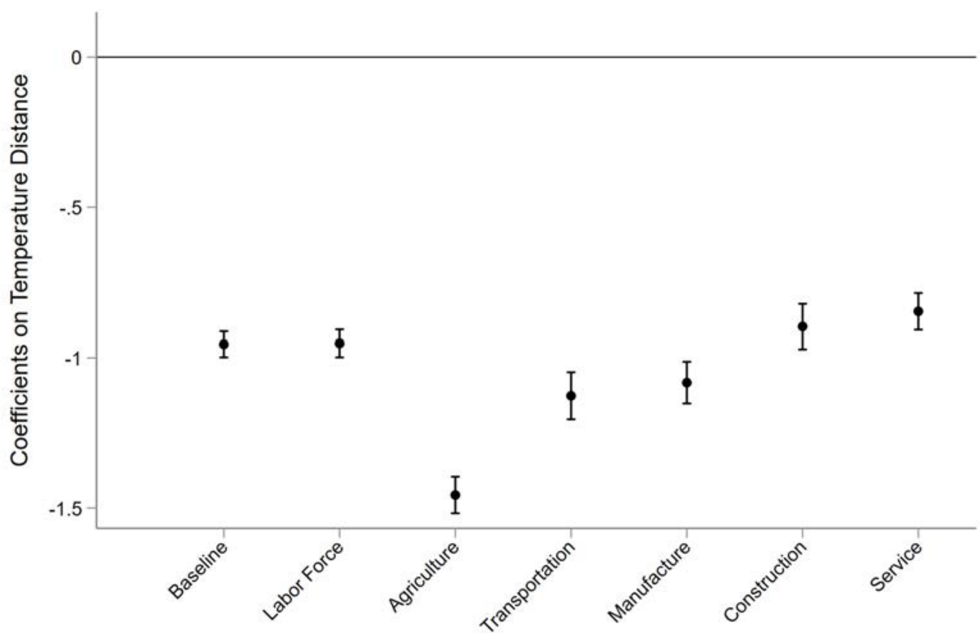
Notes: The figure plots the coefficient with corresponding 95% confidence intervals on the absolute value of the difference in temperature between (A) Norwegian municipality of origin and US destination county, (B) international country of birth and US destination county, and (C) US origin and destination counties for internal movers. Dots refer to the baseline specification, described in the notes of Table 1. Triangles refer to specifications that further control for lagged number of immigrants from a given origin to a given destination county. The dependent variable is the number of immigrants over the window reported on the x-axis. Temperature distances are standardized within the relevant sample to have zero mean and standard deviation equal to one. Standard errors are clustered at the country or state of origin by state of destination-level across columns.

Figure 6. Climate Distance and Migration: Heterogeneity by Sector

(A) International Migration

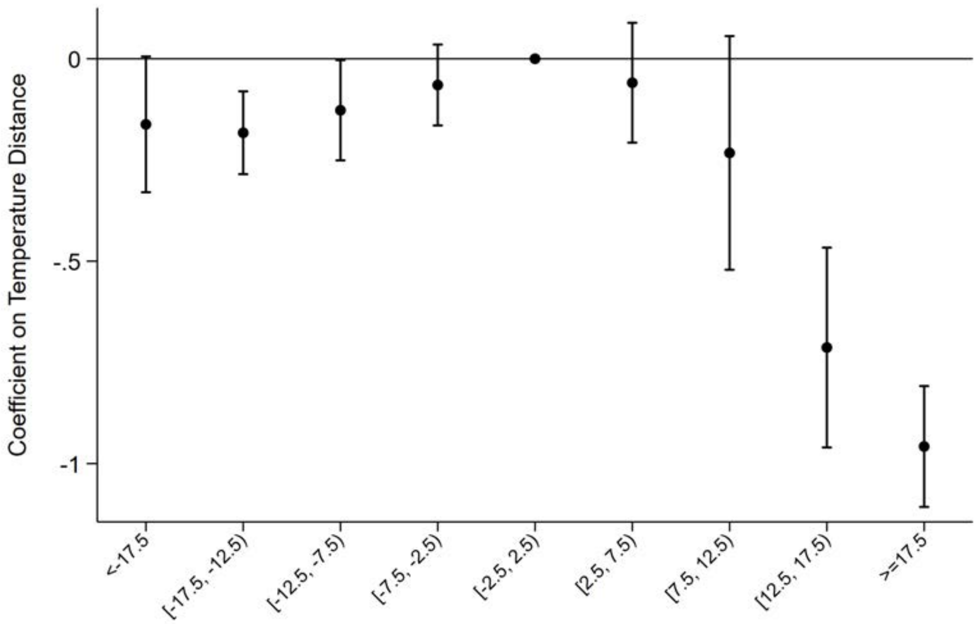


(B) Internal Migration



Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature between (A) international country of birth and US destination county, and (B) US origin and destination counties for internal movers. The first dot shows the baseline results of column 4 of Table 1. From the second dot onward, the number of migrants refers to migrant men in the sector reported on the x-axis (defined according to 3-digit IND1950 code from IPUMS). Temperature 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 7. Temperature Distance and Age at Death of Immigrants in the US



Notes: The figure plots the coefficients, with corresponding 95% confidence intervals, of the regression of age of death on 5°C bins of temperature difference between the US county of death and birth country capital. For example, an individual is placed in the [-17.5,-12.5) bin if they move to a US county colder than their origin by 12.5-17.5°C. The bin containing 0°C is omitted, while the bins for -17.5°C and 17.5°C include observations with differences beyond these thresholds. The sample comprises all people older than 65 at death, who died between 1988 and 2005, arrived after 1970, and are from outside Europe. The regression includes individual controls (sex, mover dummy, migration year, and distance). Standard errors, reported in parentheses, are clustered at the state of death by country of origin-level.

Table 1. Climate and Migration, Consolidated Regression Results

Dep. var.:	Number of Migrants				
	Norway-US *1865-1880*	Intl-US *1880-1920*	Intl-US *1970-2010*	Internal *1850-1940*	Internal *2011-2019*
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.084***	-0.139***	-0.029***	-0.258***	-0.209***
	(0.022)	(0.014)	(0.006)	(0.006)	(0.014)
Observations	178,476	2,030,556	2,643,206	30,498,928	25,071,430
Pseudo R-squared	0.330	0.910	0.894	0.704	0.781
Mean Temp. Dist.	4.720	9.978	10.00	5.038	5.126
SD Temp. Dist.	3.688	6.029	6.001	3.650	3.722
Origin o FE	Yes				
Destination d FE	Yes				
Origin o $\times$ Period FE		Yes	Yes	Yes	Yes
Destination d $\times$ Period FE		Yes	Yes	Yes	Yes

Notes: The sample in column 1 includes pairs formed by all Norwegian municipalities (origins) and all US counties (destinations) with at least one European immigrant in 1880. The number of migrants refers to Norwegian male immigrants who were at least 10 years old in 1865 and linked between the 1865 Norwegian and the 1880 US Censuses. The samples in columns 2-3 are international migration flows from Burchardi et al. (2019) sample. The sample in column 4 includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. 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. Column 5 includes all county-pairs in the contiguous US for each year from 2011-2012 to 2018-2019, with the number of migrants representing people who changed address between origin and destination counties, based on IRS Migration data from individual income tax returns. In column 1, Origin o refers to a Norwegian municipality; in columns 2-3, Origin o refers to the original country; and in columns 4-5, Origin o refers to a US county. Destination d refers to a US county. Standard errors, reported in parentheses, are clustered at the country or state of origin by state of destination by year of destination level across columns. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 2. Climate Distance and Migration in the Long Run

Dep. var.:	Number of Migrants			
	Intl-US	Intl-US	Internal	Internal
	(1)	(2)	(3)	(4)
Temperature Distance	-0.248***	-0.288***	-0.394***	-0.277***
	(0.092)	(0.083)	(0.105)	(0.075)
Observations	667,070	814,674	761,098	1,959,248
Pseudo R-squared	0.990	0.988	0.954	0.952
Mean Temp. Dist.	9.448	9.491	3.051	3.585
SD Temp. Dist.	5.974	5.927	2.792	3.104
Origin o × Period FE	Yes	Yes	Yes	Yes
Destination d × Period FE	Yes	Yes	Yes	Yes
Origin o × Destination d FE	Yes	Yes	Yes	Yes
Distance × Period FE	Yes	Yes	Yes	Yes
Migration Historical Period	1900	Avg. 1900-1920	1900-1910	Avg. 1910-1940
Migration Modern Period	2010	Avg. 1990-2010	2018-2019	Avg. 2011-2019

Notes: The samples in columns 1-2 are international migration flows to the US from the Burchardi et al. (2019) sample. The samples in columns 3-4 include all county-pairs in the contiguous US. Temperature Distance is the absolute value of the difference in temperature between the origin and the destination county. Controls include similarly constructed measures of Precipitation Distance and Physical Distance (expressed in 100 km). Historical climate refers to the average climate from 1901 to 1930, and modern climate refers to the average climate from 1991 to 2020. In column 1-2, Origin o refers to the original country; and in columns 3-4, Origin o refers to a US county. Destination d refers to a US county. Standard errors, reported in parentheses, are clustered at the country (resp., state) of origin by state of destination by period level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Climate Distance and Age at Death

Dep. var.:	Age at Death			
	(1)	(2)	(3)	(4)
Panel A: OLS				
Temperature Distance	-0.016***	-0.017***	-0.017***	-0.017***
	(0.004)	(0.004)	(0.004)	(0.004)
R-squared	0.724	0.724	0.724	0.725
Panel B: 2SLS				
Temperature Distance	-0.020***	-0.021***	-0.022***	-0.021***
	(0.006)	(0.006)	(0.006)	(0.006)
KP F-stat	114.9	116.4	115.6	112.4
Panel C: First Stage	Temperature Distance			
Pred. Temp. Distance	0.823***	0.819***	0.820***	0.814***
	(0.077)	(0.076)	(0.076)	(0.077)
R-squared	0.715	0.716	0.716	0.728
Observations	356,900	356,900	356,900	354,778
Mean Temp. Dist.	6.573	6.573	6.573	6.570
SD Temp. Dist.	5.482	5.482	5.482	5.481
Mean Death Age	80.44	80.44	80.44	80.43
SD Death Age	8.806	8.806	8.806	8.782
Cohort FE	Yes	Yes		
Continent FE	Yes	Yes		
County FE	Yes	Yes	Yes	
Continent by cohort FE			Yes	Yes
County by cohort FE				Yes
$\text{Imm}_{o,d}^{1980}$		Yes	Yes	Yes

Notes: The sample comprises all people older than 65 at death and people who died between 1988 and 2005 who arrived after 1970 and are from outside Europe. *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 over the period 1960-2000. Birth cohort span to 10 year period. All OLS regressions control for the precipitation distance, while the 2SLS for the predicted precipitation distance. From column 2 onward, we control for the number of immigrants, while the 2SLS for the predicted number of immigrants. All specifications include individual controls (sex, mover dummy, migration year, and distance). The mover dummy equals one if the individual died in a different US state than that where they applied for a Social Security Number. Standard errors, reported in parentheses, are clustered at the state of death by country of origin level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 4. Climate Distance and Infant Health Outcomes

Dep. var.:	Infant Health Outcomes			
	(1)	(2)	(3)	(4)
Panel A: Infant Mortality Rate				
Temperature Distance	0.276***	0.277***	0.277***	0.275***
	(0.015)	(0.015)	(0.015)	(0.015)
Mean Outcomes	4.494	4.494	4.494	4.484
SD Outcomes	8.997	8.997	8.997	8.801
Panel B: Low Birthweight Rate				
Temperature Distance	4.502***	4.539***	4.537***	4.552***
	(0.142)	(0.144)	(0.143)	(0.146)
Mean Outcomes	71.51	71.51	71.51	71.83
SD Outcomes	39.87	39.87	39.87	39.40
Panel C: Low Gestation Weeks Rate				
Pred. Temp. Distance	3.564***	3.611***	3.594***	3.561***
	(0.139)	(0.141)	(0.139)	(0.135)
Mean Outcomes	112.3	112.3	112.3	112.2
SD Outcomes	48.63	48.63	48.63	47.85
Observations	171,420	171,420	171,420	158,796
KP F-stat	1795	1729	1750	1673
Mean Temp. Dist.	6.247	6.247	6.247	6.316
SD Temp. Dist.	5.066	5.066	5.066	5.098
Year FE	Yes	Yes		
Continent FE	Yes	Yes		
County FE	Yes	Yes	Yes	
Continent by Year FE			Yes	Yes
County by Year FE				Yes
$\text{Imm}_{o,d}^{2000}$		Yes	Yes	Yes

Notes: The sample is restricted to non-European countries and those matching the Burchardi et al. (2019) sample. Columns 1 to 4 are weighted by the number of births. All regressions control for the predicted Precipitation Distance. Infant Mortality Rate refers to the number of deaths of infants under one year of age per 1,000 live births. Low Birthweight Rate refers to the number of births per thousand who weigh less than 2,500 grams at birth. Low Gestation Weeks Rate refers to the number of births per thousand that occur before 37 weeks of gestation. All specifications include maternal controls (the mother's age, marital status, and the number of live births she has had). From column 2 onward, we control for the predicted number of immigrants. Standard errors, reported in parentheses, are clustered at the state of death by country of origin by year level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Appendix

Table of Contents

A Data Appendix

Table A.1. Summary of data sources and their uses

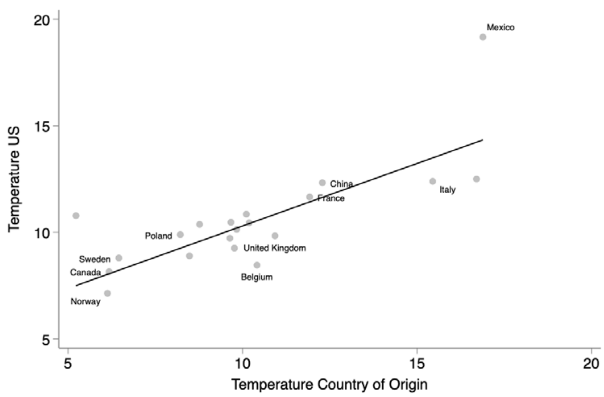
Source	Description	Area	Spatial Resolution	Period	Time Resolution
Climate					
TerraClimate (Abatzoglou et al., 2018)	Mean Temperature (°C) and Precipitation (mm) derived from monthly gridded product from terrestrial climate model; (i) Averaged across 5km radius around capital cities; (ii) Robustness where the Mexican climate is computed as a population-weighted mean across Mexican states; (iii) Robustness using the climate in Taishan county for China.	Global	4km x 4km	1960-2000	Averaged over period
Climatic Research Unit (CRU) (Harris et al., 2020)	Mean Temperature (°C) and Precipitation (mm) derived from monthly gridded product from observations (no model interpolation), averaged across 5km radius around capital cities.	Global	55km x 55km	1901-1930; 1991-2020	Averaged over period
NOAA (Vose et al., 2014)	Mean Temperature (°C) and Precipitation (mm) derived from monthly product; other climate measures for robustness.	US	US county level	1901-1930; 1960-2000; 1991-2020	Averaged over period
Immigration					
Norwegian 1865 Census (The Digital Archive, 2008)	Individual characteristics of the Norwegian population.	Norway	Individual	1865	
US Census (Ruggles et al., 2021)	Individual characteristics of the US population: Full-count data for 1860-1940 (incl. NBER restricted access data), 1% Form 1 Metro sample for 1970, 5% State sample for 1980 and 1990, 5% sample of 2000 census, and ACS 5-Year sample of the 2010 census.	US	Individual	1860-1940; 1970-2010	Decade
Census Linking Project (Abramitzky et al., 2020)	Linked samples of individuals between censuses.	US	Individual	1850-1940	Decade
IRS Statistics of Income Division	County-to-County Migration Files.	US	County	2011-2019	Yearly
Mortality					

BUNMD (Goldstein et al., 2023)	Social Security Administration death records.	US	Individual	1988-2005	Yearly
NCHS Multiple Cause of Death Data (National Center for Health Statistics, 2021)	Death certificates filed in vital statistics offices.	US	Individual	1959-1961	Yearly
WWI German casualties list (Verein für Computergenealogie e.V., 2014)	List of German Soldiers who died in WWI, with birth city information.	Germany	Individual	1914-1918	Yearly
Vital Statistics from CDC's NCHS	Infant Health Outcomes (Infant Mortality, Low Gestation Weeks, Low Birthweight).	US	Individual	2005-2021	Yearly
Miscellaneous					
US Frontier (Bazzi et al., 2020)	Total Frontier Experience; dummy for being on the frontier in a decade.	US	County	1850-1890	Decade
Market Access (Donaldson and Hornbeck, 2016)	Lowest-cost county-to-county freight routes.	US	County	1850-1930	Decade
Connection to Railroads (Atack and Margo, 2011)	Railroads connection data over time from Donaldson and Hornbeck (2016).	US	County	1850-1940	Decade
Homestead (Bureau of Land Management General Office Land records.)	Share of the county area that was homesteaded in any given decade relative to the overall area of the county.	US	County	1870-1920	Decade
Geographic Crosswalks	Bring historical county boundaries to modern ones (Eckert et al., 2020; Berkes et al., 2023).	US	Individual, County	1850-2000	Decade
Replication files from Burchardi et al. (2019)	Historical-to-modern countries and county-groups to counties crosswalks.	US, Global	County, Country	1880-2000	Decade
Historical GDP (Bolt et al., 2018)	GDP per capita.	Global	Country	1920	
Soil Characteristics	Bulk density (Hengl, 2018a), soil pH (Hengl, 2018b), water content (Hengl and Gupta, 2019).	US	County	1960-2000	Average

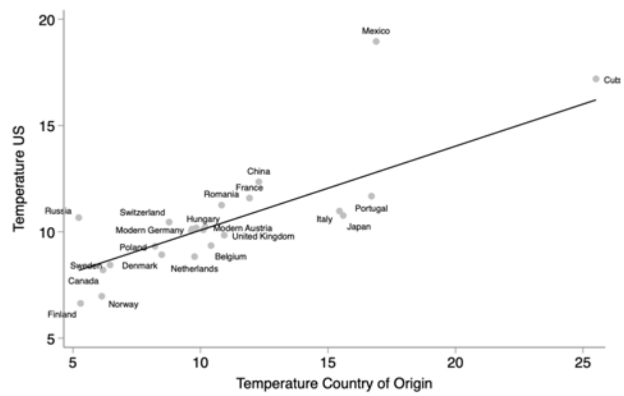
B Motivating Evidence: Figures

Figure B.1. Temperature Matching of Immigrants in the US, by Decade

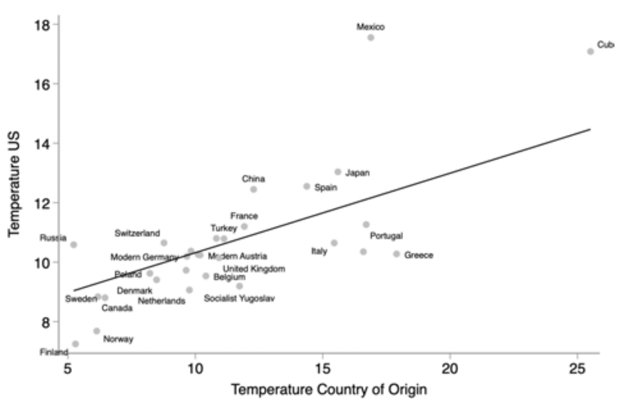
(A) 1880



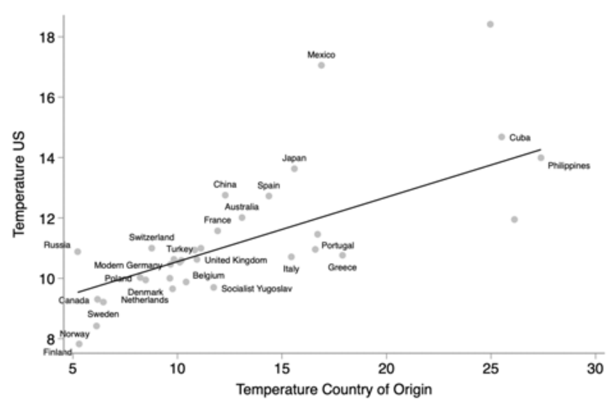
(B) 1900



(C) 1920



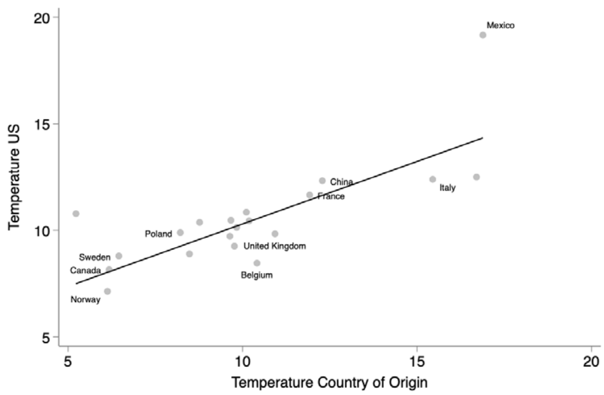
(D) 1940



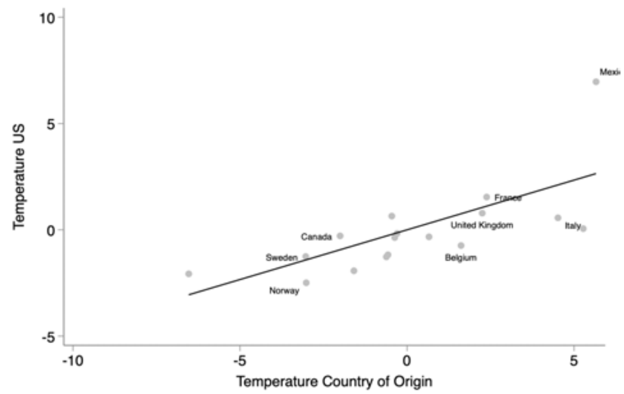
Notes: The figure displays the relationship between average temperature in degrees Celsius across US counties where immigrants from each origin lived in a given decade (y-axis) and the average temperature in the capital city of their country of origin (x-axis). Panel A reports results for 1880, replicating Figure 1. Panels B, C, and D consider, respectively: 1900, 1920, and 1940. The regression coefficients (robust standard errors) are: 0.586 (0.140) in Panel A, 0.395 (0.127) in Panel B, 0.268 (0.146) in Panel C, and 0.213 (0.106) in Panel D.

Figure B.2. Temperature Matching of Immigrants in the US (1880): Robustness

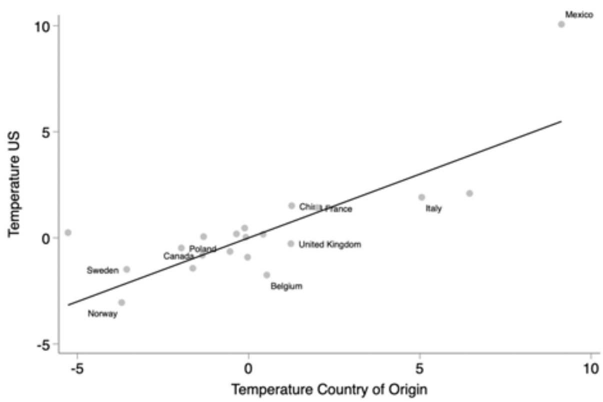
(A) Baseline



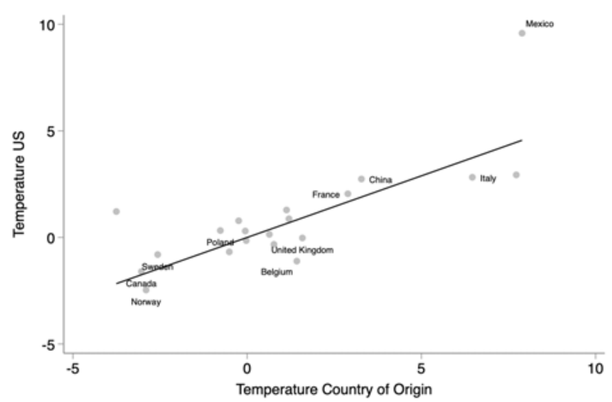
(B) GDP per capita



(C) Distance

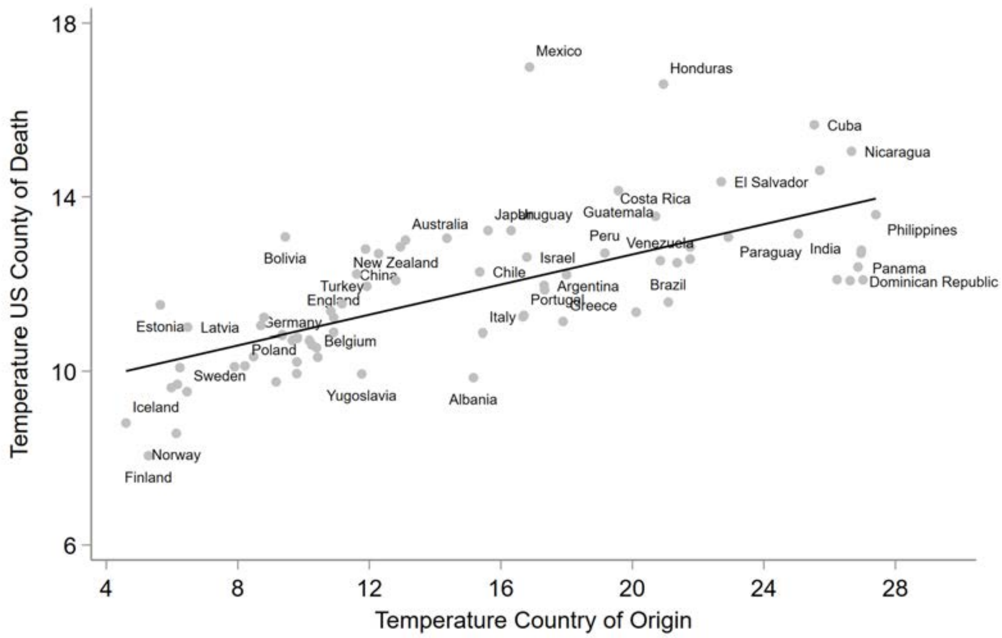


(D) Immigrants in 1870



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: 0.586 (0.140) in Panel A, 0.468 (0.130) in Panel B, 0.601 (0.197) in Panel C, and 0.578 (0.164) in Panel D.

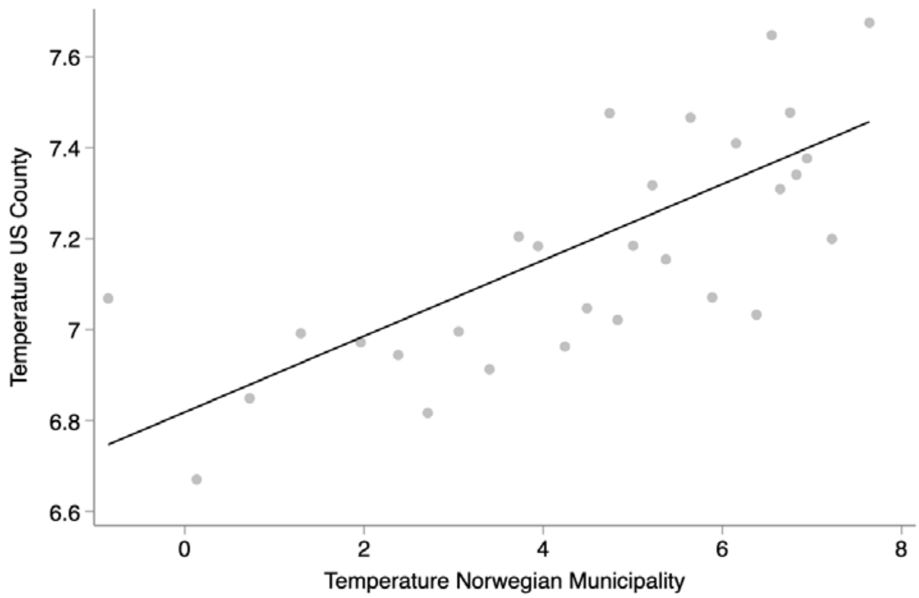
Figure B.3. Evidence from US Mortality Records: 1959-1961



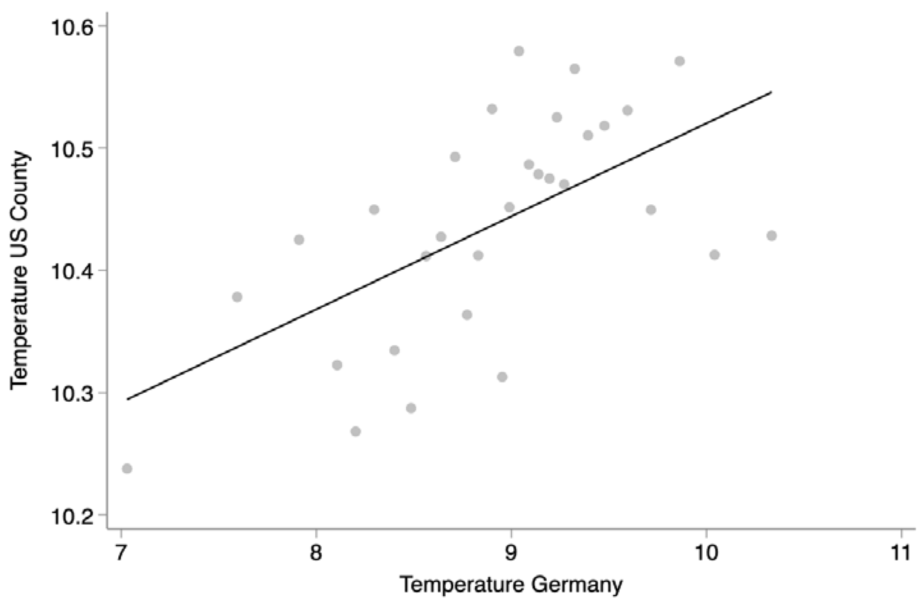
Notes: The figure displays the relationship between average temperature in degrees Celsius across US counties where immigrants from each origin were living at the time of their death (y-axis) and the average temperature in the capital city of their country of origin (x-axis).

Figure B.4. Within Country Temperature Matching

(A) Norwegian Immigrants



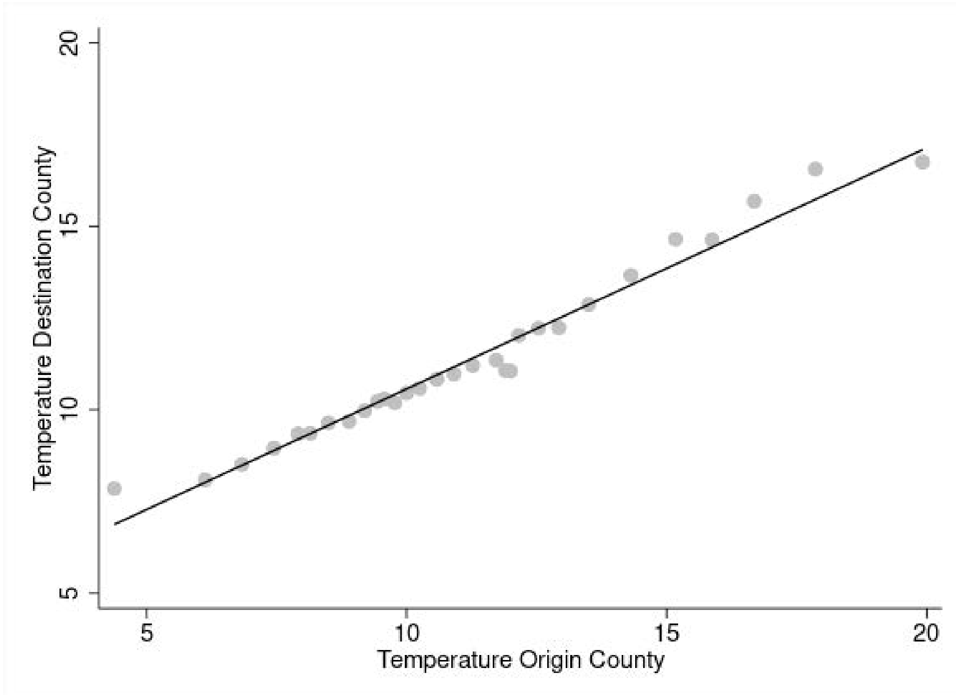
(B) German Surnames



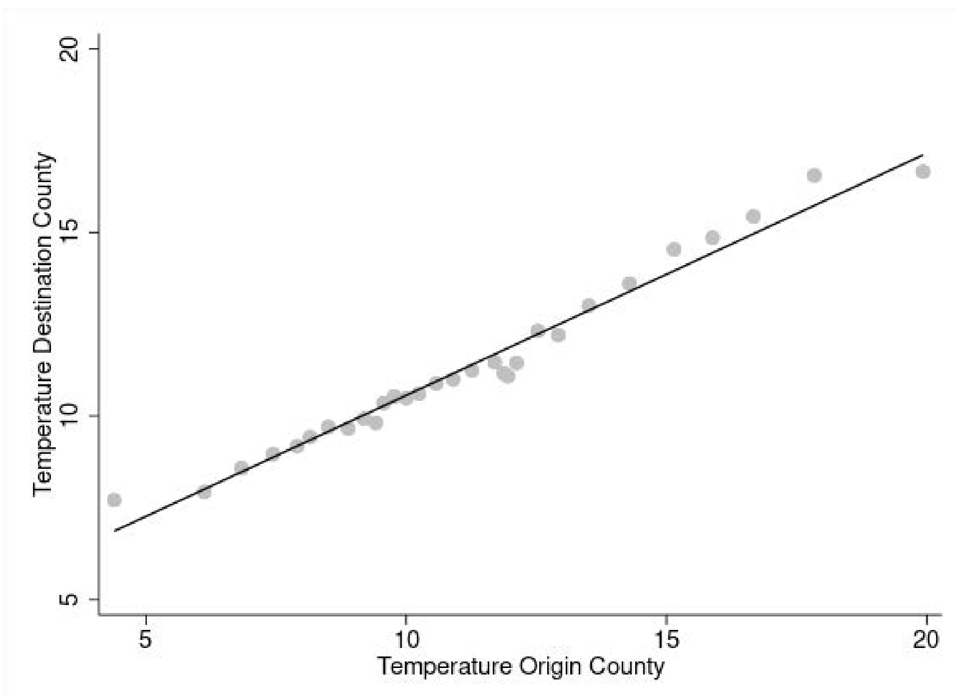
Notes: The figure displays binned scatterplots obtained from individual-level regressions of average temperature at destination against average temperature at origin. The sample in Panel A consists of Norwegian immigrants linked between the 1865 Norwegian and the 1880 US full count censuses, in Panel B, of German immigrants' surnames linked between the WWI German casualties list and the 1880 US full count censuses. In Panel A, the regression coefficients and the corresponding standard errors, clustered at the Norwegian province by US state-level, are 0.084 and 0.011; in Panel B, the regression coefficient and the corresponding robust standard errors are, respectively, 0.076 and 0.011.

Figure B.5. Temperature Matching and Migration at the Individual-Level

(A) Baseline



(B) Controlling for Distance



Notes: The figure displays binned scatterplots obtained from individual-level regressions of average temperature at destination against average temperature at origin. Sample consists 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 B, we control for the distance between the origin county and the destination county (expressed in 100 km). The regression coefficient and the corresponding standard errors, clustered at the US state of origin by US state of destination by decade-level 0.657 and 0.016 (Panel A) and 0.659 and 0.016 (Panel B).

C Norwegian Immigration to the US

C.1 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 limited. Norwegian pioneers sent letters back home praising the quality of the US soil and emphasizing the similarity between the destination and the origin climates.[34] Norwegian mass migration did not take off until the 1860s, when severe frosts led to pervasive crop failures and pushed hundreds of thousands of individuals out of the country. By 1880, Norwegian immigrants accounted for about 0.4% and 2.7% of the total and the foreign-born US population, respectively. Norwegian immigration remained high until 1920, after which it declined, partly as a result of the Immigration Acts of the 1920s. Overall, more than 800,000 Norwegian immigrants moved to the United States between 1825 and 1920 (Ulvestad, 2010).

At the onset of Norwegian mass migration, Norway was a predominantly rural country: in 1865, only 15.3% of the male population (15+) resided in urban areas, and about 50% of men (15-64) in the labor force worked in agriculture (Table C.1, columns 1-3).[35] The residential choices and the occupations of Norwegian men in the US mirrored those prevailing in the country of origin. In 1880, about 18% of Norwegian men in the US lived in urban areas, while 63% of men in the labor force worked in agriculture (Table C.2). 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 C.1). Over time, immigrant settlements expanded and, by 1920, Norwegians lived in most states, especially in the Midwest and US West (Eriksson, 2020).

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 C.1, columns 4-6). In Table C.3, 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).[36] Given east-west precipitation gradients in Norway, this explains why migrants

[34] 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).

[35] 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 C.2 for more details on the construction and the characteristics of the linked sample.

[36] Figure C.3 shows that migrants tended to live in the southeast of Norway, whereas stayers were more evenly distributed across all Norwegian municipalities.

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 C.3).

C.2 Linking Norwegian and US Censuses

Baseline sample: 1865-1880. Here, we describe the creation of a dataset of the number of immigrants who were living in a given Norwegian municipality in 1865 and who moved to a given US county in 1880. First, 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* and *exact* matches of observations between the two censuses. Matches are performed on first name, last name, 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 $\text{age}^{NOR,1865} + 15 = \text{age}^{US,1880} \pm 2$. 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 name and date of birth. The resulting dataset is a panel dataset with two observations for each individual—one in Norway in 1865 and one in the US in 1880. As shown in Table C.4, 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).[37]

We locate Norwegian immigrants in the US using data from the Census Place Project (Berkes

[37] 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.

et al., 2023). Information on Norwegians municipality of residence in Norway is obtained from the Norwegian Census.[38] Additional individual characteristics, such as labor force participation, occupation, and urban status, are retrieved from the two censuses. Table C.2 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.

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.

Using a procedure similar to the one described above, we also create a sample of "stayers," namely Norwegian men linked between the 1865 and the 1900 Norwegian Censuses. Table C.1 compares the 1865 characteristics of Norwegian men in either linked sample (migrants and stayers) to those of all men in the full count Norwegian Census of 1865. Among this linked sample, Table C.3 compares the characteristics of stayer and movers. We also present the geographic distribution of Norwegian men, both in Norway and in the US (Figures C.3 and C.1).

Alternative linking periods. We also present results considering Norwegian immigrants linked over alternative time periods. Besides the 1865 (Norway) to 1880 (US) sample, we also consider: 1865-1870, 1865-1900, 1900-1910, and 1900-1920. To this end, we adopt a procedure very similar that for the 1865 to 1880 linking. 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 procedure described earlier; 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*) over 10 years old in the Norwegian Census; *ii*) old enough to appear in the Norwegian Census (either of 1865 or 1900).[39]

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, Erik—a Norwegian immigrant living in the US in 1900 who reports an age of 37—is not part of our

[38] Few municipalities change boundaries, but in the event of a split, we assign population to the largest of the two resulting geographies.

[39] The actual constraint on age that we impose is slightly less stringent to allow for potential misreporting 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.

sample. This is because, even if Erik is old enough to appear in the 1865 Norwegian Census, he was younger than 10 at the time of that census.

Similar to Eriksson (2020), since from 1900 onward the US Census reports the year of immigration (information not available in 1880), we augment the linking procedure with this variable 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 given Norwegian Census.[40]

Table C.4 reports the number of Norwegian men linked in each time window, together with the corresponding match rate, expressed relative to the number of eligible Norwegian-born men in the US. 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.

C.3 Detailed Regression Results and Robustness Tests

Baseline specification. In this section, we present a more detailed analysis of the consolidated results contained in Table 1, column 1. We report results in Table C.5, where, for completeness, we also include the coefficients on precipitation distance as well as standardized coefficients (in square brackets). In column 1, we estimate a parsimonious specification that only includes temperature and precipitation distance. The negative and statistically significant coefficients indicate that higher temperature and precipitation distances are associated with lower migration. The transatlantic component of migration generates a "geographic break" that reduces concerns that the spatial correlation of climate may confound our estimates. 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. 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 as that between the Norwegian municipality and the closest transatlantic port.[41] Coefficients are unchanged.

In column 3, we add US state and Norwegian province fixed effects. The point estimate for temperature distance becomes smaller in absolute value, but remains negative and highly statistically significant. The drop in the magnitude of coefficients is not surprising. When adding province and state fixed effects, we are effectively comparing migration flows: *i*) between mu-

[40] Results, not reported for brevity, are virtually unchanged when dropping this additional restriction.

[41] We consider the two major ports of this period—Oslo and Kristiansand—but results are unchanged when including other ports.

nicipalities within the same Norwegian province and any US county; and, *ii*) between any Norwegian municipality and counties within the same US state.[42] 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.[43] 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, 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).[44] Coefficients remain negative and statistically significant.

Finally, in column 5 we report the preferred specification (Table 1, column 1), where we replace state and province fixed effects with county and municipality fixed effects, respectively. While the point estimate for precipitation distance is no longer statistically significant, that on temperature distance remains precisely estimated and quantitatively close to that in column 4. As discussed in the main text (Section 3.2), according to the estimates in column 5, reducing temperature distance by 4.7°C (approximately equal to the mean sample distance, or the annual temperature difference between Oslo and Brussels, Belgium (or Oslo and Detroit, MI in the US) increases migration by 0.4%. The effect of temperature distance is quantitatively relevant and comparable to that of decreasing the physical distance between New York City and a US county by about 800 km (e.g., from New York City to Detroit, MI, or Charleston, SC). The effects of climate distance on migration are quantitatively larger and more robust for temperature than for precipitation. We systematically observe this pattern across settings and time periods. For this reason, while Table C.5 presents coefficients on both temperature and precipitation distance, in the main paper and in the robustness analysis below, we focus on temperature.

As also noted in the introduction of the paper, the stronger effect of temperature distance res-

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

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

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

onates 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, 2022b).

Robustness checks. In Table C.6, we estimate more stringent specifications to examine whether forces other than climate similarity may influence our main results. To ease comparisons, we report the preferred specification in Table 1, column 1, in this column 1 as well. Then, in column 2, we add a vector of controls for the absolute value of origin-destination characteristics (X_{od}).[45] Results are unchanged. In the remainder of the table, we address the additional concern that our results may partly pick up the effect of the distance in 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 countries of origin in terms of distance from coast, elevation, temperate winters, and number of sunny days. We thus control for the origin-destination difference in elevation (column 3) and ruggedness (column 4), first separately, and then simultaneously (column 5). In all cases, the coefficient on climate distance remains in line with that from the preferred specification.

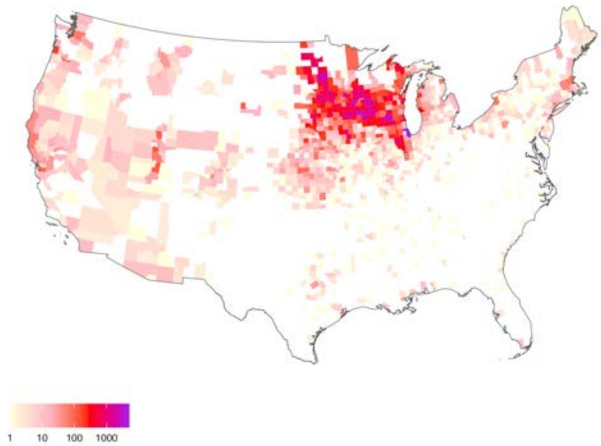
Finally, we present two additional sets of robustness checks. In Table C.7, we replicate the analysis by considering different sets of origins and destinations (reported at the bottom of the table). In Table C.8, we address the potential concern that our estimates may be biased due to false positive matches obtained in our linking procedure (Bailey et al., 2020), using different sample restrictions and linking algorithms (reported at the bottom of the table). Results are always similar to those obtained from the main sample (restated in column 1 to ease comparisons).

[45] We collect in X_{od} only those variables that can be measured for both Norway and the US, including population density, urban population share, employment share, manufacturing employment share, agriculture employment share, and sex ratios.

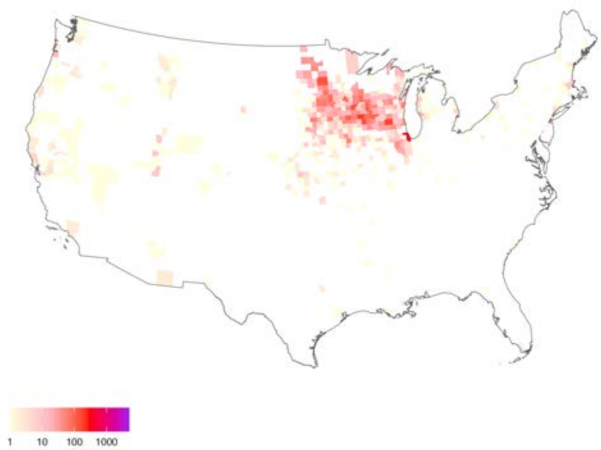
C.4 Figures and Tables

Figure C.1. Distribution of Norwegian Immigrants in the US (1880)

(A) Norwegian Immigrants in 1880 Full Count US Census



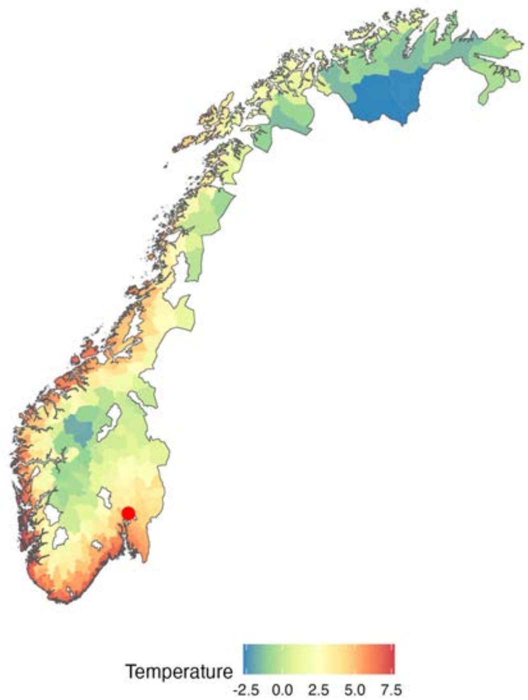
(B) Norwegian Immigrants Matched to 1865 Norwegian Census



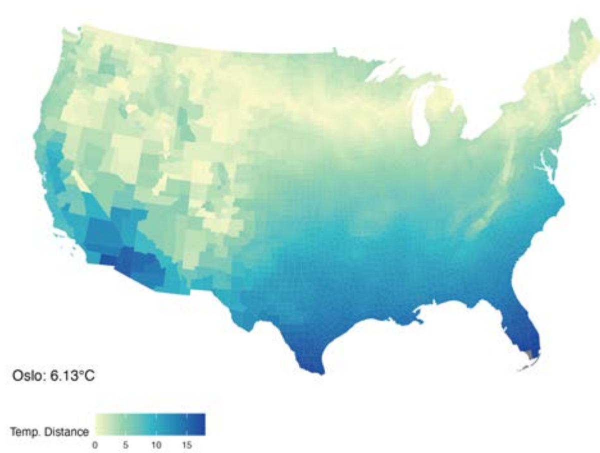
Notes: Panel A (resp., Panel B) plots the distribution of Norwegian male immigrants of age 15+ in the full count 1880 US Census (resp., who are linked between the 1865 Norwegian and the 1880 US full count censuses).

Figure C.2. Norway Climate and Temperature Distances

(A) Mean Temperature of Norwegian Municipalities



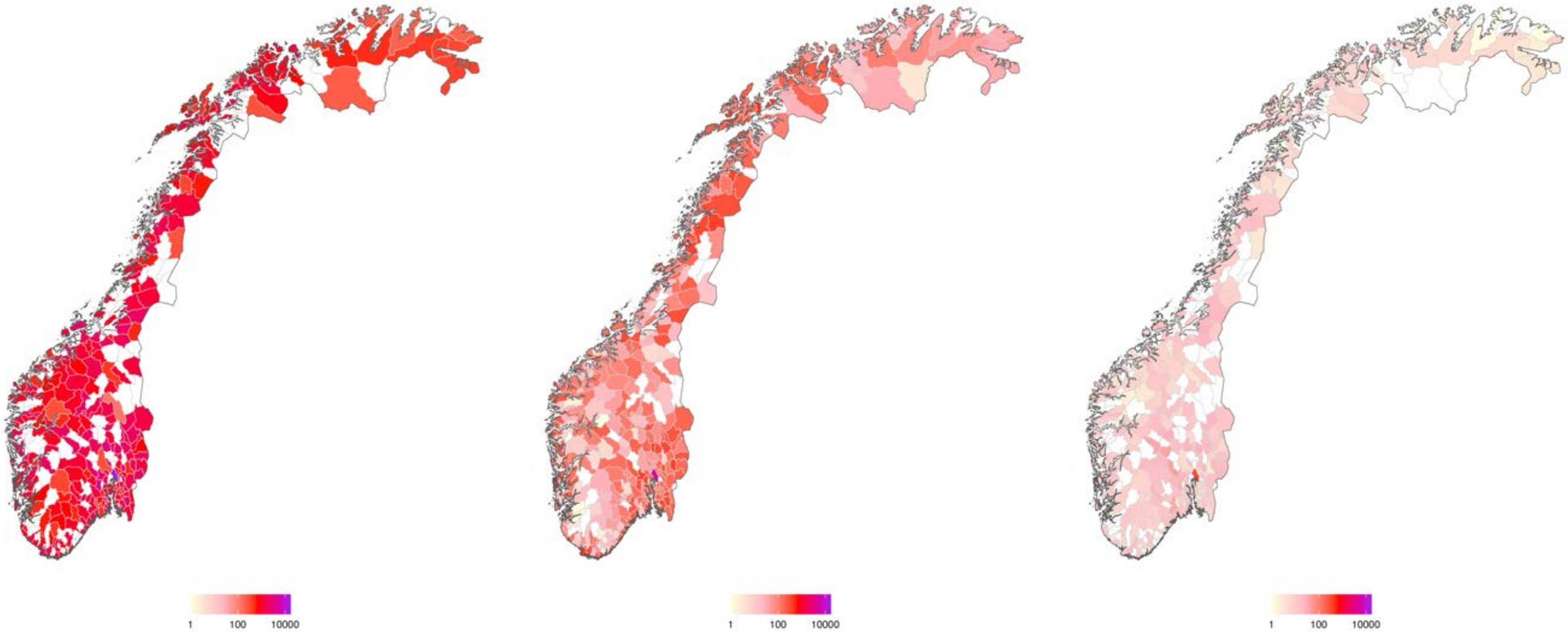
(B) Temperature Distances Between Oslo and US Counties



Notes: Panel A plots the average annual temperature in C at the Norwegian municipality-level for the period 1960-2000 based on TerraClimate monthly gridded historical data (Abatzoglou et al., 2018). The red dot is Oslo. Panel B plots the temperature distance from Oslo, which is the absolute value of the difference in average temperature between each US county and Oslo. For the US, these averages are taken over 1960-2000 using NOAA data (Vose et al., 2014).

Figure C.3. 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.

Table C.1. 1865 Norwegian Census: Full Count and Linked Sample

	Full Count (Norway 1865)			Linked to US (1880) or Norway (1900)		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Individual-level Characteristics						
Age	33.82	18.10	608,157	25.47	11.16	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	13.59	36.07	610,271	14.19	36.98	82,274
Distance from Closest Port (100km)	23.30	27.07	588,891	24.74	29.27	80,269
Average Temperature	4.329	2.257	588,891	4.528	2.156	80,269
Average Precipitation	104.5	49.39	588,891	103.5	48.84	80,269

Notes: The sample is restricted to men who were at least 10 years old as of 1865. *Linked* refers to individuals linked between the Norwegian 1865 Census and the US 1880 Census (movers), and to individuals linked between the Norwegian 1865 Census and the Norwegian 1900 Census (stayers). The dummies for being married and for living in an urban area (resp., for labor force participation) are defined only for men 15+ (resp., for men in the age range 15-64). The dummies for being a farmer and for working in manufacturing are defined for men in the age range 15-64 who were in the labor force. Population density is scaled by 100,000. Distance from closest international port is defined as the shortest between the municipality and the Port of Kristiansand and the Port of Oslo.

Table C.2. Norwegian Immigrants in the US, 1880: Full Count and Linked Sample

	Full Count (Norwegians in US 1880)			Linked (in Norway 1865 Census)		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Individual-level Characteristics						
Age	41.69	12.98	69,440	43.19	13.70	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	17.55	7.176	62,906	17.46	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 (100km)	15.95	5.486	69,440	15.76	5.106	5,747
Connected to Railroad	0.957	0.202	69,440	0.954	0.209	5,747
Average Temperature	7.241	1.989	69,440	7.187	1.908	5,747
Average Precipitation	67.11	14.11	69,440	66.72	12.88	5,747

Notes: The full count (resp., linked sample) is restricted to men who were at least 25 as of 1880 (resp., who were at least 10 as of 1865). The dummy for labor force participation is defined for men younger than 64. Occupational scores as well as the dummies for being a farmer and for working in manufacturing are defined for men younger than 64 who were in the labor force. Population density is scaled by 100,000. Norwegian and immigrant population shares are computed as a share of total county population.

Table C.3. 1865 Norwegian Census (Linked Sample): Stayers and Movers

	Linked Stayers (in Norway 1900 Census)			Linked Movers (in US 1880 Census)		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Individual-level Characteristics						
Age	25.25	10.92	76,308	28.23	13.65	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	13.61	36.35	76,308	21.63	43.63	5,966
Distance from Closest Port (100km)	24.91	29.44	74,521	22.59	26.91	5,748
Average Temperature	4.535	2.153	74,521	4.437	2.194	5,748
Average Precipitation	103.7	49.01	74,521	101.7	46.64	5,748

Notes: The sample is restricted to men who were at least 10 years old as of 1865. The dummies for being married and for living in an urban area (resp., for labor force participation) are defined only for men 15+ (resp., for men in the age range 15-64). The dummies for being a farmer and for working in manufacturing are defined for men in the age range 15-64 who were in the labor force. Population density is scaled by 100,000. Distance from closest international port is defined as the shortest between the between the municipality and the Port of Kristiansand and the Port of Oslo.

Table C.4. 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.

Table C.5. Climate Distance and Norwegian Immigration (1865-1880)

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.370***	-0.374***	-0.140***	-0.109***	-0.084***
	(0.019)	(0.020)	(0.025)	(0.017)	(0.022)
	[-15.782]	[-15.964]	[-5.413]	[-4.204]	[-1.236]
Precipitation Distance	-0.006***	-0.007***	-0.006***	-0.006***	-0.002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
	[-2.166]	[-2.485]	[-2.178]	[-2.001]	[-0.397]
Distance from NYC (in 100km)		0.010**	-0.138***	-0.052**	
		(0.005)	(0.035)	(0.025)	
		[0.801]	[-11.071]	[-4.181]	
Observations	1,002,822	1,002,822	847,894	847,894	178,476
Pseudo R-squared	0.118	0.119	0.286	0.401	0.330
Mean Temp. Dist.	7.907	7.907	7.436	7.436	4.720
SD Temp. Dist.	4.556	4.556	4.494	4.494	3.688
Mean Precip. Dist.	43.18	43.18	43.32	43.32	45.56
SD Precip. Dist.	38.84	38.84	39.98	39.98	44.27
US State FE			Yes	Yes	
Norwegian Province FE			Yes	Yes	
US Controls				Yes	
Norwegian Controls				Yes	
Norwegian Municipality FE					Yes
US County FE					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 Norwegian male immigrants who were at least 10 years old in 1865 linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* is the absolute value of the difference in temperature between the Norwegian municipality and the US county. Similarly constructed measures of *Precipitation Distance* are included in all columns. *Distance from NYC* refers to the distance between a US county and New York City (expressed in 100 km). From column 2 to 4, we also include the distance between the municipality and closest international port (either Kristiansand or Oslo). 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. 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 C.6. Climate Distance and Norwegian Immigration: Additional Controls

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.084***	-0.069***	-0.073***	-0.086***	-0.071***
	(0.023)	(0.022)	(0.023)	(0.023)	(0.023)
Observations	178,476	178,476	178,476	178,476	178,476
Pseudo R-squared	0.330	0.334	0.331	0.330	0.331
Mean Temp. Dist.	4.720	4.720	4.720	4.720	4.720
SD Temp. Dist.	3.688	3.688	3.688	3.688	3.688
Norwegian Municipality FE	Yes	Yes	Yes	Yes	Yes
US County FE	Yes	Yes	Yes	Yes	Yes
X_{od}		Yes			
Geographic Distance			Elevation	Ruggedness	Elevation & Ruggedness

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 Norwegian male immigrants who were at least 10 years old in 1865 linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* is the absolute value of the difference in temperature between the Norwegian municipality and the US county. Similarly constructed measures of *Precipitation Distance* are included in all columns. Column 1 replicates the baseline specification. Columns 2 adds the absolute value of the difference in population density; urban population share; employment share; manufacturing employment share; agriculture employment share; and sex ratios between the Norwegian municipality and the US county. Columns 3 and 4 also add the absolute value of the difference in, respectively, elevation and ruggedness, between the Norwegian municipality and the US county. Column 5 includes all variables in columns 3 to 4 simultaneously. 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 C.7. Climate Distance and Norwegian Immigration: Alternative Samples

Dep. var.:	Number of Migrants							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temperature Distance	-0.084***	-0.084***	-0.084***	-0.084***	-0.076***	-0.076***	-0.077***	-0.086***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.024)	(0.022)	(0.031)
Observations	178,476	178,476	178,476	178,476	176,391	159,444	178,410	150,332
Pseudo R-squared	0.330	0.330	0.330	0.330	0.282	0.260	0.286	0.341
Mean Temp. Dist.	4.720	4.720	4.720	4.720	4.738	4.530	4.721	5.127
SD Temp. Dist.	3.688	3.688	3.688	3.688	3.696	3.547	3.688	3.762
Norwegian Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
US County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table replicates column 5 of Table C.5, reported in column 1, using different samples. *Number of Migrants* is the number of Norwegian male immigrants who were at least 10 years old in 1865 linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* is the absolute value of the difference in temperature between the Norwegian municipality and the US county. Similarly constructed measures of *Precipitation Distance* are included in all columns. The sample includes: all US counties in column 2; 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 population in the linked sample. Column 7 drops municipality-county-pairs at the top 1% of the distribution of bilateral migration flows. Column 8 excludes Minnesota. 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 C.8. Climate Distance and Norwegian Immigration: Alternative Linking Methods

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.084***	-0.071***	-0.077***	-0.071***	-0.065**
	(0.022)	(0.023)	(0.023)	(0.021)	(0.026)
Observations	178,476	154,008	177,225	196,460	169,743
Pseudo R-squared	0.330	0.301	0.328	0.357	0.337
Mean Temp. Dist.	4.720	4.571	4.709	4.741	4.705
SD Temp. Dist.	3.688	3.627	3.682	3.656	3.742
Norwegian Municipality FE	Yes	Yes	Yes	Yes	Yes
US County FE	Yes	Yes	Yes	Yes	Yes
Restriction	Older than 10 in 1865	15-40 in 1865 (Nor)	18-65 in 1880 (US)	None	ABE-JW

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 Norwegian male immigrants linked between the 1865 Norwegian and the 1880 US Censuses. *Temperature Distance* is the absolute value of the difference in temperature between the Norwegian municipality and the US county. Similarly constructed measures of *Precipitation Distance* are included in all columns. Column 1 replicates the baseline specification in column 5 of Table C.5, 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 derive the linked sample using the Jaro-Winkler version of the ABE algorithm from Abramitzky et al. (2021). 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

D International Immigration to the US

D.1 Historical Context

Between 1850 and 1920, during the Age of Mass Migration, more than 30 million European immigrants moved to the US (Abramitzky and Boustan, 2017).[46] Until 1880, most immigrants came from Northern and Western Europe, but, as transportation costs declined, the share of Southern and Eastern European migrants steadily increased—a trend that accelerated between 1900 and 1914 (Hatton and Williamson, 1998).

These patterns are depicted in Figure D.1. In 1880, Northern and Western Europe was the source of almost 90% of the US foreign-born population; by 1920, this number had declined to 45%, while the share of Southern and Eastern Europeans had increased to 41%. As Figure D.1 makes clear, during the Age of Mass Migration, there was relatively little immigration from regions other than Europe. Approximately 300,000 Chinese immigrants arrived between 1850 and 1880, but the 1882 Chinese Exclusion Act effectively shut down Chinese immigration until 1965 (Lee, 2003).[47] Table D.1 documents that within the top 15 countries of origin in 1920, the only ones outside Europe were Canada and Mexico.

It is well-known that immigrants were clustered by region in the US (see also the quote by President Coolidge reported in the introduction of the paper). For instance, Scandinavian and German immigrants were concentrated in the upper and lower Midwest, respectively, while large Italian enclaves developed in New York, New Jersey, California, and Pennsylvania. With few exceptions (e.g., Mexicans in parts of Texas), relatively few immigrants settled in the US South during this period (see also Figure D.2, Panel A). Among the non-Europeans, Mexican immigrants were concentrated in Arizona, New Mexico, Southern California, and Texas; Canadians were predominantly found in Maine and parts of the US-Canada border (Abramitzky and Boustan, 2017).

During World War I, immigration to the US dropped (Figure D.3, Panel A). However, once the war was over, immigration flows went back to their 1910 levels. The rising number of immigrants and the change in their composition, from "old" and culturally closer to "new" and culturally more distant sending countries, triggered hostile reactions among the US-born population (Higham, 1955). After a decades-long debate, US Congress first introduced a literacy test in 1917 and then passed country-specific quotas with the Immigration Acts of 1921 and 1924 (Goldin, 1994). The quotas were designed to limit the number of immigrants from Southern and

[46] It is estimated that another 20 million Europeans moved to Latin America and Canada during this same period.

[47] The Chinese Exclusion Act—the first immigration restriction imposed by the US based on race or ethnicity—was soon followed by restrictions on Japanese immigration, formally introduced with the 1907 Gentleman's Agreement (Abramitzky and Boustan, 2017).

Eastern Europe, instead giving preferential treatment to the "more desirable immigrants" from Northern and Western Europe (King, 2002). Immigration from other parts of the world, except for Canada and Mexico, was almost completely banned. The consequences of the 1924 Johnson-Reed Act were dramatic: the yearly number of immigrants collapsed from approximately 1 million between 1900 and 1914 to 150,000 in 1924. The new regime remained in place, almost unchanged, until 1965, and the immigrant population share dropped from 14% in 1920 to about 5% in 1960 (Figure D.3, Panel B).

In 1965, Congress passed the Immigration and Nationality Act, removing the country-specific quotas, giving priority to migrants entering the US for family reunification reasons, and raising the yearly cap on overall immigration to 270,000. In practice, though, the actual yearly number of immigrants was much higher than the cap (which was later increased to 620,000). As discussed in Leonhardt (2023), legislators greatly underestimated the effects that the 1965 Immigration Act would have had on both the number and the ethnic composition of immigrants. In about 40 years, more than 31.6 million people moved to America. The immigrant population share rose from 5% in 1970 to 14% in 2010 (Figure D.3, Panel B).[48]

In stark contrast with the patterns prevailing during the Age of Mass Migration, in 2010, non-European immigrants accounted for a vast majority of the US foreign-born population. According to the 2010 US Census, 52% of immigrants came from Latin and Central America, and 27% of them came from Asia. Instead, European immigrants accounted for as little as 15% (Figure D.1, Panel A). Table D.1 reveals that, out of the top-15 immigrant-sending countries, only Germany, Russia, and England were Europeans in 2010. Not only the origins but also the destinations differed substantially between the historical and the more modern migration eras. This is evident from Figure D.2, Panel B, which shows that, in 2010, the immigrant population share was higher in many parts of the US South than elsewhere.

D.2 Detailed Regression Results and Robustness Tests

Baseline specification. As in the Norwegian and US internal migration contexts, this section presents a more detailed analysis of the consolidated results on international migration reported in columns 2 and 3 of Table 1. In Tables D.2 and D.3, we focus on historical and modern immigration to the US, respectively. In column 1, we estimate a very parsimonious regression that only includes temperature and precipitation distance. In column 2, we add geographic distance (measured in the same way as climate distance—namely, as the distance between the county centroid and the capital city of the country of origin). In column 3, we further include

[48] Since 1965, undocumented immigration also increased dramatically, and more than 11 million undocumented immigrants were living in the US starting in 2000. See statistics from the Office of Homeland Security Statistics.

year dummies. In all cases, the coefficient on temperature distance is negative and statistically significant. In columns 4 and 5, we add continent of origin and US state of destination fixed effects and country of origin and US county of destination fixed effects, respectively. Finally, in column 6, which reports the baseline specification (Table 1 columns 2 and 3), we interact year dummies with, respectively, country and county dummies.

For historical international migration to the US, Table D.2 shows that coefficients on temperature distance are always negative and statistically significant, even though they become smaller in absolute value when adding continent and US state (column 4) and country and county (column 5) fixed effects. According to the preferred specification (column 6), reducing temperature distance by 10°C (or, about the sample mean) increases migration by 1.4%. Modern migration results are shown in Table D.3. Coefficients on temperature distance are always an order of magnitude smaller (in absolute value) than for historical migration, and drop from -0.13 to -0.03 after controlling for country and county fixed effects (columns 5 and 6). This indicates that lowering temperature distance by the sample mean (about 10°C) increases migration by 0.3%. These tables also report estimates for precipitation and geographic distance, which are generally negative as well.

Robustness checks. We begin by focusing on historical migration in Table D.4. As discussed above, most immigrants during the Age of Mass Migration came from Europe. In column 2, we thus focus on European origins only. Despite the reduction in the number of observations and in the underlying variation in climate, results remain in line with those from the baseline specification (reported in column 1 to ease comparisons). In columns 3 and 4, we document that results are unchanged when conducting the analysis between 1900 and 1920 and between 1880 and 1930, respectively.[49]

Next, in Table D.4, we address potential concerns about measurement of climate (see Section 2.1 and Table A.1 for a discussion of climate datasets). First, recall that our main estimates in Table 1 are established using the climate in the capital city as a proxy for origin climate. This, however, may not be representative of the climate faced by most migrants from a specific country, especially in large countries spanning multiple climate zones that sent many immigrants to the US—such as China or Mexico. In column 5, we thus replace the capital city climate for Mexico with the population-weighed average climate and for China with the average climate of Taishan county, a major immigrant-sending region.[50]

Second, our main measure of historical climate comes from the TerraClimate data product,

[49] Note that in all decades except 1880, we consider the flow of immigrants, defined as individuals who arrived in the US during the previous decade. Since the 1880 US Census did not record arrival year, for this decade, we consider immigrant stocks. Results (not reported for brevity) are virtually unchanged when considering the immigrant stock in all other years.

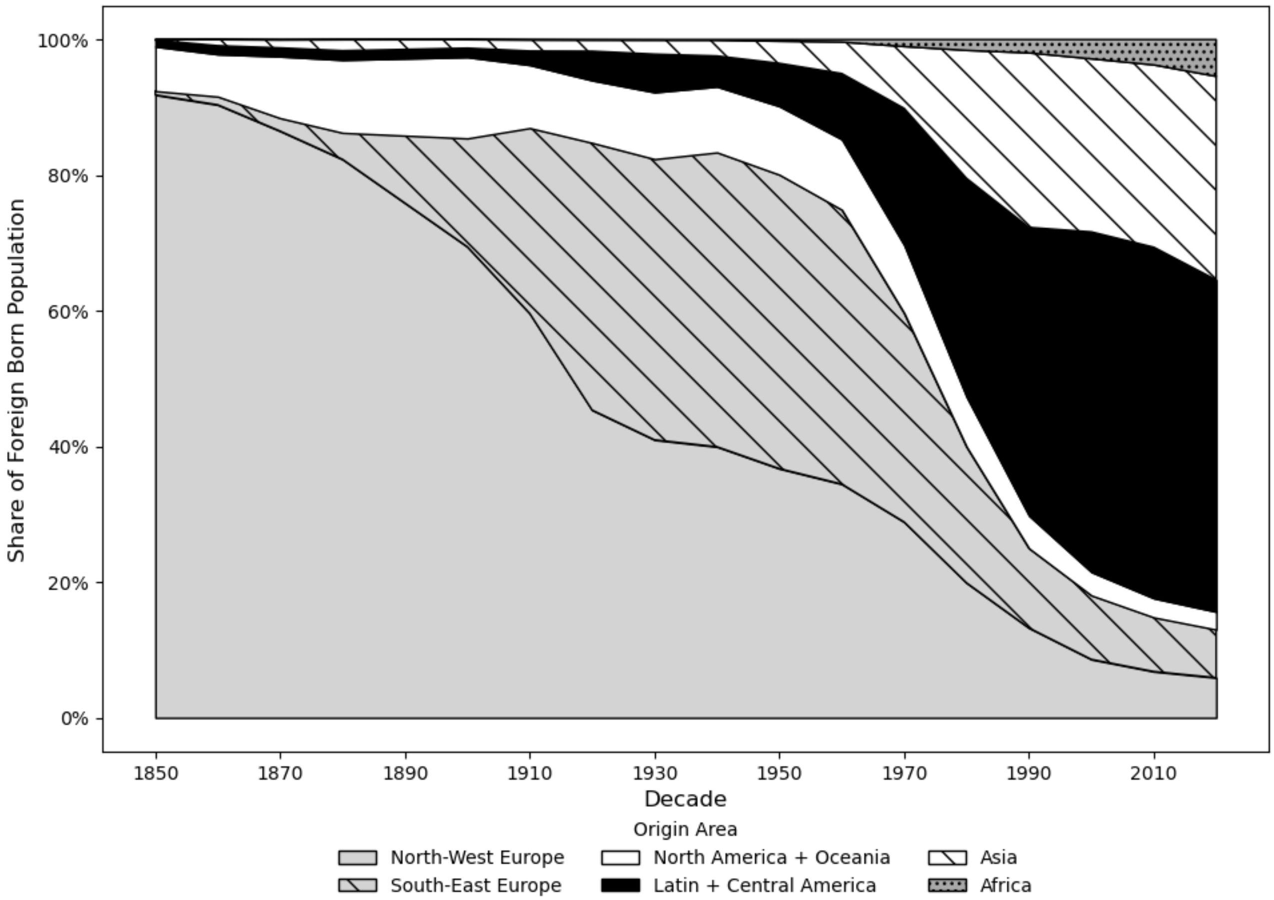
[50] Population weights come from NASA CIESEN's Gridded population of the world, version 4 (GPWv4) (Ciesin, 2016).

which begins in 1958. Even if this pre-dates the beginning of anthropogenic climate change, one may still wonder if our analysis systematically mismeasures climate distances. To address this concern, in column 6, we use two alternative sources to measure climate at origin and at destination. Specifically, we use data from the Climatic Research Unit (CRU) from 1901 and 1930, which is available at a coarser spatial resolution than TerraClimate. In all cases, results are unchanged.

In Table D.5, we turn to modern migration, reporting the baseline specification in column 1. Similar to Table D.4, column 2, focuses on the largest sending regions, dropping all European immigrants. In column 3, we verify that results are unchanged when starting the analysis in 1980, rather than in 1970. In column 4, we replace capital city climate with the alternate climate measures for Mexico and China, as discussed in the previous paragraph. In columns 5 to 7, we drop two major sending countries—Mexico and China—both individually and then at the same time. Results are always unchanged.

D.3 Figures and Tables

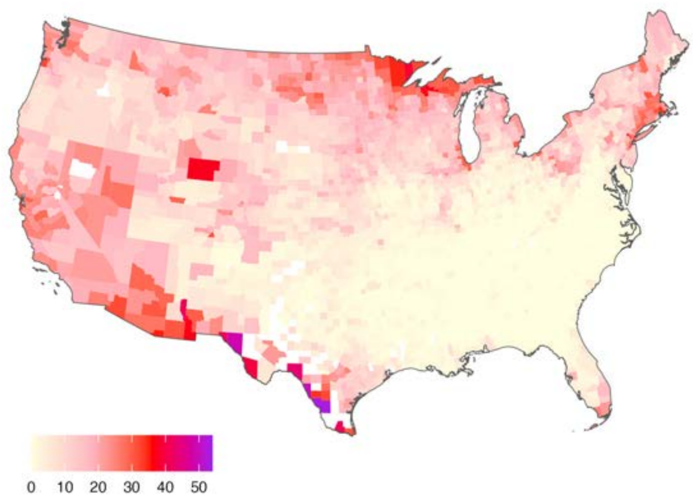
Figure D.1. US Foreign-Born Population Share, 1850-2010



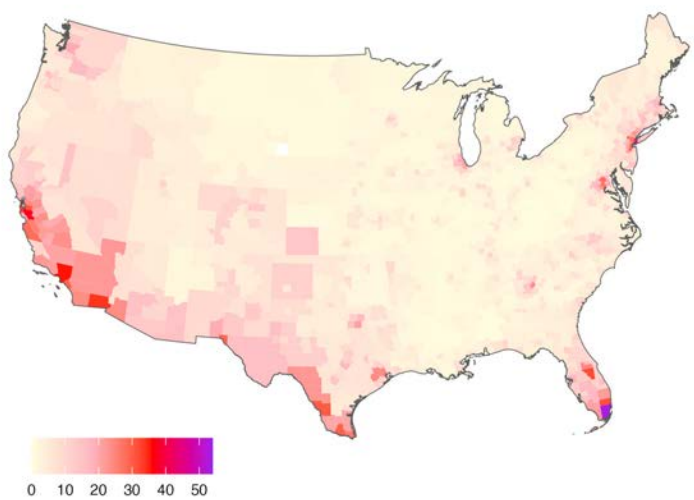
Notes: This figure shows the foreign-born population share by area in the United States from 1850 to 2010. The data comes from the IPUMS USA Census. Data for 1850-1940 is based on full-count census records, 1950-1960 comes from a 1% sample, 1970 is from a 1% metropolitan sample (fm1), 1980-1990 uses a 5% state sample, 2000 comes from a 5% sample, and 2010 is calculated from the ACS 5 year Census.

Figure D.2. Immigrant Population Share

(A) 1920



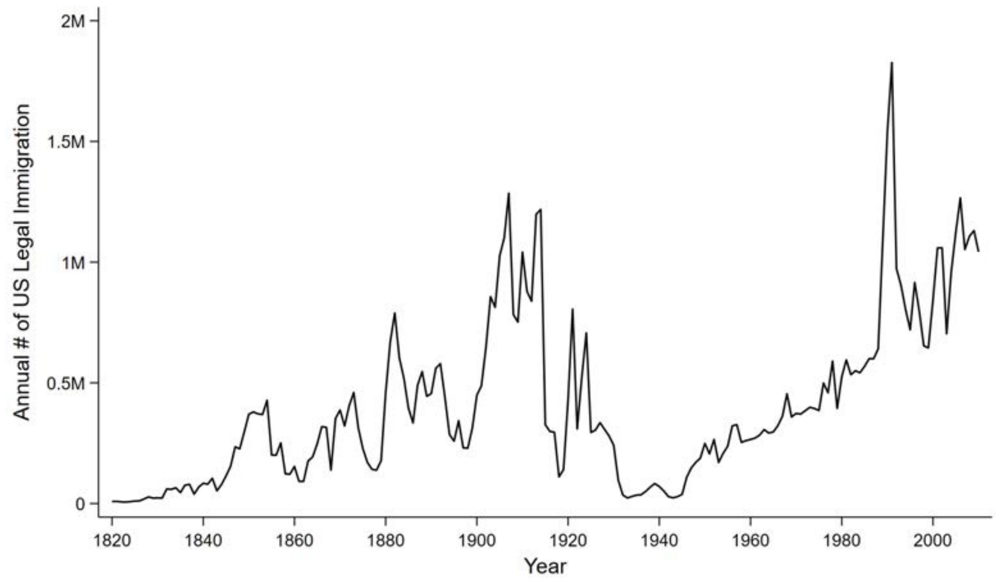
(B) 2010



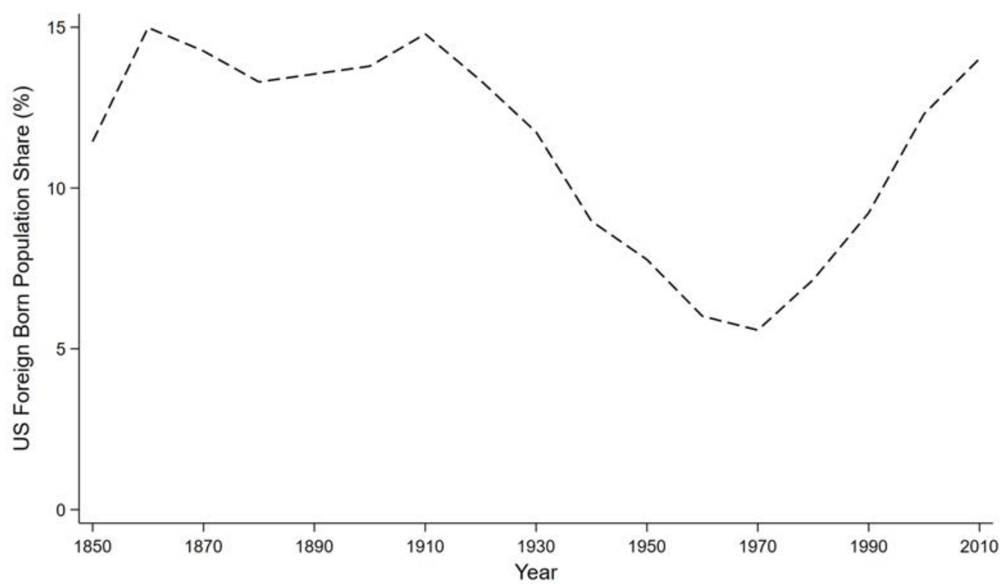
Notes: Panel A (resp., Panel B) plots the distribution of immigrant shares (%) in 1920 (resp., 2010) across US counties.

Figure D.3. International Immigration in American History

(A) Annual # of US Legal Immigration (1820-2010)



(B) US Foreign-Born Population Share (1850-2010)



Notes: Panel A shows the number of persons admitted for lawful permanent residence during the 12-month fiscal year ending on September 30, from 1820 to 2010. The data is sourced from Migration Policy Institute tabulations of the US Department of Homeland Security, Office of Immigration Statistics, Yearbook of Immigration Statistics (various years). Panel B shows the share of immigrants in the total population from 1850 to 2010. The data comes from the IPUMS USA Census. Data for 1850-1940 is based on full-count census records, data for 1950-1960 comes from a 1% sample, 1970 uses a 1% metropolitan sample (fm1), 1980-1990 uses a 5% state sample, 2000 comes from a 5% sample, and 2010 is calculated from the ACS 5-year Census.

Table D.1. Top-15 Immigrant Sending Countries, 1920 and 2010

Year	1920		2010	
Rank	Country	Immigrant Share (%)	Country	Immigrant Share (%)
1	Germany	11.65	Mexico	28.94
2	Italy	11.40	India	5.551
3	Russia	10.15	China	5.096
4	Canada	9.018	Philippines	4.302
5	Poland	7.590	Germany	2.942
6	Ireland	7.386	Vietnam	2.927
7	England	5.836	El Salvador	2.808
8	Austria	4.798	Korea	2.711
9	Sweden	4.448	Cuba	2.524
10	Mexico	3.521	Russia	2.492
11	Hungary	2.845	Canada	2.330
12	Norway	2.592	Dominican Republic	2.047
13	Czechoslovakia	2.149	Guatemala	1.924
14	Scotland	1.834	England	1.820
15	Denmark	1.304	Jamaica	1.614

Notes: The table reports the share of immigrants from each sending countries, relative to all foreign-born individuals living in the US in 1920 and 2010, respectively. The data comes from the IPUMS USA Census. Data for 1920 is based on full-count census records, and 2010 is calculated from the ACS 5 year Census.

Table D.2. Climate Distance and International Immigration: Historical Period

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.369***	-0.390***	-0.390***	-0.242***	-0.125***	-0.139***
	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)	(0.014)
	[-0.004]	[-0.004]	[-0.004]	[-0.002]	[-0.001]	[-0.001]
Precipitation Distance	-0.014***	-0.011***	-0.011***	-0.005***	0.005	0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
	[-0.001]	[-0.001]	[-0.001]	[-0.000]	[0.000]	[0.000]
Distance (in 100km)		-0.026***	-0.026***	-0.054***	-0.096***	-0.093***
		(0.001)	(0.001)	(0.006)	(0.011)	(0.007)
		[-0.002]	[-0.002]	[-0.004]	[-0.006]	[-0.006]
Observations	2,348,136	2,348,136	2,348,136	2,348,136	2,186,624	2,030,556
Pseudo R-squared	0.190	0.248	0.270	0.488	0.839	0.910
Mean Temp. Dist.	10.20	10.20	10.20	10.20	10.08	9.978
SD Temp. Dist.	5.999	5.999	5.999	5.999	6.003	6.029
Geographic Distance		Yes	Yes	Yes	Yes	Yes
Period FE			Yes	Yes	Yes	
Continent o FE				Yes		
US State d FE				Yes		
Country o FE					Yes	
US County o FE					Yes	
Country o $\times$ Period FE						Yes
US County d $\times$ Period FE						Yes

Notes: The sample is from Burchardi et al. (2019) sample. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin country and the destination county. *Distance* refers to the physical distance between the origin country and the destination county (expressed in 100 km, starting from column 2 onward). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the country of origin by state of destination by period level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table D.3. Climate Distance and International Immigration: Modern Period

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.125***	-0.136***	-0.136***	-0.130***	-0.032***	-0.029***
	(0.011)	(0.014)	(0.014)	(0.012)	(0.007)	(0.006)
	[-0.001]	[-0.001]	[-0.001]	[-0.001]	[-0.000]	[-0.000]
Precipitation Distance	-0.005***	-0.003***	-0.003***	-0.004***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[-0.000]	[-0.000]	[-0.000]	[-0.000]	[-0.000]	[-0.000]
Distance (in 100km)		-0.021***	-0.021***	-0.035***	-0.050***	-0.049***
		(0.002)	(0.002)	(0.004)	(0.003)	(0.002)
		[-0.001]	[-0.001]	[-0.002]	[-0.002]	[-0.002]
Observations	2,935,170	2,935,170	2,935,170	2,935,170	2,888,580	2,643,206
Pseudo R-squared	0.0489	0.101	0.112	0.360	0.860	0.894
Mean Temp. Dist.	10.20	10.20	10.20	10.20	10.13	10.00
SD Temp. Dist.	5.999	5.999	5.999	5.999	5.994	6.001
Geographic Distance		Yes	Yes	Yes	Yes	Yes
Period FE			Yes	Yes	Yes	
Continent o FE				Yes		
US State d FE				Yes		
Country o FE					Yes	
US County o FE					Yes	
Country o × Period FE						Yes
US County d × Period FE						Yes

Notes: The sample is from Burchardi et al. (2019) sample. *Temperature Distance* (resp., *Precipitation Distance*) is the absolute value of the difference in temperature (resp., precipitation) between the origin country and the destination county. *Distance* refers to the physical distance between the origin country and the destination county (expressed in 100 km, starting from column 2 onward). Standardized beta coefficients are reported in square brackets. Standard errors, reported in parentheses, are clustered at the country of origin by state of destination by period level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table D.4. Climate Distance and International Immigration: Historical Period Robustness

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.139***	-0.095***	-0.116***	-0.130***	-0.136***	-0.138***
	(0.014)	(0.017)	(0.015)	(0.013)	(0.013)	(0.014)
Observations	2,030,556	376,568	1,515,444	2,563,756	2,030,556	2,030,556
Pseudo R-squared	0.910	0.898	0.889	0.912	0.910	0.910
Mean Temp. Dist.	9.978	4.901	9.951	9.985	10.03	9.900
SD Temp. Dist.	6.029	3.552	6.026	6.025	6.011	6.073
Origin o × Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination d × Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Intl Mig Period	1880-1920	1880-1920	1900-1920	1880-1930	1880-1920	1880-1920
Intl Climate Source	TerraClimate	TerraClimate	TerraClimate	TerraClimate	Terraclimate	CRU
Intl Climate Period	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000	1901-1930
US Climate Period	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000	1901-1930
Origin Climate	Capital City	Capital City	Capital City	Capital City	Weighted Avg	Capital City
Countries	Any	Europe only	Any	Any	Any	Any

Notes: The sample is from Burchardi et al. (2019) sample. *Temperature Distance*is the absolute value of the difference in temperature between the origin country and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between the origin country and the destination county, expressed in 100 km) are included in all columns. US county climate based on NOAA monthly gridded historical data (Vose et al., 2014). Column 1 replicates column 6 of Table D.2. Column 2 retains only European countries. Columns 3 and 4 restrict the migration period to 1900-1920 and 1880-1930, respectively. Column 5 uses different climate measures, as opposed to the capital city average, for Mexico (population weighted climate across states) and China (average climate of Taishan county, a major immigrant-sending region). Column 6 uses CRU climate data instead of TerraClimate, and uses climate data from 1901-1930. Standard errors, reported in parentheses, are clustered at the country of origin by state of destination by period level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table D.5. Climate Distance and International Immigration: Modern Period Robustness

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.029***	-0.040***	-0.026***	-0.022***	-0.024***	-0.021***	-0.020***
	(0.006)	(0.009)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	2,643,206	2,164,882	2,155,564	2,643,206	2,627,676	2,627,676	2,612,146
Pseudo R-squared	0.894	0.902	0.894	0.893	0.869	0.893	0.863
Mean Temp. Dist.	10.00	11.13	10.04	10.06	10.07	10.05	10.07
SD Temp. Dist.	6.001	5.848	5.997	5.983	5.989	5.991	5.997
Origin o × Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination d × Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intl Mig Period	1970-2010	1970-2010	1980-2010	1970-2010	1970-2010	1970-2010	1970-2010
Intl Climate Source	TerraClimate	TerraClimate	TerraClimate	TerraClimate	TerraClimate	TerraClimate	TerraClimate
Intl Climate Period	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000
US Climate Period	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000	1960-2000
Origin Climate	Capital City	Capital City	Capital City	Weighted Avg	Capital City	Capital City	Capital City
Countries	Any	Non-Europe only	Any	Any	Drop Mexico	Drop China	Drop Mexico & China

Notes: The sample is from Burchardi et al. (2019) sample. *Temperature Distance*is the absolute value of the difference in temperature between the origin country and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between the origin country and the destination county, expressed in 100 km) are included in all columns. US county climate based on NOAA monthly gridded historical data (Vose et al., 2014). Column 1 replicates column 6 of Table D.3. Column 2 retains only non-European countries. Columns 3 and 4 restrict the migration period to 1980-2010 and 1970-2010, respectively. Column 4 uses different climate measures, as opposed to the capital city average, for Mexico (population weighted climate across states) and China (average climate of Taishan county, a major immigrant-sending region). Columns 5 and 6 drop Mexico and China, respectively, while column 7 drops both Mexico and China. Standard errors, reported in parentheses, are clustered at the country of origin by state of destination by period level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

E US Internal Migration

E.1 Historical context

Between 1860 and 1940, the geographic distribution of the US population changed dramatically. This is shown in Figure E.1, which plots population density across counties from 1860 (Panel A) to 1940 (Panel D). By 1900, the US population center of gravity had shifted from the East to the Midwest. The 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 E.1, we report the characteristics of the male population by decade: 1860 (Panel A), 1900 (Panel B), and 1940 (Panel C), for the full count (columns 1-3) and the linked sample (columns 4-6).[51] In 1860, 16.7% of men 15 or older lived in urban areas, and 52.5% (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.

E.2 Linked Samples and 1940 Full Count Census

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).[52] 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 C.2, we focus only

[51] As for our analysis of Norwegian immigrants, we restrict attention to the male population to facilitate the comparison between the full count and the linked sample (see Appendix E.2). 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 more likely to migrate. All results are unchanged when using different age or sample restrictions.

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

on men. We also exclude any individual who was younger than 15 in the baseline year.[53]

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 yields a dataset with 32,943,677 observations. Tables E.1 and E.2 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.[54]

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 over age 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).

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. Table E.2 pools the census data into broader periods: the entire period from 1860-1940 (Panel A), the early period from 1860-1900 (Panel B), and the latter period from 1910-1940 (Panel C). In the earlier sample period, linked individuals are 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.

In Table E.3, we focus on the linked sample and compare the characteristics of stayers (columns 1-3) and migrants (columns 4-6) over different periods. Stayers are defined as those living in the same county as the previous census, and migrants are those living in a different county. Between 1860 and 1900, relative to stayers, 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%). However, consistent with the evidence documented in Zimran (2023), the urban-rural gradient of internal migration changed during the 20$^{\text{th}}$ century: between 1910 and 1940, 50.4% and 52.8% of migrants and stayers lived in urban areas, respectively, and about 31% of men in both groups were farmers.[55]

To address any remaining concerns about linking processes, we take advantage of the fact that in 1940, for the first time, the US Census asked individuals their place of residence five years before. We use this piece of information to construct a migration matrix that tracks moves of

[53] 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.

[54] Table E.3 also compares the characteristics of individuals in the linked sample who did not move to those of internal migrants.

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

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). 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 were 20 years or older in 1940 (and were thus at least 15 years old in 1935 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.

E.3 Detailed Regression Results and Robustness Tests

Baseline specification. We start by presenting a more detailed analysis of the consolidated results reported in columns 4 and 5 of Table 1, which show the internal migration estimates for the historical (1850-1940) and modern (2011-2019) periods, respectively. Table E.4 focuses on the historical period. For completeness, we report also coefficients on precipitation distance and geographic distance, together with standardized beta coefficients in square brackets. Column 1 estimates the most parsimonious specification, including controls only for temperature and precipitation distance. Coefficients are negative and precisely estimated. In column 2, we control for physical distance to reduce concerns that the spatial correlation of climate may bias our results. The point estimate on temperature distance declines from -0.375 to -0.302, but remains 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—which is identical to column 4 of our main Table 1. 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 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. Coefficients are statistically significant and quantitatively relevant. According to the point estimate in column 5, lowering temperature distance by 5°C (the average distance in our sample) increases migration by 1.3%. This is similar to the effect of reducing the physical distance between counties by 1,000 km (e.g., Boston to Detroit). This is also in line with the elasticity of migration with respect to wages estimated in other contexts (Tombe and Zhu, 2019; Caliendo et al., 2021; Morten and Oliveira, 2024).

Results presented thus far were obtained using linked samples. To address census linking concerns (Bailey et al., 2020), we used an approach similar to Hornbeck (2023). We derive a county-pair migration matrix from the full count 1940 US Census, relying on a question that asked individuals where they lived five years before. For comparison, column 6 first replicates our baseline specification (column 5) using the linked sample, restricting attention to the 1930-40 decade. In column 7, we turn to the county-pair migration flows obtained from the 1940 full count census. Coefficients remain close to those obtained with the linked sample.[56] 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.

Next, in Table E.5 we focus on modern internal migration years (2011-2019) using IRS county-level mover data. In column 1, we only include temperature and precipitation distance. In column 2, we add geographic distance. In column 3, we add county of origin, county of destination, and calendar year fixed effects. In column 4, we further interact decade dummies with state of origin and state of destination dummies. Finally, in column 5, we present our preferred specification (column 5 of Table 1), which interacts county of origin and county of destination fixed effects with calendar year fixed effects. In all cases, the coefficient on temperature distance is negative, precisely estimated and quantitatively large. In the baseline analysis, we omit moves occurring after 2019 to reduce concerns that the COVID-19 pandemic may be influencing our results. In column 6, we document that results are virtually unchanged when extending the sample to moves occurring in 2019-2020 and 2020-2021.

Spatial correlation of climate. Because climate is spatially correlated, one may be worried that our estimates pick up the influence of physical—rather than climate—distance. This issue is more prescient in the US internal migration context, where we lack the Atlantic Ocean "break" that characterizes the international setting described earlier. Figure E.2 shows that in the US it gets colder as one goes north, but that geographic features such as mountains and water bodies lead to significant within-country variation in climate. To illustrate this point, Figure E.3 plots the distribution of climate distance for two example locations within the US.

Nevertheless, the higher the correlation between physical and climate distance, the more concrete the possibility that our internal migration results might capture migrants' distaste for traveling long distances. To explore this possibility, in Table E.6, we flexibly regress temper-

[56] 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.

ature distance on non-linear functions of physical distance. We regress temperature distance between the universe of US counties on physical distance polynomials of increasing degrees, from 1 to 6. Interestingly, we find that physical distance can explain at most 22% of the variation in bilateral temperature distance between US counties (based on the R-squared). The relatively low predictive power of these models suggests that spatial correlation does not absorb all the variation in our regressors of interest.

In Table E.7, we explore the possibility that climate distance might be highly correlated within short physical distances. This might pose a threat to our analysis, since internal migrants in our linked sample tend to travel relatively short distances.[57] Column 1 replicates column 2 of Table E.6. In subsequent columns, we regress temperature and precipitation distances on polynomials of degree 2 of geographic distance, restricting the sample to county-pairs whose physical distance falls within the band (in 100s of km) reported at the top of each panel. Column 3, for example, shows that even for short-range migration (e.g., 0 to 500km), a quadratic function of physical distance can only explain 20% of the temperature variation between counties. Taken together, while there exist general climate gradients within the US, geographic distance turns out to be a relatively weak predictor of climate, and that there remains large variation in climate distance after controlling for physical distance.

While we find these patterns reassuring, we now seek to address any remaining concerns about the spatial correlation of climate. In our main specification, we control for the geographic distance 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. In all the proceeding tables, we report our baseline results in column 1 to ease comparisons (which are identical to column 4 of our main Table 1):

1. Table E.8 controls non-linearly for geographic distance, allowing climate to vary across space in a non-linear way.

2. Table E.9 considers each direction of move (e.g., east-west; north-south) separately.

3. Table E.10 separately drops neighboring or next-to-neighboring counties, counties that belong to the same state or to adjacent states, counties within 100 km of geographic distance, and California.[58]

4. Table E.11 allows physical distance to have a time-varying impact on migration by inter-

[57] The median distance traveled by migrants in our sample is 245km.

[58] 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).

acting it with decade dummies. It also replaces geographic distance with measures that takes into account time and cost of travel, which varied over time due to the expansion of the railroad network.[59]

5. Table E.12 considers additional geographic attributes, which may be correlated with climate distance: elevation, ruggedness, and coastal access, and which may independently influence migration patterns (Albouy et al., 2021).

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.

Latitude and soil characteristics. So far, we focused on temperature distance. However, other geographic features correlated with climate—beyond elevation, ruggedness, and coastal access as shown in Table E.12—might partly drive our results. An important variable discussed in the historical literature to explain US internal migration during the 19$^{\text{th}}$ century is latitude (Steckel, 1983).[60] Table E.13 separately controls for latitude and longitude distance (and thus for the horizontal or vertical spatial correlation in precipitation and temperature, respectively), presenting the coefficients on these additional variables for completeness. Consistent with findings in Steckel (1983), the coefficient on latitude distance is negative and statistically significant. Moreover, the coefficient on temperature distance is about 50% smaller. While somewhat expected given the correlation between temperature and latitude, these patterns indicate that latitude distance is another important correlate of migration. At the same time, this additional analysis showcases the robustness of our results that temperature distance is a key predictor of migration.

Another feature that might be correlated with climate and that might affect migration, especially for farmers, is soil type. Soil formation is a long term geological process that is partly influenced by a place's biome and vegetation history, which is in turn influenced by historical climate. Table E.14 augments 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 content (column 5). In column 6, we include all variables together.[61] Coefficients

[59] See Appendix J.2 for a description of the railroad and transportation cost data.

[60] The key insight is that latitude is a proxy for daylight exposure. For farmers contemplating a move, this is important because for a crop like corn, a given seed stock is locally to its latitude (i.e., photoperiodic adaptation).

[61] 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. Soil water content refers to the water holding capacity of the soil, which depends on soil texture and structure. Spatial averages of each soil characteristic for each county were computed using the OpenLandMap product available on Google Earth Engine with data from Hengl (2018a), Hengl and Wheeler (2018), Hengl (2018b), and Hengl and Gupta (2019).

on temperature distance remain unchanged, suggesting that our results are not capturing the influence of soil characteristics. Interestingly, the coefficients on soil distance are negative and statistically significant. 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.[62]

Alternate climate measures. We next consider additional measures of climate distance. First, we consider seasonality. Mean annual temperature is the basis for the primary climate distance measure used in all of our analyses. While this measure is one of the simplest and most commonly used climate metrics, it masks a large degree of variation: Washington DC and San Francisco, for example, have similar annual mean temperatures, but very different seasonal patterns. To this end, Table E.15 replicates our preferred specification controlling for the distance in climate seasonality using two measures: the standard deviation of temperature (and precipitation) across a year (column 2) and the difference between the average yearly maximum and yearly minimum temperature and precipitation-level (column 3).[63]

In both cases, similarity in climate variability increases migration, suggesting that migrants take seasonality into account. However, the coefficient on average temperature remains negative and statistically significant. Moreover, its size is an order of magnitude larger (in absolute value) than that of climate variability.

Relatedly, we also verify that results are unchanged when measuring climate distance in other ways. For example, population growth patterns have been linked to a location's seasonal extremes measured by their January and July temperature (Glaeser and Tobio, 2007). Table E.16 shows results when origin-destination temperature distance is computed based on the mean of yearly maximum temperature (column 2), and the average temperature in the summer from April to September (column 3) and winter from October to March (column 4). These estimates can be compared to results using annual mean annual temperature in column 1, which is our baseline measure and identical to column 4 of our main Table 1.

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. To address these and similar concerns, Figure E.4 replicates our preferred specification (shown in both Table 1, column 4, and Table E.4, column 5), displayed in the first dot from the left, by including the county-pair

[62] Steckel (1983) also 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 Mollisol-type prairie soils (see `https://www.loc.gov/item/sn88064676/`).

[63] Using the above example, San Francisco would have a lower values for its temperature seasonality given its mild climate year round compared to the summer and winter extremes in DC.

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.[64] 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).[65] The very last dot includes all variables simultaneously. Coefficients on climate distances always remain negative, precisely estimated, and close to those from the baseline specification.

Additional robustness checks. Finally, we show that our results are robust to imposing alternative sample restrictions (Table E.17) and adjusting standard errors for spatial correlation (Table E.18).

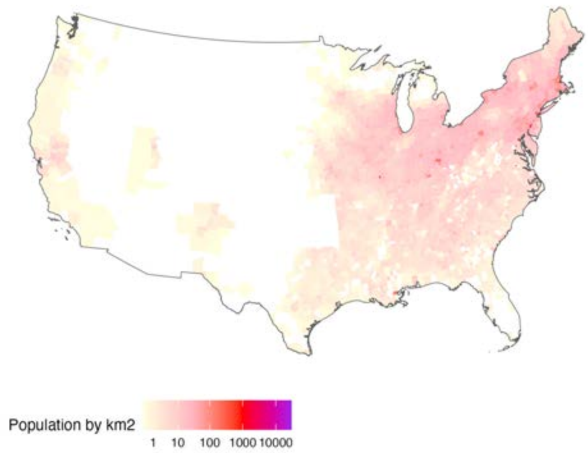
[64] In all cases, the variables are defined for men 15–64 in the baseline decade. Results are unchanged when using different age thresholds, extending the sample to women, or when considering a larger set of economic outcomes.

[65] 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.

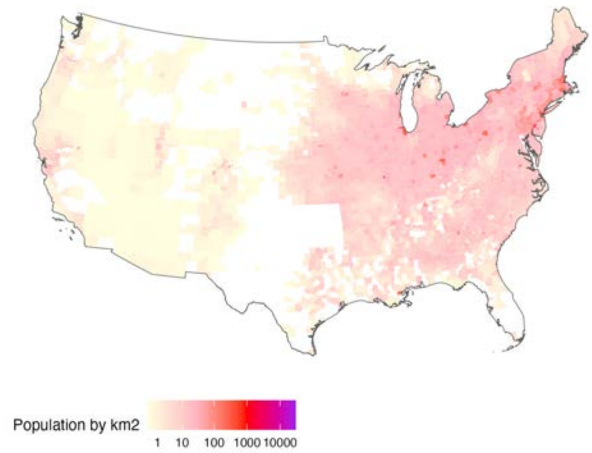
E.4 Figures and Tables

Figure E.1. Distribution of US Population, by Decade

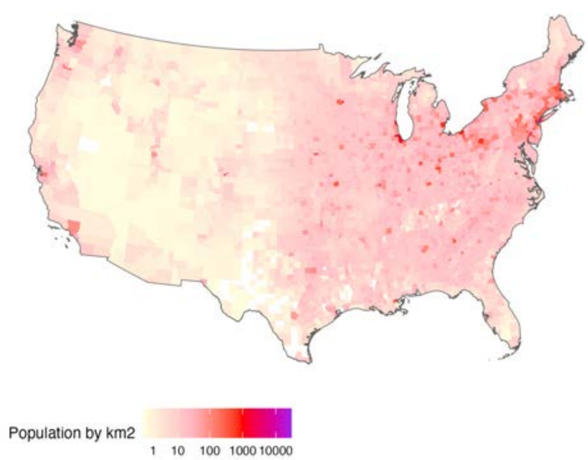
(A) 1860



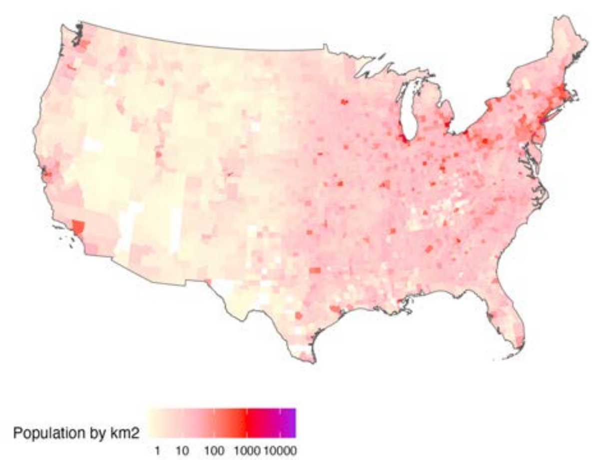
(B) 1880



(C) 1920

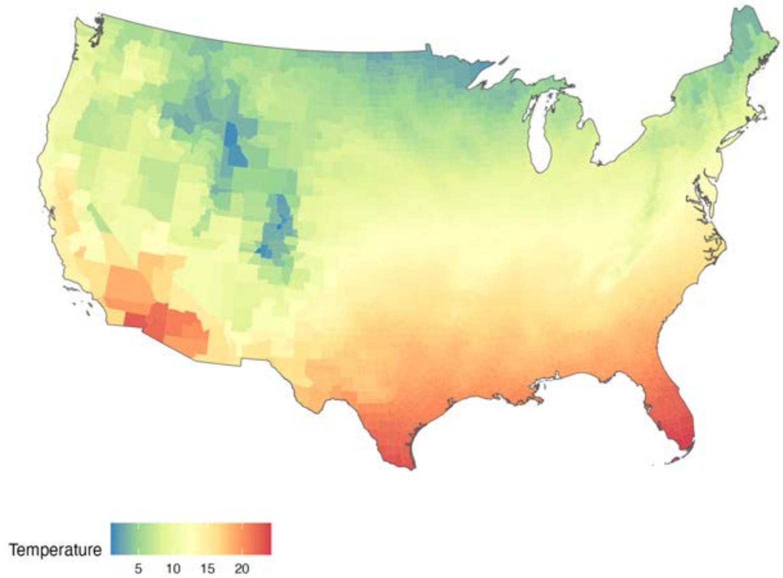


(D) 1940



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

Figure E.2. US County Temperature (1960-2000)

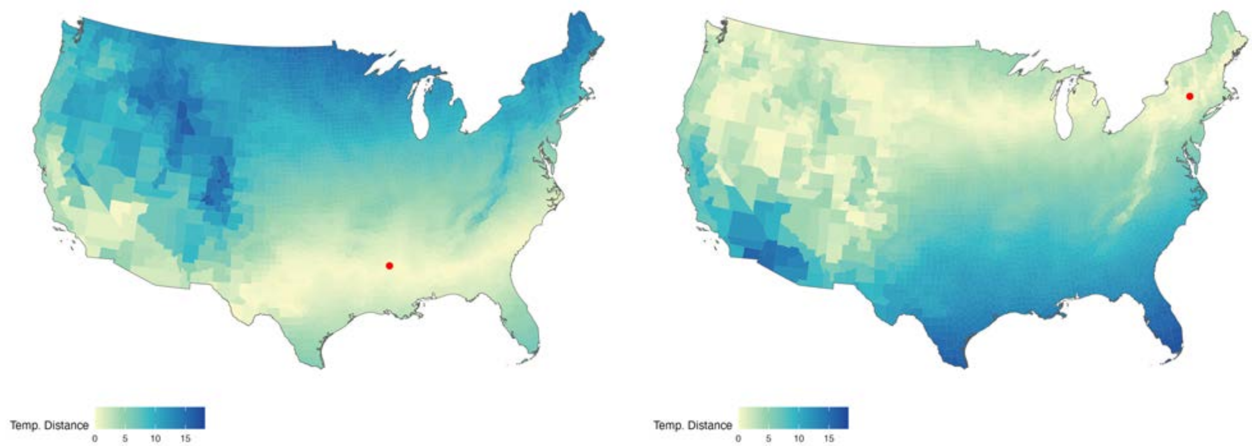


Notes: The figure displays the average temperature (in Celsius degrees) in each US county. All year averages at the US county-level for the period 1960-2000 based on NOAA monthly gridded historical data (Vose et al., 2014).

Figure E.3. Climate Distances Between US Counties

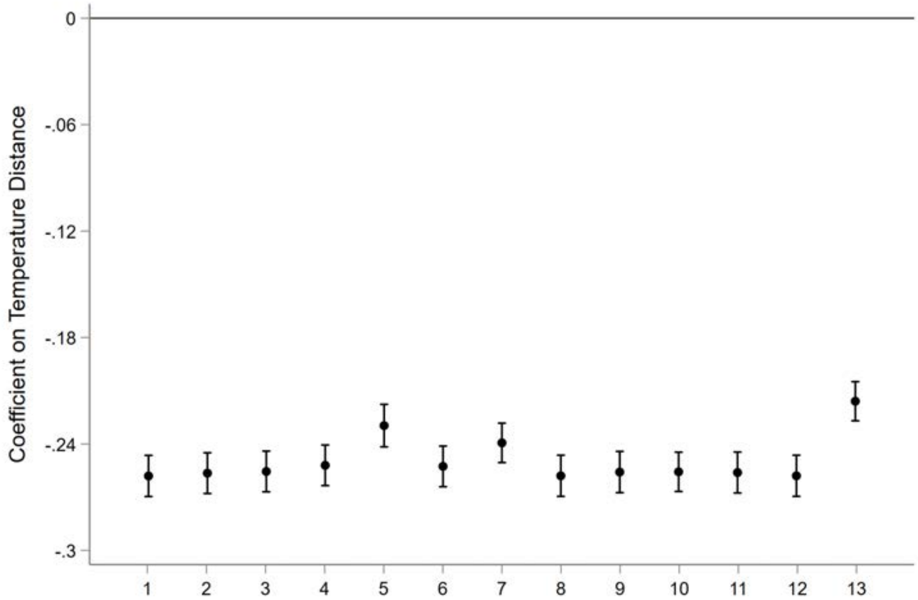
(A) Temp. Distance, Delta Region

(B) Temp. Distance, US Northeast



Notes: The figure plots the temperature distances between each US county and the centroid of the considered region. The red dot is the centroid of the considered region. Temperature is an average of all-year climate in degrees C. These averages are taken over the period 1960-2000 using NOAA data (Vose et al., 2014). Temperature distance is defined as the absolute value of the difference in average temperature between each US county and the centroid of the considered region. To interpret the figure, note that the distance in temperature between the Delta region and New York City is +6° C, even though New York City is 6° C colder than the Delta region.

Figure E.4. Temperature Distance and Migration: Including County-Pair Controls



Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature between origin and destination counties. *Temperature Distance* is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all regressions. The first dot represents the baseline specification (Table E.4, column 5). Dots 2 through 11 represent results obtained by including, one at a 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 final dot 13 reports results obtained by including all variables simultaneously. Frontier exposure is time-invariant; 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.

Table E.1. 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 US Censuses and linked samples by decade. All characteristics are measured in the year of origin. 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 who were in the labor force. All remaining dummies are defined for men 15+.

Table E.2. 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 US Censuses and linked samples by period. All characteristics are measured in the year of origin. 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 who were in the labor force. All remaining dummies are defined for men 15+.

Table E.3. 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 stayers and migrants in the linked samples by period. Stayers are defined as those living in the same county as the previous census, and migrants are those living in a different county. All characteristics are measured in the year of origin. 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 who were in the labor force. All remaining dummies are defined for men 15+.

Table E.4. Climate Distance and US Internal Migration

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.363***	-0.284***	-0.262***	-0.260***	-0.258***	-0.224***	-0.234***
	(0.011)	(0.008)	(0.006)	(0.006)	(0.006)	(0.016)	(0.021)
	[-0.102]	[-0.080]	[-0.073]	[-0.072]	[-0.066]	[-0.042]	[-0.032]
Precipitation Distance	-0.041***	-0.018***	-0.011***	-0.011***	-0.011***	-0.009***	-0.010***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
	[-0.076]	[-0.034]	[-0.021]	[-0.020]	[-0.018]	[-0.011]	[-0.009]
Distance (in 100km)		-0.138***	-0.152***	-0.152***	-0.152***	-0.176***	-0.236***
		(0.012)	(0.005)	(0.005)	(0.005)	(0.014)	(0.018)
		[-0.084]	[-0.093]	[-0.092]	[-0.087]	[-0.067]	[-0.069]
Observations	35,657,978	35,657,978	35,355,361	34,524,700	30,498,928	4,687,984	4,822,063
Pseudo R-squared	0.165	0.203	0.681	0.690	0.704	0.763	0.747
Mean Temp. Dist.	5.142	5.142	5.132	5.131	5.038	5.460	5.193
SD Temp. Dist.	3.703	3.703	3.697	3.700	3.650	3.917	3.749
Mean Precip. Dist.	32.38	32.38	32.32	32.18	30.74	31.14	32.52
SD Precip. Dist.	24.48	24.48	24.46	24.44	24	23.77	24.69
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). *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 (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. Standardized beta coefficients are reported in square brackets. 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 E.5. Climate Distance and US Internal Migration (2011-2019)

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.371***	-0.187***	-0.209***	-0.215***	-0.209***
	(0.020)	(0.018)	(0.014)	(0.011)	(0.014)
	[-0.021]	[-0.011]	[-0.011]	[-0.012]	[-0.009]
Precipitation Distance	-0.046***	-0.009***	-0.019***	-0.022***	-0.019***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
	[-0.017]	[-0.003]	[-0.006]	[-0.008]	[-0.006]
Distance (in 100km)		-0.287***	-0.213***	-0.223***	-0.214***
		(0.024)	(0.009)	(0.009)	(0.009)
		[-0.035]	[-0.024]	[-0.027]	[-0.022]
Observations	38,576,520	38,576,520	31,786,968	38,576,520	25,071,430
Pseudo R-squared	0.124	0.189	0.787	0.418	0.781
Mean Temp. Dist.	5.193	5.193	5.151	5.193	5.126
SD Temp. Dist.	3.749	3.749	3.731	3.749	3.722
Mean Precip. Dist.	32.52	32.52	31.65	32.52	30.77
SD Precip. Dist.	24.69	24.69	24.52	24.69	24.35
County o FE			Yes		
County d FE			Yes		
Year FE			Yes		
State o × Year FE				Yes	
State d × Year FE				Yes	
County o × Year FE					Yes
County d × Year FE					Yes

Notes: The sample includes all county-pairs in the contiguous US for each year from 2011-2012 to 2018-2019. *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 (expressed in 100 km). 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 E.6. 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

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

Table E.7. 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

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

Table E.8. Climate Distance and Migration: Controlling Non-Linearly for Geography

Dep. var.:	Number of Migrants			
	(1)	(2)	(3)	(4)
Temperature Distance	-0.258***	-0.324***	-0.342***	-0.347***
	(0.006)	(0.007)	(0.008)	(0.008)
Observations	30,498,928	30,498,928	30,498,928	30,498,928
Pseudo R-squared	0.704	0.669	0.658	0.655
Mean Temp. Dist.	5.038	5.038	5.038	5.038
SD Temp. Dist.	3.650	3.650	3.650	3.650
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. *Temperature Distance* is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* are included in all columns. *Distance* refers to the physical distance between counties (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. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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 E.9. Climate Distance and Migration, by Direction of Move

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.258***	-0.208***	-0.201***	-0.217***	-0.124***	-0.147***	-0.099***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.010)	(0.010)	(0.012)
Observations	30,498,928	20,101,047	19,843,728	19,130,029	10,140,378	9,679,821	9,644,442
Pseudo R-squared	0.704	0.686	0.627	0.673	0.789	0.756	0.754
Mean Temp. Dist.	5.038	4.046	4.036	4.023	6.939	6.883	6.863
SD Temp. Dist.	3.650	3.224	3.216	3.215	3.652	3.639	3.623
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. *Temperature Distance* is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* are included in all columns. *Distance* refers to the physical distance between counties (expressed in 100 km). Column 1 replicates the baseline specification (Table E.4, column 5). Columns 2-7 retain only specific directions of county pairs. All regressions control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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 E.10. Climate Distance and Migration: Dropping Selected Counties

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.258***	-0.203***	-0.184***	-0.184***	-0.177***	-0.253***	-0.253***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.007)	(0.006)
Observations	30,498,928	30,411,051	30,263,634	30,274,060	29,375,894	26,615,033	29,293,259
Pseudo R-squared	0.704	0.682	0.663	0.663	0.650	0.724	0.711
Mean Temp. Dist.	5.038	5.046	5.063	5.065	5.137	5.415	5.058
SD Temp. Dist.	3.650	3.647	3.644	3.643	3.639	3.686	3.664
County o $\times$ Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County d $\times$ 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. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. Column 1 replicates the baseline specification (Table E.4, column 5). Column 2 drops adjacent county pairs. Column 3 drops adjacent and next-to-adjacent county pairs. Column 4 removes county pairs with a migration distance of less than 100 km. Column 5 retains only county pairs across different states. Column 6 drops county pairs from adjacent states. Column 7 removes county pairs where both origin and destination are in California. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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 E.11. Climate Distance and Migration: Controlling for Time-Varying Travel Distance

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.258***	-0.258***	-0.249***	-0.269***	-0.250***	-0.294***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)
Distance (in 100km)	-0.152***		-0.069***	-0.140***		
	(0.005)		(0.006)	(0.005)		
Transportation Cost			-0.018***	-0.005***	-0.029***	-0.063***
			(0.001)	(0.002)	(0.001)	(0.003)
Observations	30,498,928	30,498,928	23,133,351	23,133,351	23,133,351	23,133,351
Pseudo R-squared	0.704	0.705	0.699	0.692	0.695	0.677
Mean Temp. Dist.	5.038	5.038	5.061	5.061	5.061	5.061
SD Temp. Dist.	3.650	3.650	3.664	3.664	3.664	3.664
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. *Temperature Distance* is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* are included in all columns. *Distance* refers to the physical distance between counties (expressed in 100 km). Column 2 replicates column 1, which estimates the baseline specification (Table E.4, 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., dollars). See Appendix J.2 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. 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 E.12. Climate Distance and Migration: Controlling for Geography

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.258***	-0.252***	-0.257***	-0.256***	-0.248***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	30,498,928	30,498,928	30,498,928	30,498,928	30,498,928
Pseudo R-squared	0.704	0.705	0.705	0.705	0.706
Mean Temp. Dist.	5.038	5.038	5.038	5.038	5.038
SD Temp. Dist.	3.650	3.650	3.650	3.650	3.650
County o × Decade FE	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes
Geographic Control		Elevation	Ruggedness	Coastal	All

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. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. In column 2, we control for the distance in elevation between county o and county d. In column 3, we control for the distance in the standard deviation in elevation between county o and county d. In column 4, we control for a dummy equal to one if both counties are coastal. Column 5 controls for the three variables mentioned above. All regressions control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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 E.13. Climate Distance and Migration: Controlling for Latitude and Longitude

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.258***	-0.114***	-0.145***	-0.111***	-0.118***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Distance (latitude)		-0.176***		-0.090***	-0.263***
		(0.008)		(0.012)	(0.009)
Distance (longitude)			0.170***	0.111***	-0.103***
			(0.010)	(0.014)	(0.003)
Observations	30,498,928	30,498,928	30,498,928	30,498,928	30,498,928
Pseudo R-squared	0.704	0.709	0.709	0.709	0.706
Mean Temp. Dist.	5.320	5.320	5.320	5.320	5.320
SD Temp. Dist.	3.829	3.829	3.829	3.829	3.829
County o × Decade FE	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	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. *Temperature Distance* is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measure of *Precipitation Distance* is included in all columns. *Distance* refers to the physical distance between counties (expressed in 100 km). Columns 2 and 3 replicate the baseline specification of Table E.4, 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. 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 E.14. Climate Distance and Migration: Controlling for Soil Types

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.258***	-0.249***	-0.246***	-0.258***	-0.253***	-0.241***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Soil: Bulk Density Distance		-0.032***				-0.021***
		(0.002)				(0.003)
Soil: Organic Matter Distance			-0.180***			-0.126***
			(0.017)			(0.018)
Soil: pH Distance				-0.023***		-0.020***
				(0.002)		(0.002)
Soil: Water Content Distance					-0.077***	-0.050***
					(0.005)	(0.005)
Observations	30,498,928	30,498,928	30,498,928	30,498,928	30,498,928	30,498,928
Pseudo R-squared	0.704	0.706	0.706	0.705	0.706	0.708
Mean Temp. Dist.	5.038	5.038	5.038	5.038	5.038	5.038
SD Temp. Dist.	3.650	3.650	3.650	3.650	3.650	3.650
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. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. All regressions control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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 E.15. Climate Distance and Migration: Average vs Seasonal Variability

Dep. var.:	Number of Migrants		
	(1)	(2)	(3)
Temperature Distance	-0.258***	-0.249***	-0.244***
	(0.006)	(0.006)	(0.006)
Precipitation Distance	-0.011***	-0.010***	-0.009***
	(0.001)	(0.001)	(0.001)
Distance (in 100km)	-0.152***	-0.147***	-0.145***
	(0.005)	(0.005)	(0.005)
Temp. SD Distance		-0.068***	
		(0.016)	
Temp. Range Distance			-0.037***
			(0.005)
Observations	30,498,928	30,498,928	30,498,928
Pseudo R-squared	0.704	0.704	0.705
Mean Temp. Dist.	5.038	5.038	5.038
SD Temp. Dist.	3.650	3.650	3.650
County o × Decade FE	Yes	Yes	Yes
County d × Decade FE	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. *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 (expressed in 100 km). *Temp. SD Distance* (resp., *Precip. SD Distance*) is the absolute value of the difference in the standard deviation of temperature (resp., precipitation) between the origin and the destination county. For each county, the standard deviation of both temperature and precipitation is first computed at the yearly level between 1895 and 1920, and then averaged over the entire period. *Temp. Range Distance* (resp., *Precip. Range Distance*) is the absolute value of the difference in the range of temperature (resp., precipitation) between the origin and the destination county. For each county, the range of temperature (resp., precipitation) is first computed as the difference between the highest and lowest daily temperature (resp., monthly precipitation) at the yearly level between 1895 and 1920, and then averaged over the entire period. Column 2 also controls for Precipitation SD Distance, while Column 3 controls for Precipitation Range Distance. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. Standard errors are clustered at the state of origin by state of destination by decade of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table E.16. Climate Distance and Migration: Alternative Climate Definitions

Dep. var.:	Number of Migrants			
	(1)	(2)	(3)	(4)
Temperature Distance	-0.258***	-0.253***	-0.278***	-0.195***
	(0.006)	(0.006)	(0.007)	(0.005)
Observations	30,498,928	30,498,928	30,498,928	30,498,928
Pseudo R-squared	0.704	0.705	0.698	0.703
Mean Temp. Dist.	5.038	5.227	4.125	6.184
SD Temp. Dist.	3.650	3.750	3.030	4.517
County o × Year FE	Yes	Yes	Yes	Yes
County d × Year FE	Yes	Yes	Yes	Yes
Temperature Measurement	Yearly Avg.	Yearly Max	Summer	Winter

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. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. *Yearly Avg.* is the baseline measure of climate, obtained by taking the 1960-2000 mean of yearly average climate. *Yearly Max* refers to the 1960-2000 mean of yearly maximum temperature. *Summer* (resp., *Winter*) corresponds to the 1960-2000 mean of summer (resp., winter) average climate. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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 E.17. Climate Distance and Migration: Alternative Samples

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.258***	-0.192***	-0.162***	-0.274***	-0.276***	-0.176***
	(0.006)	(0.003)	(0.003)	(0.007)	(0.007)	(0.006)
Observations	30,498,928	30,178,007	29,242,793	14,020,506	13,389,801	3,103,053
Pseudo R-squared	0.704	0.344	0.194	0.747	0.746	0.606
Mean Temp. Dist.	5.038	5.071	5.151	4.504	4.452	2.894
SD Temp. Dist.	3.650	3.648	3.650	3.261	3.227	2.701
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample in column 1 includes all county-pairs in the contiguous US for each decade from 1850-1860 to 1930-1940. 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. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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 E.18. Climate Distance and Migration: Alternative Standard Errors

Dep. var.:	Number of Migrants									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Temperature Distance	-0.258***	-0.258***	-0.258***	-0.258***	-0.258***	-0.258***	-0.258***	-0.258***	-0.258***	-0.258***
	(0.006)	(0.013)	(0.009)	(0.012)	(0.019)	(0.025)	(0.012)	(0.019)	(0.019)	(0.025)
Observations	30,504,354	30,504,354	30,504,354	30,504,354	30,504,354	30,504,354	30,504,354	30,504,354	30,504,354	30,504,354
Pseudo R-squared	0.704	0.704	0.704	0.704	0.704	0.704	0.704	0.704	0.704	0.704
Mean Temp. Dist.	5.142	5.142	5.142	5.142	5.142	5.142	5.142	5.142	5.142	5.142
SD Temp. Dist.	3.703	3.703	3.703	3.703	3.703	3.703	3.703	3.703	3.703	3.703
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
Latitude & Longitude			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. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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.

F Long-Run Change in Climate Distance

F.1 Detailed Regression Results and Robustness Tests

F.1.1 Additional Robustness Checks on Internal Migration

In Table F.1, we report additional results from the internal long-run analysis shown in Table 2. In column 1, we present a more parsimonious specification that only includes origin county-by-period and destination county-by-period fixed effects as well as physical distance. The negative and statistically significant coefficient is in line with that from the baseline specification, reported in column 2 to ease comparisons. Subsequent columns of Table F.1 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 E.4). This specification addresses the possibility that climate distance changed differentially across county-pairs in a way that 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 census sample and from 2011-2012 to 2018-2019 for the modern period from IRS county mover data. Coefficients on temperature distance are stable and highly robust.

F.1.2 Long-Difference Regressions

To more intuitively illustrate the variation exploited in Table 2, we replicate the analysis by taking variables in long differences, as in Burke and Emerick (2016). Using OLS, we estimate the following cross-sectional model:

$$\Delta_t \log(M_{odt}) = \alpha_o + \gamma_d + \beta \Delta_t \log(\text{Dist}_{odt}^{\text{Climate}}) + \epsilon_{odt} \tag{7}$$

where $\Delta_t \log(M_{odt})$ and $\Delta_t \log(Dist_{odt}^{Climate})$ are the long-run change in the log of migration and climate (temperature and precipitation) 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 F.2, which follows the same structure as Table 2, focusing on international and internal migration in columns 1-2 and columns 3-4, respectively. In column 1, we only include country of origin and county of destination fixed effects. In column 2, we present our preferred specification, further controlling for geographic distance. In columns 3 and 4, we turn to internal migration. Reassuringly, coefficients are negative and statistically significant. Perhaps not surprisingly, given that we are only leveraging variation along the intensive margin of migration, the implied magnitudes are smaller (in absolute value) than in Table 2, especially for the internal migration sample (columns 3 and 4).

F.2 Figures and Tables

Table F.1. Climate Distance and Migration in the Long Run: Internal Robustness

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.250***	-0.394***	-0.458***	-0.450***	-0.240***	-0.141***
	(0.029)	(0.105)	(0.096)	(0.114)	(0.069)	(0.045)
Observations	7,421,619	761,098	761,098	708,492	723,640	2,192,972
Pseudo R-squared	0.771	0.954	0.956	0.945	0.952	0.964
Mean Temp. Dist.	5.316	3.051	3.051	3.032	3.390	4.124
SD Temp. Dist.	3.835	2.792	2.792	2.776	3.118	3.410
Origin o × Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination d × Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin o × Destination d FE		Yes	Yes	Yes	Yes	Yes
Distance × Period FE		Yes	Yes	Yes	Yes	Yes
Controls × Period FE			Yes			
Migration Historical Period	1900-1910	1900-1910	1900-1910	1900-1910	1930-1940	1935-1940
Migration Modern Period	2018-2019	2018-2019	2018-2019	2014-2015	2018-2019	2018-2019

Notes: The samples include all county-pairs in the contiguous US. *Temperature Distance* is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* are included in all columns. *Distance* refers to the physical distance between counties (expressed in 100 km) and is included in column 1, as well as interacted with the period dummies in the subsequent columns. 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. Historical climate refers to the average climate from 1901 to 1930, and modern climate refers to the average climate from 1991 to 2020. Migration Period refers to the endline year of the migration flow. Origin o refers to a US county. Destination d refers to a US county. 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 F.2. Climate Distance and Migration: Long Differences

Dep. var.:	Change in Log Migrants			
	Intl-US	Intl-US	Internal US	Internal US
	(1)	(2)	(3)	(4)
Change Log Temp. Distance	-0.224**	-0.211**	-0.041**	-0.057***
	(0.110)	(0.098)	(0.017)	(0.018)
Observations	96,029	143,471	12,888	38,585
R-squared	0.573	0.534	0.778	0.664
Origin o FE	Yes	Yes	Yes	Yes
Destination d FE	Yes	Yes	Yes	Yes
Log Distance	Yes	Yes	Yes	Yes
Log Bilateral Controls			Yes	Yes
Intl Climate source	CRU	CRU		
Migration Historical Period	1900	Avg. 1900-1920	1900-1910	Avg. 1910-1940
Migration Modern Period	2010	Avg. 1990-2010	2018-2019	Avg. 2011-2019

Notes: The sample includes all county-pairs in the contiguous US (or all country-county pairs) with strictly positive migration flows in both the historical and the modern period. *Change in Log Migrants* is the difference between the log number of modern and historical migrants. *Change Log Temp. Distance* is the difference between the log temperature of modern and historical period. Similar controls are included for log precipitation distance and log geographic distance, which refers to the log of physical distance between counties (expressed in 100 km). In columns 1-4, historical climate is the average climate from 1901 to 1930, and modern climate is the average climate from 1991 to 2020. In column 1-2, Origin o refers to the origin country; and in columns 3-4, Origin o refers to a US county. Destination d refers to a US county. *Bilateral Controls* include the absolute value of the difference in the following variables: labor force participation, the manufacturing employment share, the agriculture employment share, the Black population share, the urban population share, the immigrant population share, sex ratios, population density, exposure to the frontier (Bazzi et al., 2020), market access (Donaldson and Hornbeck, 2016), and a dummy equal to one if the two counties were both connected to the railroad. All variables are measured in 1930, except for market access, which is measured in 1920. Standard errors, reported in parentheses, are clustered at at the country (resp., state) of origin by state of destination level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

G Mechanisms

G.1 Controlling for Past Migration

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. As also noted in Section 3.4, Figure 5 replicates our preferred specification for each of the three settings—Norwegian, international, and internal migration—over time, from the first to the last year available in the corresponding sample.[66] The black dots depict the coefficients for the effect of temperature distance on migration for each time period. The grey triangles plot these same coefficients after controlling for the previous period's number of migrants between the origin (e.g., a Norwegian municipality or a foreign country) and the US county of destination. In all panels, they are indistinguishable from the black dots, suggesting that climate matching influences migration independently of historical enclaves.[67]

Except for the Norwegian context (Panel A), where coefficients fluctuate across sample periods without a clear pattern, the estimates become smaller in absolute value after 1900. In the overall international migration setting (Panel B), for instance, coefficients for 1880 and 1900 are twice as large, in absolute value, as those for 1910 and 1920.[68] Point estimates for the modern migration era are about half those corresponding to 1910 and 1920, but remain quantitatively sizeable and statistically significant, either at the 5 or at the 10% level. The effects of climate distance on internal migration (Panel C) also decline over time, but, interestingly remain roughly the same in 1940 and in the post-2010 years.

G.2 US Frontier Destinations

In this section, we focus on historical US internal migration, and consider a setting where past networks should have limited scope to influence migration: 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 E.3). 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). In a first step, Bazzi et al. (2020) trace out the boundary at which population

[66] See Appendix C.2 and notes to Figure 5 for more details. To ease comparisons, climate distances are standardized in each sample to have zero mean and standard deviation equal to one.

[67] For 1865-1870, 1880, and 1860, respectively, we cannot perform this exercise, since we have no data on past migration. In a previous version of the paper (Obolensky et al., 2024), we also show that, consistent with the existing literature (Card, 2001; Munshi and Rosenzweig, 2016), past networks have a positive impact on migration.

[68] Note that the timing is consistent with the one for Norwegian immigrants, since 1910 refers to migration that occurred between 1901 and 1910.

density falls below two people per square mile—over time. Then, for each decade, they assign to the frontier all counties that are in close proximity (100 km) to the frontier line and that have a population density below six people per square mile (a threshold first indicated by Porter et al., 1890).[69]

We present results in Table G.1, reporting our full-sample baseline specification (Table 1, column 4) in column 1. In column 2, we replicate the baseline specification for the selected set of county pairs where the origin was never on the frontier and the destination was on the US frontier over the period 1850-1890. Despite the drastic reduction in sample size, the effect of climate distance remains in line with estimates derived using the full sample over the same time period. Subsequent columns verify this when considering each decade separately. These results provide additional support to the idea that past networks are unlikely to explain the strong, positive relationship between climate similarity and migration documented in the paper.

G.3 Heterogeneity by Sector and Occupation

In Section 3.4 of the paper, we explore the heterogeneity of results for historical international and internal migration by sector (Figure 6) and occupation (Figure G.1) of individuals. Since we use the linked sample for internal migration (see also Appendix E.2), we observe the occupation or the sector of the person at baseline—prior to migration. We thus use this, rather than the industry or occupation of the individual after the move.[70] Instead, since for international immigrants we rely on repeated cross-sections, we only observe information reported after migration has taken place.

To construct sector-specific migration flows, we restrict attention to men 15-64 in the labor force, and classify sectors based on the *IND1950* variable provided by IPUMS.[71] Based on the classifications in the 1950 Census Bureau industrial classification system, we extract data for five major sectors: manufacturing, agriculture, construction, transportation, and services.[72] We then aggregate the data at the origin-destination-decade-level, for each of the five sectors. This process yields sector-specific migration flows consistent with our overall migration dataset construction.

To define occupations according to their climate intensity and as to whether they are mostly

[69] 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).

[70] However, results are virtually unchanged when using industry or occupation at endline.

[71] *IND1950* recodes industry information into the 1950 Census Bureau industrial classification system, enhancing comparability across all years included in IPUMS. See U.S. Bureau of the Census, *Alphabetic Index of Occupations and Industries: 1950* (Washington, D.C., 1950).

[72] For detailed correspondence between *IND1950* codes and industries, see `https://usa.ipums.org/usa-action/variables/IND1950#codes_section`.

performed indoor or outdoor, we proceed in a similar way. Restricting attention to men 15-64 in the labor force, we rely on the detailed occupation code available in the US Census (*OCC1950*), and implement two different strategies.[73] First, we use an AI language model (ChatGPT) to classify occupation titles associated with each *OCC1950* code as: *i*) "climate intensive," *ii*)"not climate intensive," *iii*) "indoor," and *iv*) "outdoor."[74] Second, and relying on the same categories, we manually classify occupations based on our own judgment and that of a team of research assistants. We acknowledge that neither approach is perfect, and we view this analysis as suggestive, rather than definitive. In the main text, we present results using the classification obtained from ChatGPT (Figure G.1); reassuringly, they remain very similar when using our own manual classification (Figure G.2).[75]

[73] As for sectors, we use pre-migration occupation for internal migrants, and rely on the occupation reported in the US Census (after the move) for international immigrants.

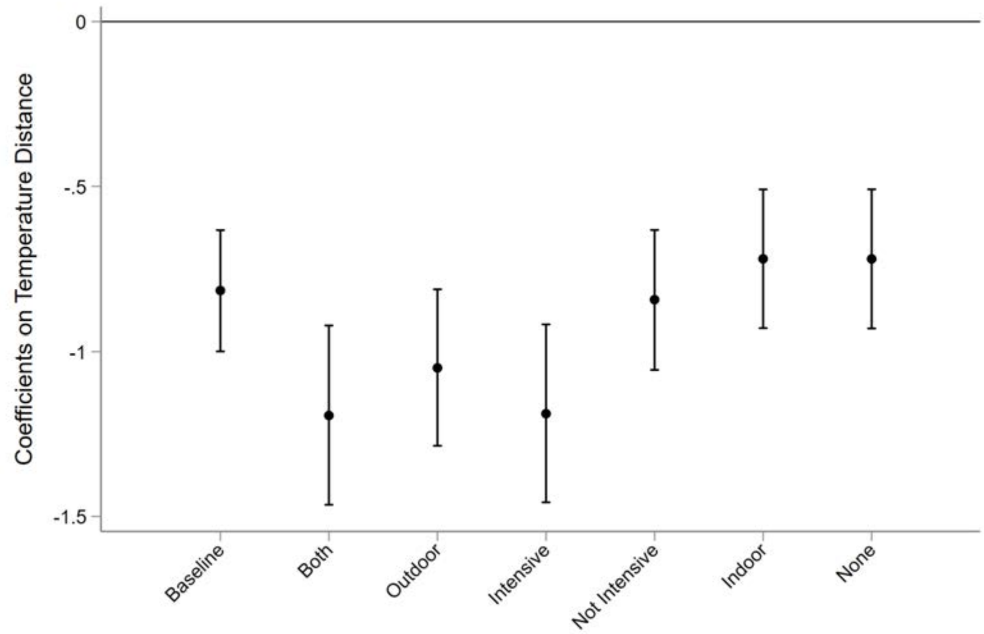
[74] Note that categories *i*) and *ii*) and *iii*) and *iv*) are mutually exclusives. However, the same occupation can be both outdoor and not climate intensive (or, outdoor and climate intensive). In the analysis, we also consider occupations that are both climate intensive and outdoor as well as those that are both indoor and not climate intensive.

[75] Examples of climate intensive (resp., not climate intensive) occupations are: farmers, gardeners, and truck or tractor drivers (resp., carpenters, managers and proprietors, and private household workers). Examples of outdoor (resp., indoor) occupations are: farmers, machinists, and laborers not otherwise classified (resp., tailors, clerical workers, and private household workers).

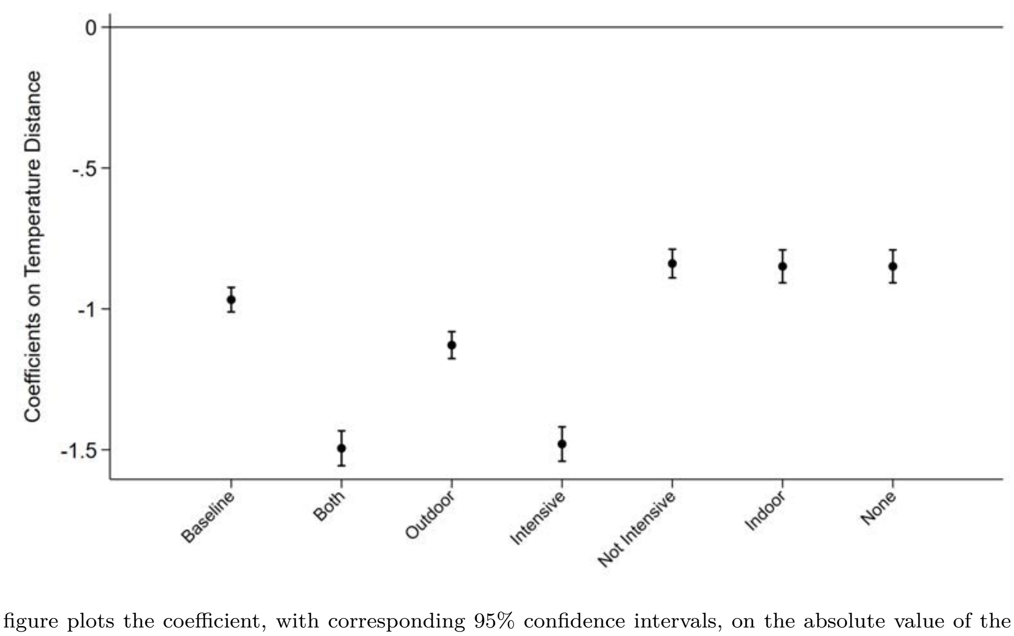
G.4 Figures and Tables

Figure G.1. Climate Distance and Migration: Heterogeneity by Occupation

(A) International Migration



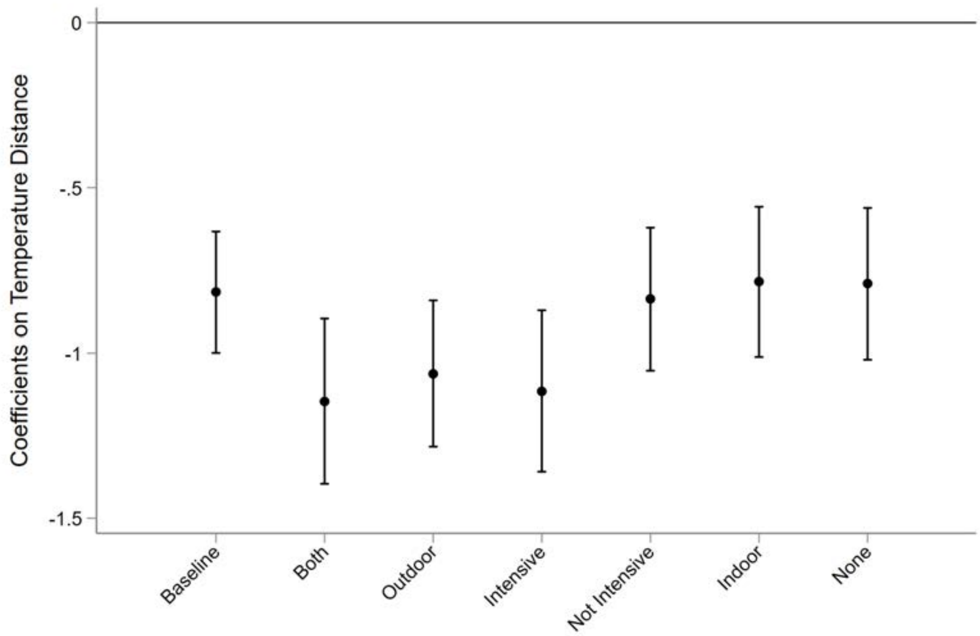
(B) Internal Migration



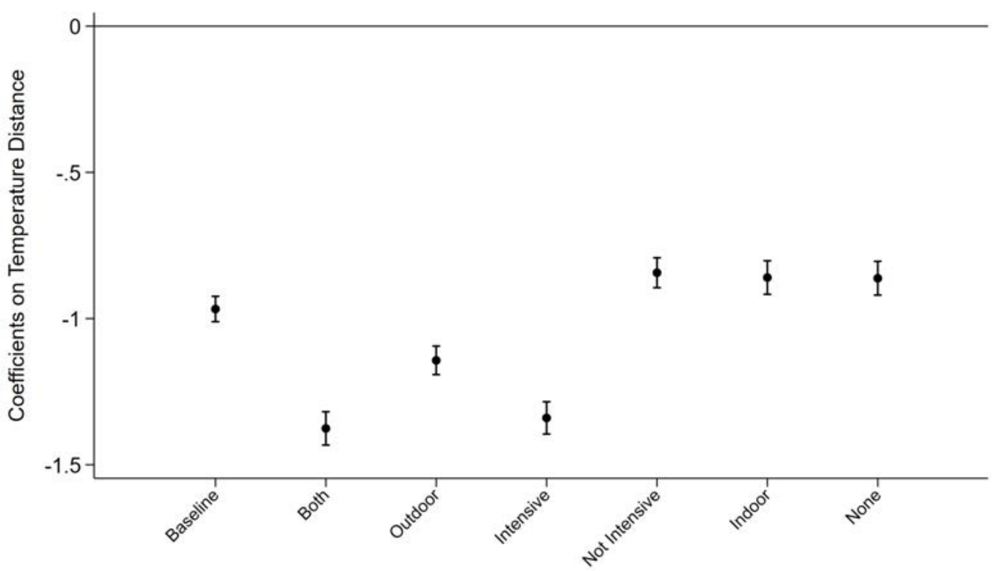
Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature between (A) international country of birth and US destination county, and (B) US origin and destination counties for internal movers. The first dot shows the baseline results of column 4 of Table 1. From the second dot onward, the number of migrants refers to migrant men in the sector reported on the x-axis (defined according to 3-digit OCC1950 code from IPUMS). The classification of whether an occupation is climate-intensive or categorized as indoor or outdoor comes from ChatGPT. Temperature distances are standardized within the relevant sample to have zero mean and standard deviation equal to one. Standard errors are clustered at the country (resp., state) of origin by state of destination by decade-level.

Figure G.2. Climate Distance and Migration: Heterogeneity by Occupation, Robustness

(A) International Migration to the US



(B) Internal Migration in US



Notes: The figure plots the coefficient, with corresponding 95% confidence intervals, on the absolute value of the difference in temperature between (A) international country of birth and US destination county, and (B) US origin and destination counties for internal movers. The first dot shows the baseline results of column 4 of Table 1. From the second dot onward, the number of migrants refers to migrant men in the occupation type reported on the x-axis (defined according to 3-digit OCC1950 code from IPUMS). The classification of whether an occupation is climate-intensive or categorized as indoor or outdoor was done manually. Temperature distances are standardized within the relevant sample to have zero mean and standard deviation equal to one. Standard errors are clustered at the country (resp., state) of origin by state of destination by decade-level.

Table G.1. Climate Distance and Migration: Evidence from the US Frontier

Dep. var.:	Number of Migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature Distance	-0.312***	-0.200***	-0.248***	-0.140***	-0.177***	-0.200***
	(0.008)	(0.011)	(0.021)	(0.019)	(0.020)	(0.020)
Observations	19,994,393	1,227,965	309,380	305,085	307,040	306,460
Pseudo R-squared	0.523	0.365	0.378	0.320	0.385	0.348
Mean Temp. Dist.	4.731	5.233	5.328	5.088	5.106	5.406
SD Temp. Dist.	3.445	3.526	3.561	3.440	3.449	3.639
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	1900	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, and is derived from the linked sample described. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. All regressions also control for county of origin (county o) by decade and county of destination (county d) by decade fixed effects. 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.

H Mortality

H.1 Deriving an Instrument for Climate Mismatch

As discussed in Section 4, to identify the causal effect of climate distance on mortality (of both immigrants and their US-born children), we derive an instrument for climate mismatch. To this end, we build on previous work by Burchardi et al. (2019) and Terry et al. (2024), adapting their methodology to our context. We proceed in three steps. First, we predict the 1970 number of non-European immigrants (our proxy for ancestry) from a given origin country to a given US county, $A_{o,d,1970}$, by combining two variables: i) the number of immigrants arrived in the US from other countries within the same continent in decade τ ($I_{o,-r(d),\tau}$), a proxy for push shocks that induced individuals to leave their country of origin; and, ii) the share of European immigrants who settled in each US county in decade τ $\left(\frac{I_{Europe,d,\tau}}{I_{Europe,\tau}}\right)$, a proxy for pull shocks shifting the relative attractiveness of each US destination.

Formally, we estimate the following equation:

$$A_{o,d,1970} = \delta_{o,r(d)} + \delta_{c(o),d} + X'_{o,d}\zeta + \sum_{\tau=1880}^{1970} a_{r(d),\tau} I_{o,-r(d),\tau} \frac{I_{Europe,d,\tau}}{I_{Europe,\tau}} + \upsilon_{o,d,t} \quad (8)$$

where, in addition to the variables mentioned before, $\delta_{o,r(d)}$ and $\delta_{c(o),d}$ are country of origin-by-region of destination and continent of origin-by-county of destination fixed effects, and $X'_{od}\zeta$ is a vector of controls including geographic distance and latitude distance between the capital city of the country of origin and the county of destination.[76] We then compute predicted ancestry as the interaction of the coefficients $\hat{a}_{r(d),\tau}$ and the $\left(I_{o,-r(d),\tau}\frac{I_{-c(o),d,\tau}}{I_{-c(o),\tau}}\right)^{\perp}$ components, residualized with respect to all the controls and fixed effects in (8).

In formulas, predicted immigration in 1970 is defined as:

$$\hat{A}_{o,d,1970} = \sum_{\tau=1880}^{1970} \hat{a}_{r(d),\tau} \left(I_{o,-r(d),\tau}\frac{I_{-c(o),d,\tau}}{I_{-c(o),\tau}}\right)^{\perp} \quad (9)$$

Second, we isolate exogenous non-European immigration shocks by interacting predicted ancestry in a given county with contemporaneous nationwide immigration from that origin. We do so by estimating by PPML the following model:

$$I_{o,d,1980} = \exp\left[\delta_{o,r(d)} + \delta_{c(o),d} + X'_{o,d}\zeta + b_t[\hat{A}_{o,d,1970} \times \tilde{I}_{o,-r(d),1980}]\right] u_{o,d,1980} \quad (10)$$

[76] Regions are defined according to the US Census Division classification: see this link for more details. We group countries into continents following Burchardi et al. (2019).

where $\tilde{I}_{o,-r(d),1980} = I_{o,-r(d),1980}(I_{Europe,r(d),1980}/I_{Europe,-r(d),1980})$ as in Terry et al. (2024). Predicted immigration is then defined as

$$\hat{I}_{o,d,1980} = \exp\{\hat{b}_t \cdot [\hat{A}_{o,d,1970} \times \tilde{I}_{o,-r(d),1980}]\} \tag{11}$$

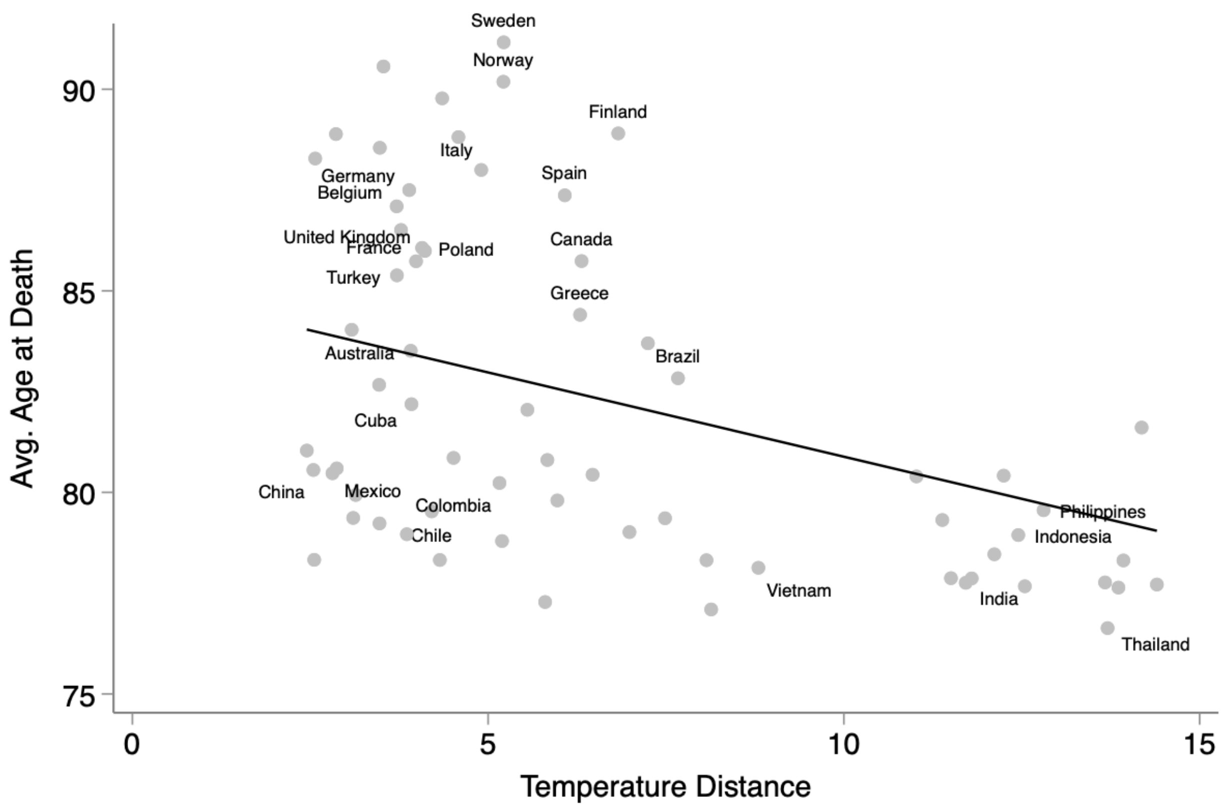
Finally, we use these matrices of country-county immigration flows to construct plausibly exogenous weights to obtain predicted origin-specific average temperature distance. Our instrument is thus:

$$\text{Clim}\tilde{\text{a}}\text{Dist}_o = \sum_d \frac{\hat{I}_{o,d,1980}}{\sum_d \hat{I}_{o,d,1980}} \times \text{ClimaDist}_{od} \tag{12}$$

where ClimaDist_{od} is the absolute value of the difference in climate between the destination county temperature and that of the origin country's capital city; both variables are computed as the average over 1960-2000.

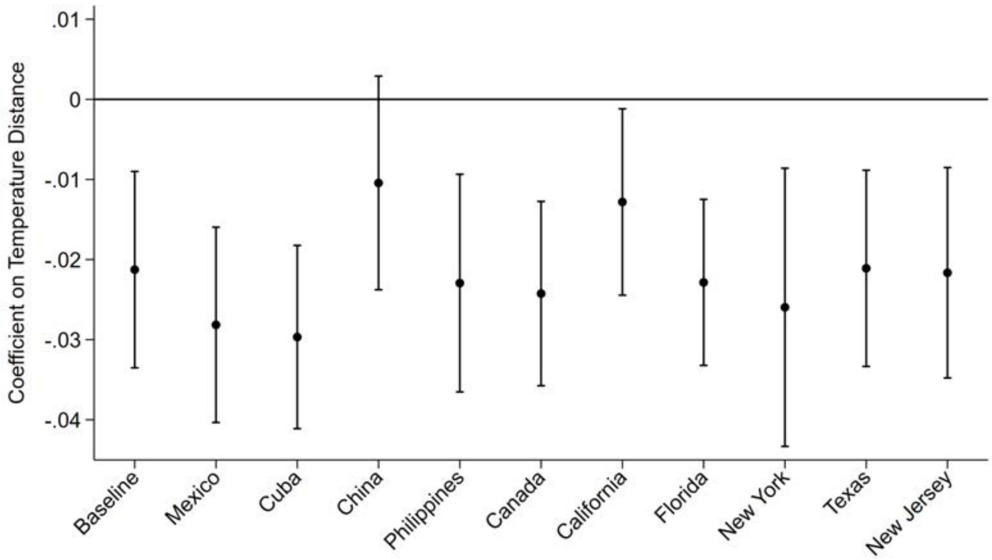
H.2 Figures and Tables

Figure H.1. Temperature Matching and Age at Death (1988-2005)



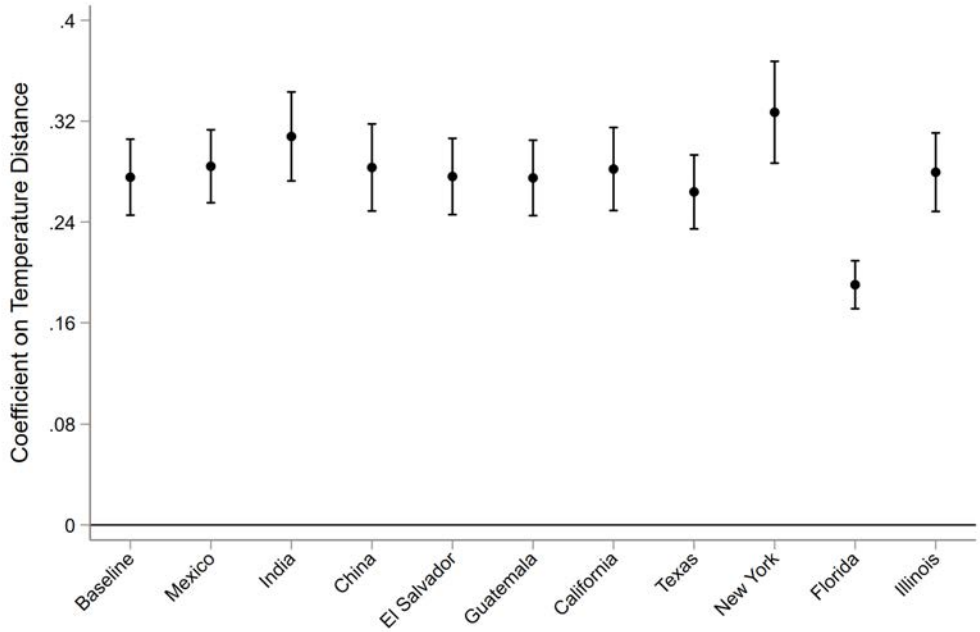
Notes: Plot shows the relationship between the average age at death of immigrants to the US and the average climate distance between their county of death and their origin country from BUNMD individual-level mortality records from 1988 to 2005, restricted to immigrants who arrived in the US after 1970 and who died after the age of 65. The regression coefficient and the corresponding robust standard errors are, respectively, -0.418 and 0.132.

Figure H.2. Climate Distance and Age at Death: Excluding Selected Samples



Notes: The sample includes all individuals aged 65 or older at the time of death, who died between 1988 and 2005, arrived after 1970, and were from outside Europe. The first point replicates column 4 of Panel B from Table 3. Points 2 to 6 successively exclude the following countries: Mexico, Cuba, China, Philippines, and Canada. Points 7 to 11 exclude the following US states: California, Florida, New York, Texas, and New Jersey. All regressions control for the predicted Precipitation Distance, the predicted number of immigrants, and individual controls (sex, mover status, year of migration, and migration distance). Standard errors are clustered at the state of death by country of origin.

Figure H.3. Climate Distance and Infant Health Outcomes: Excluding Selected Samples



Notes: The sample is restricted to non-European countries and those matching the Burchardi sample. The first point replicates column 4 of Panel A from Table 4. Points 2 to 6 successively exclude the following countries: Mexico, India, China, El Salvador, and Guatemala. Points 7 to 11 exclude the following US states: California, Texas, New York, Florida, and Illinois. All regressions control for the predicted Precipitation Distance, maternal controls (including the mother's age, marital status, and the number of live births), and the predicted number of immigrants. Standard errors are clustered at the state of death by country of origin and year-level.

Table H.1. Top 10 Countries in Mortality Data and Infant Mortality Data

Panel A. Age at Death Data	Individuals (k)	Share (%)	Age at Death
Mexico	55.48	15.52	80.59
Cuba	46.21	12.93	82.19
China	40.59	11.36	80.56
Philippines	39.81	11.14	79.56
Canada	20.93	5.850	85.74
South Korea	16.58	4.640	79.37
Vietnam	13.13	3.670	78.13
India	11.53	3.230	77.76
Jamaica	9.590	2.680	79.31
Iran	8.810	2.460	81.04
Panel B. Infant Mortality Data	**Births (k)**	**Deaths (k)**	**IMR (‰)**
Mexico	4,971	23.37	4.700
India	655.2	2.340	3.570
China	495.4	1.030	2.080
El Salvador	488.8	2.250	4.600
Guatemala	430.7	2.050	4.760
Philippines	333.1	1.560	4.680
Honduras	311.0	1.360	4.370
Vietnam	281.3	1.010	3.590
Dominican Republic	270.2	1.190	4.400
Haiti	202.1	1.720	8.510

Notes: The Age at Date sample (Panel A) comprises all people older than 65 at death and people who died between 1988 and 2005 who arrived after 1970 and are from outside Europe. The Infant Mortality sample (Panel B) is restricted to non-European countries and those matching the Burchardi et al. (2019) sample. Excluding European countries, the Age at Date sample includes a total of 152 countries, and the Infant Mortality sample includes a total of 127 countries.

Table H.2. Summary Statistics from Mortality Records

	Obs	Mean	SD	Min	Median	Max
Panel A. Age at Death Data						
Age at Death	357,399	80.44	8.808	65	80	122
Temperature Distance	357,399	6.573	5.481	0.001	4.517	24.48
Precipitation Distance	357,399	46.03	44.60	0.001	29.55	384.4
Distance(in 100 km)	357,399	66.12	48.23	1.142	62.52	184.7
Pred. Temperature Distance	357,399	9.307	4.580	3.759	8.094	17.33
Pred. Precipitation Distance	357,399	47.21	27.67	23.31	35.14	317.5
Year of Arrival	357,399	1980	7.225	1970	1980	2005
Male	357,399	0.427	0.495	0	0	1
Panel B. Infant Mortality Data						
IMR (‰)	171,577	4.494	8.999	0	3.593	600
Temperature Distance	171,577	6.247	5.066	0.001	4.906	24.47
Precipitation Distance	171,577	40.57	36.13	0.003	31.30	384.5
Distance(in 100 km)	171,577	53.66	43.95	1.142	29.49	184.7
Pred. Temperature Distance	171,577	8.58	3.937	3.721	5.853	17.88
Pred. Precipitation Distance	171,577	40.04	27.28	22.65	27.83	322.4
Mother's Age	171,577	29.64	1.817	19.13	29.01	37.33
Mother Married	171,577	0.597	0.201	0	0.557	1
Mother's Live Births	171,577	2.275	0.431	1.091	2.382	6.535
Low weight Ratio (%)	171,577	7.151	3.988	0	6.515	100
Low Gestation Weeks Ratio (%)	171,577	11.24	4.866	0	11.03	100

Notes: The Age at Date sample (Panel A) comes from BUNMD and comprises all people older than 65 at death and people who died between 1988 and 2005 who arrived after 1970 and are from outside Europe. The Infant Mortality sample comes from CDC's NVSS records (Panel B) and is restricted to non-European countries and those matching the Burchardi et al. (2019) sample. Panel B are weighted by the number of births.

Table H.3. Climate Distance and Age at Death: OLS Robustness

Dep. var.:	Age at Death				
	(1)	(2)	(3)	(4)	(5)
Panel A: Non-Europe post 1970					
Temperature Distance	-0.014***	-0.015***	-0.014***	-0.013***	-0.012***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Observations	356,897	356,897	356,810	354,685	340,719
R-squared	0.724	0.724	0.725	0.726	0.736
Mean Temp. Dist.	6.573	6.573	6.572	6.570	6.547
SD Temp. Dist.	5.482	5.482	5.481	5.481	5.474
Mean Death Age	80.44	80.44	80.44	80.43	80.38
SD Death Age	8.806	8.806	8.803	8.779	8.665
Panel B: Non-Europe post 1960					
Temperature Distance	-0.016***	-0.016***	-0.016***	-0.016***	-0.015***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	504,998	504,998	504,920	502,543	486,756
R-squared	0.712	0.712	0.712	0.713	0.722
Mean Temp. Dist.	6.394	6.394	6.394	6.391	6.378
SD Temp. Dist.	5.474	5.474	5.474	5.474	5.473
Mean Death Age	79.33	79.33	79.33	79.31	79.26
SD Death Age	8.633	8.633	8.631	8.610	8.511
Panel C: Full Sample					
Temperature Distance	-0.011***	-0.012***	-0.012***	-0.012***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	1,376,698	1,376,698	1,376,614	1,374,066	1,349,016
R-squared	0.692	0.692	0.693	0.694	0.702
Mean Temp. Dist.	5.459	5.459	5.459	5.460	5.474
SD Temp. Dist.	4.729	4.729	4.729	4.730	4.739
Mean Death Age	79.97	79.97	79.97	79.97	79.94
SD Death Age	8.579	8.579	8.578	8.566	8.499
Cohort FE	Yes	Yes			
Country FE	Yes	Yes			
County FE	Yes	Yes	Yes		
Country by cohort FE			Yes	Yes	Yes
County by cohort FE				Yes	
ZIP by cohort FE					Yes
$Imm_{o,d}^{1980}$		Yes	Yes	Yes	Yes

Notes: The sample comprises all people older than 65 at death and people who died between 1988 and 2005. Temperature distance is defined in the main text. Birth cohort refers to 10 years ones. All regressions control for Precipitation Distance and individual controls, which comprise sex, mover dummy, migration year, and the distance of the country from the capital of the origin country. Standard errors, reported in parentheses, are clustered at the state of death by country of origin level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table H.4. Climate Distance and Age at Death: 2SLS Robustness

Dep. var.:	Age at Death								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Temperature Distance	-0.021***	-0.018***	-0.012*	-0.023***	-0.016***	-0.014***	-0.014**	-0.021***	-0.021***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Observations	354,778	354,769	340,822	342,611	354,778	353,087	344,821	354,778	354,778
KP F-stat	112.4	106	103.7	112.4	67.27	110.6	104.5	112.5	112.4
Mean Death Age	80.43	80.43	80.38	80.56	80.43	80.42	80.39	80.43	80.43
SD Death Age	8.782	8.782	8.667	8.304	8.782	8.768	8.719	8.782	8.782
Continent by Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County by Cohort FE	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes
Out State Avg Death Age		Yes							
ZIP by Cohort FE			Yes						
Trimm. 1%				Yes					
Alt. Climate					Yes				
5 Year Cohort						Yes			
1 Year Cohort							Yes		
Alt. Shares								1930	1970

Notes: The sample comprises all people older than 65 at death, who died between 1988 and 2005, arrived after 1970 and are from outside Europe. Temperature distance is defined in the main text. Birth cohort refers to 10 years ones. Column 1 replicates Panel B, column 4 of Table 3 . Column 2 controls for Out State Avg Death Age, which is defined as the average age at death in the origin country, excluding individuals in the state of the county of death. Column 3 replaces County by Cohort FE with ZIP by Cohort FE. Column 4 removes observations in the top and bottom 1% of death age. Column 5 uses an alternative climate distance for Mexico and China. Columns 6 and 7 redefine Cohort in County by Cohort FE as a 5-year cohort and a 1-year cohort, respectively. Columns 8 and 9 use immigrant data from 1930 and 1970, respectively, to construct the IV. All regressions control for Precipitation Distance and individual controls, which comprise sex, mover dummy, migration year, and the distance of the country from the capital of the origin country. Standard errors, reported in parentheses, are clustered at the state of death by country of origin level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table H.5. Climate Distance and Infant Health Outcomes: 2SLS Robustness

Dep. var.:	IMR								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Temperature Distance	0.275***	0.092***	0.050***	0.275***	0.253***	0.292***	0.288***	0.275***	0.275***
	(0.015)	(0.014)	(0.002)	(0.019)	(0.015)	(0.017)	(0.017)	(0.015)	(0.015)
Observations	158,796	158,761	131,705	158,796	157,021	158,796	135,887	158,796	158,796
KP F-stat	1673	1786		12517	1665	1391	1386	1673	1673
Mean IMR	4.484	4.484	0.378	4.827	3.020	4.827	4.536	4.827	4.827
SD IMR	8.801	8.799	2.360	21.83	12.18	21.83	8.705	21.83	21.83
Continent by Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County by Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weight	Yes	Yes			Yes	Yes	Yes	Yes	Yes
Out State IMR		Yes							
Poisson			Yes						
Trimm. 1%					Yes				
Alt. Climate						Yes			
Drop 2000-2021							Yes		
Alt. Shares								1930	1970

Notes: The sample is restricted to non-European countries and those matching the Burchardi et al. (2019) sample. Column 1 replicates Panel A, column 4 of Table 4. Column 2 controls for Out State IMR, which represents the average IMR of the origin country *o* in other states within the U.S. Column 3 uses Poisson regression, with the number of infant deaths as the dependent variable and the number of births as the exposure. Column 4 is unweighted, while the other 2SLS columns are weighted by the number of births. In column 5, we remove the counties with the top 1% highest IMR. Column 6 uses an alternative climate distance for Mexico and China. Column 7 excludes observations from 2000-2021. Columns 8 and 9 use immigrant data from 1930 and 1970, respectively, to construct the IV. All specifications include maternal controls (the mother's age, marital status, and the number of live births she has had) and the predicted Precipitation Distance. Standard errors, reported in parentheses, are clustered at the state of death by country of origin by year level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

I The Homestead Act

I.1 Historical Context

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).[77] We obtain information on the universe of land patents from the Bureau of Land Management General Office Land records. 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 who arrived in a place five years before.[78] For example, we assign contracts signed between 1915 and 1925 to the 1910-1920 migration wave. As the vast majority of contracts were signed between 1870 and 1920, we focus on this time period.

Figure I.1 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 the West, but that the Homestead Act also made land available across southern states, such as Florida, Alabama, and Mississippi. We calculate the 80$^{\text{th}}$ percentile of county homesteaded shares, considering all US counties and all years from 1860 to 1920. We then construct a dummy equal to one if the homesteaded share of the destination county in a given decade surpasses this calculated 80$^{\text{th}}$ percentile.[79] Figure I.3 maps the counties that meet this threshold across decades and documents the substantial variation across both time and space.

I.2 Regression Results and Robustness Tests

We augment the baseline specification for (historical) internal migration in equation (3) by interacting the Homestead dummy with both climate and geographic distances.[80] As in the

[77] 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 of the application and provide evidence of permanent and continuous residence as well as 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.

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

[79] Figure I.2 shows graphically the definition of the threshold.

[80] 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 as those that are never homesteaded,

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 I.1, column 1. Consistent with our baseline estimates, the coefficient on temperature distance remains negative and statistically significant. More importantly for our purposes, 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.

Since the Homestead Act was explicitly designed to attract farmers, one might expect a stronger pull effect for this population—and thus, less climate matching on the margin. On the other hand, climate matching is more important to farmer migrants than non-farmer migrants (Figure 6). 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.[81] In particular, the temperature distance elasticity for farmers settling in a Homestead county is 51% smaller than that of farmers settling in a non-Homestead county. Meanwhile, the temperature distance elasticity for non-farmers settling in a Homestead county is 35% of the elasticity observed for the same migrants settling outside of a Homestead county.[82]

In the remainder of Table I.1, we document that results are robust to using different ways to define the set of origin and destination counties. In Table I.2, we also verify that the estimates are not sensitive to using different thresholds to define the Homestead dummy.

Overall, results suggest that the Homestead Act induced migrants to match *less* on climate than they would have otherwise. This finding relates to recent work showing that the Homestead Act slowed down economic development in the Western US. Allen and Leonard (2021) stress the role of agricultural path dependence due to the Homestead Act's farming obligation, while

while leaving the set of destinations unrestricted. As discussed below, all results are robust to using alternative time windows or origin-destination samples, and different thresholds to define the Homestead dummy.

[81] 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.

[82] We obtain these numbers by dividing the total elasticity of climate distance by the elasticity of climate distance net of the Homestead for each sample.

Mattheis and Raz (2019) posit that homesteaders were negatively selected.[83] Our findings offer a complementary explanation: the Homestead Act may have reduced farmers' productivity by increasing their climate mismatch. More broadly, our results are relevant for the literature on migration and labor misallocation (Young, 2013; Bryan et al., 2014; Munshi and Rosenzweig, 2016). In particular, they showcase the potential labor (mis)allocation impacts of public programs and place-based policies (Glaeser and Gottlieb, 2008; Kline and Moretti, 2014), with relevance to discussions of managed retreat and climate change adaptation.

I.3 Estimating the Value of Climate

The Homestead Act also allows us to estimate the value of climate for migrants, complementing the analysis conducted in Section 4.3. As shown throughout the paper, migrants seek to minimize the climate distance between origin and destination. However, when offered cheap land through the Homestead Act, migrants deviated from the baseline—that is, climate matching became 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 large valuation of climate similarity.[84]

To estimate the marginal value of climate similarity, we proceed as follows. First, we compute counterfactual county-to-county migration flows under two alternative scenarios. In the *baseline* scenario, we predict farmer migration based on climate and geographic distance and the roll-out of the Homestead Act between 1860 and 1920. In the *no Homestead Act* counterfactual scenario, we predict migration based on climate and physical distance only, setting the Homestead dummy to zero for all counties and all decades.[85] These flows are used to obtain the migration-weighted average climate and geographic distance traveled under the two scenarios. Under the *baseline* scenario, a migrant would travel to counties that were, on average, 3.56°C apart from their origin. Under the *no Homestead Act* scenario, the average traveled temperature distance drops to 3.48°C. The 0.08°C difference between the two scenarios represents the increase in climate mismatch attributable to free land under the Homestead Act. Let us denote by ΔDist Temp the difference in traveled distances between the *baseline* and *no Homestead Act* scenarios.

[83] Findings in Smith (2020) indicate that these negative effects were partly offset by the Homestead Act's promotion of smaller-scale land cultivation relative to the less productive use of the large plots distributed through railroads grants.

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

[85] We use the coefficients from column 2 in Table I.1 to derive counterfactual migration flows.

Second, we obtain the migration-weighed average price of land paid in both scenarios. In the *baseline* scenario, we assume that all farmers migrating to a Homestead county pay an amount equal to the administrative fee of $345 (in 2020 dollars) for 160 acres,[86] whereas in the *no Homestead Act* scenario, we assume that those migrating to non-Homestead counties purchase 160 acres of land directly from the federal government at a price of $4,763 (in 2020 dollars)[87] Combining these land prices with predicted migration flows, a migrant pays an average of $4,373 for 160 acres of land in the *baseline* scenario, versus $4,763 in the *no Homestead Act* scenario.[88] We denote the price difference between the *baseline* and *no Homestead Act* scenarios with Δp. Finally, we recover an estimate for the marginal valuation of climate and geographic distance for farmers using the following formula:

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

The marginal valuation of temperature is estimated at $4,875 in current dollars per 1°C. For comparison, farmers would be willing to pay this amount to avoid the cost of traveling an extra 188 km. Note that this estimate may be an upper bound, since homesteaded land was likely of lower quality than nearby alternatives (Mattheis and Raz, 2019). However, we find it reassuring that the value is in the ballpark of that obtained in Section 4.3, relying on a very different context (and assumptions used to estimate the value of climate).

I.4 Figures and Tables

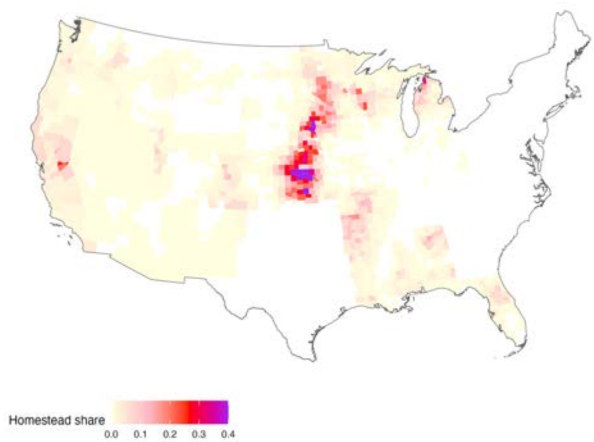
[86] For Homestead land, we assume a cost of $22 for 160 acres that includes average filing fees and commission ($345 in 2020 dollars).

[87] 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 in 2020 dollars.

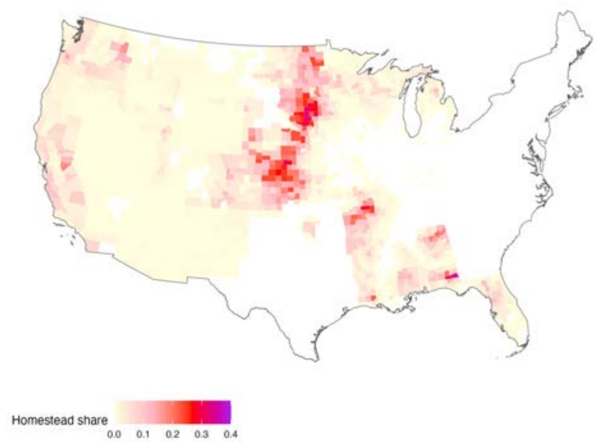
[88] The baseline scenario number is obtained by computing the average land price paid by newcomers, weighted by predicted migration flows and Homestead county dummy. In the no Homestead Act scenario, the same procedure recovers the price of direct land purchase.

Figure I.1. Homestead Shares, by County and Decade

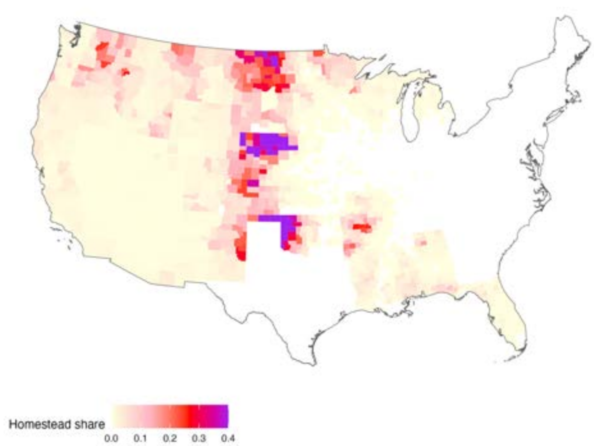
(A) 1870



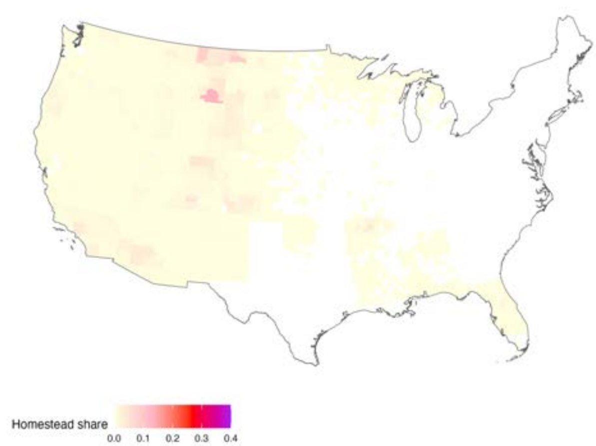
(B) 1880



(C) 1900

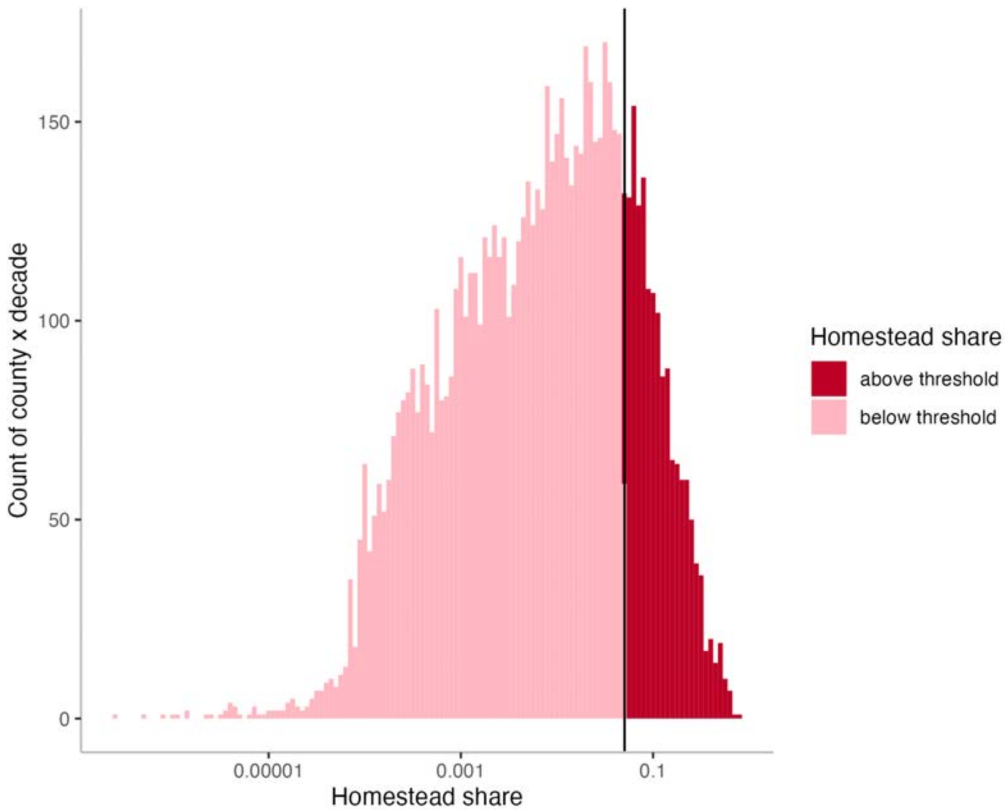


(D) 1920



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.

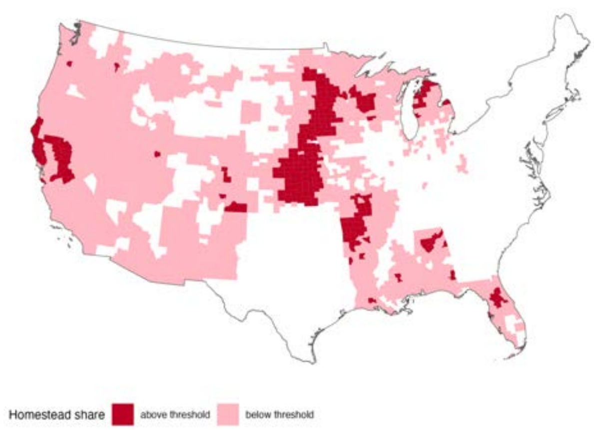
Figure I.2. Distribution of Homestead Shares (1860-1920)



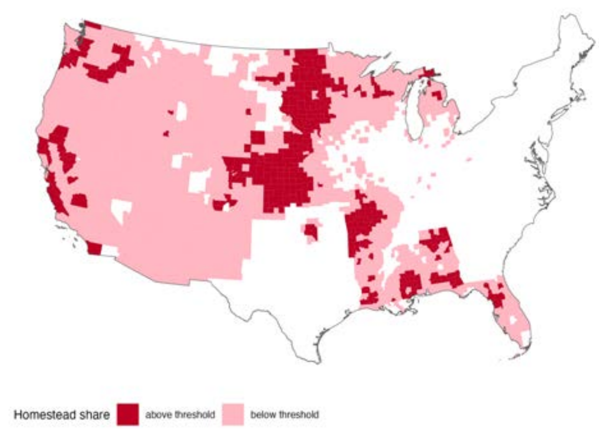
Notes: The figure plots the distribution of county-level Homestead shares for all years from 1860 to 1920. Because a Homestead patent was signed five years after the arrival of the homesteader on the land, we attribute patents signed over a given decade to the previous five years. For instance, patents 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 I.3. Homestead Counties, by Decade

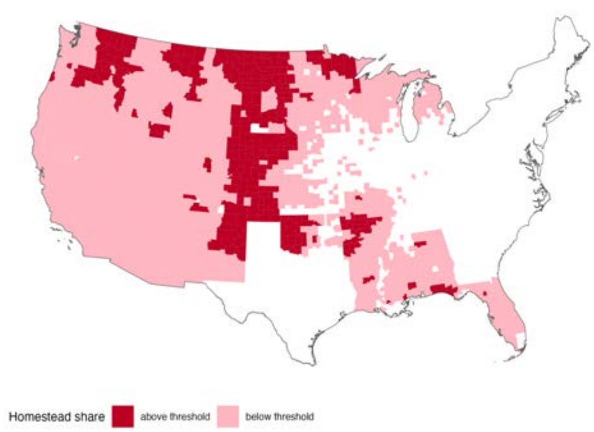
(A) 1870



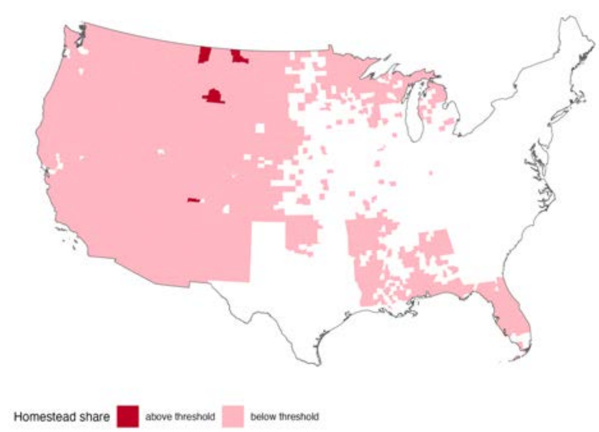
(B) 1880



(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.

Table I.1. Climate, Migration, and the Homestead Act

Dep. var.:	Number of Migrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Temperature Distance	-0.290***	-0.384***	-0.246***	-0.264***	-0.293***	-0.221***	-0.179***
	(0.007)	(0.009)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)
Temperature Distance × Homestead	0.149***	0.187***	0.159***	0.167***	0.125***	0.080***	0.040***
	(0.012)	(0.015)	(0.012)	(0.014)	(0.012)	(0.012)	(0.012)
Observations	27,678,550	26,956,182	26,937,930	19,546,989	35,164,852	13,337,957	6,345,932
Pseudo R-squared	0.597	0.387	0.717	0.637	0.578	0.534	0.499
Mean Temp. Dist.	4.850	4.809	4.828	4.608	4.948	5.320	5.361
SD Temp. Dist.	3.536	3.505	3.522	3.357	3.596	3.767	3.753
County o × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	Any	Farmer	Non-Farmer	Any	Any	Any	Any
Origins	Never Htd	Never Htd	Never Htd	Never Htd States	No Htd t-10, t	Never Htd	Never Htd
Destinations	Any	Any	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 (columns 1-3); *ii*) origin counties belong to US states that never offered any Homestead land (column 4); *iii*) origin counties did not offer any Homestead land in the previous decade (column 5); *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 6 and 7). *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. We also control for *Precipitation Distance* × *Homestead* and *Distance* × *Homestead* in all columns. In columns 1 to 3 *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. Columns 1 and 4-7 include all migrants, while columns 2 and 3 restrict attention to men 15+ (in the baseline year) who were farmers and non-farmers, respectively. County o (resp., county d) refers to county of origin (resp., of destination). 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 I.2. Climate, Migration, and the Homestead Act: Robustness to Area Threshold

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.281***	-0.289***	-0.288***	-0.290***	-0.289***
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Temperature Distance × Homestead	0.171***	0.180***	0.177***	0.149***	0.091***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)
Observations	23,031,689	24,027,360	25,577,474	27,678,550	30,637,624
Pseudo R-squared	0.625	0.615	0.607	0.597	0.586
Mean Temp. Dist.	4.664	4.711	4.773	4.850	4.920
SD Temp. Dist.	3.402	3.439	3.485	3.536	3.579
County o × Decade FE	Yes	Yes	Yes	Yes	Yes
County d × Decade FE	Yes	Yes	Yes	Yes	Yes
Threshold (percentile)	50	60	70	80	90

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, and is derived from the linked sample. *Temperature Distance*is the absolute value of the difference in temperature between the origin and the destination county. Similarly constructed measures of *Precipitation Distance* and *Distance* (physical distance between counties, expressed in 100 km) are included in all columns. We also control for *Precipitation Distance* × *Homestead* and *Distance* × *Homestead* in all columns. *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 distribution 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. 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.

J Other Items

J.1 Surname-Level Analysis for German Immigrants

In Panel B of Figure B.4, we focus on German immigration to the US in the late 19$^{\text{th}}$ 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 census data for late 19$^{\text{th}}$ 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 (weighted) average German temperature, based on the geographic distribution of that surname in Germany. Likewise, leveraging the distribution of German immigrants across US counties, we compute the (weighted) 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.[89] Each entry contains information about the soldier's first and last name, and town of birth. The lists allow us to infer the geographic 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). However, the lists represent a rare source of information on the geographic distribution of the population in a country that experienced mass migration to the US at the turn of the 20$^{\text{th}}$ 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's geolocation tool. 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 characters, accents, and *umlauts*. These steps leave us with a list of $132,910$ surnames. Next, we select geographically clustered surnames, in order to ensure that the

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

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.[90] 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 misspellings. Second, surnames associated with only one individual would automatically get selected based on our definition of geographically clustered surnames. Given 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 exclude them.[91] This leaves us with $9,633$ surnames.

The average German temperature associated with 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 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 0.1.[92] 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:

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

[90] The threshold was selected as follows. First, for each surname, we computed the distribution of distances from the centroid of each individual with that surname. Then, for each surname we considered the 85^{th} percentile (the 85^{th} percentile of the distance for the average surname is 338km). Next, within the distribution of the 85^{th} percentiles, we picked the 10% of surnames with the lowest 85^{th} percentile. Such 10% defines an implicit threshold of 30km (in terms of 85^{th} percentile).

[91] 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.

[92] 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 C.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 with surname s.[93] To see this, note that $\text{Temp}^{GER}\left(s^{US}\right)$ is the average of the temperature associated to each matched surname weighted 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.

In the last step, 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 the US by a German immigrant with surname s.

We use this dataset to estimate surname-level regressions that correlate the US and the German average temperatures associated with the same surname. Results are presented in Panel B of Figure B.4, in Section 2.2.

[93] 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. To allow for perfect matches, whenever $d\left(s^{GER}, s^{US}\right) = 0$ we set $d\left(s^{GER}, s^{US}\right) = 10^{-8}$.

J.2 Railroads and Transportation Costs

Data on the US railroad network expansion are taken from Donaldson and Hornbeck (2016). Figure J.1 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 stagecoach connections (Figure J.2). 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 \$0.03 for all other edges.[94] Then, with this weighted 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.

[94] 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 J.1. Railroads Expansion over Time

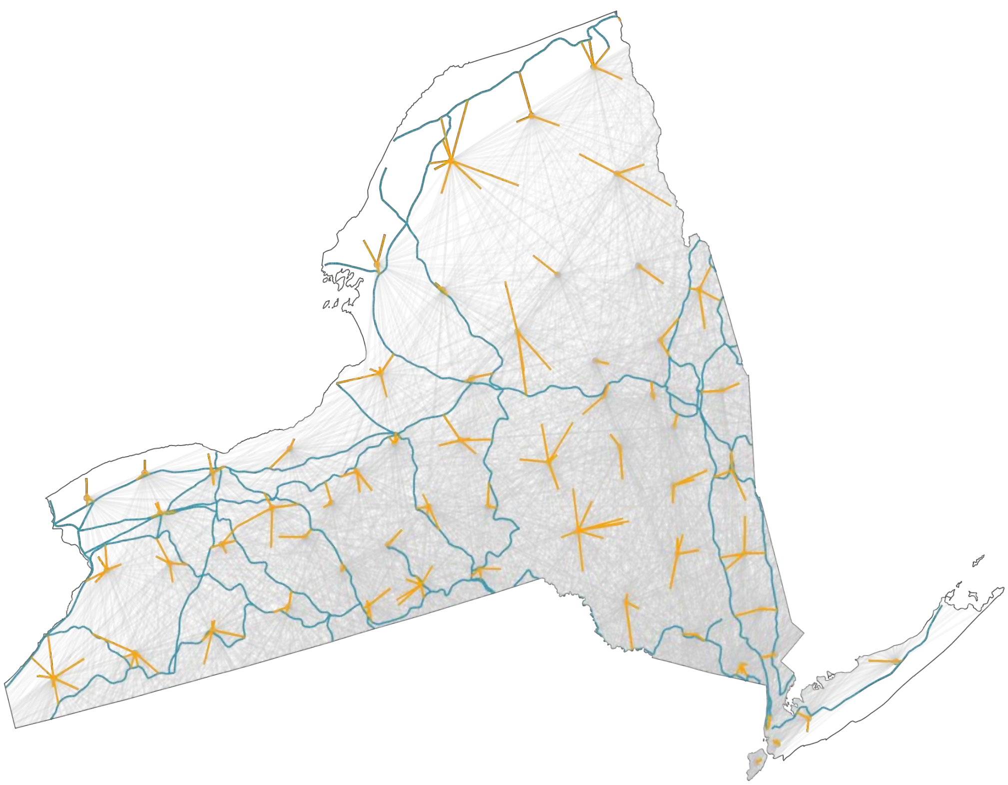
(A) 1860

(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 J.2. 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.